Received 26 April 2016; | Revised 6 June 2016; | Accepted 6 June 2016

DOI: xxx/xxxx

ARTICLE TYPE

Power Grid Frequency Prediction Using Spatiotemporal Modeling

Amanda Lenzi*[,1] | Julie Bessac[1] | Mihai Anitescu[1]

[1]Mathematics and Computer Science Division, Argonne National Laboratory, Illinois, USA

Correspondence
*lenzi.amanda88@gmail.com

Present Address
9700 S. Cass Avenue, Lemont, IL 60439

Summary

Understanding power system dynamics is essential for interarea oscillation analysis and the detection of grid instabilities. The FNET/GridEye is a GPS-synchronized wide-area frequency measurement network that provides an accurate picture of the overall real-time operational condition of the power system dynamics, giving rise to new and intricate spatiotemporal patterns of power loads. We propose to model FNET/GridEye grid frequency data from the U.S. Eastern Interconnection with a spatiotemporal statistical model. We predict the frequency data at unobserved locations without observations, a critical need during disruption events where measurement data are inaccessible. Spatial information is accounted for either as neighboring measurements in the form of covariates or with a spatiotemporal correlation model captured by a latent Gaussian field. The proposed method is useful in estimating power system dynamic response from limited phasor measurements and holds promise for predicting instability that may lead to undesirable effects such as cascading outages.

KEYWORDS:
Hierarchical models, phasor measurement unit, power systems, spatiotemporal correlations

1 | INTRODUCTION

Ensuring the United States (U.S.) electrical system's reliability is a significant concern for the U.S. Department Of Energy because it impacts the distribution of electricity to homes, businesses, and industries. Power grids in North America use alternating current, with the flow of electricity switching direction with a frequency of 60 cycles per second (Hertz, abbreviated Hz). If the frequency drops beyond a certain point, protection systems are designed to automatically disconnect components of the grid to prevent their damage, at the cost of sometimes disconnecting load. Because of the highly interconnected nature of the grid, a shutdown in one sector has a high probability of rapidly propagating to other sectors. Therefore, the ability to anticipate these frequency drops and provide early warnings of the grid instability to system operators is important for the reliability of the electrical system, from preventing a transmission line failure to triggering a systemwide power outage.

Power system frequency data is essential for understanding the power system dynamics and improving its operation and control. Recently, phasor measurement units (PMUs) and frequency disturbance recorders (FDRs) were introduced and installed at critical locations in the grid transmission system, providing a massive amount of high-accuracy data. With the rapid increase in the number of installed PMUs, extensive research using such data has been carried out. For instance, frequency prediction for early situation awareness was performed using multivariate time series models in Li et al. [21] and a space-state approach in Dong et al. [10]. Studies of the dynamics and stability analysis of the power system include Yamashita and Kameda [30], Diao et al. [8], and Kamwa et al. [17], using autoregressive models, decision trees, and artificial intelligence methods, respectively. Evidence showed that system instability results in a substantial increase in the autocorrelation, and variability of frequency response was found in Cotilla-Sanchez et al. [6].

A limitation of the methods described above is that each installed PMU is treated independently, making it hard to generalize the conclusions to unobserved locations without any measurements. Having a technique that provides an overview of the process throughout the entire grid is vital to enhancing system stability and reliability. In practice, however, only a small subsample of the sensors (also denoted as buses) in the power grid network is monitored in real time; and even for the locations where the sensors are installed, measurements may be lost because of communication failure between the devices and the central operator or because of cyberattacks. Furthermore, spatial predictions are critical during disruption events where measurement data are inaccessible, compromising the entire power system dynamic stability.

Cyberattack effects on smart grids are usually mitigated by using techniques for reconstructing missing data [13, 20, 7]. Since the spatial aspect of the problem is disregarded, however, such techniques can restore the data only at monitored locations. To extend the findings to new locations, Aminifar et al. [1] proposed the use of a power system dynamic state estimation, where the measurements at locations without PMUs were estimated via voltage and current measurements of adjacent buses. Instead of relying on the uncertainties of the circuit model, a different approach to the same problem is to use only a measurement-based model. Using FDR data for dynamic frequency, angle estimation, and voltage estimation, Bai et al. [2] proposed a purely data-driven approach that first describes the correlation between all the different locations using past data during a generation trip (also called an event) and then fits an autoregressive model with exogenous covariates (ARX) to predict the response at unobserved locations during a second event. This approach has the shortcoming that it relies on data from a similar event that has happened in the past, in order to calculate the correlations, and these events are not always available for every location.

Motivated by a data set with ten observations per second from 70 PMUs in the U.S. Eastern inter-connection (EI), we propose and compare two statistical models for simultaneous probabilistic prediction of the frequency response at new buses during an event. The first is an ARX model where the exogenous predictors are chosen according to the geographical distance from the target location. This model is similar to the ARX proposed in Bai et al. [2] to estimate the frequency and angle in a missing power system dynamics. Instead of selecting the exogenous variables based on the correlations, we choose the buses with the smallest

distances to the target. As a result, we can predict even at locations that have never had PMUs installed. Because of its compute-intense nature, however, the ARX model can afford information from only a small number of predictors, thus limiting the spatial information to short ranges. Our second model overcomes this restriction by incorporating neighboring spatial information through a Gaussian random field with a Matérn covariance function, enabling longer spatial ranges dictated by the data features. Both of our models produce predictions that are based only on a subset of the spatial locations and can be calculated in real-time at several places simultaneously, once the model fitting and parameter estimation is performed offline. Another application of our techniques is related to reconstructing post-event analysis. For instance, spatial predictions are essential for forensic studies, which requires a precise comprehension of the power grid at any given location and time. We take advantage of the Integrated nested Laplace approximation (INLA) [24] to perform fully Bayesian inference and probabilistic prediction of the frequency response.

While probabilistic predictions have proven to be a prerequisite for optimal decision-making under uncertainty, classical methods for analyzing PMU data provide only point predictions, that is, a single prediction value of some future observation. Point predictions have the advantage of the intuitive interpretation that a nonexpert can communicate with relative ease. However, because they provide no information as to how confident one can be, they are not informative enough for several applications that are subject to the inherent uncertainties in the data. The Bayesian approach taken in our analysis provides a predictive probability distribution instead of only a point prediction. With the full distribution, we can quantify the uncertainty in the predictions and give a summary of the model reliability in terms of quantiles, which is essential to secure the system stability and guide operational decisions.

A significant challenge that arises from modeling such a complex network as the U.S. EI is the high sampling resolution in time. For instance, a large spurious relationship between time series may occur when both are internally highly correlated. In other words, the significant correlations between two variables are often due to the history of the individual variables. Therefore, when the behavior across the PMUs is similar and the variability in the internal frequency response is minimal, strong correlations are observed even at long distances. To deal with this issue, our models contain a linear time trend common to all sites. As a result, the residuals from fitting the linear trend to the data encode mainly the spatial correlation that is left between the PMUs. Information about the power grid network and the generators' location can be used to understand and quantify the remaining spatial correlations between the different buses. For security and privacy reasons, however, this information is rarely available; and models that rely only on the available data to define the correlations have to be used in practice.

The remainder of this paper is organized as follows. In Section 2 we first present the FNET data and then describe the two hierarchical models used for fitting and predicting these data. The prediction performances of the two models are compared in Section 3, followed by an analysis of the spatial correlations in the FNET data. Conclusions and discussion about further improvements in our procedures are given in Section 4.

2 | METHODS

We start by describing the main features of the FNET data and by providing exploratory graphical support for our modeling choices, which are detailed in Section 2.1. In Section 2.2, we describe two different hierarchical models for predicting the frequency response from the FNET data. We begin with a simpler ARX model, where neighboring information is introduced through exogenous variables chosen according to geographical proximity. This model intends to provide a benchmark for the subsequent spatiotemporal model, in which a latent Gaussian field describes the spatial correlation with a Matérn covariance function and the temporal correlation with an autoregressive structure in time. The Bayesian estimation approach used to obtain probabilistic predictions of the frequency response is summarized in Section 2.3.

