International Journal of Forecasting 39 (2023) 967–980

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Physics-informed Gaussian process regression for states estimation and forecasting in power grids

Alexandre M. Tartakovsky [a,b,*], Tong Ma [b], David A. Barajas-Solano [b], Ramakrishna Tipireddy [b]

[a] University of Illinois at Urbana-Champaign, Urbana, IL, United States
[b] Pacifc Northwest Natonal Laboratory, Richland, WA, United States

ARTICLE INFO

Keywords:
Power grid
Parameter estimation
Forecast
Gaussian process regression

ABSTRACT

Real-time state estimation and forecasting are critical for the efficient operation of power grids. In this paper, a physics-informed Gaussian process regression (PhI-GPR) method is presented and used for forecasting and estimating the phase angle, angular speed, and wind mechanical power of a three-generator power grid system using sparse measurements. In standard data-driven Gaussian process regression (GPR), parameterized models for the prior statistics are fit by maximizing the marginal likelihood of observed data. In the PhI-GPR method, we propose to compute the prior statistics offline by solving stochastic differential equations (SDEs) governing the power grid dynamics. The short-term forecast of a power grid system dominated by wind generation is complicated by the stochastic nature of the wind and the resulting uncertainty in wind mechanical power. Here, we assume that the power grid dynamics are governed by swing equations, with the wind mechanical power fluctuating randomly in time. We solve these equations for the mean and covariances of the power grid states using the Monte Carlo simulation method.

We demonstrate that the proposed PhI-GPR method can accurately forecast and estimate observed and unobserved states. For the considered problem, PhI-GPR has computational advantages over the ensemble Kalman filter (EnKF) method: In PhI-GPR, ensembles are computed offline and independently of the data acquisition process, whereas for EnFK, ensembles are computed online with data acquisition, rendering real-time forecast more challenging. We also demonstrate that the PhI-GPR forecast is more accurate than the EnKF forecast when the random mechanical wind power is non-Markovian. In contrast, the two methods produce similar forecasts for the Markovian mechanical wind power.

For observed states, we show that PhI-GPR provides a forecast comparable to the standard data-driven GPR; both forecasts are significantly more accurate than the autoregressive integrated moving average (ARIMA) forecast. We also show that the ARIMA forecast is more sensitive to observation frequency and measurement errors than the PhI-GPR forecast.

1. Introduction

Real-time monitoring and short-term forecasting of power grid states are important for control and planning, including power flow optimization and anomaly

* Corresponding author.
E-mail address: amt1998@illinois.edu (A.M. Tartakovsky).

https://doi.org/10.1016/j.ijforecast.2022.03.007
0169-2070/

detection. Other applications requiring real-time monitoring and short-term forecasting include efficient operation of controllers and the determination of necessary corrective actions against possible failures in power grid systems (Huang, Werner, Huang, Kashyap, & Gupta, 2012). Although modern power grids are heavily instrumented, it remains a challenge to measure all the power grid states due to the inherent high-frequency oscillations of the power grid dynamics and increasing penetration of renewable energy sources. Hence, it is necessary to develop new algorithms for immediate forecasting of observed and unobserved states so that power grid systems can operate efficiently and safely.

Here, we distinguish between forecasting (extrapolation) and state estimation, which, for this work, we define as computing unobserved states in the (recent) past from the values of observed states. There are two general types of forecasting methods, including machine learning (ML) forecasting techniques (e.g., fuzzy regression models (Hong & Wang, 2014; Song, Baek, Hong, & Jang, 2005), support vector machine (Fentis et al., 2016), deep neural networks (Grant, Eltoukhy, & Asfour, 2014; Thiyagarajan & Kumar, 2016; Yeung, Kundu, & Hodas, 2017), gradient boosting machines (Huang & Perry, 2016; Landry, Erlinger, Patschke, & Varrichio, 2016; Lloyd, 2014; Taieb & Hyndman, 2014; Xenochristou, Hutton, Hofman, & Kapelan, 2020), and Gaussian process regression (GPR) (Cai et al., 2020; Quiñonero-Candela & Rasmussen, 2005; Snelson & Ghahramani, 2006; Williams & Rasmussen, 2006)), and statistical techniques (e.g., Markov chains (Yoder, Hering, Navidi, & Larson, 2014), data mining (Ahmad & Chen, 2018; Kusiak, Zheng, & Song, 2009), multiple linear regression models (Charlton & Singleton, 2014; Hong, 2012; Hong, Wilson, & Xie, 2013; Luo, Hong, & Fang, 2018; Sobhani, Hong, & Martin, 2020; Wang, Liu, & Hong, 2016), semi-parametric additive models (Fan & Hyndman, 2011; Goude, Nedellec, & Kong, 2013; Hyndman & Fan, 2009; Nedellec, Cugliari, & Goude, 2014), auto-regressive integrated moving average (ARIMA) models (Brockwell & Davis, 2016; de Oliveira & Oliveira, 2018; Hyndman & Athanasopoulos, 2018; Singh, Mohapatra, et al., 2019), and exponential smoothing models (Gardner, 2006; Hong, Pinson, & Fan, 2014; Hyndman & Athanasopoulos, 2018)). Various hybrid approaches combine some elements of ML and statistics (Catalao, Pousinho, & Mendes, 2011; Dantas & Oliveira, 2018; Liu, Shi, Yang, & Lee, 2012; Smyl, 2020). All ML and statistics methods work better for interpolation than extrapolation. Therefore, forecasting (which is an extrapolation problem) is more challenging than parameter estimation for ML or statistical models.

If fully known, physics-based models should be able to accurately forecast the dynamics of complex systems because the conservation laws on which these models are based hold both in the past and future. However, physics-based models are not fully known for complex systems such as power grid systems. For example, in a model of an electrical grid dominated by wind energy, the wind mechanical power is uncertain and, in general, impossible to predict in the future with absolute certainty. Therefore, physics-based models alone cannot be used for accurate deterministic forecasting of the states of the power grid.

This work proposes a physics-informed GPR (PhI-GPR) method for real-time forecast and estimation of power grid states. In GPR, a state of the power grid is represented as a linear combination of measured values of this state. The coefficients are found as a function of this state's prior mean and covariance functions. In standard "data-driven" GPR, prior statistics are chosen by fitting parameterized models. The hyperparameters of these models are found by maximizing a pseudo-likelihood function of the observations (Genton & Kleiber, 2015; Williams & Rasmussen, 2006). In PhI-GPR, we assume that the power grid dynamics are governed by a known set of equations (e.g., swing equations), and we treat the unknown forcing terms (e.g., the wind mechanical power) in these equations as a random process, which turns these equations into stochastic differential equations (SDEs). We solve these equations for the mean and covariance of the power grid's state using the Monte Carlo (MC) simulation method. We demonstrate that for the considered power grids with states oscillating around equilibrium due to random wind power fluctuations, PhI-GPR can accurately estimate and forecast unobserved states.

We note that our approach is different from the GPR methods, where the physics is enforced by requiring the covariance kernels to satisfy a set of deterministic equations (Chen, Chen, Zhang, & Wu, 2020; Raissi & Karniadakis, 2018; Swiler, Gulian, Frankel, Safta, & Jakeman, 2020). In general, such equations can be obtained in closed form for linear problems or problems that allow accurate linearization. The proposed approach is more general as it will enable estimating the kernel for any highly non-linear system if its accurate mathematical model is available.

There are some similarities between PhI-GPR and the ensemble Kalman filter (EnKF) method (Gillijns et al., 2006; Takeda, Tamura, & Sato, 2016). Both methods use physics-based models to construct distributions in the form of ensembles and construct forecasts of observed and unobserved states in the form of distributions conditioned on the measurements of observed states. However, there are two significant differences between the two methods. In PhI-GPR, the physics-based model is used to compute the prior distribution of the states (i.e., the distribution that is not conditioned on data) and can be done offline, that is, separate from the data acquisition process. The online component of the PhI-GPR forecast only involves regression, which is computationally less expensive than the ensemble generation. In EnKF, the conditional ensembles are computed in real-time with data acquisition. It uses the latest measurements of states to construct initial conditions for solving the physics-based equations. This gives EnKF some advantage as the EnKF statistical model is better centered on data than the one in Ph-GPR. However, in our proposed method, the conditioning is directly performed using all historical observations of the observed states. In EnKF, the conditioning is directly done only on the latest observations of the observed states (past observations are indirectly reflected on the prior statistics). The numerical experiments presented in this work show that the PhI-GPR forecast is more accurate than the EnKF forecast for the considered power grid driven by a non-Markovian process. For

a Markovian driver, we expect the difference between the two conditioning methods on past data to be less pronounced. Indeed, we find that both methods provide a comparable forecast.

For observed states, we find that the accuracy of the PhI-GPR forecast is comparable to the accuracy of the standard "data-driven" GPR forecast. We also compare the PhI-GPR method against the ARIMA method, a commonly used forecasting technique for forecasting observed variables. We find that ARIMA is sensitive to observation frequency and measurement noise. We show that PhI-GPR and ARIMA have comparable accuracy for forecasting observed states with noiseless observations and short times between observations. Nevertheless, in the presence of noise and/or large times between observations, the accuracy of ARIMA deteriorates faster than that of PhI-GPR. Most importantly, the PhI-GPR method can forecast and estimate both observed and unobserved states, while data-driven GPR and ARIMA can only be used for forecasting and estimation of observed states as the observation data are needed to "train" these methods.

This paper is organized as follows. We introduce the PhI-GPR method in Section 2 and the stochastic model of the power grid in Section 3. In Section 4, we apply the PhI-GPR method to the forecasting and state estimation of a synthetic system and compare it to the GPR method. In Sections 5 and 6, we compare the PhI-GPR method with the ARIMA and EnKF methods, respectively. Conclusions are presented in Section 7.

2. PhI-GPR method

This section presents the PhI-GPR method for forecasting observed and unobserved states of dynamical systems. The formulation of multivariate GPR for forecasting is described in Section 2.1. In Subection 2.2, we describe how the physics-based priors for multivariate GPR and PhI-GPR are evaluated.

