Graph Neural Networks for Electric and Hydraulic Data Fusion to Enhance Short-term Forecasting of Pumped-storage Hydroelectricity

Raffael Theiler[1], Olga Fink[2]

[1,2] *EPFL, Lausanne, Vaud, 1015, Switzerland*
raffael.theiler@epfl.ch
olga.fink@epfl.ch

ABSTRACT

Pumped-storage hydropower plants (PSH) actively participate in grid power-frequency control and therefore often operate under dynamic conditions, which results in rapidly varying system states. Predicting these dynamically changing states is essential for comprehending the underlying sensor and machine conditions. This understanding aids in detecting anomalies and faults, ensuring the reliable operation of the connected power grid, and in identifying faulty and miscalibrated sensors. PSH are complex, highly interconnected systems encompassing electrical and hydraulic subsystems, each characterized by their respective underlying networks that can individually be represented as graph. To take advantage of this relational inductive bias, *graph neural networks* (GNNs) have been separately applied to state forecasting tasks in the individual subsystems, but without considering their interdependencies. In PSH, however, these subsystems depend on the same control input, making their operations highly interdependent and interconnected. Consequently, hydraulic and electrical sensor data should be fused across PSH subsystems to improve state forecasting accuracy. This approach has not been explored in GNN literature yet because many available PSH graphs are limited to their respective subsystem boundaries, which makes the method unsuitable to be applied directly. In this work, we introduce the application of *spectral-temporal graph neural networks*, which leverage self-attention mechanisms to concurrently capture and learn meaningful subsystem interdependencies and the dynamic patterns observed in electric and hydraulic sensors. Our method effectively fuses data from the PSH's subsystems by operating on a unified, system-wide graph, learned directly from the data, This approach leads to demonstrably improved state forecasting performance and enhanced generalizability.

1. INTRODUCTION

In power grids, pumped-storage hydropower plants (PSH) are well-established for large-scale energy storage due to their efficiency, scalablilty and flexibility. In this role, these plants dynamically respond to potentially large fluctuations in grid demand. In the transition towards smart grids, PSH sensor data is collected via *wide area measurement systems (WAMS)* (Pagnier & Chertkov, 2021a) and is stored in centralized *energy management systems (EMS)*. By processing the aggregated WAMS data, modern EMS provide crucial functionalities for PSH operators such as load forecasting, real-time monitoring, distribution and demand-side management, and various decision support tools aimed at increasing efficiency and sustainability. To enhance system reliability, EMS implement anomaly and sensor fault detection based on short-term forecasting and state estimation, playing a pivotal role in preventing failures that could lead to widespread power grid outages and significant economic losses. However, the dynamic operation of the PSH and the vast amounts of data transmitted by the WAMS significantly complicate the task. In the PSH environment, conventional state estimation is often ineffective because the computation can take several minutes (Li, Pandey, Hooi, Faloutsos, & Pileggi, 2022). This delay leads to a rapid divergence between the most recent and the previously used system state for the estimation, resulting in numerous false-positives when applied to anomaly detection. Consequently, it becomes challenging to maintain a comprehensive overview of the system's health and performance. As a solution, deep-learning-based short-term state forecasting has recently been applied, which offers significantly faster processing times and holds the potential to benefit from the additional data increasingly collected at a high sampling rate (Kundacina, Cosovic, & Vukobratovic, 2022).

Developing deep-learning-based state forecasting for PSH is particularly challenging. These challenges stem from the necessity to accurately represent two distinct physical domains within the PSH: the hydraulic and electrical systems. Al-

though these domains are mechanically interconnected by electromagnetic generators, they are traditionally modeled independently in mechanical engineering. This division mainly stems from the distinct dynamics governing each subsystem. It is, therefore, challenging to model hydraulic and electrical domains simultaneously. Nonetheless, considering the direct causal relationship between the systems – wherein kinetic energy is transformed to electric energy – we hypothesize that fusing data from both subsystems, which operate under a unified control input, can significantly enhance state forecasting.

To address the challenge of fusing electric and hydraulic data, we posit that both subsystems of the PSH consist of extensive networks, which are coarsely monitored with sensors that can be represented in the non-Euclidean graph domain. By operating on this more effective graph representation, which can capture biases given by the PSH system architecture and homophily biases, addressing the phenomenon that sensor measurements tend to be connected with "similar" or "alike" others (Ma, Liu, Shah, & Tang, 2023), *graph neural networks* (GNNs) have recently gained significant attention. When a graph is available, GNNs have been effectively applied in key applications to (hydro) power plants (Liao, Bak-Jensen, Radhakrishna Pillai, Wang, & Wang, 2022) and in the broader power grid environment. However, these methods depend on the availability of apriori graphs, derived from PSH's electrical and hydraulic network diagrams. Therefore, their applicability is limited by the fact that, although the underlying network structure of both electric and hydraulic subsystems of a PSH can be modeled as a graph, network diagrams for PSH exist typically only separately for each subsystem. Consequently, most graph-based methods are confined within the boundaries of their respective systems. To overcome this limitation, we propose learning a PSH sensor graph from latent dependencies in the data. While it has previously been demonstrated that graph structures can be efficiently learned from data, this approach remains unexplored in the context of hydropower plants. In light of this, inspired by (Cao, Li, Ma, & Tomizuka, 2021), we propose using *spectral-temporal graph neural networks* (STGNN) to learn a latent correlation graph structure across the entire PSH asset for the fusion of electric and hydraulic data, leveraging the self-attention mechanism.

Compared to numerical simulation, our data-driven GNN-based methodology is computationally inexpensive and does not require expert knowledge while maintaining interpretability, due to the accessibility of the learned graph. Our proposed approach can be easily transferred to different PSH assets without any calibration. To the best of our knowledge, there is no other work on data-driven electric and hydraulic data fusion for PSH using graph neural networks.

To summarize, in this work, we introduce the application of attention-based graph neural networks to effectively learn

intra- and interdependencies between the subsystems' sensors to enhance the short-term state forecast in the pumped storage power plant (PSH) environment. We tackle several challenges when applying state forecasting to PSH:

- In line with the dynamic operation of the PSH, the dynamic behavior of sensors adds complexity to the forecasting. We propose a *spectral-temporal graph neural network* (STGNN) that effectively captures these patterns by incorporating the PSH' underlying structural and homophily biases, such as load patterns that are reflected across sensor measurement sites.

- State forecasting in the PSH environment depends strongly on environmental parameters, such as temperature, daily load profiles, and power grid customer-related factors, which cannot be modeled in numerical simulations (Lin, Wu, & Boulet, 2021). In contrast, our STGNN is able to learn these factors from data.