2.1 | The FNET data

The power system Frequency Monitoring Network (FNET) is a wide-area and high-precision phasor measurement system operated by the Power Information Technology Laboratory at the University of Tennessee. The FNET data uses high dynamic accuracy Frequency Disturbance Recorders (FDRs) to produce synchronized measurements of frequency, phase angle, and voltage of the power system at ordinary 120 V outlets. The synchronization is performed by the Global Positioning System (GPS), and these measurements are transmitted through an Ethernet connection, giving accurate real-time monitoring of the smart grid. The data is essential for power system researchers and operators to understand the power grid dynamics, such as event detection, grid control input signals, and early warning of grid instability, which is critical for reducing blackout. Compared with model-based simulation, data from FDRs tend to be more reliable because they directly measure voltage and current waveforms instead of relying on model accuracy.

We consider measurements of the frequency response from the FNET data with a sampling rate of 10 samples per second at 70 locations in the EI in the United States. The period covered ranges from 2017-08-01 01:00:53 to 2017-08-02 01:00:52 Coordinated Universal Time (UTC), resulting in a total of 663,598 measurements for each bus. We focus on 200 measures (i.e., 20 seconds) from 2017-08-01 21:47:71 to 2017-08-01 22:07:71 UTC, which is the most significant frequency drop observed in the available data and when the power consumption peaks along the east coast of North America. The data are illustrated in Figure 1 (left), where each curve corresponds to a location in the maps in Figure 2 . A sudden drop in frequency is usually a result of a local disturbance that then propagates to other locations, potentially leading to a blackout. The measurements at the different buses are clearly related. We see that the changes in frequency are similar everywhere after the event has started, with a significant drop at every location during the 20 seconds.

Figure 1 (middle) shows the empirical spatial correlations (time lag equal to zero) of the frequency response, which is calculated by using the 20-seconds of data shown in Figure 1 (left) and at all pairs of locations (see map in Figure 2 (left)).

While the spatial correlations at time lag zero between any two places are more than 0.92 even at a distance of 3000 km apart, a simple linear time trend removal (common to all locations) captures most of these high correlations, as indicated by Figure 1 (right). More important, even after the temporal trend removal, the residuals in Figure 1 (right) show a clear spatial pattern, with the strongest correlations still coming from smaller distances. The analysis of the correlations shown in this figure will motivate the choices of our models in Section 2.2.

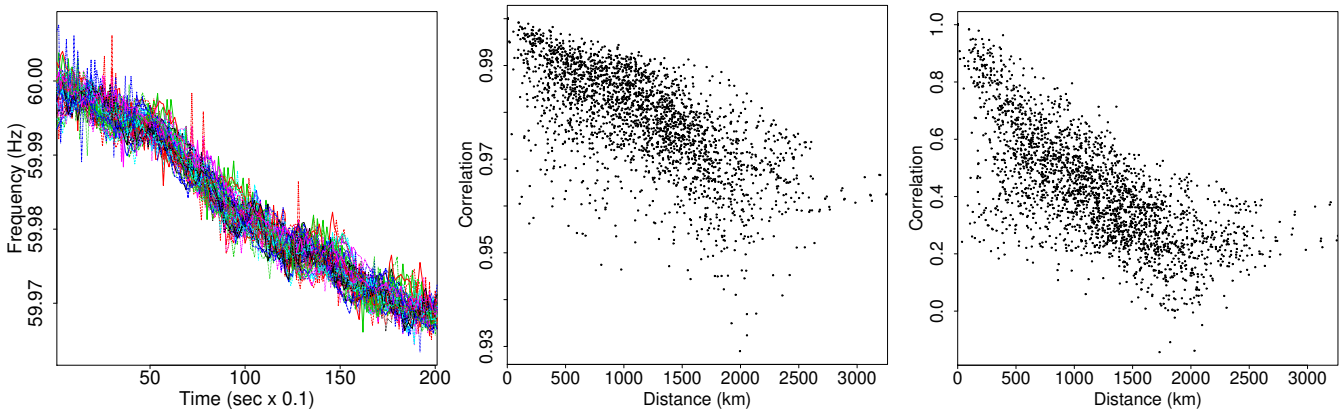

FIGURE 1 Left: 20 seconds of the frequency response at the 70 locations in Figure 2 . Middle: Spatial correlations of the frequency response between every pair of locations at temporal lag equal to zero. Right: Spatial correlations after removing a common temporal linear trend to the 70 locations.

Next, we investigate the overall relationship between the frequency data at different locations and identify possible clusters of regions with similar patterns. Figure 2 shows the spatial distribution of the mean (left) and standard deviation (right) of the frequency data from each time series displayed on the left side of Figure 1 . Examples of clusters with high frequency on average are located near Oklahoma, Ohio, Wisconsin, and Florida. Locations in the Northeast are the ones with the lowest mean values. We notice a heterogeneous cluster in the south of Florida, with somewhat different mean and standard deviation values even at short distances.

2.2 | Hierarchical models

Hierarchical modeling provides a practical and straightforward approach for constructing complex spatiotemporal models by defining dependence through marginalization and conditioning, therefore allowing for fast computations. In this section, we formulate the hierarchical statistical models for the FNET data described in Section 2.1. We denote by $Y(\mathbf{s}, t)$ the frequency at location $\mathbf{s} \in \mathcal{D}_s \subset \mathbb{R}^2$ and time $t \in \mathcal{D}_T \subset \{1, \ldots, T\}$, where $| \mathcal{D}_s | = 70$ and $T = 200$ with a temporal resolution of 0.1 seconds.

For a spatiotemporal process, where a series of measurements are made at multiple locations, dependencies between series may exist on a range of scales. We propose using a stationary Gaussian dependence model for describing the temporal and

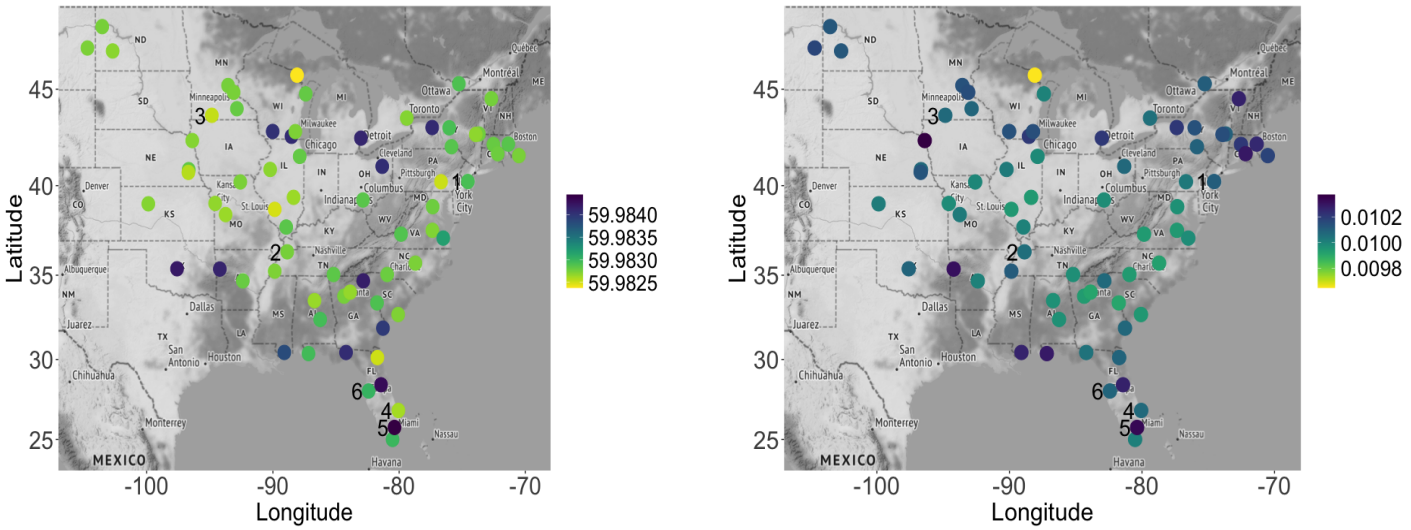

FIGURE 2 Maps of the mean (left) and standard deviation (right) calculated using the 20 seconds from the time series of the frequency response in the left plot of Figure 1 .

spatial dynamics of the stochastic process with two different latent structures. We start with an ARX, which is a purely temporal model, and spatial information is added from neighbors in the form of covariates. Choosing the location and the number of neighbors in an ARX model is not trivial. Our second approach captures these dependencies through a spatiotemporal dynamic latent process instead of independent temporal models. The resulting separable space-time covariance enables longer spatial ranges and requires less training data than the ARX model does.