2.1. Multivariate GPR for forecasting observed and unobserved states

We assume that the system is composed of N observed states, $x_i(t)$ $(i = 1, \dots, N)$, and M unobserved states $y_i(t)$, $(i = 1, \dots, M)$, and that there are measurements $x_{i,j}$ of the observed states at the N_{t^o} discrete times t_j^o $(j = 1, \dots, N_{t^o})$ over the observation window $[0, T^o]$, contaminated by observation noise. We assume that the observation errors are normal, identically distributed, and uncorrelated across times and states; therefore, we model the observations as

$$x_{i,j} = x_i(t_j^o) + \epsilon_{i,j}, \quad \epsilon_{i,j} \sim \mathcal{N}(0, \sigma_n),$$

where σ_n is the standard deviation of the observation errors $\epsilon_{i,j}$. We organize these observations into the vector

$$X^o = [x_{1,1}, x_{1,2}, \dots, x_{1,N_{t^o}}, \dots, x_{N,1}, x_{N,2}, \dots, x_{N,N_{t^o}}]^\top.$$

First, we are interested in forecasting the observed states for the discrete times t_j^f $(j = 1, \dots, N_{t^f})$ over the forecast window $(T^o, T^f]$; that is, we want to estimate the vector of values

$$X^f = [x_1(t_1^f), \dots, x_1(t_{N_{t^f}}^f), \dots, x_N(t_1^f), \dots, x_N(t_{N_{t^f}}^f)]^\top.$$

For this purpose, we employ multivariate GPR regression, in which we model the vectors X^o and X^f as realizations of the random vector $[(X^o)^\top, (X^f)^\top]^\top$ with the distribution

$$\begin{bmatrix} X^o \\ X^f \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{X}^o \\ \bar{X}^f \end{bmatrix}, \begin{bmatrix} K_{X^oX^o} & K_{X^oX^f} \\ K_{X^oX^f}^T & K_{X^fX^f} \end{bmatrix} \right), \tag{1}$$

where $\bar{X}^o$ and $\bar{X}^f$ are the so-called prior (or unconditional) means of X^o and X^f, respectively, and $K_{X^oX^o}$, $K_{X^oX^f}$, and $K_{X^fX^f}$ are the prior covariance matrices between X^o and X^o, X^o and X^f, and X^f and X^f, respectively. The covariance matrix $K_{X^oX^f}$ has the block structure

$$K_{X^oX^f} = \begin{bmatrix} K_{x_1,x_1}(t^o, t^f) & \cdots & K_{x_1,x_N}(t^o, t^f) \\ \vdots & \ddots & \vdots \\ K_{x_N,x_1}(t^o, t^f) & \cdots & K_{x_N,x_N}(t^o, t^f) \end{bmatrix}, \tag{2}$$

where each component $K_{x_i,x_j}(t^o, t^f)$ is given by

$$K_{x_i,x_j}(t^a, t^b) = \begin{bmatrix} \langle x_i(t_1^a) x_j(t_1^b) \rangle & \cdots & \langle x_i(t_1^a) x_j(t_{N_{t^b}}^b) \rangle \\ \vdots & \ddots & \vdots \\ \langle x_i(t_{N_{t^a}}^a) x_j(t_1^b) \rangle & \cdots & \langle x_i(t_{N_{t^a}}^a) x_j(t_{N_{t^b}}^b) \rangle \end{bmatrix}, \tag{3}$$

with $a = o$ and $b = f$, and where $\langle \cdot \rangle$ denotes the expectation operator. The covariance matrix $K_{X^fX^f}$ has a similar structure, with blocks given by (3) with $a = f$ and $b = f$. Finally, the covariance matrix $K_{X^oX^o}$ has the structure

$$K_{X^oX^o} = \begin{bmatrix} K_{x_1,x_1}(t^o, t^o) & \cdots & K_{x_1,x_N}(t^o, t^o) \\ \vdots & \ddots & \vdots \\ K_{x_N,x_1}(t^o, t^o) & \cdots & K_{x_N,x_N}(t^o, t^o) \end{bmatrix} + \sigma_n^2 I, \tag{4}$$

where I denotes the $(N_{t^o}N) \times (N_{t^o}N)$ identity matrix, and σ_n is again the standard deviation of the observation noise. The addition of the term $\sigma_n^2 I$ accounts for observation noise.

Given the state observations, the prior mean vectors, and the covariance matrices, the conditional (or posterior) estimate of the forecast vector X^f is given by

$$\hat{X}^f = \bar{X}^f + K_{X^oX^f}^T K_{X^oX^o}^{-1} \left(X^o - \bar{X}^o \right), \tag{5}$$

and its posterior covariance is given by

$$\hat{K}_{X^fX^f} = K_{X^fX^f} - K_{X^oX^f}^T K_{X^oX^o}^{-1} K_{X^oX^f}. \tag{6}$$

The posterior covariance provides a measure of uncertainty or credibility for the forecast of (5).

We now consider the forecasting of the unobserved states $y_i(t)$. Our goal is to estimate these states over the observation window $[0, T^o]$ and to forecast them over the forecast window $(T^o, T^f]$. For simplicity, we consider forecasting unobserved states at the discrete times t_j^f, $j = 1, \dots, N_{t^f}$ over the time window $(T^o, T^f]$, which we perform again using multivariate GPR. For this purpose, we introduce the vector

$$Y^f = [y_1(t_1^f), y_1(t_2^f), \dots, y_1(t_{N_{t^f}}^f), \dots, y_N(t_1^f), y_N(t_2^f), \dots, y_N(t_{N_{t^f}}^f)]^\top$$

and the random vector $[(X^o)^\top, (Y^f)^\top]^\top$ with the distribution

$$\begin{bmatrix} X^o \\ Y^f \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{X}^o \\ \bar{Y}^f \end{bmatrix}, \begin{bmatrix} K_{X^oX^o} & K_{X^oY^f} \\ K_{X^oY^f}^T & K_{Y^fY^f} \end{bmatrix} \right), \tag{7}$$

where $\bar{X}^o$ and $\bar{Y}^f$ are the prior mean vectors of X^o and Y^f, respectively, and $K_{X^oX^o}$, $K_{X^oY^f}$, and $K_{Y^fY^f}$ are the prior covariance matrices between X^o and X^o, X^o and Y^f, and Y^f and Y^f, respectively. Here, $K_{X^oX^o}$ is given by (4). The remaining covariances, $K_{X^oY^f}$ and $K_{Y^fY^f}$, have the structure of (2) but with blocks $K_{x_i,y_j}(t^o, t^f)$ and $K_{y_i,y_j}(t^f, t^f)$ given by (3). The posterior estimate and covariance of the forecast vector Y^f are then given by

$$\begin{aligned} \hat{Y}^f &= \bar{Y}^f + K_{X^oY^f}^T K_{X^oX^o}^{-1} \left(X^o - \bar{X}^o\right), \\ \hat{K}_{Y^fY^f} &= K_{Y^fY^f} - K_{X^oY^f}^T K_{X^oX^o}^{-1} K_{X^oY^f}. \end{aligned}$$

Finally, we note that multivariate GPR can be used to estimate "missing" observations in incomplete time series, which is a standard regression or interpolation task. This can be accomplished by adding the time of the missing observation $t_j^e \in [0, T^o]$ $(j = 1, \ldots, N_{t^e})$ to the vector of forecast values X^f defined above, where N_{t^e} is the number of missing observation instances to be estimated.

2.2. Prior statistics for the PhI-GPR method

Selecting prior statistics (prior mean and covariance) for the multivariate GP model (1) is one of the main challenges in GPR. In standard data-driven GPR, prior statistics are often selected from parametric models by maximizing the so-called marginal likelihood of the observations X^o (Williams & Rasmussen, 2006). For the case of forecasting observed states, this approach consists of assuming parametric models for $\bar{X}^o$, $\bar{X}^f$, and for the block components $K_{x_i,x_j}(t^a, t^b)$ of $K_{X^oX^o}$, $K_{X^oX^f}$ and $K_{X^fX^f}$, with hyperparameters γ. A point estimate of these hyperparameters, γ^*, is computed as the value that maximizes the marginal likelihood of the observations, that is,

$$\begin{aligned} \gamma^* = \arg\max_{\gamma} &-\frac{1}{2}(X^o - \bar{X}^o)^T K_{X^oX^o}^{-1}(X^o - \bar{X}^o) \\ &- \frac{1}{2}\log\det K_{X^oX^o} - \frac{N_{t^o}}{2}\log 2\pi, \end{aligned} \tag{8}$$

while assuring that the covariance of the joint process $[(X^o)^\top, (X^f)^\top]^\top$ is positive definite. For a review of prior models for multivariate GPR, see Genton and Kleiber (2015).

This data-driven GPR approach cannot be employed when no observations are available for a subset of states to be forecasted. This is because the marginal likelihood of (8) does not include block terms of the form $K_{x_iy_j}(\cdot,\cdot)$ or $K_{y_iy_j}(\cdot,\cdot)$. Therefore, marginal likelihood maximization cannot estimate the covariance of unobserved states and the cross-covariance between observed and unobserved states.

To address this challenge, in PhI-GPR we assume that X^o are the observations of the stochastic processes $x_i(t;\psi)$ $(i = 1, \ldots, N)$ and $y_i(t;\psi)$ $(i = 1, \ldots, M)$ (ψ is a coordinate in the outcome space) that are governed by a known stochastic model (such as stochastic difference equations, stochastic differential equations, etc.).

In PhI-GPR, we employ this stochastic model to compute the prior statistics for multivariate GPR via simple MC simulation. For the multivariate GP models (1) and (7), the MC estimates of $\bar{X}^o$, $\bar{X}^f$, and $\bar{Y}^f$ are given by

$$\begin{aligned} \bar{X}^o &= [\bar{x}_1(t_1^o), \ldots, \bar{x}_1(t_{N_{t^o}}^o), \ldots, \bar{x}_N(t_1^o), \ldots, \bar{x}_N(t_{N_{t^o}}^o)]^\top, \\ \bar{X}^f &= [\bar{x}_1(t_1^f), \ldots, \bar{x}_1(t_{N_{t^f}}^f), \ldots, \bar{x}_N(t_1^f), \ldots, \bar{x}_N(t_{N_{t^f}}^f)]^\top, \\ \bar{Y}^f &= [\bar{y}_1(t_1^f), \ldots, \bar{y}_1(t_{N_{t^f}}^f), \ldots, \bar{y}_N(t_1^f), \ldots, \bar{y}_N(t_{N_{t^f}}^f)]^\top, \end{aligned}$$

with each component given by the MC estimate

$$\bar{\alpha}(t) = \frac{1}{N_{\mathrm{MC}}} \sum_{n=1}^{N_{\mathrm{MC}}} \alpha(t;\psi^{(n)}),$$

where $\alpha(t,\psi^{(n)})$ denotes the nth simulated realization of the state α and N_{MC} is the number of random simulations of the stochastic model. Similarly, the components of the covariance matrices in the multivariate GP models (1) and (7) are estimated using the MC estimate

$$\begin{aligned} \langle \alpha(t)\beta(\tau)\rangle = &\frac{1}{N_{\mathrm{MC}} - 1} \\ &\times \sum_{n=1}^{N_{\mathrm{MC}}} \left\{ [\alpha(t;\psi^{(n)}) - \bar{\alpha}(t)][\beta(\tau;\psi^{(n)}) - \bar{\beta}(\tau)] \right\}. \end{aligned} \tag{9}$$

where $\alpha(t)$ and $\beta(\tau)$ are any two states of the system at times t and τ $((t,\tau) \in [0, T_f])$, respectively.

