Forecasting high-frequency spatio-temporal wind power with dimensionally reduced echo state networks

Huang Huang1 | Stefano Castruccio2 | Marc G. Genton1

1Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
2Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, Indiana, USA

Correspondence
Huang Huang, Statistics Program, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia.
Email: huang.huang@kaust.edu.sa

Funding information
King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR), Grant/Award Number: OSR-2018-CRG7-3742

Abstract
Fast and accurate hourly forecasts of wind speed and power are crucial in quantifying and planning the energy budget in the electric grid. Modelling wind at a high resolution brings forth considerable challenges given its turbulent and highly nonlinear dynamics. In developing countries, where wind farms over a large domain are currently under construction or consideration, this is even more challenging given the necessity of modelling wind over space as well. In this work, we propose a machine learning approach to model the nonlinear hourly wind dynamics in Saudi Arabia with a domain-specific choice of knots to reduce spatial dimensionality. Our results show that for locations highlighted as wind abundant by a previous work, our approach results in an 11% improvement in the 2-h-ahead forecasted power against operational standards in the wind energy sector, yielding a saving of nearly one million US dollars over a year under current market prices in Saudi Arabia.

KEYWORDS
- echo state network
- Gaussian random field
- machine learning
- reservoir computing
- space-time model

1 | INTRODUCTION

The wide consensus of the increasingly negative effects of a warming climate caused by greenhouse gas emissions (Intergovernmental Panel on Climate Change, 2014) has prompted the global
community to seek alternative, carbon-free energy sources. While heterogeneous across countries, the increase in the usage of renewable resources worldwide has been steady over the past decades, and the aftermath of the current COVID-19 pandemic is foreseen to further accelerate this trend (Organisation for Economic Co-operation and Development, 2020).

Saudi Arabia has the second-largest oil reserve in the world, and its internal energy portfolio is almost exclusively focused on fossil fuels. Worldwide, the country currently ranks sixth in oil consumption and has one of the highest per capita energy consumptions (British Petroleum, 2020). In an effort to align the country’s targets with the Paris Agreement (Kinley, 2017) and reduce its carbon footprint, the Saudi Arabian government recently outlined ‘Vision 2030’, a blueprint plan to promote an economically viable transition to renewable energies to reduce its dependence on oil. While solar energy has the greatest potential given the country’s geographical position, the plan aims to also install 16 GW of wind capacity by 2030. Wind investment is strategic for the diversification and increased penetration of renewable energies given its high abundance, especially during night hours when solar power is not available. For a country with little to no installed capacity, the development of ‘Vision 2030’ requires a preliminary understanding of where the turbines should be built and what factors are more influential in determining their sites. Very recent work provided some initial evidence. Giani et al. (2020) investigated various types of installation and maintenance costs and identified the optimal wind farm build-out under the present wind regimes; Zhang et al. (2021) studied the resilience of the plan under a changing climate over the next 30 years; and Chen et al. (2021) investigated the probability of disruption of wind operations due to extreme wind. One fundamental question must be addressed to provide a pathway for its practical implementation: What is the benefit of a reliable forecasting system, and what would be its economic advantage against standard operational forecasting systems? Such information would be instrumental to policymakers in their final decision regarding the optimal construction sites.

Modelling hourly wind fields (the resolution of interest in the energy market) for a country of one-fourth the size of the United States, with nonlinear dynamics implied by diverse geography comprising two coastal areas, mountain ranges and sandy and rocky deserts, is a considerable challenge. Standard statistical approaches such as the autoregressive integrated moving average (ARIMA) models are effective in capturing a short-range (approximately) linear dependence; however, more flexible alternatives are necessary for nonlinear dynamics, and simple modifications such as fractional integration (Taqqu et al., 1995) have been shown to be largely insufficient in capturing a more structured dependence in time.