- The PSH is a spatially distributed complex system that spans across the hydraulic and electric domains. We propose a graph learning module that learns a unified graph representation across the hydraulic and electrical subsystems from latent dependencies in the data.

- We assess the performance of our method on a multivariate PSH dataset containing 58 signals, showcasing the dynamic operation of the asset.

The reminder of this paper is organized as follows: Sec. 2 reviews relevant literature that focuses on graph-based deep learning and data fusion. Sec. 3 introduces our STGNN approach. In Sec. 4, we discuss the case study conducted on a Swiss PSH plant, including the experimental and training setups. In Sec. 5, we present our results. Finally, Sec. 6 concludes this work and outlines future steps.

2. BACKGROUND AND RELATED WORK

Conventional machine learning applied to power systems have primarily focussed on linear regression models and recurrent neural networks (Zheng, Xu, Zhang, & Li, 2017). These methodologies continue to be effective and provide competitive results, particularly in areas like short-term load forecasting (Guo, Che, Shahidehpour, & Wan, 2021) and daily peak-energy demand forecasting (Kim, Jeong, & Kim, 2022). Since their introduction, GNNs (Bronstein, Bruna, LeCun, Szlam, & Vandergheynst, 2017) have been applied to many tasks in power systems. By now, graph-based deep learning has become a well-established method for analyzing power system data, thanks to its ability to include structural and homophily biases that cannot be modeled conventionally. Message-passing GNNs have been successfully applied to state estimation (Kundacina et al., 2022), and power flow estimations (Ringsquandl et al., 2021). The same tasks have also been addressed using *graph convolutional neural networks*

(GCN) (Fatah, Claessens, & Schoukens, 2021). In the broader power-grid environment, GNNs are also used for wind speed forecasting in renewable energy (Liao, Yang, Wang, & Ren, 2021). Additionally, GNN-based state forecasting was used in a range of downstream tasks in several previous research studies, including graph-based early fault detection for IIoT systems (Zhao & Fink, 2024), anomaly detection in the electrical grid (Li et al., 2022), fault diagnosis for three-phase flow facility (Chen, Liu, Hu, & Ding, 2021), predicting dynamical grid stability (Nauck et al., 2022), and physics-informed parameter and state estimations (Pagnier & Chertkov, 2021b, 2021a).

Other works have used spatial-temporal extensions of GNNs in the electrical domain for residential load-forecasting (Lin et al., 2021), fault diagnostics in power distribution systems (Nguyen, Vu, Nguyen, Panwar, & Hovsapian, 2022), and with complex-value extensions (T. Wu, Scaglione, & Arnold, 2022) for state forecasting. In another lie of research, (Wang et al., 2022) propose spatial-temporal graph learning for power flow analysis, where the graph is dynamically created from thresholded normalized mutual information. At the component feature level, attention-based graph learning (GAT) has been applied to power flow analysis (Jeddi & Shafieezadeh, 2021) and in a different work for probabilistic power flow to quantify uncertainties of distribution power systems (H. Wu, Wang, Xu, & Jia, 2022). To the best of our knowledge, the approach of employing a self-attention mechanism at the graph level to learn a graph structure that integrates electrical and hydraulic data using *graph neural networks* has not yet been addressed in previous research.

In the context of power systems, data fusion represents a crucial technique for enhancing the accuracy and reliability of forecasting algorithms by integrating diverse data sources. In the electrical domain, the fusion of diverse electrical system information was utilized to estimate the voltage in distribution networks (Y. Zhu, Gu, & Li, 2020), using cross-correlations between individual transformers. Another research study achieved state-of-the-art multi-site photovoltaic (PV) power forecasting (Simeunović, Schubnel, Alet, & Carrillo, 2022), fusing spatially distributed PV data by exploiting the intuition that PV systems provide a dense network of virtual weather stations. Integrating weather station data was also explored for anomaly detection for the industrial internet of things (Y. Wu, Dai, & Tang, 2022). Given its importance, modeling the interaction between the PSH subsystems has previously been explored using higher-order numerical simulation (SIMSEN) (Simond, Allenbach, Nicolet, & Avellan, 2006). However, operating numerical simulators in practice requires precise calibration and, consequently, extensive documentation of the components, which is typically not readily available. This calibration step is indispensable due to the components exhibiting highly non-linear characteristics (Nicolet et al., 2007). Given the significant varia-

tions in designs across different PSH assets, and the necessity for expert knowledge (which is often unavailable), applying this simulator-based methodology is often infeasible in real-world applications. For PSH, these limitations shift the focus to data-driven interaction modelling with GNNs, which is computationally affordable and does not necessitate expert knowledge, yet remains unexplored.

3. METHODOLOGY

This section introduces the Spectral-Temporal Graph Neural Network (STGNN) that we propose for the fusion of electric and hydraulic data in PSH state forecasting. In Section 3.1, we define the forecasting problem. From Section 3.3 onwards, we decompose the forecasting problem into learning the underlying latent graph structure from time series data (Sec. 3.4). Subsequently, on the learned graph, we introduce graph-spectral and time-spectral filtering (Sec. 3.6). An overview of the methodology is provided in Figure 1.

Notation: In this work, we use slicing notation denoted by the colon (:) symbol. Given a matrix $A \in \mathbb{R}^{m \times n}$, where m and n represent the number of rows and columns respectively, slicing is expressed as $A[i:j, k:l]$ or $A^{i:j,k:l}$. This notation represents the selection of rows i through $j - 1$ and columns k through $l - 1$ of matrix A. If i or k is omitted, it implies starting from the first row or column, respectively. Similarly, if j or l is omitted, it implies selection until the last row or column, respectively. We use $\otimes$ to denote element wise multiplication and $\oplus$ for concatenation. The Frobenius norm is denoted as $\| \bullet \|_F$.