Suppose the stochastic model is biased with respect to data (i.e., there is a bias in the MC estimate of $\bar{X}^o$ and the data X^0). In that case, this bias can be removed by modeling the difference $\bar{X}^o$ and X^0 with a data-driven GPR model (Yang, Barajas-Solano, Tartakovsky, & Tartakovsky, 2019). In this work, we assume that the stochastic model is correct (unbiased) by considering the synthetic case where data is generated as a realization of the solution of the stochastic model, as described in the following section.

3. Power grid model

We consider a power transmission network with power generators modeled as classical generators driven by wind mechanical power. We assume that the dynamics of the system can be fully described by swing equations for each generator in the network, together with a constant impedance model for the loads (Nishikawa & Motter, 2015). Furthermore, we assume that wind mechanical power is not known deterministically but follows a Langevin equation.

In this work, we consider the dynamics of the network shown in Fig. 1, consisting of three classical generators driven by wind mechanical power and one load, described by the equations (Wang et al., 2015)

$$\dot{\theta}_k = \omega_B(\omega_k - \omega_s), \quad k = 1, 2, 3, \tag{10}$$

$$2H_k\dot{\omega}_k = -D_k(\omega_k - \omega_s) - P_k^e(\boldsymbol{\theta}) + P_k^m, \tag{11}$$

$$P_k^e(\theta) = \sum_{i=1}^{N} E_k E_i \left[G_{ki}\cos(\theta_k - \theta_i) + B_{ki}\sin(\theta_k - \theta_i)\right]. \tag{12}$$

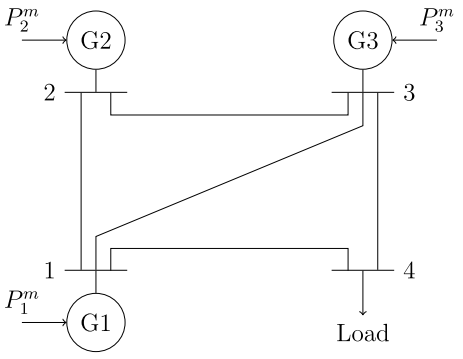

Fig. 1. Schematic of a power system composed of three synchronous generators and four buses (3 generator buses and 1 load bus). Wind mechanical power driving synchronous generators indicated by P_k^m, $k \in [1, 3]$.

Here, ω_k and θ_k are the angular velocity [rad s^{-1}] and angle [rad] of the kth generator, H_k [s] and D_k [p.u.] are the generators' inertia and damping constants, ω_B [rad s^{-1}] is the base speed, ω_s [rad s^{-1}] is the synchronization speed, E_k is the phasor internal electromotive force (emf) of the kth generator [p.u.], and G_{ki} [p.u.], B_{ki} [p.u.], and $k, i \in [1, 3]$ are the transfer conductances and susceptances, respectively. Finally, P_k^e [p.u.] and P_k^m [p.u.] are the active generated power and wind mechanical power injection of the kth generator, respectively.

Unless mentioned otherwise, we assume that we have frequent (every 0.05 s) measurements of ω_k and θ_k, and that neither measurements nor an accurate deterministic model of P_k^m are available. Given the available observations, we are interested in forecasting the grid's dynamics subject to the initial conditions $\theta_k(0) = \theta_{k,0}$ and $\omega_k(0) = \omega_{k,0}$, $k \in [1, 3]$. Since $P_k^m(t)$ is an unknown function of time, without additional data or assumption about the wind power, Eqs. (10) to (12) cannot be used to forecast the power grid states.

In PhI-GPR, we treat $P_k^m(t)$ as a stochastic process, which turns Eqs. (10)–(12) into SDEs. Then, we use these stochastic equations to estimate the mean and covariances of the power grid states and wind mechanical power injections. These means and covariances are employed as the prior statistics in the GPR Eqs. (5) and (6).

Due to the stochastic nature of wind, we model $P_k^m(t)$ as the random process (Rosenthal, Tartakovsky, & Huang, 2018)

$$P_k^m(t) = \overline{P}_k^m(t) + P_k'^m(t), \tag{13}$$

with the mean $\overline{P}_k^m(t) > 0$ and zero-mean Gaussian fluctuations $P_k'^m(t)$ with the covariance

$$\left\langle P_k'^m(t) P_k'^m(s) \right\rangle = \sigma_k^2 \exp\left(-\frac{|t-s|}{\lambda_k}\right), \tag{14}$$

$$\left\langle P_k'^m(t) P_l'^m(s) \right\rangle = 0 \quad k \neq l, \tag{15}$$

where σ_k^2 and λ_k are the variance and correlation time of the fluctuations, respectively. We assume that the wind prior mean and variance are known (e.g., from meteorological observations). Following Rosenthal et al. (2018), we model the fluctuations $P_k'^m(t)$ as the Ornstein–Uhlenbeck (O-U) process (Arnold, 1974)

$$dP_k'^m = -\frac{1}{\lambda_k} P_k'^m dt + \sqrt{\frac{2}{\lambda_k}} \sigma_k dW_k, \tag{16}$$

subject to the initial condition

$$P_k'^m(0) = P_{k,0}'^m, \tag{17}$$

where W_k is the standard Wiener process, and $P_{k,0}'^m$ is the random initial condition drawn from the stationary distribution of $P_k'^m$. We employ a second-order strong Runge–Kutta scheme (Milshtein & Tret'yakov, 1994) for integrating the system of Eqs. (10)–(12) and (16). This scheme, including all parameters and the time step, is described in Appendix.

4. Numerical experiments

We consider the 3-generator system shown in Fig. 1, with generators 1 and 2 powered by wind and generator 3 by traditional power with known constant mechanical power P_3^m. To compute covariances in the PhI-GPR forecast model, we treat P_1^m and P_2^m as random processes. The 3-generator system parameters used in our simulations are shown in Table A.7.

We compute 10^4 MC realizations over the simulation domain $[0, 25]$ s with time step $h = 0.0025$ s. The initial conditions for angular velocities are set to $\omega_{1,0} = \omega_{2,0} = \omega_{3,0} = 0$. A power grid described by Eqs. (10) to (12) has several stable equilibrium points, each corresponding to possible safe operating conditions (e.g., the system remains stable against sufficiently small fluctuations in wind) and each with its basin of attraction (Hasegawa & Ueda, 1999). We have numerically determined that the system considered here with the initial condition for θ_i in the interval [0, 0.5] operates in the same basin of attraction. Therefore, we set the initial conditions for θ_i to $\theta_{1,0} = 0.0431$, $\theta_{2,0} = 0.4584$, $\theta_{3,0} = 0.2372$ by drawing them randomly from the uniform distribution with the 0 and 0.5 bounds. Due to the angular indeterminacy of the classical model with constant impedance, we only present phase angles and angular velocities relative to those of the first generator.

The mean and standard deviation of $\theta_2(t) - \theta_1(t)$, $\theta_3(t) - \theta_1(t)$, $\omega_2(t) - \omega_1(t)$, and $\omega_3(t) - \omega_1(t)$, computed from MC simulation, are shown in Fig. 2, which demonstrates that these states converge to a statistical steady state after approximately 15 s. In this work, we focus on the forecast of states for times less than 15 s, i.e., the forecast of states with non-stationary statistics. Such predictions are especially challenging for the standard GPR, which often relies on the assumption of stationary statistics for covariance estimation. Specifically, we set T_0 to 8.3375 s and T_f to 12.5 s.

In Section 4.1, we assume that measurements are only available for θ_k ($k = 1, 2, 3$) and in Section 4.2 we assume that only ω_k ($k = 1, 2, 3$) measurements are available. In both cases, we forecast all states $\theta_2(t) - \theta_1(t)$, $\theta_3(t) - \theta_1(t)$, $\omega_2(t) - \omega_1(t)$, $\omega_3(t) - \omega_1(t)$, $P_1'^m(t)$ and $P_2'^m(t)$. In Section 4.3,

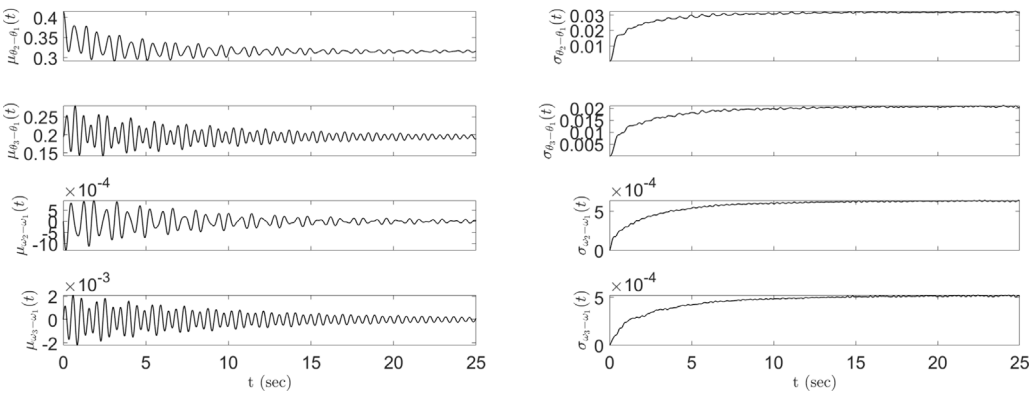

Fig. 2. Mean (left) and standard deviation (right) of $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$, $\omega_2(t)-\omega_1(t)$, $\omega_3(t)-\omega_1(t)$.

we compare the physics-informed GPR and the standard data-driven GPR when measurements of both θ_k and ω_k ($k = 1, 2, 3$) are available. In Section 5, we compare the PhI-GPR and ARIMA forecasts of θ_k and ω_k using both noiseless and noisy measurements of θ_k and ω_k. Finally, in Section 6 we compare the PhI-GPR and EnKF forecasts.

4.1. PhI-GPR state estimation and forecasting using measurements of θ_k ($k = 1, 2, 3$)

In this case, we assume that measurements of θ_k ($k = 1, 2, 3$) are available for $t < 8.3375$ s every 0.05 s. Our goal is to forecast $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$ for $t > 8.3375$ s and $\omega_2(t)-\omega_1(t)$, $\omega_3(t)-\omega_1(t)$, $P_1'^m(t)$ and $P_2'^m(t)$ for the entire time interval $t \in [0, 12.5]$ s. The measurements are taken from the ground truth solution for $t < 8.3375$ s. The ground truth solution is also used to validate the PhI-GPR forecast of $\theta_2(t)-\theta_1(t)$ and $\theta_3(t)-\theta_1(t)$ for $t > 8.3375$ s and $\omega_2(t)-\omega_1(t)$, $\omega_3(t)-\omega_1(t)$, $P_1'^m(t)$, and $P_2'^m(t)$ for $t \in [0, 12.5]$ s. In the PhI-GPR forecast, the prior mean and covariance of states are computed as described in Section 2.2 using the stochastic model presented in Section 3. Of the 10^4 realizations of the stochastic model, one is employed as the ground truth, while the remaining $10^4 - 1$ realizations are used to compute the prior mean and covariances.