The first layer of the hierarchical structure is the data model, in which $Y(\mathbf{s}, t)$ is assumed to be Gaussian and independent *conditional* on the latent process $\eta(\mathbf{s}, t)$ and the precision parameter τ_y. The second layer specifies the model for the process $\eta(\mathbf{s}, t)$ given the parameters $\boldsymbol{\theta}_\eta$. The final layer defines a model for the parameters. The different parts of the hierarchical formulation are

$$\begin{aligned} \mathbf{Y}|\boldsymbol{\eta}, \boldsymbol{\theta}_y &\sim \prod_{\mathbf{s}\in\mathcal{D}_s, t\in\mathcal{D}_T} \mathcal{N}\{Y(\mathbf{s}, t)|\eta(\mathbf{s}, t), \tau_y^{-1}\}, \\ \boldsymbol{\eta} \mid \boldsymbol{\theta}_\eta &\sim \mathcal{N}(\boldsymbol{\mu}_\eta, \mathbf{Q}_\eta^{-1}), \\ (\tau_y, \boldsymbol{\theta}_\eta) &\sim \pi(\tau_y, \boldsymbol{\theta}_\eta), \end{aligned} \tag{1}$$

with $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_T)^\top$, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_T)^\top$, where $\mathbf{Y}_t = \{Y(\mathbf{s}_1, t), \ldots, Y(\mathbf{s}_n, t)\}^\top$, and $\boldsymbol{\eta}_t = \{\eta(\mathbf{s}_1, t), \ldots, \boldsymbol{\eta}(\mathbf{s}_n, t)\}^\top$, for $t = 1, \ldots, T$, are the n-dimensional data and latent process vectors, respectively. The distribution of the underlying process $\boldsymbol{\eta}$ is assumed to be Gaussian, with mean vector $\boldsymbol{\mu}_\eta$ and precision matrix $\mathbf{Q}_\eta$. The prior distribution for the parameters τ_y and $\boldsymbol{\theta}_\eta$ is denoted by π. Hierarchical representations based on latent Gaussian models, such as in (1), are widely used to model spatiotemporal data and encompass a broad class of models, including non-Gaussian and nonstationary Rue et al. [25]. In what follows, we detail the parts of this hierarchical representation for our problem.

For location $\mathbf{s}$ at time t, the observations are connected to the linear predictor $\boldsymbol{\eta}$ as

$$Y(\mathbf{s},t) = \eta(\mathbf{s},t) + \epsilon(\mathbf{s},t), \tag{2}$$

where $\epsilon(\mathbf{s},t) \sim \mathcal{N}(0,\tau_y^{-1})$ accounts for measurement error. The term $\boldsymbol{\eta}$ is where most of the modeling effort is placed in the hierarchical formulation, since it characterizes the spatial and temporal dynamics of the process. For each $\mathbf{s}$ and t, we define

$$\eta(\mathbf{s},t) = \beta_0 + \beta_1 t + W(\mathbf{s},t), \tag{3}$$

where the intercept β_0 is the same for all locations and time steps and β_1 is the coefficient of a linear time trend, which is common to all sites. The inclusion of the linear temporal trend is motivated by the descriptive analysis in Figure 1 , which shows that most of the correlation between the frequency data at different locations is due to a common temporal dynamic. The residuals from removing this trend from the data (see the right plot in this figure) still have some space-time interaction that is captured by $W(\mathbf{s},t)$, as it will be described in the following.

Autoregressive with exogenous input latent process (ARX)

This model is similar to the approach used in Bai et al. [2] to estimate the frequency observations at missing locations from disturbances in the FNET data described in Section 2.1. The exogenous variables in the ARX model were chosen according to the empirical Pearson correlation between the target location and the remaining ones. Bai et al. [2] showed results for the two and the four most correlated neighbors. Although that approach performed well in the scenarios considered, it relies on having a similar event in the past with data collected at the same locations where prediction is desired in order to derive the correlations. We propose an ARX model where neighboring information is selected based on geographical proximity to the target. Thus, in contrast to Bai et al. [2], it is applicable to new locations where data from a previous event have never been collected.

We assume here that each location receives a common signal, which is denoted by W. Therefore, we use the notation $W(t)$ instead of $W(\mathbf{s},t)$ to define the interaction in (3). We focus on the first-order Markov in time, also called an autoregressive process. This representation is widely used in time-series approaches and parametrized by a single parameter denoted by ρ, which represents the correlation of the process with its previous value. A higher-order Markov model can be used, with the price of increasing the dimensionality of the problem. Temporal information from the closest neighbors is included in the form of covariates. We write

$$W(t) = \sum_{i=1}^{|N_{\mathbf{s}}|} \alpha_i Y(\mathbf{s}_i,t) + \rho W(t-1) + \xi(t), \tag{4}$$

where $t = 2,\ldots,T$, $|\rho| < 1$ is common to all locations and $W(1) \sim \mathcal{N}(0, \{\tau_\xi(1-\rho^2)^{-1}\}$, which is the stationary distribution of the process. The frequency value at the ith neighbor of the target location $\mathbf{s}$ is denoted by $Y(\mathbf{s}_i,t)$, whereas $\left|N_{\mathbf{s}}\right|$ is the cardinality

of the set of neighbors, and α_i is the linear coefficient for the ith neighbor at time t. The measurement error term $\xi(t) \sim N(0, \tau_\xi^{-1})$ is assumed uncorrelated with $W(t-1)$.

With the first-order-Markov property of $W(t)$, the joint distribution of $\pi(\mathbf{W} \mid \rho, \tau_\xi) = \pi\{W(1)\}\pi\{W(2) \mid W(1)\}, \ldots, \pi\{W(T) \mid W(T-1)\}$ is given by

$$\mathbf{W} \mid \rho, \tau_\xi \sim \mathcal{N}(\boldsymbol{\mu}_W, \mathbf{Q}_1^{-1}), \tag{5}$$

where $\mathbf{W} = (W_1, \ldots, W_T)^\top$, $\boldsymbol{\mu}_W$ is the mean vector with each element equal to $\sum_{i=1}^{|N_s|} \alpha_i Y(\mathbf{s}_i, t)$, and $\mathbf{Q}_1$ is the tridiagonal precision matrix of an autoregressive process of order 1 [23]:

$$\mathbf{Q}_1 = \tau_\xi \begin{pmatrix} 1 & -\rho & & & & \\ \rho & 1+\rho^2 & -\rho & & & \\ & \ldots & \ldots & \ldots & & \\ & & & -\rho & 1+\rho^2 & -\rho \\ & & & & -\rho & 1 \end{pmatrix}. \tag{6}$$

Recall that we select the exogenous variables among the set $N_\mathbf{s}, \mathbf{s} \in \mathcal{D}_s$ in (4) based on the spatial separation to the target location $\mathbf{s}$. A small cross-validation study showed little improvement after the two most correlated neighbors are used; therefore we fix $|N_\mathbf{s}| = 2$.

Spatiotemporal dynamic latent process (ST)

Whereas the ARX model in (4) accounts for neighboring information using lagged exogenous variables, here we present a model that explicitly considers the spatiotemporal dependence among different buses. Compared with the ARX previously described, this model not only enables longer spatial ranges for the correlations but also has the advantage that the fit is done in just one step. The first step of ARX is not needed because neighboring information is dictated by the estimated spatial correlations directly, and pairwise geographical distances or correlations do not have to be precalculated. Therefore, predictive distributions for the ST model can be derived from multiple locations simultaneously.