3.1. Problem Formulation

The specific objective of this work is to provide accurate state forecasting for the electrical subsystem of the PSH. Our approach utilizes learnable graph-spectral and time-spectral filtering to compute state predictions (the forecast). Let $\bar{\mathbf{X}} \in \mathbb{R}^{T \times H + E}$ represent the smoothed time series data $\bar{\mathbf{X}} = S(\mathbf{X})$ computed from unsmoothed time series $\mathbf{X}$ using the smoothing function S for E sensors in the electric and H sensors in the hydraulic subsystem, respectively, over a time period T. We define our forecasting model as a function $M : \mathbb{R}^{w \times H + E} \rightarrow \mathbb{R}^{h \times E}$ for a specific point in time $w < t < T$, that operates on input windows of the data of length w sliced as $\bar{\mathbf{X}}[t - w : t]$. The goal is to forecast the subset of electrical sensors $\bar{\mathbf{X}}[t : t + h, : E]$ for a horizon of size h. Model M operates on a graph $\mathcal{G}$ that is either inferred from the input data by a parameterized function $\mathcal{G}_\phi(\bar{\mathbf{X}}[t - w : t])$, which is trained alongside M, or may be provided apriori. We introduce two sets of learnable parameters: θ for the filtering M_θ and ϕ for the graph learning $\mathcal{G}_\phi$. Thus, state forecasting and state reconstruction estimates denoted by $\widetilde{\bullet}$ at a selected timepoint t, are computed by the model M as follows:

$$\widetilde{\mathbf{X}}_{\text{elec}}^{t:t+h}, \widetilde{\mathbf{X}}_{\text{elec}}^{t-w:t} = M_\theta(\bar{\mathbf{X}}^{t-w:t} \mid \mathcal{G}_\phi) \tag{1}$$

Input Window $X^{t-w:t}$

Attention-based Graph Learning

Graph- and Time-spectral Filtering
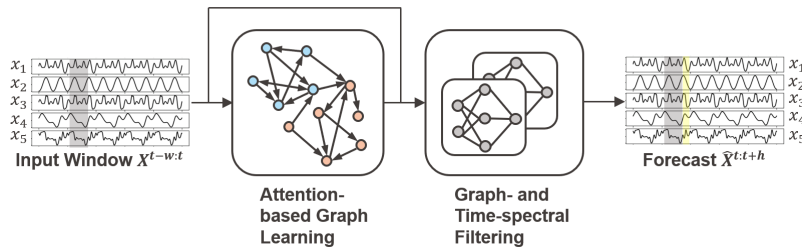
Forecast $\widetilde{X}^{t:t+h}$

Figure 1. An overview of the two processing steps of our *spectral-temporal graph neural network* to fuse data from the electrical and hydraulic subsystems of a PSH. Utilizing *attention-based graph learning*, our method dynamically constructs a graph based on an input window $\bar{\mathbf{X}}^{t-w:t}$. Subsequently, by operating on this graph, the *graph- and time-spectral filtering* module efficiently extracts information from the hydraulic and electrical sensor data to forecast the subset of electrical sensors $\widetilde{\mathbf{X}}_{\mathrm{elec}}^{t:t+h}$.

3.2. Training Objective

By employing a sliding window approach, we construct a training dataset $\mathcal{X}^{\mathrm{Train}}$ of length N_{train}, where N_{train} depends on the training data split. We also construct analogous validation and test datasets $\mathcal{X}^{\mathrm{Val}}$ and $\mathcal{X}^{\mathrm{Test}}$, respectively:

$$\mathcal{X}^{\mathrm{Train}} = \{\bar{\mathbf{X}}_{\mathrm{elec}}^{t:t+h}, \bar{\mathbf{X}}_{\mathrm{elec}}^{t-w:t}, \bar{\mathbf{X}}^{t-w:t}\}_{t=w}^{N_{\mathrm{train}}}$$

We optimize the parameters of the model by minimizing the forecasting error $\mathbf{E}_{\mathrm{f}}^t = \widetilde{\mathbf{X}}_{\mathrm{elec}}^{t:t+h} - \bar{\mathbf{X}}_{\mathrm{elec}}^{t:t+h}$. To learn meaningful and compact representations, we introduce an optional reconstruction error $\mathbf{E}_{\mathrm{r}}^t = \widetilde{\mathbf{X}}_{\mathrm{elec}}^{t-w:t} - \bar{\mathbf{X}}_{\mathrm{elec}}^{t-w:t}$ for regularization. The final training objective is expressed as:

$$\mathcal{L}(\widetilde{\mathbf{X}}, \bar{\mathbf{X}}) = \sum_{t=w}^{N_{\mathrm{train}}} \left(\lambda_{\mathrm{f}} \|\mathbf{E}_{\mathrm{f}}^t\|_F^2 + \lambda_{\mathrm{r}} \|\mathbf{E}_{\mathrm{r}}^t\|_F^2 \right)$$

During the model's training process, we identify the optimal parameters:

$$\theta^*, \phi^* = \arg\min_{\theta,\phi} \mathcal{L}(\widetilde{\mathbf{X}}, \bar{\mathbf{X}}; \phi, \theta) \qquad (2)$$

3.3. Node Features and Graph Representation

For the forecasting problem, we consider spatially distributed sensor sites modeled as nodes (vertices) $v \in V$ of a graph that spans the pumped storage hydropower plant. Due to differences in raw sensor sampling rates, we use resampled time series based on simple moving averages, taking into account the true sensor sampling rate of the j-th sensor S_j. This step smooths the time series:

$$\bar{\mathbf{X}}[i,j] = S(\mathbf{X}[i,j]) = \frac{1}{S_j} \sum_{\tau=1}^{S_j} s_\tau^j \qquad (3)$$

We model each measurement site, containing one or more sensors as an individual node. Each node is associated with a feature vector $\mathbf{x}_v^{t-w:t} \in \mathbb{R}^{w \times d}, \quad \forall v \in V$, containing a window w of the smoothed sensor data and additional $d-1$ covariate dimensions such as a temporal encoding. This strategy

is uniformly applied to both the electrical and hydraulic components within the pumped-storage power plant environment. Nodes are exclusively assigned to one of the sets: $\mathbb{1}_{\mathrm{el}}(v) = 1$ for electrical components or $\mathbb{1}_{\mathrm{hyd}}(v) = 1$ for hydraulic components, ensuring $\mathbb{1}_{\mathrm{hyd}}(v) + \mathbb{1}_{\mathrm{el}}(v) = 1$. As input for the subsequent model M, we consider the joint graph:

$$\mathcal{G}_\phi = \left(V_{\mathrm{el}} \cup V_{\mathrm{hyd}}, E_\phi(\bar{\mathbf{X}}[t-w:t]) \right)$$

where the edges may be learned by a parameterized function E_ϕ.

3.4. Attention-based Graph Learning

We define a trainable function that implements self-attention among the sensor nodes to infer the edges $E_\phi(\mathbf{X})$ of the graph $\mathcal{G}$. To compute the self-attention, we first map the time series to an embedding space $\mathbf{E} = \mathrm{GRU}(\bar{\mathbf{X}})$ using a gated recurrent unit (GRU). We then proceed by computing the self-attention of the embedded time series. For this purpose, we define a query sequence Q and a key sequence K to compute the attention scores W:

$$Q = EW^Q, K = EW^K, W = \mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right) \qquad (4)$$

by linear projection with the trainable matrices $\phi = (W^Q, W^K)$. Unlike in *graph attention networks* (GAT) (H. Wu et al., 2022), which define attention over features of a pre-existing graph, we directly compute the symmetrically normalized graph Laplacian L from the attention scores W, which we convert into a symmetrical adjacency matrix $A = \frac{1}{2}(W + W^T)$. We compute the Laplacian as $L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, where L is the Laplacian matrix, I is the identity matrix, D is the diagonal degree matrix of A, the adjacency matrix. L is then used for the graph spectral filtering in Section 3.6.