In Table 1, we give the log predictive probabilities for the states $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$, $\omega_2(t)-\omega_1(t)$, $\omega_3(t)-\omega_1(t)$, $P_1'^m(t)$ and $P_2'^m(t)$. The log predictive probability (lpp) is a quantitative measure of the accuracy of predictions from statistical models, which corresponds to the sum of the pointwise log probabilities of reference values being observed given the statistical model (Williams & Rasmussen, 2006). For a certain estimated or forecasted state $\alpha(t)$, the lpp is given by

$$\text{lpp} = -\sum_{k=1}^{N^{f/e}} \left\{ \frac{[\mu^{f/e}(t_k) - \alpha(t_k)\,]^2}{2\,[\sigma^{f/e}(t_k)\,]^2} + \frac{1}{2}\log 2\pi\,[\sigma^{f/e}(t_k)\,]^2 \right\},$$

where $N^{f/e}$ denotes the number of forecast or estimation times, $\mu^{f/e}(t_k)$ and $\sigma^{f/e}(t_k)$ are the posterior mean and standard deviation of the forecast or estimation at time t_k, respectively, and $\alpha(t_k)$ is the reference value at time t_k. The larger the lpp, the more accurate is the model estimation or forecast.

Table 1
Log predictive probabilities for $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$, $\omega_2(t)-\omega_1(t)$, $\omega_3(t)-\omega_1(t)$, $P_1'^m(t)$, and $P_2'^m(t)$ when measurements of θ_k ($k = 1, 2, 3$) are available for $t < T_0 = 8.3375$ s every 0.05 s.

k	$\theta_k-\theta_1$		$\omega_k-\omega_1$		$P_k'^m$	
	$t<T_0$	$t>T_0$	$t<T_0$	$t>T_0$	$t<T_0$	$t>T_0$
1	NA	NA	NA	NA	663.302	136.435
2	NA	194.642	1359.04	550.43	690.176	124.035
3	NA	247.571	1359.29	573.736	NA	NA

Fig. 3 shows the states estimation and forecast results, together with the associated uncertainties (two posterior standard deviations corresponding to 95% credibility intervals). The forecast of $\theta_2(t)-\theta_1(t)$ and $\omega_2(t)-\omega_1(t)$ for $t \in [8.3, 12.5]$ s is satisfactory, with the ground truth staying within two standard deviations of the forecast. For the first two seconds, the forecasts closely match the ground truth. The forecast of $P_1'^m$ is less satisfactory, but the ground truth still mostly stays within two standard deviations of the forecast. The challenges with forecasting $P_k'^m$ are to be expected due to its stochastic nature. The estimation of the unobserved states $\omega_2(t)-\omega_1(t)$ and $P_1'^m$ for $t < 8.3$ s is very accurate, indicating that there is a very strong correlation between the observed and unobserved states. Another outcome of the strong correlation is that the uncertainty (posterior standard deviation) for parameter estimation (for $t < 8.3$ s) is much smaller than that for forecasting (for $t > 8.3$ s). The forecast and estimation performance for the remaining states shows similar behavior and is summarized in Table 1.

4.2. PhI-GPR forecasting using measurements of ω_k ($k = 1, 2, 3$)

Here, we assume that measurements of ω_k ($k = 1, 2, 3$) are available for $t < 8.3375$ s every 0.05 s. Our goal is to predict $\omega_2(t)-\omega_1(t)$, $\omega_3(t)-\omega_1(t)$ for $t > 8.3375$ s and $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$, $P_1'^m(t)$ and $P_2'^m(t)$ for the entire time interval $t \in [0, 12.5]$ s. As before, the ground truth solution provides measurements of observed states and is used to validate forecasted and estimated states. The

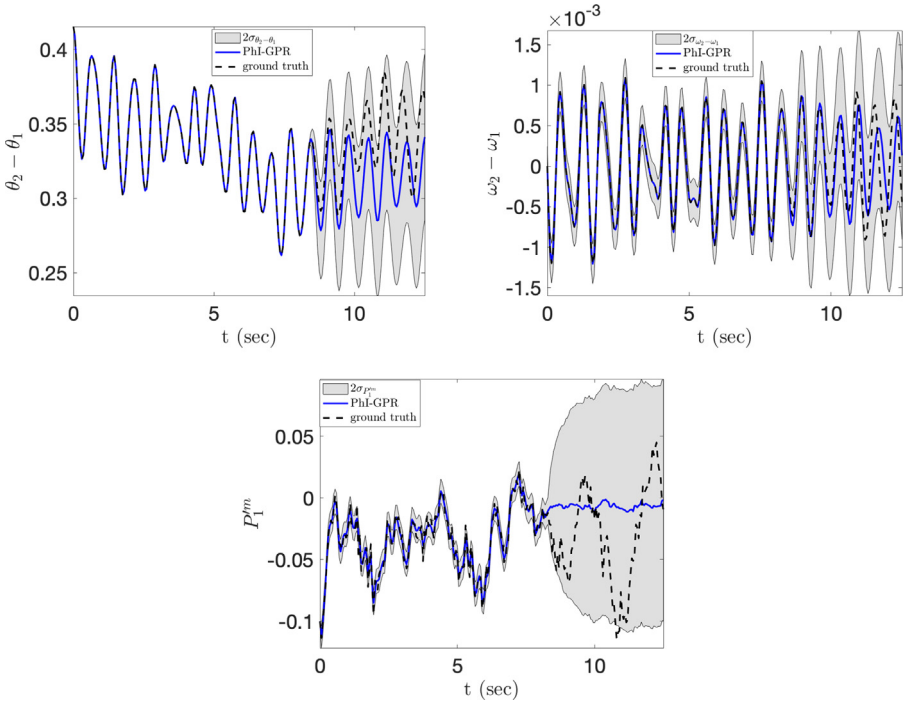

Fig. 3. Forecast of $\theta_2(t) - \theta_1(t)$ (top left), $\omega_2(t) - \omega_1(t)$ (top right), and $P_1'^m(t)$ (bottom) when measurements of θ_k $(k = 1, 2, 3)$ are available for $t < 8.3375$ s every 0.05 s. Grey areas indicate the 95% credibility intervals.

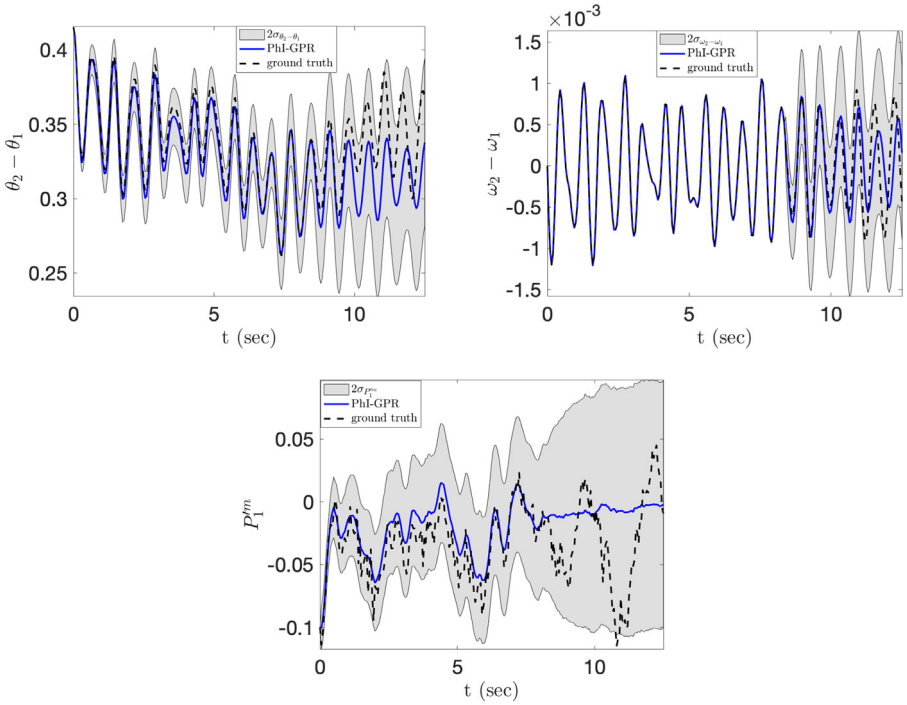

Fig. 4. Forecast of $\theta_2(t) - \theta_1(t)$ (top left), $\omega_2(t) - \omega_1(t)$ (top right), and $P_1'^m(t)$ (bottom) when measurements of ω_k $(k = 1, 2, 3)$ are available for $t < 8.3375$ s every 0.05 s. Grey areas indicate the 95% credibility intervals.

log predictive probabilities for $\theta_2(t) - \theta_1(t)$, $\theta_3(t) - \theta_1(t)$, $\omega_2(t) - \omega_1(t)$, $\omega_3(t) - \omega_1(t)$, $P_1'^m(t)$ and $P_2'^m(t)$ are given in Table 2.

Fig. 4 shows the states estimation and forecast results, together with the associated uncertainties (two posterior standard deviations, corresponding to 95% credibility

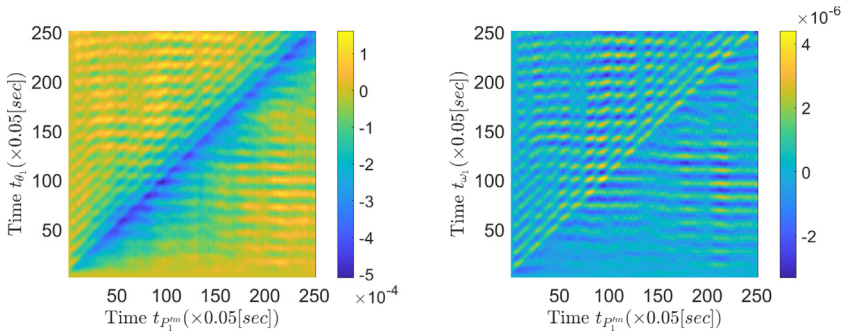

Fig. 5. Covariance matrices $K_{P'^m,\theta}$ (left) and $K_{P'^m,\omega}$ (right) for the case shown in Fig. 4.