A convenient way of introducing space-time correlation is through temporal hierarchical dynamical models. Storvik et al. [27] showed how to build a time autoregressive spatial model through a convolution kernel

$$W(\mathbf{s}, t) = \int_{\mathbb{R}^2} h(\mathbf{x}) W(\mathbf{s}+\mathbf{x}, t) \mathrm{d}\mathbf{x} + \xi(\mathbf{s}, t), \quad \mathbf{s}, \mathbf{x} \in \mathcal{D}_s, t \in \mathcal{D}_T, \tag{7}$$

where $h(\mathbf{x})$ is usually assumed to be a Gaussian kernel. The innovation ξ is a stationary zero-mean Gaussian field that accounts for the small-scale dependence. It is assumed to be independent of $W(\mathbf{s}, t)$, uncorrelated in time but correlated in space. A separable space-time model can be formulated as a special case of this representation when $h(\mathbf{x}) = \rho\delta(\mathbf{x})$, where $\delta(\mathbf{x})$ is the Dirac

delta. An attractive property of separability is its representation in a dynamic, autoregressive form. Assuming that $W(\mathbf{s},t) \sim \mathcal{N}\{0, \mathbf{Q}_1^{-1}(|\ t-t'\ |) \otimes \mathbf{Q}_2^{-1}(\|\mathbf{s}-\mathbf{s}'\|)\}$, where $\otimes$ denotes the Kronecker product, $\mathbf{Q}_1$ is the precision matrix for the temporal model, and $\mathbf{Q}_2$ is the precision matrix for the purely spatial process, then one can show that

$$W(\mathbf{s},t) = \rho\eta(\mathbf{s},t-1) + \xi(\mathbf{s},t), \tag{8}$$

where $\xi(\mathbf{s},t) \sim \mathcal{N}\{0,(1-\rho^2)\mathbf{Q}_2^{-1}\}$ and $\mathbf{Q}_1$ is of the form $\mathbf{Q}_1(|t-t'|) = \rho^{-|t-t'|}$ (see Equation (6)). In terms of the spatial dependence, we assume a Matérn covariance function for $\mathbf{Q}_2^{-1}$, parametrized in terms of three parameters

$$\mathbf{Q}_2^{-1}(\|\mathbf{s}-\mathbf{s}'\|) = \sigma_\xi^2 \frac{1}{\Gamma(\nu)2^{\nu-1}}(\kappa\|\mathbf{s}-\mathbf{s}'\|)^\nu \mathcal{K}_\nu(\kappa\|\mathbf{s}-\mathbf{s}'\|), \tag{9}$$

where σ_ξ^2 denotes the marginal variance. Whereas in the R-INLA parametrization, the scaling parameter κ is used; a more natural interpretation of this parameter is the range r, defined as the distance at which the observations are nearly independent. The empirically derived definition $r = \sqrt{8\nu}/\kappa$ gives an explicit formula connecting the scaling and the range, which corresponds to a correlation of 0.13 at the distance r for all $\nu > 1/2$ [22]. We denote by $\mathcal{K}_\nu$ the modified Bessel function of the second kind of order ν, where ν is a smoothness parameter. The parameter ν determines the mean-square differentiability of the field, and it is difficult to identify from data [9]. In R-INLA, the default value is $\nu = 1$, and its accuracy for different 2-dimensional data problems has been extensively tested, showing to capture a broad class of spatial variation [22]. Examples of fixing $\nu = 1$ in spatial and spatiotemporal modeling include Cameletti et al. [5] and Ingebrigtsen et al. [16].

The full covariance matrix is separable and given by the Kronecker product $\mathbf{Q}_1^{-1} \otimes (1-\rho^2)\mathbf{Q}_2^{-1}$. Its Kronecker product structure makes it efficient to handle from a computational perspective [14]. While realistic representations of the spatiotemporal process should respect the fact that interactions between space and time are necessary for dynamic propagation, separable models are widespread since they decrease the number of parameters in the covariance matrix, resulting in a substantial reduction of the computation time [14]. Besides the computational aspect, recent studies have showed significant chances of model overfitting when a high degree of flexibility is imposed in the covariance structure when the information in the data is weak [11].

Prior specification

The last step before performing inference with the models described previously consists of defining the prior distributions for the parameters (see the last layer of (1)). We use the concept of Penalized Complexity (PC) to construct informative priors reflecting our belief on the data [26]. This approach prevents model overfitting by penalizing complex models over a parsimonious one. The PC concept is used for the autoregressive coefficient ρ; the Matérn covariance parameters range r, and the marginal variance σ_ξ. We define the prior such that $\mathrm{P}(\rho > 0) = 0.9$, which is motivated by the highly unlikely negative temporal correlation. The joint PC priors introduced in [12] are used for r and σ_ξ, with probabilities $\mathrm{P}(r < r_0) = 0.5$ and $\mathrm{P}(\sigma_\xi > \sigma_0) = 0.5$, respectively.

We fix $r_0 = 1700$ km, which is half of the domain size and $\sigma_0 = 0.01$.We assume a zero-mean prior Gaussian distribution with a small precision for the intercept β_0 and the exogenous variables' coefficients α_i's. A moderately informative prior distribution is used for the precision τ_y, such that the probability of observing a standard deviation larger than 0.01, the empirical standard deviation in our data, is 10%.

2.3 | Bayesian inference and prediction

In this section, we set up the Bayesian framework used for performing inference and prediction with the models described in Section 2.2. Let $\boldsymbol{\theta} \equiv (\tau_y, \boldsymbol{\theta}_\eta)^\top$ denote the vector of all model parameters in (1). The joint posterior distribution of the latent process $\boldsymbol{\eta}$ and the parameters $\boldsymbol{\theta}$ given observations $\mathbf{Y}$ (with components $Y(i)$, $i \in \mathcal{I}$, for some index set $\mathcal{I}$) is given by

$$\begin{aligned}\pi(\boldsymbol{\eta}, \boldsymbol{\theta} \mid \mathbf{Y}) &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta}) \prod_{i \in \mathcal{I}} \pi\{Y(i) | \eta(i), \boldsymbol{\theta}\} \\ &\propto \pi(\boldsymbol{\theta})|\mathbf{Q}_\eta|^{1/2}\tau_y^{1/2}\exp\left[-\frac{1}{2}\left\{(\boldsymbol{\eta}^\top - \boldsymbol{\mu}_\eta)^\top \mathbf{Q}_\eta(\boldsymbol{\eta} - \boldsymbol{\mu}_\eta) + \tau_y(\mathbf{Y} - \boldsymbol{\eta})^\top(\mathbf{Y} - \boldsymbol{\eta})\right\}\right],\end{aligned} \tag{10}$$

where $\boldsymbol{\mu}_\eta$ and $\mathbf{Q}_\eta$ are the mean and precision matrix of $\boldsymbol{\eta}$, respectively (1). The posterior distribution of the parameters given the data, that is, $\pi(\boldsymbol{\theta} \mid \mathbf{Y})$, before normalization can be written as

$$\pi(\boldsymbol{\theta} \mid \mathbf{Y}) \propto \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{Y} \mid \boldsymbol{\eta}, \boldsymbol{\theta})\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta})}{\pi(\boldsymbol{\eta} \mid \mathbf{Y}, \boldsymbol{\theta})}, \tag{11}$$

where the priors for the parameters $\pi(\boldsymbol{\theta})$ are given in the end of Section 2.2, whereas both $\pi(\boldsymbol{\eta} \mid \boldsymbol{\theta})$ and $\pi(\mathbf{y} \mid \boldsymbol{\eta}, \boldsymbol{\theta})$ are Gaussian distributed (see (1)). The denominator in (11) can be easily computed from the joint distribution (10), by ignoring the terms that do not depend on $\boldsymbol{\eta}$, since $\mathbf{Y}$ and $\boldsymbol{\theta}$ are considered fixed in $\pi(\boldsymbol{\eta} \mid \mathbf{Y}, \boldsymbol{\theta})$.

The main objective is to obtain predictions by computing the posterior marginal distributions for the elements i of the latent field given the data, that is, $\pi(\boldsymbol{\eta} \mid \mathbf{Y})$. Starting from the joint posterior in (10), this task is achieved by averaging over the posterior distribution:

$$\pi\{\eta(i) \mid \mathbf{Y}\} = \int \pi(\eta(i) \mid \mathbf{Y}, \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \mathbf{Y})\mathrm{d}\boldsymbol{\theta}, \tag{12}$$

where $\pi(\boldsymbol{\theta} \mid \mathbf{Y})$ on the right-hand side of (12) is computed as in (11). The posteriors in (12) and (11) for the ST and ARX models described in Section 2.2 are computed through the integrated nested Laplace approximation approach implemented in the R-INLA package (see Krainski et al. [19] for details).

After the inference is performed, samples of the predictive posterior distribution of the frequency response at location $\mathbf{s}$ and time t can be obtained by first sampling a set of parameters and then, conditional on these parameter values, sampling from the Gaussian distributions arising from the latent process. For each model, we have implemented this simulation procedure to obtain 5,000 samples of the frequency response at unobserved locations. Values of point prediction are obtained by taking either the mean or the median of the posterior samples, depending on the loss function to be minimized [3].

3 | RESULTS

We assess the ARX and ST models' adequacy to predict the frequency response at new unobserved locations. A comparison of the performance from the models in Section 2.2 is performed in terms of both point- and probabilistic-prediction skills. Moreover, we highlight the cases where spatial information improves prediction. We also discuss the specific instances when the models fail to capture the data characteristics, and we suggest how this issue can be addressed when knowledge of the power grid is available. Details of the evaluation scheme are described in Section 3.1, and results from fitting our models to the FNET data are described in Section 3.2.