Machine learning approaches to time series, such as artificial neural networks, have been more successful in capturing nonlinear dynamics, owing to their flexibility and the possibility of defining a recursive structure (recurrent neural networks, RNNs) suitable for temporally resolved data. However, the inference for RNN is very challenging given the high dimensionality of the parameter space and the impossibility to directly apply the iterative gradient computations of traditional neural networks, which would lead to numerical instabilities (Doya, 1992). To achieve more stable and computationally affordable inference, Jaeger (2001) proposed a new approach to RNN, that is, the echo state networks (ESNs), which is predicated on the use of sparse, random matrices instead of dense unknown ones (Lukoševičius & Jaeger, 2009). Maas et al. (2002) independently developed a similar approach with the name liquid state machines but with more sophisticated topological constraints motivated by neuroscience. This paradigm is now commonly referred to as reservoir computing when inputs of the neural networks are mapped to high-dimensional hidden reservoir states through fixed nonlinear dynamics. These hidden reservoir states consider sequential linkages and thus allow for a nonlinear transformation of the input
history. Despite the use of reservoir states being ideally suited as a baseline for statistical modelling, the machine learning community has mostly focused on this class of methods as a means to ease the computational burden.

Recently, a series of seminal works (McDermott & Wikle, 2017, 2019a,b) on the ESN framework was proposed in the statistical literature to forecast Pacific sea surface temperature and the United States soil moisture. The empirical orthogonal function (EOF) method, or equivalently, the principal component analysis (PCA), is used to reduce spatial dimensionality. The leading orthogonal basis functions, which are the leading eigenvectors of the spatial covariance matrix of the space-time data, capture the main modes of spatial variability. Therefore, they reduce the dimensionality and preserve the primary spatial variability structure when using the leading orthogonal basis functions with a much smaller number than the original number of locations to represent the entire spatial field. In this work, however, we propose a new ESN framework that describes the dynamics of a spatial field through a collection of knots, sampled at both a grid and fixed points over areas of complex patterns such as rugged terrains (see Figure 3). The entire field is then reconstructed with a spatial interpolation approach, thereby allowing for a full spatio-temporal forecast without a computationally prohibitive ESN. Comparison with the EOF-based ESN, ARIMA, a spatio-temporal random process model, and the persistence forecasting (i.e. assuming the variable of interest does not change from the last observed value) shows that the prediction under our proposed ESN framework is more accurate. Furthermore, we offer a comprehensive simulation study investigating how and to what extent the ESN can capture the data dynamics and yield predictive improvements against other methods when the underlying simulated process has a controlled level of nonlinearity.

The manuscript proceeds as follows: Section 2 presents the wind data; Section 3 introduces the ESN and spatial model; Section 4 discusses the inferential and forecasting approach; Section 5 compares the forecasting from our model to several operational alternatives in the wind energy market; Section 6 presents the computation of the wind power forecast and discusses and quantifies the economic advantages of our method; and Section 7 provides the conclusions of the study.

Implementation of our proposed ESN for the wind speed forecasts in Saudi Arabia can be found at https://github.com/hhuang90/KSA-wind-forecast.