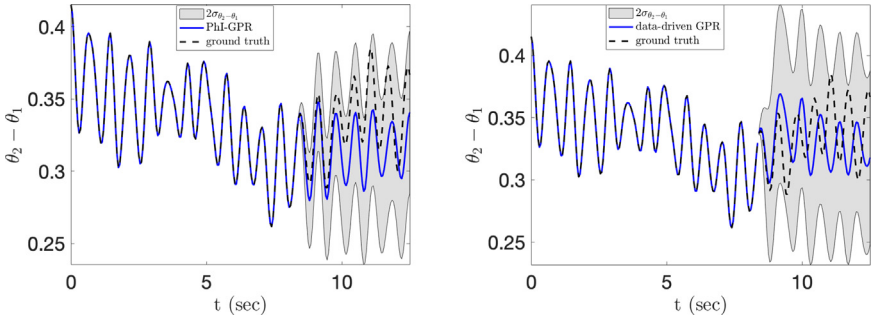

Fig. 6. PhI-GPR (left) and data-driven GPR (right) forecast of $\theta_2(t) - \theta_1(t)$ with θ_k and ω_k $(k = 1, 2, 3)$ measurements available for $t < 8.3375$ s every 0.05 s.

intervals). It can be seen that the estimation and forecasting of $\theta_2 - \theta_1$ and $\omega_2 - \omega_1$ for $t < 8.3375$ s is very accurate, with the ground truth laying between two posterior standard deviations from the predictions. It can also be seen that the estimation of $P_1'^m(t)$ for $t < 8.3375$ s, based on ω_k observations, is not as good as the one based on θ_k (Section 4.1). This is because $P_1'^m(t)$ is more strongly correlated to θ_k than to ω_k. To illustrate the correlation between $P_1'^m(t)$ and θ_1 and ω_1, in Fig. 5 we present the covariance matrices $K_{P_1'^m,\theta_1}$ and $K_{P_1'^m,\omega_1}$. The $K_{P_1'^m,\theta_1}$ matrix has negative diagonal terms and the absolute values of these terms are much larger than the absolute values of the off-diagonal terms that indicates a strong (negative) correlation between $P_1'^m$ and θ_1. On the other hand, we do not see any significant correlation between $P_1'^m$ and ω_1 as the diagonal and off-diagonal terms in the corresponding matrix are of the same order.

The accuracy of $P_1'^m(t)$ forecasting for $t > 8.3375$ s, based on ω_k measurements is approximately the same as the one based on θ_k measurements, with the ground truth being within two posterior standard deviations of the PhI-GPR prediction.

4.3. Comparison between physics-informed and data-driven GPR methods

In this section we present a comparison between PhI-GPR and standard data-driven GPR for a case where the measurements of both θ_k and ω_k $(k = 1, 2, 3)$ are available for $t < 8.3375$ s every 0.05 s. Here, our goal is to predict $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$, $\omega_2(t)-\omega_1(t)$ and $\omega_3(t)-\omega_1(t)$ for $t > 8.3375$ s. For standard data-driven GPR, we employ covariance models given by a combination of a squared exponential function, rational quadratic function, periodic function, and Kronecker delta function as follows:

$$K_{\alpha\alpha}(t,\tau) = \gamma_1^2 \exp\left[-\frac{(t-\tau)^2}{2\gamma_2^2}\right] + \gamma_3^2\left[1+\frac{(t-\tau)^2}{2\gamma_4\gamma_5^2}\right]^{-\gamma_4} + \gamma_6^2 \exp\left\{-\frac{2\sin^2\left[\frac{\pi}{24}(t-\tau)\right]}{\gamma_7^2}\right\} + \gamma_8^2\delta(t-\tau).$$

The means $\overline{\omega}^o(t) = \overline{\omega}^f(t)$ and $\overline{\theta}^o(t) = \overline{\theta}^f(t)$ and parameters γ_i $(i = 1, \ldots, 8)$ are determined by minimizing the negative marginal likelihood of the observed data (Williams & Rasmussen, 2006), as described in Section 2.2.

The log predictive probabilities for the forecast of $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$, $\omega_2(t)-\omega_1(t)$ and $\omega_3(t)-\omega_1(t)$, computed via PhI-GPR and data-driven GPR, are summarized in Table 3. Fig. 6 presents the forecast of $\theta_2(t)-\theta_1(t)$

Table 2
Log predictive probabilities for $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$, $\omega_2(t)-\omega_1(t)$, $\omega_3(t)-\omega_1(t)$, $P_1'^m(t)$, and $P_2'^m(t)$ when measurements of ω_k $(k = 1, 2, 3)$ are available for $t < T_o = 8.3375$ s every 0.05 s.

k	$\theta_k - \theta_1$		$\omega_k - \omega_1$		$P_k'^m$	
	$t < T_0$	$t > T_0$	$t < T_0$	$t > T_0$	$t < T_0$	$t > T_0$
1	NA	NA	NA	NA	463.01	140.327
2	624.416	183.523	NA	549.362	489.118	115.266
3	735.93	240.512	NA	570.322	NA	NA

Table 3
Log predictive probabilities for the forecast of $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$, $\omega_2(t)-\omega_1(t)$, and $\omega_3(t)-\omega_1(t)$ for $t < T_0 = 8.3375$ s, computed using both PhI-GPR and data-driven GPR when measurements of θ_k and ω_k $(k = 1, 2, 3)$ are available for $t < T_0 = 8.3375$ s every 0.05 s.

k	$\theta_k - \theta_1$		$\omega_k - \omega_1$	
	PhI-GPR	GPR	PhI-GPR	GPR
2	198.858	211.542	551.258	555.884
3	246.031	261.263	565.729	515.667

using both PhI-GPR and data-driven GPR, respectively. The log predictive probabilities of the data-driven GPR forecast are slightly larger (less than 10%) than those of the PhI-GPR forecast. However, Fig. 6 clearly shows that the PhI-GPR forecast is significantly more accurate than the data-driven GPR forecast for the first 2 s. The PhI-GPR forecast closely matches the ground truth for approximately 2 s, after which the ground truth stays within two posterior standard deviations from the forecast. The data-driven GPR forecast deviates from the exact forecast after approximately 1 s but stays within two standard deviations of the ground truth. We also note that the data-driven GPR results in a less certain forecast, i.e., the data-driven GPR produces a larger standard deviation of forecasted variables than PhI-GPR.

5. Comparison with the ARIMA method

In this section, we compare PhI-GPR against the univariate ARIMA method. Unlike PhI-GPR, the univariate ARIMA only allows for forecasting observed variables, e.g., θ_k measurements are needed for forecasting θ_k. As in GPR, the ARIMA forecast of θ_k^f and ω_k^f is given as a linear combination of the N_o measurements of θ_k^o and ω_k^o, respectively. Using the notation introduced in Section 2.1, the ARIMA forecast can be expressed as

$$\omega_k^f(t_{N_o+1}) - \alpha_{k,N_o}\omega_k^o(t_{N_o}) - \cdots - \alpha_{k,p_1}\omega_k^o(t_{N_o-p_1}) = e_{k,N_o+1} + \beta_{k,N_o}e_{k,N_o} + \cdots + \beta_{k,N_o-q_1}e_{k,N_o-q_1} \quad (18)$$

$$\theta_k^f(t_{N_o+1}) - \gamma_{k,N_o}\theta_k^o(t_{N_o}) - \cdots - \gamma_{k,p_2}\theta_k^o(t_{N_o-p_2}) = \epsilon_{k,N_o+1} + \lambda_{k,N_o}\epsilon_{k,N_o} + \cdots + \lambda_{k,N_o-q_2}\epsilon_{k,N_o-q_2} \quad (19)$$

where $\omega_k^f(t_{N_o+1})$ and $\theta_k^f(t_{N_o+1})$ are the forecasted values of ω_k and θ_k, and p_1, p_2, q_1, q_2, $\alpha_{k,i}$, $\beta_{k,i}$, $\gamma_{k,i}$, $\lambda_{k,i}$, $e_{k,i}$ and $\epsilon_{k,i}$ are the parameters of ARIMA. Non-seasonal ARIMA models are generally denoted as ARIMA(p, d, q), where d is the degree of differencing used to remove a trend in data. We assume that the data does not have a trend and set d to zero. The order parameters are selected using Akaike's information criterion (Hyndman & Athanasopoulos, 2018).

As before, we assume that the measurements of θ_k, ω_k $(k = 1, 2, 3)$ are available for $t < 8.3375$ s and our objective is to forecast these states for $t > 8.3375$ s. We test the performance of PhI-GPR and ARIMA for three cases when the data are collected every 0.05 s, 0.125 s, and 0.25 s, respectively. The ARIMA model for forecasting is ARIMA$(15, 0, 1)$ for all three cases.

Fig. 7 compares ARIMA and PhI-GPR performance when data is available every $\delta t = 0.05$, 0.125, and 0.25 s. We can see that the ARIMA forecast is sensitive to δt, while PhI-GPR prediction is practically independent of δt for the considered values of δt. We also note that the accuracy of the ARIMA forecast increases with decreasing δt. For the smallest tested δt, we find that the accuracy of PhI-GPR is higher than ARIMA for the first two seconds and then comparable with ARIMA.

The observed superior performance of the Phi-GPR method in the considered examples is attributed to the following factors. The accuracy of the GPR method depends on the accuracy of the estimated covariance kernel and δt. The dependence on δt is weaker for δt smaller than the correlation time of the system λ. For a Markov process (as one considered in this section), the forecasting is only dependent on the last measurement of the state. In PhI-GPR, the covariance is estimated from the stochastic swing equations. Therefore, the accuracy of the covariance estimation does not depend on δt. For state estimation (i.e., estimating states at time $t < 8.3375$ s), we observe a very small dependence of GPR results on δt because the correlation time of the system $\lambda = 1.8$ s (in our case, the correlation time of the system is defined by the correlation time of $P^{'m}$) is larger than the considered values of δt. We also note that the ARIMA's predictions start deteriorating even for $\delta t \ll \lambda$.

Finally, we consider the effect of measurement noise on the PhI-GPR and ARIMA forecasting. We study two cases where we add 1% and 5% noise to ω_k and θ_k measurements. We assume that (noisy) measurements of θ_k and ω_k are available for $t < 8.3375$ s every 0.25 s. The ARIMA models for $\theta_2(t) - \theta_1(t)$ and $\omega_2(t) - \omega_1(t)$ forecasting are both ARIMA$(15, 0, 1)$. The PhI-GPR and ARIMA forecasts of $\theta_2(t) - \theta_1(t)$ and $\omega_2(t) - \omega_1(t)$ based on data with 1% and 5% noise are shown in Figs. 8 and 9, respectively. These figures also show the PhI-GPR states estimation every 0.025 s for $t < 8.3375$ s. We can see that the PhI-GPR forecast is more accurate than ARIMA, especially for the first two seconds. Also, the PhI-GPR forecast is considerably less sensitive to the measurement noise than the ARIMA forecast, which significantly worsens as the noise increases. The PhI-GPR estimation of $\theta_2(t)-\theta_1(t)$ and $\omega_2(t) - \omega_1(t)$ for $t < 8.3375$ s is in good agreement with the ground truth for both noise levels.