3.5. Spectral-temporal Graph Neural Network

To predict the sensor dynamics, the model processes the input data $\bar{\mathbf{X}}^{t-w:t}$ on the learned graph $\mathcal{G}$ (obtained as introduced in Section 3.4) by mapping the input data from the spatial-temporal vertex domain of the sensor signals to a spectral latent representation. This mapping is achieved through the sequential application of graph-spectral and time-spectral transformations, as introduced in Sec. 3.6. The corresponding inverse transformations are utilized to reconstruct the sensor signal in the spatial-temporal domain. To address the problem of vanishing gradients and performance degradation with increasing network depth, we introduce skip connections to compute the final forecast. We denote the output of the residual blocks as s_k. Thus, model M can be expressed with spectral filtering (F), and a bypass layer f_b as follows in the recursive equation:

$$(s_k^f, s_k^b) = \begin{cases} F\left(\bar{\mathbf{X}}\right) \mid \mathcal{G}_\phi(\bar{\mathbf{X}})\right), s_0^b = \bar{\mathbf{X}} & k = 0 \\ F\left(\sigma(s_{k-1}^b - f_s(s_{k-1}^b)) \mid \mathcal{G}_\phi(\bar{\mathbf{X}})\right), & k > 0 \end{cases}$$

(5)

where the final forecast is computed as $\widetilde{\mathbf{X}}_{\text{elec}}^{t:t+h} = \Omega_f\left(\sum_{i=1}^k s_i^f\right)$ with an application-specific feedforward head function (Ω), and the backcast as $\widetilde{\mathbf{X}}_{\text{elec}}^{t-w:t} = \sum_{i=1}^k s_i^b$.

3.6. Graph- and Time-spectral Filtering

For the spectral filtering module F, we use graph convolutional filtering and the trainable *Spe-Seq Cell* S_θ introduced in (Cao, Wang, et al., 2021). We denote the graph Fourier transform as $\mathcal{GF}$, and its inverse as $\mathcal{IGF}$. The j-th channel y_j in the graph-spectral domain is therefore computed as follows:

$$y_j = \mathcal{IGF}\left(\sum_i g_{\theta_{ij}}(\Lambda_i) S_\theta(\mathcal{GF}(\mathbf{X}_i))\right).$$

(6)

In the graph-spectral domain, we implement the parameterized filtering $g_\theta(\Lambda_i)$ on the eigenvalues Λ. Instead of computing y_j directly, we compute the Chebyshev polynomials of L to efficiently approximate the graph Fourier transform without performing the costly eigenvalue decomposition of the Laplacian matrix $L = U\Lambda U^T$, where U is the graph's eigenvector matrix. We obtain the i-th Chebyshev polynomial $T_i(\bullet)$ with the recurrence relation: $T_0(x) = 1, \quad T_1(x) = x, \quad T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$. Thus, we implement graph-spectral filtering with the graph spectral operator $g(L)$ as follows:

$$g(\widetilde{L})\mathbf{X}_j \approx \sum_{n=0}^{N} c_n S_\theta(T_n(\widetilde{L})\mathbf{X}_j)$$

(7)

where c_n are the learnable parameters and $\widetilde{L} = 2L/\lambda_{\max} - I_N$ is the normalized Laplacian matrix. The *Spe-Seq Cell* enhances the output of $\mathcal{GF}$, treating it as a multivariate time-series in the graph-spectral domain. It then elevates this output into the time-spectral domain to learn feature representations. To achieve this, the *Spe-Seq Cell* uses the Discrete Fourier Transform (DFT) and gated linear units (GLU) for element-wise modulation of the signal in time-spectral domain as follows:

$$\text{GLU}(\mathbf{x}) = \mathbf{x} \otimes \sigma(\mathbf{W}^g\mathbf{x} + \mathbf{b}^g)$$

(8)

This approach effectively implements convolution on the multivariate time-series in the graph-spectral domain.

4. CASE STUDY AND EXPERIMENTAL SETUP

The dataset in this case study was obtained in collaboration with the Swiss Federal Railways (SBB). SBB maintains a separate railway traction current network (RTN) that operates at a frequency of 16.66 Hz to power rolling stock across Switzerland. The power plant operators of SBB use *supervisory control and data acquisition (SCADA)* protocols to transmit sensor data to a centralized *energy management system* (EMS). This setup allows for real-time monitoring of assets, ensuring timeliness and synchronicity between sensor signals, making it a technically sound environment to evaluate the proposed methodology.

Objective: For this case study, our aim is to forecast the currents measured by the electrical sensor network of the PSH. From an operator's perspective, forecasting currents is particularly compelling when dealing with rolling stock, given their highly dynamic current profile that is vastly different to residential power grids. While the residential sub-grid of Zurich, the largest City of Switzerland, is subject to transient load changes within 15 minutes intervals of up to 35MW, the RTN of SBB experiences load changes up to 250 MW whithin the same time interval due to the orchestrated and periodic timetable of the Swiss railway network (Halder, 2018). The importance of accurate current forecasts is additionally heightened because electrical components in power systems, like transformers and conductors, have thermal limitations that depend on the amount of current flowing through them. Unlike voltage levels, current levels in the PSH dynamically react to transient loads changes. Anomalies such as sudden increases or decreases in power demand, or failures in equipment, are therefore more immediately reflected in current fluctuations. We therefore focus on phasor current forecasting in this study.