3.1 | Evaluation framework

We show the results to several windows of the dataset described in Section 2.1. Specifically, we compare two different scenarios. We first divide the entire day worth of data into 4120 non-overlapping windows of 20-seconds. The first scenario comprises the windows with the most extreme drops. These windows are found by fitting a linear model $\beta_0 + \beta_1 t$ to each window, and selecting the 20 the most negative estimated β_1's. The second scenario is a random sample of 50 windows. While the first scenario intends to compare the models in the case of a disturbance, the second deals with the bias in pre-selected samples. The randomness captures how the models perform overall throughout the dataset. The scores for each model will be shown separately for the two scenarios in Section 3.2.

For each scenario and each selected window, we perform location-specific leave-one-out cross-validation. The data in which window are split such that the training set consists of the 20-second time series data from all the locations except the bus where prediction is desired, which is the testing set. Both the training and testing sets comprise 20 seconds of the frequency response, equivalent to 200 time points, which is a typical length of a generation trip [2]. In total we fit each model 4900 times, with data from 70 windows and 70 different training sets, and in each case we predict the entire 20-second window at the location left out during estimation. The forecast performance is obtained by combining the estimates from the 4900 datasets.

To speed up computations, we parallelize the code using the PARDISO sparse library for fast numerical matrix computations [18] with 10 INLA threads on 10 cores. The optimization relies on different processors that calculate different parameter sets. The parallelization speeds up computation considerably, so that each evaluation takes less than a minute.

3.2 | Prediction comparison

In this section, we show the results from fitting the ARX and the ST models described in Section 2.2 to predict the frequency response at a new location from the FNET using the framework outlined in Section 3.1. In addition, we introduce a benchmark

method, which we denote by *Neighbor Mean*, such that the prediction of the frequency at location $\mathbf{s}_i$ and time t, i.e. $Y(\mathbf{s}_i, t)$, is defined as $\widehat{Y}_A = \frac{1}{N-1} \sum_{j \neq i} Y(\mathbf{s}_j, t)$, the average of the frequency among the all but the i-th location at time t.

The benchmark *Neighbor Mean* can only provide point predictions, which is assessed using the root mean squared error (RMSE) and mean absolute error (MAE). Probabilistic predictions are essential in decision-making problems, and we show results in terms of predicted quantiles with reliability diagrams. The MAE, RMSE, and reliability diagrams are the combined results from all the 70 sites used in our leave-one-out spatial cross-validation at each window.

Figure 3 shows examples of four predicted frequency time series from the ARX and ST models at different locations, respectively numbered 1, 2, 3 and 4 in Figure 2 . Location 1 is in New York state (upper left plot in Figure 3), location 2 in Florida (upper right plot in Figure 3), location 3 in Tennessee (lower left plot in Figure 3) and location 4 in Minnesota (lower right plot in Figure 3). The black lines in this figure show the actual frequency data, whereas the shaded areas correspond to the 95% central interval of the predictions from the ARX model (green) and the ST model (red). The ST model outperforms the ARX for all four locations, and especially at peaks at roughly 5 and 15 seconds for the disturbance in New York, with the 95% confidence interval always covering the actual observations for model ST. For both models, the 95% confidence intervals for the predictions in Florida do not cover the true value for most time points. Surprisingly, the models are not able to predict well at this location, and, in fact, the predictions are shifted by around one or two seconds from the actual values. In the next section, we explore further the reasons for and the implications of the uncalibrated predictions generated by our models at this specific bus, which we call the target location.

Calibration is critical in practical applications and is assessed here by using a reliability diagram [15]. To construct the diagram, we compute the proportion of times that the nominal coverage from the predictive distributions of both models matches the actual observations. For a calibrated prediction, it is expected that a nominal proportion α covers the observations α% of the time, which would result in points in the reliability diagram aligned to the diagonal. Reliability diagrams with nominal quantiles from 5% to 95% in steps of 5% for the ARX and ST models are presented in Figure 4 . This figure's left plot corresponds to the first scenario, where the windows with the most massive drops are selected, and the right side is for the second scenario with the randomly selected windows. In both scenarios, the predictions by model ST are better calibrated than those of the ARX model, with points close to the diagonal for most quantiles. The average time for obtaining inference and predictions in a single processor Intel Core 8-Core i9/2.4 GHz machine with the ARX and ST models are 12 and 59 seconds for the first scenario and 9 and 63 seconds for the second scenario, respectively. Moreover, because INLA uses the mesh nodes for fitting spatio-temporal models, the spatial dimension has less impact on the models' running time. A small experiment showed that our models scale reasonably well with the time window length, inference, and prediction using 3000-time points (i.e., 5-minutes of frequency data) and the ST model, taking around 17 minutes on the same machine.

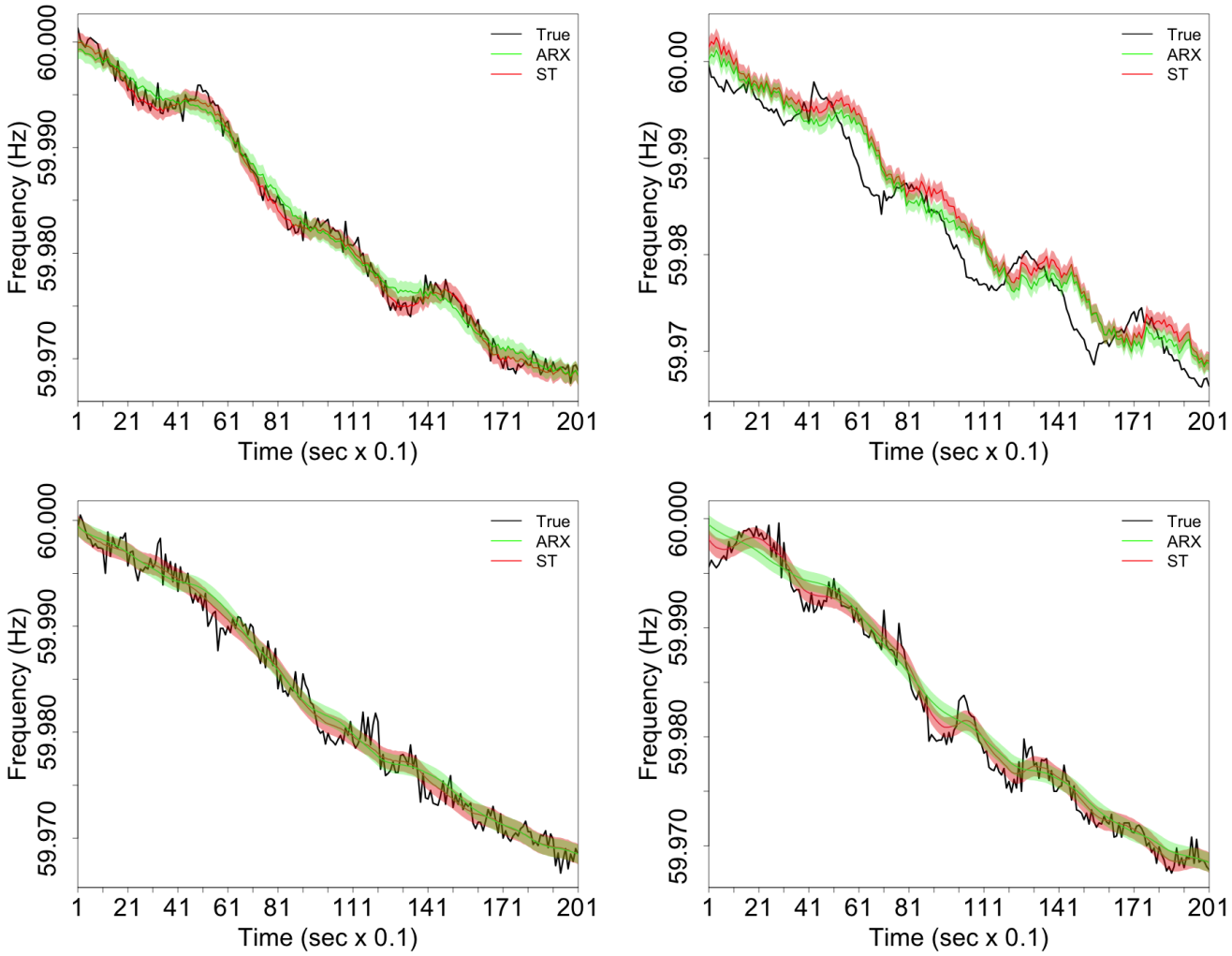

FIGURE 3 Examples of the predicted frequency response during a 20-seconds disturbance at four different locations: New York (left, upper) and Florida (right, upper), Tennessee (left, lower) and Minnesota (right, lower). The RMSE and MAE at these four locations are 0.0008, 0.0016, 0.0012, 0.0011 and 0.0004, 0.0010, 0.0009, 0.0008, respectively, for the ST model and 0.0011, 0.0026, 0.0014, 0.0016 and 0.0004, 0.0014, 0.0011, 0.0013 for the ARX model. Point-wise 5% and 95% quantiles of the predictive distributions are shown as the green shaded area for the ARX model and red for the ST model. The solid black curve is the true frequency disturbance.