6. Comparison with the EnKF method

In the EnKF method (Gillijns et al., 2006; Takeda et al., 2016) and other ensemble forecasting approaches, the forecast is given as a distribution (or its mean and standard deviation) computed from an ensemble of simulations with the initial condition computed from the last measurement of the system of interest, with unknown parameters (driving forces, boundary conditions, etc.) treated as random variables. With application to the forecasting of the 3-generator power grid for time $t > 8.3375$ given noiseless measurements at $t = 8.3375$, such an ensemble approach is realized by solving the swing equations with the initial condition given by the measurements of states and $P_i'^m(t)$ at $t = 8.3375$ and treating $P_i'^m(t)$ for $t >$

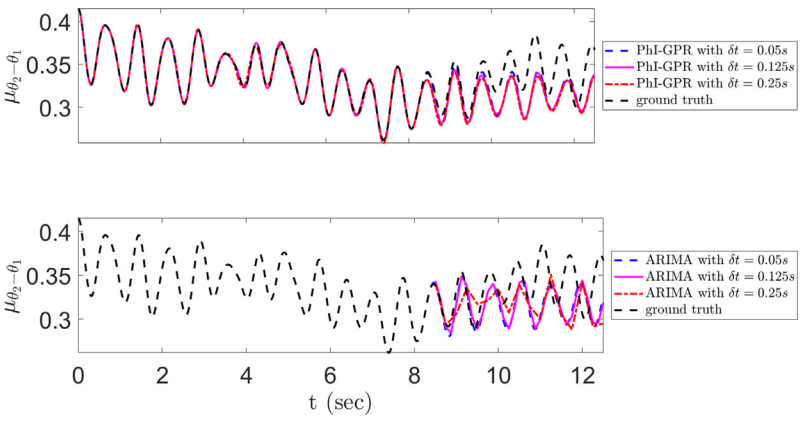

Fig. 7. Forecasting of $\theta_2(t) - \theta_1(t)$ using PhI-GPR and ARIMA when measurements of θ_k, ω_k $(k = 1, 2, 3)$ are available for $t < 8.3375$ s.

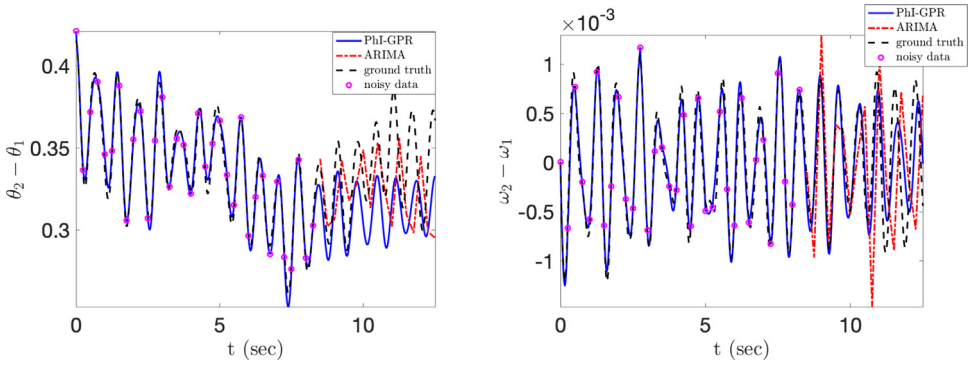

Fig. 8. Forecasting of $\theta_2(t) - \theta_1(t)$ (left) and $\omega_2(t) - \omega_1(t)$ (right) using GPR and ARIMA when measurements of θ_k, ω_k $(k = 1, 2, 3)$ are available for $t < 8.3375$ s every 0.25 s with 1% measurement noise.

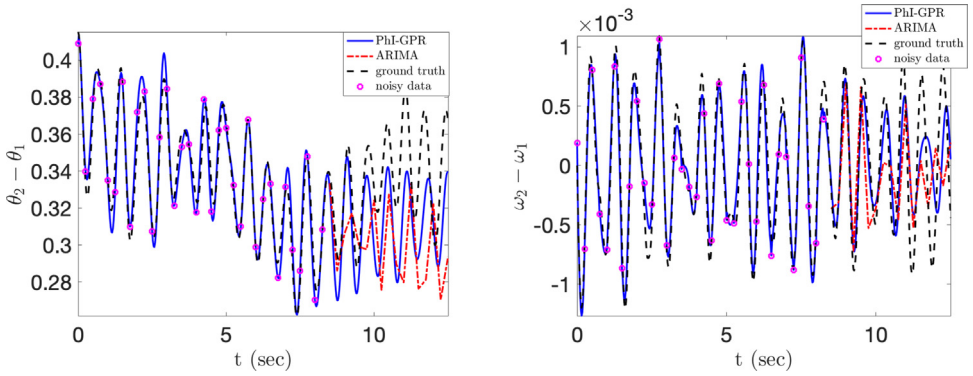

Fig. 9. Forecasting of $\theta_2(t) - \theta_1(t)$ (left) and $\omega_2(t) - \omega_1(t)$ (right) using GPR and ARIMA when measurements of θ_k, ω_k $(k = 1, 2, 3)$ are available for $t < 8.3375$ s every 0.25 s with 5% measurement noise.

8.3375 as a random process. The resulting distribution of power grid states is a distribution conditioned on the measurements of states at $t = 8.3375$. In comparison, the PhI-GPR approach gives the distribution of states conditioned on the multiple measurements of states at times $t \leq 8.3375$.

It should be noted that if the states of the power grid form a multivariate Markov process (for the considered system, the states are a Markov process if $P_i'^m(t)$ are modeled as a Markov process), then the conditional distribution of these states will depend only on the last measurement. Therefore, the ensemble forecasting described above will be optimal for a Markov process, as conditioning on historic measurements other than the last measurements would not produce additional information. However, for non-Markovian $P_1'^m(t)$, the states of the power grid will be non-Markovian. Conversely, for

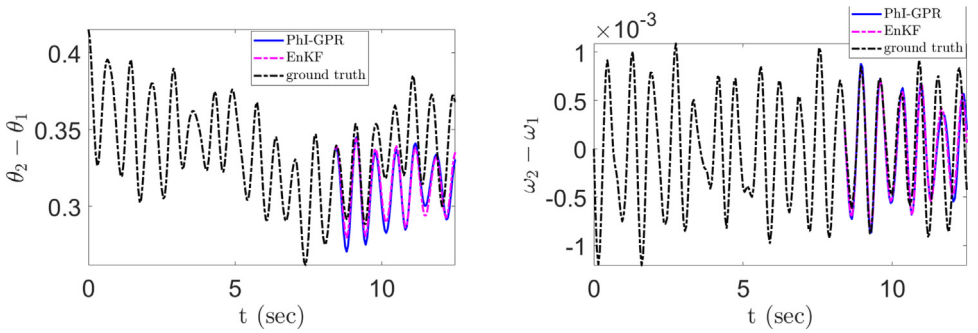

Fig. 10. Forecasting of Markovian $\theta_2(t) - \theta_1(t)$ (left) and $\omega_2(t) - \omega_1(t)$ (right), using PhI-GPR and EnKF, when noiseless measurements of θ_k and ω_k $(k = 1, 2, 3)$, $P_k'^m$ $(k = 1, 2)$ are available for $t < 8.3375$ s every 0.05 s.

non-Markovian systems, the conditional distribution depends on past measurements (at least within several correlation times range). We expect the PhI-GPR approach proposed here to be more accurate than EnKF. We also reiterate that the PhI-GPR approach (as any GPR-based approach) assumes that the considered system is Gaussian, which is not true for systems described by nonlinear ODEs. Therefore, the PhI-GPR approach only gives an approximation of the conditional distribution of the states.

In Section 3 we presented the power grid model with Markovian $P_i'^m(t)$, and in Sections 4 and 5 we presented forecasts of the power grid driven by Markovian $P_i'^m(t)$. In the remainder of this section, we compare the PhI-GPR and EnKF methods for forecasting the considered power grid driven by Markovian and non-Markovian $P_i'^m(t)$. Our results, summarized in Figs. 10 and 11 and Tables 4 and 5, show that for the power grid with Markovian $P_i'^m(t)$, the EnKF and PhI-GPR methods have a comparable performance, while PhI-GPR is more accurate than EnKF for non-Markovian $P_i'^m(t)$.

Finally, we note that both the PhI-GPR and EnKF methods require solving the power grid equation multiple times. In the considered example, the EnKF ensemble is constructed for $t \in [8.33s, 12.5s]$ while in the PhI-GPR method, the ensemble (to compute the covariances of the power grid states) is computed for $t \in [0, 12.5s]$ (that is the time interval that includes both the observation and forecast windows). While the computational cost of the PhI-GPR is higher than that of EnKF (because the time domain on which ensembles are computed is larger in PhI-GPR), the ensemble in PhI-GPR is computed independently of measurements and, therefore, can be constructed offline. As a result, the PhI-GPR forecast can be produced faster than the EnKF forecast. Updating a forecast after a new measurement is collected requires recomputing the ensemble with the latest measurement used as an initial condition. Also, the computational cost of PhI-GPR could be reduced by only considering data that are strongly correlated to the forecast values, i.e., data collected within several (two to three) correlation times from the beginning of the forecasting window. Data further away from the forecasting window should not significantly contribute to the forecast. This is because the components of the covariance matrix $K_{X^oX^f}$ in the GPR Eq. (5) corresponding to such "remote" data are much smaller than those corresponding to the data in the vicinity of the forecasting window.

Table 4
Log predictive probabilities for Markovian $\theta_2(t) - \theta_1(t)$, $\theta_3(t) - \theta_1(t)$, $\omega_2(t) - \omega_1(t)$, $\omega_3(t) - \omega_1(t)$, $P_1'^m(t)$, and $P_2'^m(t)$, using PhI-GPR and EnKF, when noiselss measurements of θ_k and ω_k $(k = 1, 2, 3)$, $P_k'^m$ $(k = 1, 2)$ are available for $t < T_0 = 8.3375$ s every 0.05 s.

Forecasted states	PhI-GPR	EnKF
$\theta_2(t) - \theta_1(t)$	181.032	194.112
$\theta_3(t) - \theta_1(t)$	240.272	175.847
$\omega_2(t) - \omega_1(t)$	548.604	555.03
$\omega_3(t) - \omega_1(t)$	568.349	579.122

Table 5
Log predictive probabilities for non-Markovian $\theta_2(t) - \theta_1(t)$, $\theta_3(t) - \theta_1(t)$, $\omega_2(t) - \omega_1(t)$, $\omega_3(t) - \omega_1(t)$, $P_1'^m(t)$, and $P_2'^m(t)$, using PhI-GPR and EnKF, when noiseless measurements of θ_k and ω_k $(k = 1, 2, 3)$, $P_k'^m$ $(k = 1, 2)$ are available for $t < T_0 = 8.3375$ s every 0.05 s.