Dataset & Data Preparation: We collected data spanning four months from a PSH in Switzerland, consisting of readings from 58 sensors that monitor pressures, flow rates, and lake levels of the hydraulic subsystem, as well as electrical currents from seven generating units, including connected substations in the electrical subsystem. The time series are averaged to a 1-minute resolution and were collected from

January to March in 2021. We maintained the temporal ordering of training (70% of the data), validation (15%) and test (15%) datasets to ensure that the validation and test indices are sequentially higher than the training indices. We normalize the data using feature-wise min-max scaling. To provide a comprehensive understanding of the dataset, we show a detailed segment of the sensor data in Figure 2.
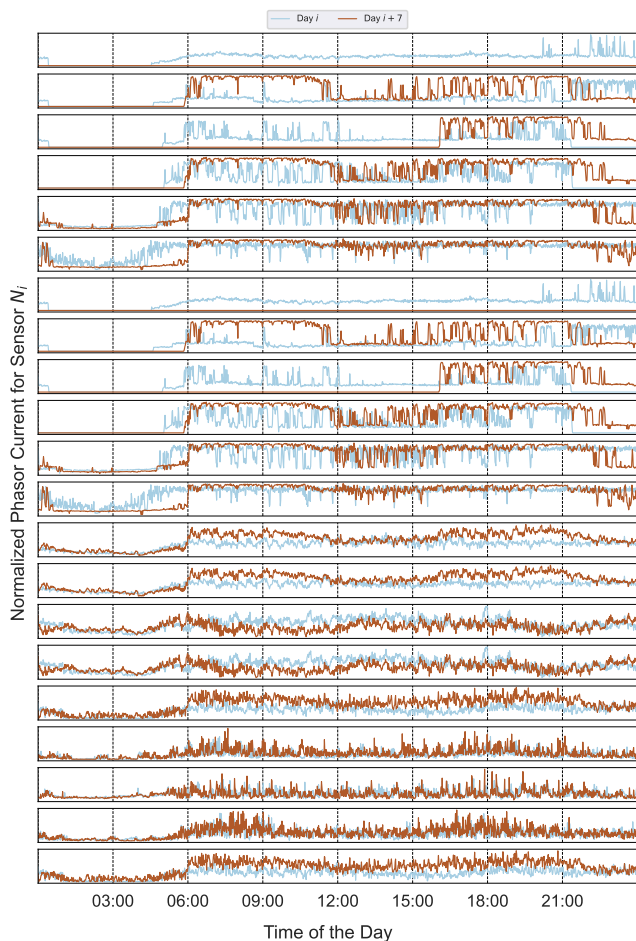

Figure 2. Segment of the dataset, displaying all 21 normalized phasor current sensors (the forecasting target of our case study), indicating the dynamic nature of the sensor measurements. We show the same day of the week (i and $i + 7$) for two consecutive weeks.

Model & Training: The experiments were conducted on an NVIDIA RTX3060 using PyTorch 2.0 and CUDA 11.8 for the development and training of the models. The proposed model utilizes a window size (w) of 24 and a horizon size of 1, meaning that it predicts the currents for the next minute. This configuration is tailored to the synchronized operation of the Swiss railway network, which organizes its periodic timetable in half-hourly intervals. Our selected model's input window takes this operational profile into account, thereby reducing the influence of the previous interval. During model

fine-tuning, we truncate the Chebychev polynomial expansion to $k = 4$ for both the graph- and time-spectral filtering. We set the number of residual blocks to two and configure the *Spe-Seq Cells* to five layers. We adjust the negative slope of *LeakyReLU* α to 0.2, and dropout to 0.5 for regularization. We use Adam for optimization.

Table 1. Overview of trainable parameters in the neural network models.

Model	Parameter
MLP (4-layer, el+hy)	2.8 M
STF (el+hy)	366 K
Ours (el)	481k
Ours (el+hy)	481k

4.1. Baseline Approaches for Comparison

We compare our approach with relevant baselines for time series forecasting in the power grid domain, such as a linear model with trend decomposition, a fully connected neural network (FNN) and a *recurrent neural network*, specifically the LSTM. Given the recent increase in attention towards transformer-based energy forecasting, we also consider the *time series transformer* model (*Spacetimeformer*, (Grigsby, Wang, & Qi, 2022), denoted as STF) Additionally, we include an Attention-based GNN (A3-GCN, (J. Zhu, Song, Zhao, & Li, 2020)) that has been demonstrated to outperform the classical GCN on similar forecasting tasks. All baseline models were trained and validated on the same dataset and with the same input window size. Each model was trained to reach convergence on the validation dataset.

5. RESULTS

In this section, we summarize the numerical results to evaluate the proposed method and compare it to the baselines. Additionally, we study the output from the graph learning and compare it to the connectivity of the PSH asset. Furthermore, we ablate the hydraulic information from the model input to assess its benefits. All results are based on the dataset introduced in Section 4.

In the initial step, we compare the performance of the adopted *spectral-temporal graph neural network* to the baseline models introduced in Section 4.1 based on normalized mean squared error (NMSE). We summarize model performances in Table 2. Our model surpasses all conventional baselines, including LSTM (by 28%) and the, in terms of parameters, much larger FNN (by 14%), as evaluated by the NMSE. A3-GCN is unable to integrate hydraulic information due to the lack of a graph. Spacetimeformer (STF) can learn from hydraulic signals but does not outperform our method.

For the evaluation of A3-GCN, we translate the PSH's electrical network diagram into a processable graph, as having apri-

Table 2. Average (normalized) model performance across nodes, comparing six different methods. We indicate whether the PSH network diagrams were translated into a processable graph for the computation (Network Diagram) and epmhasize if hydraulic (Hyd.) or electric (El.) information was used for training.

Method	El.	Hyd.	Network Diagram	Type	NMSE
Linear	✓	✓	✗	-	1.11e-1
A3-GCN	✓	✗	✓	GCN	8.74e-3
LSTM	✓	✓	✗	RNN	7.51e-3
MLP (3-layer)	✓	✓	✗	FNN	6.84e-3
MLP (4-layer)	✓	✓	✗	FNN	6.21e-3
STF	✓	✗	✗	Transformer	5.84e-3
STF	✓	✓	✗	Transformer	5.83e-3
Ours	✓	✗	✗	Att. GCN	5.71e-3
Ours	✓	✓	✗	Att. GCN	**5.34e-3**

ori graph is a computational requirement for the method. Surprisingly, we found that the A3-GCN is outperformed by the much simpler LSTM by 14% in terms of NMSE. This finding highlights that the intuitive approach of applying GNN directly to a graph derived from schematic diagrams, does not always yield acceptable results. In this context, since our proposed STGNN is also based on GCN, the 34.7% improvement in NMSE illustrates that, beyond choosing the right model, finding a suitable computational graph is crucial for processing PSH data. Our results provide further support for the observation in (Ringsquandl et al., 2021) that statistical properties of graphs derived from network diagrams of power grids may be unsuitable for direct graph processing. Graphs derived from such network diagrams significantly differ from those typically discussed in the graph-theoretical literature, with statistical properties like lower clustering-coefficients, lower node degrees, and higher graph diameters, which could explain the subpar performance of A3-GCN in the state forecasting task. From a message-passing perspective, the specific properties of these graphs hinder effective message propagation unless the GNN comprises many layers. Unfortunaltely, this model choice, in turn, significantly boosts oversmoothing, which is already a prevalent challenge in the power grid environment due to the high similarity of the electrical sensor data.