To complement the probabilistic prediction evaluation in Figures 3 and 4 , we assessed the point prediction with RMSE and MAE values. The RMSE values are obtained as $\{Y(\mathbf{s},t) - \hat{Y}(\mathbf{s},t)\}^2$, where $Y(\mathbf{s},t)$ is the frequency response at location $\mathbf{s}$ and time t, whereas $\hat{Y}(\mathbf{s},t)$ is taken as the mean of the predicted posterior distribution of $Y(\mathbf{s},t)$, since the RMSE is a quadratic loss function [4]. Similarly, the MAE values are computed as $\text{MAE}(t) = \mid Y(\mathbf{s},t) - \hat{Y}(\mathbf{s},t) \mid$, where $\hat{Y}(\mathbf{s},t)$ is the median of the predicted posterior distribution, as the cost function of the MAE is symmetric and linear [4].

Figure 5 presents boxplots of RMSE in the left plot and MAE in the right plot from the predictions given by the ARX and the ST model in each scenario (drops and random). Windows with the most significant drops are shown in red, and the random samples of the windows are in green. When comparing the three models, we can see that Model ST has a superior performance

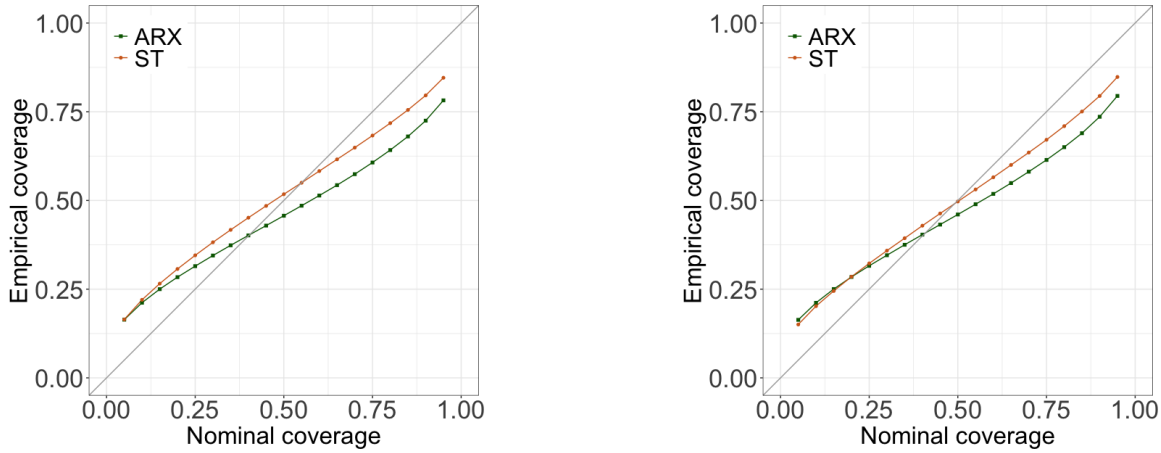

FIGURE 4 Reliability diagrams of the predicted frequency response from the ARX (green) and the ST (orange) models for the 20 windows with the largest drops (left) and 50 randomly selected windows (right). In both scenarios, the diagrams are calculated using the combined prediction results from the 70 sites used in our leave-one-out spatial cross-validation at each window.

concerning RMSE and MAE values, as well as in both scenarios. We find that the random samples have a lower variance for all three models, indicating that extreme drops' predictions are less consistent than for other frequency conditions. The largest MAE and RMSE values for the ST and ARX models, correspond to the predictions in Florida (see upper left plot in Figure 3). The conclusion for point predictions is in agreement with the conclusion for the reliability diagram on probabilistic predictions, with better results for the ST model. The south of the U.S. is known to suffer from under-generation. Therefore, it depends highly on generators from nearby regions, explaining the high spatial dependence at longer distances that are not well captured by classical spatial statistical models. Table 1 displays the mean RMSE and MAE values for models *NeighborMean*, ARX, and ST according to the two considered scenarios. The relative differences (in %) in RMSE and MAE values from the ARX and ST models with respect to the *NeighborMean* are shown between brackets. For instance, in terms of RMSE and for the scenario with the largest frequency drops, the ST model improves the *NeighborMean* by 9.5%, whereas the ARX worsens it by 2.3%.

Scenario		NeighborMean	ARX	ST
Drop	RMSE	1.262	1.291 (-2.3%)	**1.142** (9.5%)
	MAE	0.917	0.932 (-1.6%)	**0.804** (12.3%)
Random	RMSE	1.223	1.231 (-0.7%)	**1.092** (10.7%)
	MAE	0.889	0.888 (0.1%)	**0.772** (13.1%)

TABLE 1 Mean skill scores (RMSE and MAE ×1000) across all sites and broken down by scenario. The best scores are highlighted in bold. Between brackets are the relative differences (in %) in RMSE and MAE values with respect to the *NeighborMean* model. Negative values indicate a worsening from the *NeighborMean* model.

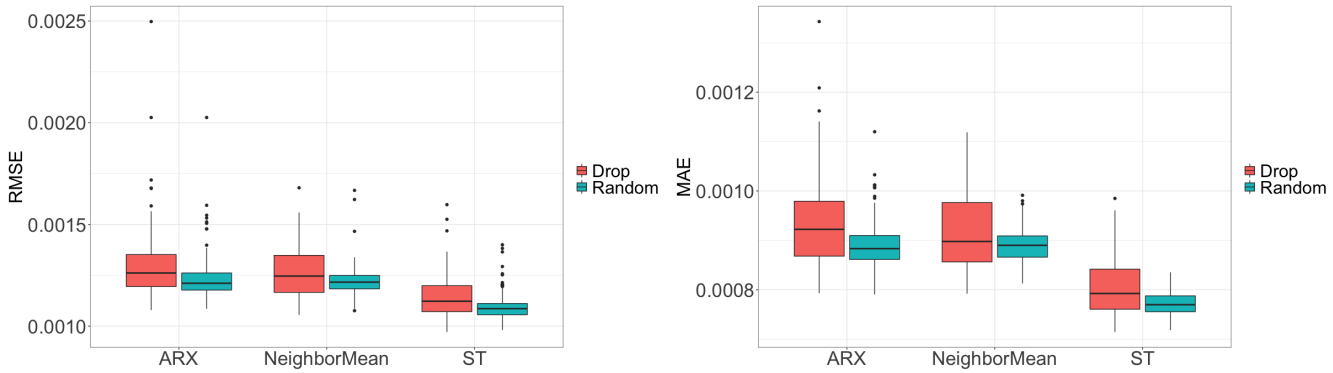

FIGURE 5 Boxplots of the RMSE (left) and MAE (right) for the 20 windows with the most extreme drops (orange) and 50 randomly selected windows (blue) from models ARX, mean of the neighbors, and ST. Each boxplot is composed of the results from the 70 sites used in our leave-one-out spatial cross-validation, times the number of windows. The ST model produces the smallest RMSE and MAE values.

3.3 | Analysis of the spatial correlation

In this section we investigate further the predictions from Florida (see the right plot of Figure 3), since this is the worst fit from all models. Whereas the predictions from New York, on the left plot of Figure 3 are close to the truth, predictions for Florida are shifted from the true values. Whereas Florida example shows that the models described in Section 2.2 do have limitations for predicting power grid frequency, it is clear that the ST model is able to provide sensible results and it overall outperforms both the simple neighbor average and the ARX models (see Section 3.1. Furthermore, if we had further information about the grid, we could improve our method.