Forecasted states	PhI-GPR	EnKF
$\theta_2(t) - \theta_1(t)$	423.09	306.863
$\theta_3(t) - \theta_1(t)$	416.776	320.461
$\omega_2(t) - \omega_1(t)$	616.979	604.035
$\omega_3(t) - \omega_1(t)$	621.11	603.621

Fig. 10 shows the forecast of Markovian $\theta_2(t) - \theta_1(t)$ and $\omega_2(t) - \omega_1(t)$, respectively, obtained with the PhI-GPR and EnKF methods. For the PhI-GPR forecast, measurements are available every 0.05 s.

Table 4 lists the log predictive probabilities of the PhI-GPR and EnKF forecasts for Markovian $\theta_2(t) - \theta_1(t)$, $\theta_3(t) - \theta_1(t)$, $\omega_2(t) - \omega_1(t)$ and $\omega_3(t) - \omega_1(t)$. For all states except $\theta_3(t) - \theta_1(t)$, the lpp of EnKF is higher than that of PhI-GPR on average by 3%. For $\theta_3(t) - \theta_1(t)$, the lpp of PhI-GPR is larger by 36%.

Fig. 11 and Table 5 compare the PhI-GPR and EnKF methods for forecasting the states of the power grid driven by non-Markovian $P_i'^m(t)$. The non-Markovian processes $P_1'^m(t)$ and $P_2'^m(t)$ are generated as realizations of the Gaussian distribution with the squared-exponential correlation function

$$\left\langle P_i'^m(t) P_i'^m(s) \right\rangle = \sigma_i^2 \exp\left(-\frac{(t-s)^2}{2\lambda_i^2} \right), \tag{20}$$

$$\left\langle P_k'^m(t) P_l'^m(s) \right\rangle = 0 \quad k \neq l \tag{21}$$

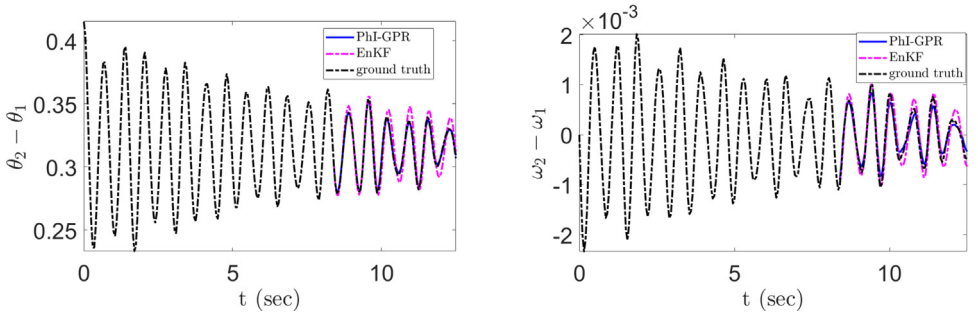

Fig. 11. Forecasting of non-Markovian $\theta_2(t)-\theta_1(t)$ (left) and $\omega_2(t)-\omega_1(t)$ (right), using PhI-GPR and EnKF, when noiseless measurements of θ_k and ω_k $(k = 1, 2, 3)$, $P_k'^m$ $(k = 1, 2)$ are available for $t < 8.3375$ s every 0.05 s.

Table 6
Statistics of log predictive probabilities for non-Markovian $\theta_2(t)-\theta_1(t)$, $\theta_3(t)-\theta_1(t)$, $\omega_2(t)-\omega_1(t)$, $\omega_3(t)-\omega_1(t)$, $P_1'^m(t)$, and $P_2'^m(t)$, using PhI-GPR and EnKF, when noiseless measurements of θ_k and ω_k $(k = 1, 2, 3)$, $P_k'^m$ $(k = 1, 2)$ are available for $t < T_0 = 8.3375$ s every 0.05 s.

Forecasted states	Mean		Standard deviation	
	PhI-GPR	EnKF	PhI-GPR	EnKF
$\theta_2 - \theta_1$	420.9528	300.1570	10.8858	22.5248
$\theta_3 - \theta_1$	409.6049	323.8245	7.0277	20.8429
$\omega_2 - \omega_1$	611.8038	601.1384	11.7366	11.9252
$\omega_3 - \omega_1$	615.2695	608.8694	12.6511	10.1212

with the same values of the variance σ_i and the correlation time λ_i as in the case of Markovian $P_i'^m(t)$. Fig. 11 shows the forecasts of $\theta_2(t) - \theta_1(t)$ and $\omega_2(t) - \omega_1(t)$, while the log predictive probabilities of the $\theta_2(t) - \theta_1(t)$, $\theta_3(t) - \theta_1(t)$, $\omega_2(t) - \omega_1(t)$, and $\omega_3(t) - \omega_1(t)$ forecasts are listed in Table 5. On average, PhI-GPR has an 18% larger lpp than EnKF does. For the phase angles, the lpp of PhI-GPR is higher than that of EnKF by more than 30%.

To verify that these conclusions are not specific to this data set, we generate nine more synthetic data sets with the initial conditions for θ_i randomly drawn from the uniform distribution with the bounds 0 and 0.5 and the rest of the initial conditions set as in the above cases. Table 6 lists the means and variances of the log predictive probabilities for the forecasted states. We can see that PhI-GPR is, on average, more accurate than the EnKF method for the forecast of all considered states.

7. Discussion and conclusions

The PhI-GPR method for short-term forecasting and state estimation of the phase angle, angular speed, and wind mechanical power of an N-generator power grid system with partial measurements is presented. The standard data-driven GPR method estimates the prior mean and covariance functions from measurements by maximizing a pseudo-likelihood function of the observations. In contrast, in PhI-GPR, we compute the prior mean and covariance from a stochastic swing equation describing the power grid dynamics, where the deterministically unknown variables (in our case, the mechanical wind power) are treated as random processes. Therefore, unlike data-driven GPR, the PhI-GPR method can be used to estimate and forecast even unobserved variables. For example, given observations of the angular velocity of generators, PhI-GPR can forecast and estimate all system states, including the angular velocity, phase angle, and mechanical power of these generators. For the considered power grid system consisting of two wind generators and one traditional generator, we find that PhI-GPR provides accurate estimation and forecast of the unobserved states for times smaller than the correlation time of the system. The PhI-GPR forecast stays within two standard deviations of the ground truth for larger times.

The PhI-GPR method resembles the EnKF method, which also uses ensemble statistics in its ability to forecast both observed and unobserved variables. However, there is a significant difference between the two methods. In the case of EnKF with noiseless observations, the ensemble statistics are conditioned on the latest measurements only, while in PhI-GPR, the ensemble is conditioned on all historical measurements. Because of this, PhI-GPR is more accurate than EnKF for forecasting the states of the considered power grid when wind mechanical power is modeled as a non-Markovian process. The two methods give similar results for Markovian wind mechanical power as the conditional distribution of Markov processes depends only on the last measurement.

In both methods, the construction of ensembles involves solving the power grid equations, which is the main computational cost of these methods. In EnKF, these calculations must be conducted online with data acquisition. On the other hand, in PhI-GPR, the ensemble calculation is performed separately from data acquisition, and the forecast can be computed efficiently online with data acquisition using GPR.

We also compare PhI-GPR and ARIMA, a standard forecasting method. For observed variables, we demonstrate that the PhI-GPR is at least as accurate as ARIMA when the time between observation δt is sufficiently small. As δt increases, the accuracy of the ARIMA forecast deteriorates, while PhI-GPR remains accurate. While both PhI-GPR and ARIMA forecasts deteriorate in the presence of measurement noises, PhI-GPR remains more accurate than ARIMA.

The accuracy of the PhI-GPR method depends on the fidelity of the physics-based model used to compute the covariance functions. In general, this is not a problem for modeling power grids since the equations describing

Table A.7
3-generator system parameters.

H_1 [s]	13.64	H_2 [s]	6.4	H_3	3.01	ω_B [rads^{-1}]	120
D_1 [p.u.]	9.6	D_2 [p.u.]	2.5	D_3 [p.u.]	1.0	ω_s [rads^{-1}]	0
E_1 [p.u.]	1.0156	E_2 [p.u.]	1.0359	E_3 [p.u.]	1.0053	λ_1 [s]	1.8
G_{11} [p.u.]	0.8815	G_{21} [p.u.]	0.3083	G_{31} [p.u.]	0.2258	λ_2 [s]	1.8
G_{12} [p.u.]	0.3083	G_{22} [p.u.]	0.4357	G_{32} [p.u.]	0.2247	σ_1 [p.u.]	0.05
G_{13} [p.u.]	0.2258	G_{23} [p.u.]	0.2247	G_{33} [p.u.]	0.2860	σ_2 [p.u.]	0.05
B_{11} [p.u.]	−3.0273	B_{21} [p.u.]	1.4904	B_{31} [p.u.]	1.2088	$\bar{P}_1^m$ [p.u.]	0.7195
B_{12} [p.u.]	1.4904	B_{22} [p.u.]	−2.7397	B_{32} [p.u.]	1.0764	$\bar{P}_2^m$ [p.u.]	1.6300
B_{13} [p.u.]	1.2088	B_{23} [p.u.]	1.0764	B_{33} [p.u.]	−2.3770	$\bar{P}_3^m$ [p.u.]	0.8500

the power grid behavior are well established. The main computational cost of the PhI-GPR is associated with computing the covariance functions, which requires solving the governing equations multiple times (for different realizations of unknown parameters). In our future work, we will investigate the use of linearized equations, which allow deriving simple (deterministic) equations for covariance functions that could be solved at a reduced cost compared to MC.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially supported by the U.S. Department of Energy (DOE) Office of Science, Office of Advanced Scientific Computing Research (ASCR) as part of the Multifaceted Mathematics for Rare, Extreme Events in Complex Energy and Environment Systems (MACSER) project. Pacific Northwest National Laboratory is operated by Battelle for the DOE under Contract DE-AC05-76RL01830. A. Tartakovsky was also supported by the NSF grant OIA 2033607.