We assess the performance of time-series transformers (Spacetimeformer), which are structurally similar to our attention-based GNN approach, because they incorporate a self-attention layer across the one-dimensional timeseries. However, the experiment with Spacetimeformer displays a 9.2% reduction in performance in terms of NMSE compared to the STGNN. Additionally, we found it difficult to scale Spacetimeformer to the problem without overfitting. Altrough showcasting respectable performance, compared to the STGNN, STF did not benefit from the additional information from hydraulic systems (resulting in a 0.1% improvement).

An advantage of our STGNN, compared to conventional methodology suited for multivariate time series analysis such as LSTM, is that we have access to the learned graph topology. To derive insights from the inferred graph, we calculate the mean attention $a_{ij} = \frac{1}{N} \sum_{i=1}^{N} a_{ij}$ over the test data set. We expect our method to converge to the same graph for randomized training initialization when learning physical relationships between the sensors. To verify our expectation, we visually compare the average attention across random seeds in Figure 4. Additionally, our analysis reveals that the inferred attention graph's minimal parameterization yields temporally stable graphs, accurately reflecting the situation in the PSH, which usually has stable topology across time.

Interestingly, the attention graph recovers casual relations given by the functioning of the hydropower plant and therefore shows similarity to the underlying physical network of the hydropower plant. In Figure 6, we depict the relationships between the water inflow (flow-rates, pressures), the generator units (groups and transformers, indicated by TRF), and the PSH outlets (substations, denoted by SP), overall making the model's predictions more interpretable. Leveraging this interpretability, we compare Figures 5 and 6. Surprisingly, our model focuses on the PSH outlets to predict the power plant input's phasor currents in the absence of hydraulic information from the forecast, which could explain the more severe outliers in the forecast (Figure 3). When adding hydraulic information to the forecast, we observe that the model makes additional use of penstock flow-rate and pressure sensor data, thereby improving the prediction quality.

5.1. Ablation Study

In an ablation study, we exclude the hydraulic sensor data from the forecast to validate the effectiveness of fusing electric and hydraulic sensor data for improving the electrical state forecasting. We find that incorporating the hydraulic subsystem leads to an 6.5% reduction in NMSE. The NMSE absolute forecasting performance from the ablation experiment is included in Table 2. Figure 7 and 8 evaluate the relative improvement on a per-node basis, demonstrating that the benefits of our methodology are distributed across most sen-
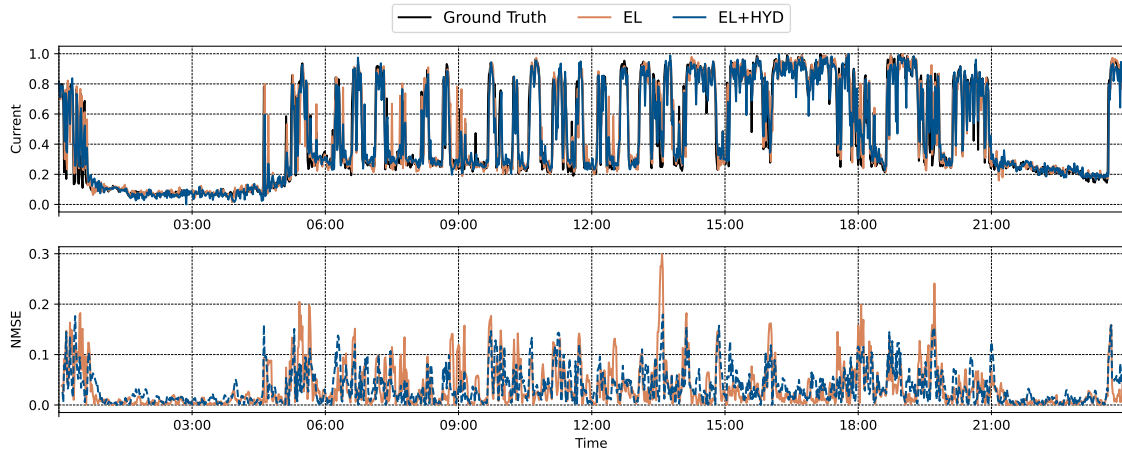
Figure 3. Comparison of normalized phasor current forecasts with (EL+HYD) and without the hydraulic information (EL) for our proposed STGNN model. We show the forecast for single node N_i $(i = 1)$ across a randomly selected day including ground truth. In the upper Figure, we display the dynamic range of the forecast. In the lower Figure, we display the normalized MSE of both approaches with respect to the ground truth. Removing hydraulic information results in heightened discrepancies and more pronounced outliers in the predictions. First, we select the data based on the above criteria. Then, we normalize the selected data using min-max scaling.
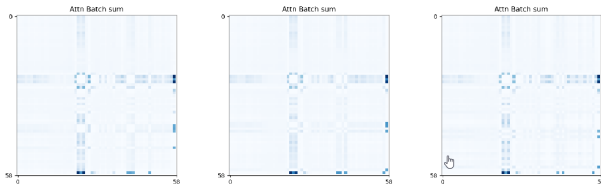
Figure 4. The averaged learned attention across the test set of the *attention-based graph learning* module over all 58 signals from the electrical and hydraulic subsystems. We show three random seeds. The learned attention is stable for different random initializations.
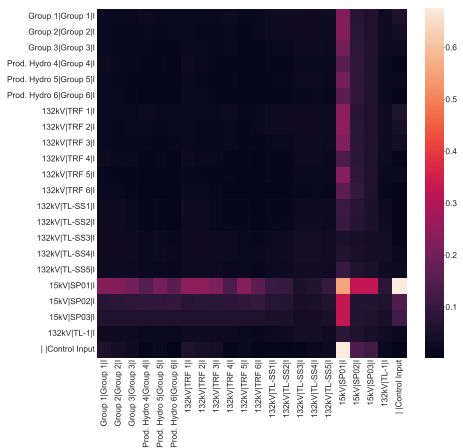

Figure 5. The heatmap represents the averaged learned attention by the *attention-based graph learning* module across the test set as for the model processing only electrial information (EL). Counterintuitively, the model focuses on the PSH output (SP) when forecasting the phasor currents of the electromagnetic generators, which represent the PSH input.

sor forecasts (nodes), thereby ensuring that no sensor forecast experiences a major decline in performance from including hydraulic sensor information.