We investigated the reason for this shift by looking at the time series of the frequency response during the event at the target location as well as its closest neighbors. The target location is represented with the number 4 on the maps in Figure 2 , and its corresponding frequency response is shown as an orange dashed line in the time series in Figure 6 . The nearest and most correlated neighbors to the target are labeled with 5 and 6 in Figure 2 and have time series of frequency data plotted with dotted dark blue and solid yellow lines, respectively.

A shift of about one second in predictions from both models on the right-hand side of Figure 3 is also seen in the time series with the dotted dark blue curve in Figure 2 . The covariance structure in (9) can explain this shift in predictions, where locations close by are assumed to have a higher correlation than those farther apart. In other words, the proposed models give more weight to the sites nearby and ignore the unobserved generator network from the correlation structure, which can be unsuitable for certain cases such as the one in Florida shown in Figure 2 .

The empirical correlations between the frequency data at the target location and these neighbors are calculated after removing the linear temporal component (see Figure 1). We notice that the two closest neighbors have a relatively weak relationship to

the target, with values equal to 0.66 and 0.62. On the other hand, the frequency values from the two farther neighbors are more similar to the target location, with correlations values around 0.95.

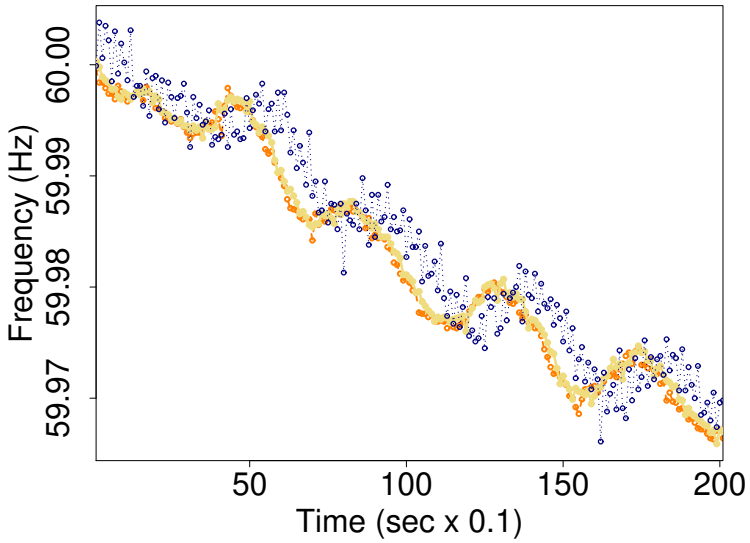

FIGURE 6 Time series of the frequency response at three locations in the south of Florida marked with numbers 4, 5, and 6 on the maps in Figure 2 . The time series in orange dashed line is number 2 in Figure 2 , the dotted dark blue line is number 5, and the solid yellow line is number 6.

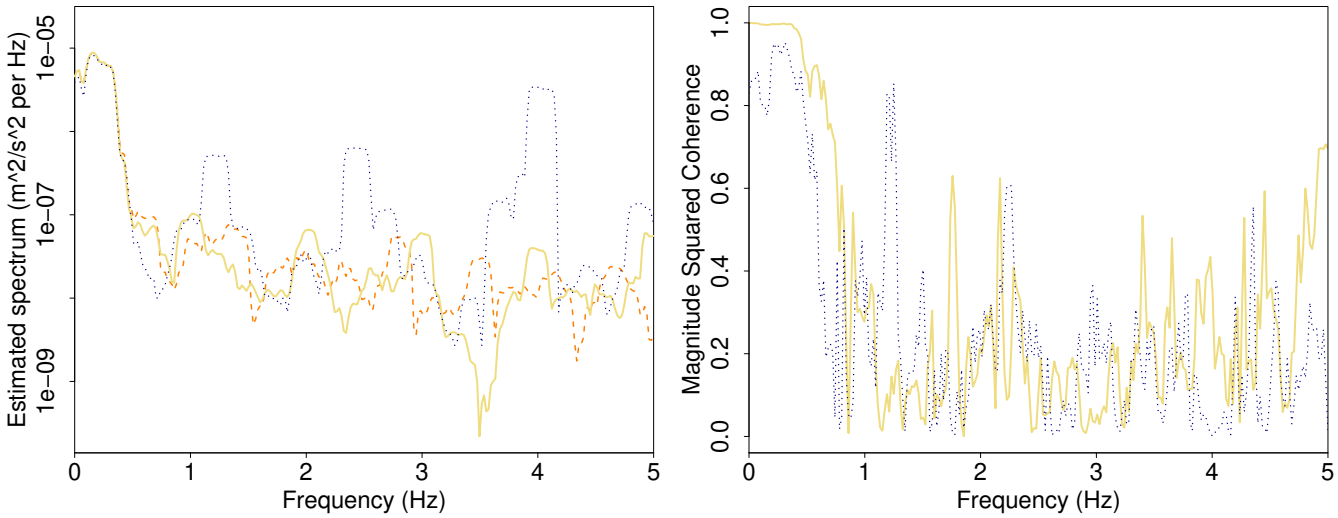

FIGURE 7 Left: Multitaper spectrum estimates of the residuals from removing a linear trend from the three time series in Figure 6 . Right: Multitaper magnitude-squared coherence (MSC) curves between the curves on the left plot. The solid yellow represents the MSC values between the dashed orange curve and the solid yellow, and the dotted dark blue is the MSC curve between the dashed orange and the dotted dark blue lines. The MSC curves show an example where the geographically closest neighbors are not the most correlated with the target.

.

To substantiate these differences, Figure 7 (left) shows the multitaper spectrum estimates of the residuals after removing a linear trend from the three time series in Figure 6 . In contrast to the commonly used Fourier transform, this method performs spectral density estimation of a time series without assuming that the coefficients are well described by the amplitude and the phase of the frequency component. Instead, the multitaper method reduces estimation bias by averaging over several tapered spectra, with each taper being pairwise orthogonal to all other tapers and therefore independent estimates of the underlying spectrum [28].

Recall that our sampling frequency in seconds is 10. The Nyquist rate is the highest frequency used to visualize a spectrum estimate, and it is half of the sampling frequency, namely, 5 Hz (see x-axis in Figure 7). The estimated spectrum is shown as dark blue dots in the left plot of Figure 7) and contains numerous harmonic oscillations with peaks at different frequencies that stand out from the other series. Peaks in multitaper spectra are known to be approximately square-shaped, with a particular width connected with the number of tapers used [28], as also shown in the spectrum estimates in Figure 7 . The right plot in this figure shows the magnitude squared coherence (MSC) values. The solid yellow line is the MSC between the dashed orange and the solid yellow curves on the left-hand side, and the dotted dark blue line is the MSC between the dashed orange and the solid yellow time series. The MSC values are computed by using a jackknife technique, and they are always a value between zero and one, equivalent to the cross-correlation between two time series [29]. This figure shows strong oscillations at low frequencies common to all series, as well as oscillations at 1.2 Hz, 2.4 Hz, and 3.0 Hz for the dotted dark blue line. In terms of seconds, the peak at 1.2 Hz corresponds to an oscillation with period $1/1.2 = 5/6 = 0.83$ seconds, for example. From these three instances, we notice that the two series representing farther-away locations (dashed orange and solid yellow) have larger MCS values than those that are placed closer together (dashed orange and dotted dark blue).

Because the closest neighbors are not the most correlated with the location we want to predict, we have evidence that models based solely on geographical proximity are not the most appropriate to describe power grid data. In a power grid system, the measurements at different locations have underlying relationships that depend on how the network is designed and on the distribution of generators. Ideally, information about the grid network would be used to learn how to define a more suitable distance metric. Unfortunately such information in not available to us at this time. A necessary condition for building a new model is that the Matérn covariance in our ST model (see (9)) must result in a valid autocovariance function concerning this new metric. Nonetheless, grid information is often unavailable, and it is worth pursuing distance-based approaches as we did in this work. We also note that the Florida example we highlighted was the exception and not the rule; for most sites we observed a behavior close to the New York example in Figure 3 where the space-time model did a good job in predicting the unobserved response.

4 | DISCUSSION

We presented model-building strategies for predicting the frequency response during a disturbance in power grid systems at new locations, which takes into account the correlations between different buses across space and time. Our case study is based on data from the FNET/GridEye, a wide-area frequency measurement network. We used a Bayesian framework where the model components are specified hierarchically so that conditional independence applies and computations were feasible. The Bayesian framework makes it possible to simulate from the posterior distribution of unobserved data points and obtain not only point prediction but also quantiles, which are essential for assessing calibration.