Appendix. Simulation of the power grid stochastic swing equations

To simulate the system of Eqs. (10)–(12) and (16), we discretize these equations in time using a second-order strong Runge–Kutta scheme (Milshtein & Tret'yakov, 1994). To simplify the notation, we introduce the vectors $y_k = [\theta_k, \omega_k]^T$ and $y = [y_1^\top, y_2^\top, y_3^\top]^\top$. Using this notation, Eqs. (10), (11), and (16) read

$$dy_k = f_k(y)dt + g_k(y)P_k'^m dt,$$
$$dP_k'^m = a_k P_k'^m dt + b_k dW, \quad k = 1, 2, 3,$$

where

$$f_k(y) = \begin{bmatrix} \omega_B(\omega_k - \omega_s) \\ [\bar{P}_k^m - P_k^e(\boldsymbol{\theta}) - D_k(\omega_k - \omega_s)]/2H_k \end{bmatrix},$$
$$g_k(y) = \begin{bmatrix} 0 \\ 1/2H_k \end{bmatrix},$$

$$a_k = -\frac{1}{\lambda_k}, \quad b_k = \sqrt{\frac{2}{\lambda_k}}\sigma_k, \quad k = 1, 2, 3.$$

The Runge–Kutta discretization of these equations is (Milshtein & Tret'yakov, 1994)

$$\begin{aligned} y_{k,i+1} &= y_{k,i} + \frac{h}{2}\{(f_k + g_k P_k'^m)_i + (f_k + g_k P_k'^m)_{\bar{i}}\} \\ &\quad + \frac{1}{\sqrt{12}} g_{k,i} b_k h^{3/2} \eta_i, \\ P_{k,i+1}'^m &= P_{k,i}'^m + b_k \xi_i h^{1/2} + \frac{h}{2} a_k (P_{k,i}'^m + P_{k,\bar{i}}'^m) \\ &\quad + \frac{1}{\sqrt{12}} a_k b_k h^{3/2} \eta_i, \end{aligned} \tag{A.1}$$

where h is the time step in the Runge–Kutta scheme satisfying both the stability constraints in Milshtein and Tret'yakov (1994) and the condition $h = \delta/m$ (δ is the time between measurements and m is an integer) that is needed to model the measurements, ξ_i and η_i are the realizations of independent standard Gaussian random variables ξ and η at time step i, $f_{k,\bar{i}} = f_k(y_{\bar{i}})$, $g_{k,\bar{i}} = g_k(y_{\bar{i}})$, $y_{k,\bar{i}} = y_{k,i} + (f_k + g_k P_k'^m)_i h$, and $P_{k,\bar{i}}'^m = P_{k,i}'^m + b_k \xi_i h^{1/2} + a_k P_{k,i}'^m h$. In our simulations, we set $\delta = 0.0025m$ s, where the integer m is defined in Section 4 for each case. To satisfy the constraints for the time step and to allow for interpolation between successive measurements for $m > 1$, in all simulations we set $h = 0.0025$ s.

As described in Section 2.2, we use simple Monte Carlo to compute the prior mean and covariances of ω_k, θ_k, and P_k^m ($k = 1, \ldots, 3$). Each of the N_{MC} realizations of the dynamics is generated by sampling (A.1). Table A.7 presents the stochastic swing equation parameters employed in this work.

References

Ahmad, T., & Chen, H. (2018). Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. *Energy and Buildings*, *166*, 460–476.

Arnold, L. (1974). *Stochastic differential equations*. New York.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting*. Springer.

Cai, H., Jia, X., Feng, J., Li, W., Hsu, Y. M., & Lee, J. (2020). Gaussian process regression for numerical wind speed prediction enhancement. *Renewable Energy*, *146*, 2112–2123.

Catalao, J., Pousinho, H., & Mendes, V. (2011). Hybrid wavelet-pso-anfis approach for short-term wind power forecasting in portugal. *IEEE Transactions on Sustainable Energy*, *2*, 50–59.

Charlton, N., & Singleton, C. (2014). A refined parametric model for short term load forecasting. *International Journal of Forecasting*, *30*, 364–368.

Chen, J., Chen, Z., Zhang, C., & Wu, C. (2020). Apik: Active physics-informed kriging model with partial differential equations. arXiv preprint arXiv:2012.11798.

Dantas, T. M., & Oliveira, F. L. C. (2018). Improving time series forecasting: An approach combining bootstrap aggregation, clusters and exponential smoothing. *International Journal of Forecasting*, *34*, 748–761.

de Oliveira, E. M., & Oliveira, F. L. C. (2018). Forecasting mid-long term electric energy consumption through bagging arima and exponential smoothing methods. *Energy*, *144*, 776–788.

Fan, S., & Hyndman, R. J. (2011). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, *27*, 134–141.

Fentis, A., Bahatti, L., Mestari, M., Tabaa, M., Jarrou, A., & Chouri, B. (2016). Short-term pv power forecasting using support vector regression and local monitoring data. In *2016 International renewable and sustainable energy conference* (pp. 1092–1097). IEEE.

Gardner, E. S., Jr. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, *22*, 637–666.

Genton, M. G., & Kleiber, W. (2015). Cross-covariance functions for multivariate geostatistics. *Statistical Science*, *30*, 147–163. http://dx.doi.org/10.1214/14-STS487.

Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B., Bernstein, D., & Ridley, A. (2006). What is the ensemble kalman filter and how well does it work? In *2006 American control conference* (p. 6). IEEE.

Goude, Y., Nedellec, R., & Kong, N. (2013). Local short and middle term electricity load forecasting with semi-parametric additive models. *IEEE Transactions on Smart Grid*, *5*, 440–446.

Grant, J., Eltoukhy, M., & Asfour, S. (2014). Short-term electrical peak demand forecasting in a large government building using artificial neural networks. *Energies*, *7*, 1935–1953.

Hasegawa, Y., & Ueda, Y. (1999). Global basin structure of attraction of two degrees of freedom swing equation system. *International Journal of Bifurcation and Chaos*, *09*, 1549–1569. http://dx.doi.org/10.1142/S0218127499001085.

Hong, T. (2012). *Short term electric load forecasting*. (Ph.D. thesis), North Carolina State University.

Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012.

Hong, T., & Wang, P. (2014). Fuzzy interaction regression for short term load forecasting. *Fuzzy Optimization and Decision Making*, *13*, 91–103.

Hong, T., Wilson, J., & Xie, J. (2013). Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, *5*, 456–462.

Huang, J., & Perry, M. (2016). A semi-empirical approach using gradient boosting and k-nearest neighbors regression for gefcom2014 probabilistic solar power forecasting. *International Journal of Forecasting*, *32*, 1081–1086.

Huang, Y. F., Werner, S., Huang, J., Kashyap, N., & Gupta, V. (2012). State estimation in electric power grids: Meeting new challenges presented by the requirements of the future grid. *IEEE Signal Processing Magazine*, *29*, 33–43.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Hyndman, R. J., & Fan, S. (2009). Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, *25*, 1142–1153.

Kusiak, A., Zheng, H., & Song, Z. (2009). Short-term prediction of wind farm power: A data mining approach. *IEEE Transactions on Energy Conversion*, *24*, 125–136.

Landry, M., Erlinger, T. P., Patschke, D., & Varrichio, C. (2016). Probabilistic gradient boosting machines for gefcom2014 wind forecasting. *International Journal of Forecasting*, *32*, 1061–1066.

Liu, Y., Shi, J., Yang, Y., & Lee, W. J. (2012). Short-term wind-power prediction based on wavelet transform–support vector machine and statistic-characteristics analysis. *IEEE Transactions on Industry Applications*, *48*, 1136–1141.

Lloyd, J. R. (2014). Gefcom2012 hierarchical load forecasting: Gradient boosting machines and gaussian processes. *International Journal of Forecasting*, *30*, 369–374.

Luo, J., Hong, T., & Fang, S. C. (2018). Benchmarking robustness of load forecasting models under data integrity attacks. *International Journal of Forecasting*, *34*, 89–104.

Milshtein, G., & Tret'yakov, M. (1994). Numerical solution of differential equations with colored noise. *Journal of Statistical Physics*, *77*, 691–715.

Nedellec, R., Cugliari, J., & Goude, Y. (2014). Gefcom2012: Electric load forecasting and backcasting with semi-parametric models. *International Journal of Forecasting*, *30*, 375–381.

Nishikawa, T., & Motter, A. E. (2015). Comparative analysis of existing models for power-grid synchronization. *New Journal of Physics*, *17*, Article 015012.

Quiñonero-Candela, J., & Rasmussen, C. E. (2005). A unifying view of sparse approximate gaussian process regression. *Journal of Machine Learning Research*, *6*, 1939–1959.

Raissi, M., & Karniadakis, G. E. (2018). Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, *357*, 125–141.

Rosenthal, W. S., Tartakovsky, A. M., & Huang, Z. (2018). Ensemble kalman filter for dynamic state estimation of power grids stochastically driven by time-correlated mechanical input power. *IEEE Transactions on Power Systems*, *33*, 3701–3710.

Singh, S., Mohapatra, A., et al. (2019). Repeated wavelet transform based arima model for very short-term wind speed forecasting. *Renewable Energy*, *136*, 758–768.

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, *36*, 75–85.

Snelson, E., & Ghahramani, Z. (2006). Sparse gaussian processes using pseudo-inputs. In *Advances in neural information processing systems* (pp. 1257–1264).

Sobhani, M., Hong, T., & Martin, C. (2020). Temperature anomaly detection for electric load forecasting. *International Journal of Forecasting*, *36*, 324–333.

Song, K. B., Baek, Y. S., Hong, D. H., & Jang, G. (2005). Short-term load forecasting for the holidays using fuzzy linear regression method. *IEEE Transactions on Power Systems*, *20*, 96–101.

Swiler, L. P., Gulian, M., Frankel, A. L., Safta, C., & Jakeman, J. D. (2020). A survey of constrained gaussian process regression: Approaches and implementation challenges. *Journal of Machine Learning for Modeling and Computing*, *1*.

Taieb, S. B., & Hyndman, R. J. (2014). A gradient boosting approach to the kaggle load forecasting competition. *International Journal of Forecasting*, *30*, 382–394.

Takeda, H., Tamura, Y., & Sato, S. (2016). Using the ensemble kalman filter for electricity load forecasting and analysis. *Energy*, *104*, 184–198.

Thiyagarajan, K., & Kumar, R. S. (2016). Real time energy management and load forecasting in smart grid using compactrio. *Procedia Computer Science*, *85*, 656–661.

Wang, P., Barajas-Solano, D. A., Constantinescu, E., Abhyankar, S., Ghosh, D., Smith, B., et al. (2015). Probabilistic density function method for stochastic odes of power systems with uncertain power input. *SIAM/ASA Journal on Uncertainty Quantification*, *3* 873–896.

Wang, P., Liu, B., & Hong, T. (2016). Electric load forecasting with recency effect: A big data approach. *International Journal of Forecasting*, *32*, 585–597.

Williams, C. K., & Rasmussen, C. E. (2006). *Gaussian processes for machine learning. Vol. 2*. Cambridge, MA: MIT Press.

Xenochristou, M., Hutton, C., Hofman, J., & Kapelan, Z. (2020). Water demand forecasting accuracy and influencing factors at different spatial scales using a gradient boosting machine. *Water Resources Research*, *56*, e2019WR026304.

Yang, X., Barajas-Solano, D., Tartakovsky, G., & Tartakovsky, A. M. (2019). Physics-informed cokriging: A Gaussian-process-regression-based multifidelity method for data-model convergence. *Journal of Computational Physics*, *395*, 410–431.

Yeung, E., Kundu, S., & Hodas, N. (2017). Learning deep neural network representations for koopman operators of nonlinear dynamical systems. arXiv preprint arXiv:1708.06850.

Yoder, M., Hering, A. S., Navidi, W. C., & Larson, K. (2014). Short-term forecasting of categorical changes in wind power with markov chain models. *Wind Energy*, *17*, 1425–1439.