6. CONCLUSIONS

In this paper, we demonstrate that integrating information across the electrical and hydraulic subsystems is beneficial for state forecasting in pumped-storage hydropower plants (PSH). Our proposed *spectral-temporal graph neural network* is the first approach to integrate information across the PSH's subsystems by applying *attention-based graph learning*, which effectively represents PSH states for short-term phasor current forecasting. Compared to numerical simulation, our method requires neither knowledge of the underlying network and sensor connectivity graph nor a tedious calibration step. Through a real world case study, we demonstrate that relying exclusively on graphs derived from network diagrams for state forecasting does not always yield the best performance. We highlight the advantages of learning a PSH-wide graph, complementing the critical perspective on network-diagram-derived graphs introduced in (Ringsquandl et al., 2021). Moreover, we show that our method remains interpretable, unlike other deep-learning methods that process electrical and hydraulic data simultaneously. Future work looks to reintegrate the underlying network diagram while maintaining the flexibility of attention-based graph learning, thereby harnessing the strengths of both approaches. This could also allow for the incorporation of physics-informed losses, such as electromagnetic generator efficiency or power flow, which may reduce the volume of training data required.
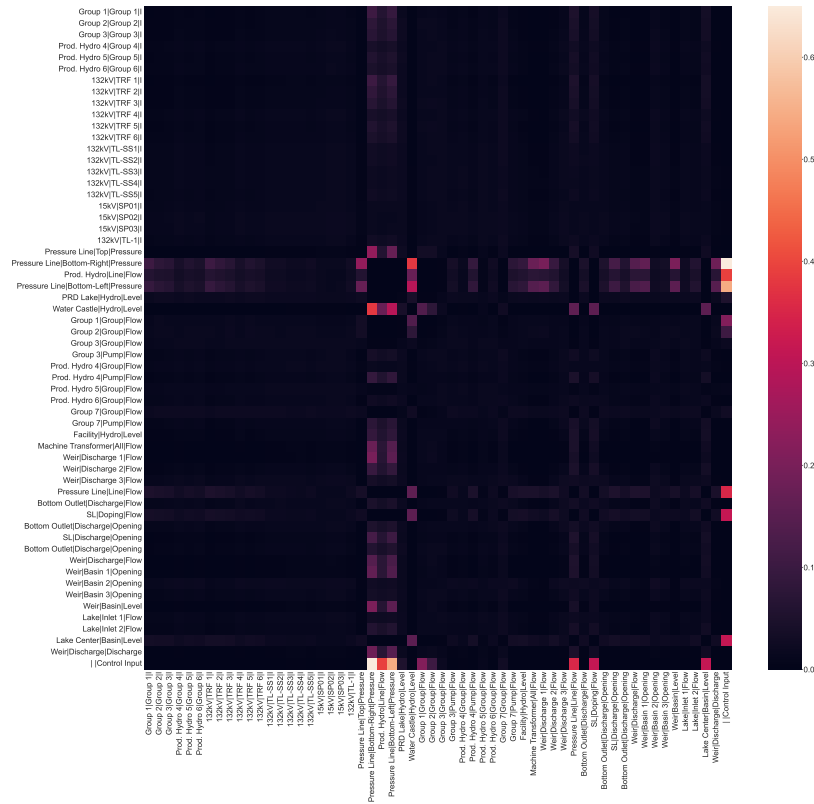
Figure 6. The heatmap visualizes the averaged learned attention of the *attention-based graph learning* module across the test set visualized for the model that processes both electrial and hydraulic information (EL+HYD). Notably, the model focuses on the hydraulic subsystem (the PSH input) when forecasting the phasor currents of the electromagnetic generators.
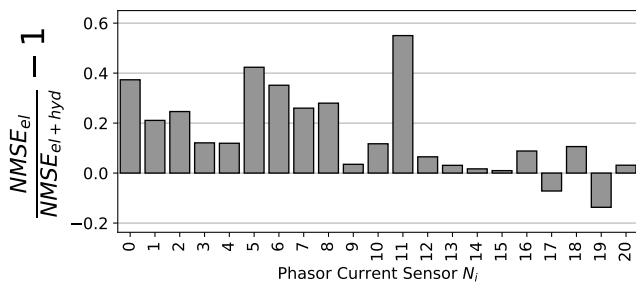

Figure 7. Relative improvements across the test set (normalized MSE, averaged) for our STGNN models with (EL+HYD) and without hydraulic information (EL). for each of the 21 phasor currents sensors of the electric subsystem. 19 out of 21 phasor current forecasts are improved by the additional hydraulic information.
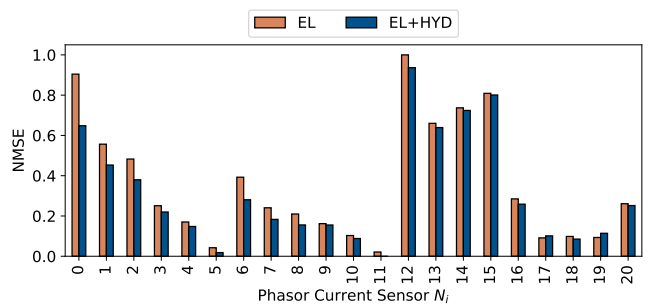

Figure 8. The normalized MSE averaged across the test set is displayed, comparing our STGNN models with (EL+HYD) and without hydraulic information (EL). The performance across the 21 phasor currents sensors of the electric subsystem is shown. The additional hydraulic information improves the phasor current forecasts in 19 out of the 21 cases.

REFERENCES

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., & Vandergheynst, P. (2017, July). Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, *34*(4), 18–42. doi: 10.1109/MSP.2017.2693418

Cao, D., Li, J., Ma, H., & Tomizuka, M. (2021, May). Spectral Temporal Graph Neural Network for Trajectory Prediction. In *2021 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1839–1845). doi: 10.1109/ICRA48506.2021.9561461

Cao, D., Wang, Y., Duan, J., Zhang, C., Zhu, X., Huang, C., ... Zhang, Q. (2021, March). *Spectral Temporal Graph Neural Network for Multivariate Time-*

series Forecasting (No. arXiv:2103.07719). arXiv. doi: 10.48550/arXiv.2103.07719

Chen, D., Liu, R., Hu, Q., & Ding, S. X. (2021). Interaction-Aware Graph Neural Networks for Fault Diagnosis of Complex Industrial Processes. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14. doi: 10.1109/TNNLS.2021.3132376

Fatah, M. G. A., Claessens, B. J., & Schoukens, M. (2021, August). Integrating Power Grid Topology in Graph Neural Networks for Power Flow. *Eindhoven University of Technology*, 11.

Grigsby, J., Wang, Z., & Qi, Y. (2022, May). *Long-Range Transformers for Dynamic Spatiotemporal Forecasting* (No. arXiv:2109.12218). arXiv.

Guo, W., Che, L., Shahidehpour, M., & Wan, X. (2021, January). Machine-Learning based methods in short-term load forecasting. *The Electricity Journal*, *34*(1), 106884. doi: 10.1016/j.tej.2020.106884

Halder, M. (2018, September). *Power Demand Management – Smart Grid @ SBB*. Innotrans.