We proposed two types of interactions within the hierarchical formulation describing the spatiotemporal dynamics of the frequency response. In both cases, the linear predictor is composed of an intercept that is independent across location and time and a linear time trend that is common to all sites. Two types of interactions are introduced through latent Gaussian processes to model the correlations that are left. We started with an autoregressive with exogenous input model with neighboring information in terms of lagged frequency response. This model was inspired by the work in Bai et al. [2], where an ARX based on correlations from previous data was used to predict disturbances at unobserved locations. This approach avoids the issue caused by using distances to model correlations but relies on the assumption that there was a similar event in the past and therefore cannot predict at new unobserved locations.

In the second model, space-time correlations were introduced for the underlying process, which varies in time with first-order autoregressive dynamics and has spatially correlated innovations given by a zero-mean Gaussian process with Matérn covariance. Although including exogenous input in the autoregressive model enables strength from neighbors to influence the predictions, the spatiotemporal model improved the predictive performance by letting the data dictate the length of the spatial correlation. We believe that a more dense spatial data set could substantially benefit from having the explicit spatiotemporal correlations in the model.

Our approaches result in calibrated predictions; however, the ST model outperformed the ARX model in terms of point and probabilistic predictions. We noticed that both models do not always capture the underlying space-time correlation; this is especially true when there is a weak dependence with locations nearby compared with those farther apart. We conjecture that geographical distances are not sufficient to represent the dependencies across locations, given the complex power grid structure driving the underlying process. The case study shows that locations with the smallest distance to the target are not necessarily the most correlated. A more appropriate distance metric could be used if the information on the unobserved grid network is available.

It has been demonstrated that the proposed ST model produces predictions that are a non-negligible improvement on competitive benchmarks in terms of several skill scores during both events and normal operations. The conventional benchmark

consists of the average frequency response among the neighboring location. However, it does not offer probabilistic information required for optimal decision-making under uncertainty, hence the move to more sophisticated approaches.

Building on the work presented here, future studies may involve incorporating information about the grid structure, such as distance to generators, into the spatiotemporal correlations in our model. If this information is available, one could build a distance informed by the grid characteristics, or something like that a Matérn model with a different distance from the Euclidean could be used to model the frequency response at the various locations. A challenge with using a new metric is that the Matérn model should still give a valid (i.e., positive-definite) autocorrelation function concerning the new metric. Another interesting problem that can improve predictions is to consider whether the grid structure should affect the autoregressive coefficients since the evolution in time might also depend on the dynamic of the power grid system.

ACKNOWLEDGMENTS

We are grateful to Yilu Liu, Henry Yin, and Chujie Zeng for making one day of FNET data available to us as well as for technical discussions about the data and detection. We are grateful to Charlotte Haley for helping in the multitapered spectral estimation. This material was based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contract DE-AC02-06CH11347 and by NSF through award CNS-1545046.

AUTHOR BIOGRAPHY

References

[1] Aminifar, F., M. Shahidehpour, M. Fotuhi-Firuzabad, and S. Kamalinia, 2013: Power system dynamic state estimation with synchronized phasor measurements. *IEEE Transactions on Instrumentation and Measurement*, **63**, no. 2, 352–363.

[2] Bai, F., Y. Liu, Y. Liu, K. Sun, N. Bhatt, A. Del Rosso, E. Farantatos, and X. Wang, 2015: Measurement-based correlation approach for power system dynamic response estimation. *IET Generation, Transmission & Distribution*, **9**, no. 12, 1474–1484.

[3] Banerjee, A., X. Guo, and H. Wang, 2005: On the optimality of conditional expectation as a bregman predictor. *IEEE Transactions on Information Theory*, **51**, no. 7, 2664–2669.

[4] Banerjee, S., B. P. Carlin, and A. E. Gelfand, 2014: *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press.

[5] Cameletti, M., F. Lindgren, D. Simpson, and H. Rue, 2013: Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, **97**, no. 2, 109–131.

[6] Cotilla-Sanchez, E., P. D. Hines, and C. M. Danforth, 2012: Predicting critical transitions from time series synchrophasor data. *IEEE Transactions on Smart Grid*, **3**, no. 4, 1832–1840.

[7] Cui, Y., W. Wang, Y. Liu, P. Fuhr, and M. Morales-Rodriguez, 2019: Spatio-temporal synchrophasor data characterization for mitigating false data injection in smart grids. *2019 IEEE Power & Energy Society General Meeting (PESGM)*, IEEE, 1–5.

[8] Diao, R., K. Sun, V. Vittal, R. J. O'Keefe, M. R. Richardson, N. Bhatt, D. Stradford, and S. K. Sarawgi, 2009: Decision tree-based online voltage security assessment using PMU measurements. *IEEE Transactions on Power Systems*, **24**, no. 2, 832–839.

[9] Diggle, P., P. Ribeiro, and M.-b. Geostatistics, 2007: *Springer series in statistics*.

[10] Dong, J., X. Ma, S. M. Djouadi, H. Li, and Y. Liu, 2014: Frequency prediction of power systems in fnet based on state-space approach and uncertain basis functions. *IEEE Transactions on Power Systems*, **29**, no. 6, 2602–2612.

[11] Fuglstad, G.-A., D. Simpson, F. Lindgren, and H. Rue, 2015: Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics*, **14**, 505–531.

[12] — 2019: Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, **114**, 445–452.

[13] Gao, P., M. Wang, S. G. Ghiocel, and J. H. Chow, 2014: Modeless reconstruction of missing synchrophasor measurements. *2014 IEEE PES General Meeting| Conference & Exposition*, IEEE, 1–5.

[14] Genton, M. G., 2007: Separable approximations of space-time covariance matrices. *Environmetrics*, **18**, no. 7, 681–695.

[15] Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**, no. 2, 243–268.

[16] Ingebrigtsen, R., F. Lindgren, and I. Steinsland, 2014: Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, **8**, 20–38.

[17] Kamwa, I., S. Samantaray, and G. Joos, 2009: Development of rule-based classifiers for rapid stability assessment of wide-area post-disturbance records. *IEEE Transactions on Power Systems*, **24**, no. 1, 258–270.

[18] Kourounis, D., A. Fuchs, and O. Schenk, 2018: Toward the next generation of multiperiod optimal power flow solvers. *IEEE Transactions on Power Systems*, **33**, no. 4, 4005–4014.

[19] Krainski, E. T., V. Gómez-Rubio, H. Bakka, A. Lenzi, D. Castro-Camilo, D. Simpson, F. Lindgren, and H. Rue, 2019: *Advanced Spatial Modeling with Stochastic Partial Differential Equations using R and INLA*. CRC press. Github version www.r-inla.org/spde-book.

[20] Le, N. T. and W. Benjapolakul, 2018: A data imputation model in phasor measurement units based on bagged averaging of multiple linear regression. *IEEE Access*, **6**, 39324–39333.

[21] Li, C., Y. Liu, K. Sun, Y. Liu, and N. Bhatt, 2014: Measurement based power system dynamics prediction with multivariate autoregressive model. *2014 IEEE PES T&D Conference and Exposition*, IEEE, 1–5.

[22] Lindgren, F., H. Rue, et al., 2015: Bayesian spatial modelling with r-inla. *Journal of Statistical Software*, **63**, no. 19, 1–25.

[23] Rue, H. and L. Held, 2005: *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.

[24] Rue, H., S. Martino, and N. Chopin, 2009: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, no. 2, 319–392.

[25] Rue, H., A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren, 2017: Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application*, **4**, 395–421.

[26] Simpson, D., H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye, 2017: Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, **32**, no. 1, 1–28.

[27] Storvik, G., A. Frigessi, and D. Hirst, 2002: Stationary space-time Gaussian fields and their time autoregressive representation. *Statistical Modelling*, **2**, no. 2, 139–161.

[28] Thomson, D. J., 1982: Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, **70**, no. 9, 1055–1096.

[29] — 1991: Jackknife error estimates for spectra, coherences, and transfer functions, advances. *Spectral Analysis and Array Processing*, 58–113.

[30] Yamashita, K. and H. Kameda, 2004: Out-of-step prediction logic for wide-area protection based on an autoregressive model. *IEEE PES Power Systems Conference and Exposition, 2004.*, IEEE, 307–312.

How to cite this article: Lenzi, A., Bessac, J., Anitescu, M. (2020), Power grid frequency prediction using spatiotemporal modeling, *Statistical Analysis and Data Mining*.