Jeddi, A. B., & Shafieezadeh, A. (2021, December). A Physics-Informed Graph Attention-based Approach for Power Flow Analysis. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1634–1640). doi: 10.1109/ICMLA52953.2021.00261

Kim, H., Jeong, J., & Kim, C. (2022, November). Daily Peak-Electricity-Demand Forecasting Based on Residual Long Short-Term Network. *Mathematics*, *10*(23), 4486. doi: 10.3390/math10234486

Kundacina, O., Cosovic, M., & Vukobratovic, D. (2022, April). *State Estimation in Electric Power Systems Leveraging Graph Neural Networks* (No. arXiv:2201.04056). arXiv.

Li, S., Pandey, A., Hooi, B., Faloutsos, C., & Pileggi, L. (2022, September). Dynamic Graph-Based Anomaly Detection in the Electrical Grid. *IEEE Transactions on Power Systems*, *37*(5), 3408–3422. doi: 10.1109/TPWRS.2021.3132852

Liao, W., Bak-Jensen, B., Radhakrishna Pillai, J., Wang, Y., & Wang, Y. (2022). A Review of Graph Neural Networks and Their Applications in Power Systems. *Journal of Modern Power Systems and Clean Energy*, *10*(2), 345–360. doi: 10.35833/MPCE.2021.000058

Liao, W., Yang, D., Wang, Y., & Ren, X. (2021, March). Fault diagnosis of power transformers using graph convolutional network. *CSEE Journal of Power and Energy Systems*, *7*(2), 241–249. doi: 10.17775/CSEE-JPES.2020.04120

Lin, W., Wu, D., & Boulet, B. (2021, November). Spatial-Temporal Residential Short-Term Load Forecasting via Graph Neural Networks. *IEEE Transactions on Smart Grid*, *12*(6), 5373–5384. doi: 10.1109/TSG.2021.3093515

Ma, Y., Liu, X., Shah, N., & Tang, J. (2023, July). *Is Homophily a Necessity for Graph Neural Networks?* (No. arXiv:2106.06134). arXiv.

Nauck, C., Lindner, M., Schürholt, K., Zhang, H., Schultz, P., Kurths, J., ... Hellmann, F. (2022, April). Predicting Basin Stability of Power Grids using Graph Neural Networks. *New Journal of Physics*, *24*(4), 043041. doi: 10.1088/1367-2630/ac54c9

Nguyen, B., Vu, T., Nguyen, T.-T., Panwar, M., & Hovsapian, R. (2022, October). *Spatial-Temporal Recurrent Graph Neural Networks for Fault Diagnostics in Power Distribution Systems* (No. arXiv:2210.15177). arXiv. doi: 10.48550/arXiv.2210.15177

Nicolet, C., Greiveldinger, B., Herou, J. J., Kawkabani, B., Allenbach, P., Simond, J.-J., & Avellan, F. (2007, November). High-Order Modeling of Hydraulic Power Plant in Islanded Power Network. *IEEE Transactions on Power Systems*, *22*(4), 1870–1880. doi: 10.1109/TPWRS.2007.907348

Pagnier, L., & Chertkov, M. (2021a, March). *Embedding Power Flow into Machine Learning for Parameter and State Estimation* (No. arXiv:2103.14251). arXiv. doi: 10.48550/arXiv.2103.14251

Pagnier, L., & Chertkov, M. (2021b, February). *Physics-Informed Graphical Neural Network for Parameter & State Estimations in Power Systems* (No. arXiv:2102.06349). arXiv. doi: 10.48550/arXiv.2102.06349

Ringsquandl, M., Sellami, H., Hildebrandt, M., Beyer, D., Henselmeyer, S., Weber, S., & Joblin, M. (2021, October). Power to the Relational Inductive Bias: Graph Neural Networks in Electrical Power Grids. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 1538–1547). doi: 10.1145/3459637.3482464

Simeunović, J., Schubnel, B., Alet, P.-J., & Carrillo, R. E. (2022, April). Spatio-Temporal Graph Neural Networks for Multi-Site PV Power Forecasting. *IEEE Transactions on Sustainable Energy*, *13*(2), 1210–1220. doi: 10.1109/TSTE.2021.3125200

Simond, J.-J., Allenbach, P., Nicolet, C., & Avellan, F. (Eds.). (2006). Simulation tool linking hydroelectric production sites and electrical networks. *Proceedings of 27th Int. Conf. on Electrical Machines, ICEM*.

Wang, F., Chen, P., Zhen, Z., Yin, R., Cao, C., Zhang, Y., & Duić, N. (2022, October). Dynamic spatio-temporal correlation and hierarchical directed graph structure based ultra-short-term wind farm cluster power forecasting method. *Applied Energy*, *323*, 119579. doi: 10.1016/j.apenergy.2022.119579

Wu, H., Wang, M., Xu, Z., & Jia, Y. (2022, November). Graph Attention Enabled Convolutional Network for Distribution System Probabilistic Power Flow. *IEEE Transactions on Industry Applications*, *58*(6), 7068–

7078. doi: 10.1109/TIA.2022.3202159

Wu, T., Scaglione, A., & Arnold, D. (2022, September). *Complex-Value Spatio-temporal Graph Convolutional Neural Networks and its Applications to Electric Power Systems AI* (No. arXiv:2208.08485). arXiv. doi: 10.48550/arXiv.2208.08485

Wu, Y., Dai, H.-N., & Tang, H. (2022, June). Graph Neural Networks for Anomaly Detection in Industrial Internet of Things. *IEEE Internet of Things Journal*, *9*(12), 9214–9231. doi: 10.1109/JIOT.2021.3094295

Zhao, M., & Fink, O. (2024, January). *DyEdgeGAT: Dynamic Edge via Graph Attention for Early Fault Detection in IIoT Systems* (No. arXiv:2307.03761). arXiv. doi: 10.48550/arXiv.2307.03761

Zheng, J., Xu, C., Zhang, Z., & Li, X. (2017, March). Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network. In *2017 51st Annual Conference on Information Sciences and Systems (CISS)* (pp. 1–6). doi: 10.1109/CISS.2017.7926112

Zhu, J., Song, Y., Zhao, L., & Li, H. (2020, June). *A3T-GCN: Attention Temporal Graph Convolutional Network for Traffic Forecasting* (No. arXiv:2006.11583). arXiv. doi: 10.48550/arXiv.2006.11583

Zhu, Y., Gu, C., & Li, F. (2020, July). Cross-Domain Data Fusion On Distribution Network Voltage Estimation with D-S Evidence Theory. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–6). doi: 10.1109/IJCNN48605.2020.9207414