2 | WIND DATA

While developed countries can perform forecasting with publicly available, continuously updated high-resolution simulations routinely used for trading energy options (e.g. the high-resolution rapid refresh (Benjamin et al., 2016) in North America), developing countries lack such data products. To investigate the wind speed and subsequent wind power potentials in Saudi Arabia, Giani et al. (2020) performed hourly high-resolution simulations from the Weather Research and Forecasting model (WRF; Skamarock et al., 2008) driven by the operational high-resolution European Centre for Medium-Range Weather Forecast model. WRF was simulated from 2013 to 2016, given the simultaneous availability of wind speed from meteorological masts for validation. The simulation was resolved on a regular planar grid with a fixed resolution from the Lambert Conformal Conic projection. Four simulations with different model setups were performed. For this study, we choose the one closest to the observations, which relied on the Mellor–Yamada–Janjić planetary boundary layer scheme (Gómez-Navarro et al., 2015; Janjić, 2001; Mellor & Yamada, 1982) and had a 6-km resolution, and we consider the hourly near-surface wind speed. For practical
purposes, we exclude islands over the Red Sea (Tiran Island, Sanafir Island in the Red Sea Project Lagoon, and Farasan Island) for a total of 53,333 locations. Figure 1a depicts the average hourly near-surface wind speed from 2013 to 2016, showing generally higher values in the west due to mountain ranges. While abundant in wind, these regions are generally unsuitable for building wind farms due to the high installation costs implied by the rough terrain. The area in northwest Saudi Arabia near the Gulf of Aqaba is the construction site of NEOM City (Farag, 2019), and data indicate its suitability for wind energy harvesting, as also discussed in Giani et al. (2020). We select two locations with a high or low average wind to display the diurnal temporal variability both in terms of mean and standard deviation in Figure 1b and c. We observe that the wind speed annual variability is low in the Arabian Peninsula, as also highlighted in Chen et al. (2018), hence lending support for the extension of the results in this work beyond the simulation years. WRF was validated by the observations at 10 sites within the King Abdullah City for Atomic and Renewable Energy monitoring network through various assessment metrics in Giani et al. (2020).

3 | MODELS

Throughout this work, we denote by $Z_t(s_i)$ the near-surface wind speed at location $s_i$ and time $t$, where $t$ takes values from $\{0, \ldots, T\}$ indicating the number of hours after 00:00 January 1, 2013.
(February 29 in the leap year of 2016 is removed for simplicity, but could be easily incorporated with more general harmonics in Equation (1)), and \( s_i, \ i = 1, \ldots, n = 53,333 \), is the data location in Saudi Arabia.

### 3.1 Mean structure

We observe the periodic inter-daily and multi-daily patterns in the wind speed, which are explainable by both the daily land-ocean heat fluxes and mesoscale patterns of recurring winds in the Arabian Peninsula; see Figure 1 for the daily patterns. For each location \( s_i, \ i = 1, \ldots, n \), we assume that the dynamics of \( \sqrt{Z_t(s_i)} \) depend on harmonics:

\[
\sqrt{Z_t(s_i)} = \beta_0(s_i) + \sum_{k=1}^{K} \left\{ \beta_{k,1}(s_i) \cos \left( \frac{2\pi t}{T_k} \right) + \beta_{k,2}(s_i) \sin \left( \frac{2\pi t}{T_k} \right) \right\} + \gamma(s_i)Y_t(s_i), \tag{1}
\]

where \( K \) is the number of harmonics, \( \beta_0(s_i) \) is the location-wise intercept, \( \beta_{k,1}(s_i) \) and \( \beta_{k,2}(s_i), \ k = 1, \ldots, K \), are the location-wise coefficients for each harmonic with period \( T_k \), and \( \gamma(s_i) \) is the location-wise scaling parameter such that the residual \( Y_t(s_i) \) has unit variance at each location. According to the discrete Fourier transform diagnostics illustrated in Figure S2 in the Supplementary Material, there are \( K = 5 \) significant harmonics at periods of 1 year, half a year, 1 day, 12 and 8 h. We consider the square root because the wind speed is generally right skewed due to occasional wind gusts. The choice of this particular transformation to normality is justified by well-established literature on wind modelling (e.g. Gneiting, 2002), as well as our diagnostics in Figure S3 in the Supplementary Material.

The parameters \( \beta_0(s_i), \beta_{k,1}(s_i), \) and \( \beta_{k,2}(s_i), \ i = 1, \ldots, n, \ k = 1, \ldots, K \), are estimated by linear regression independently across sites using the data from 2013 to 2015 (training data), a task that can be performed in parallel across locations. The scaling parameter \( \gamma(s_i), \ i = 1, \ldots, n \), is also estimated site by site so that the resulting \( Y_t(s_i) \) is a zero-mean spatio-temporal residual process with a structure specified in the following sections.

### 3.2 Temporal dynamics

The hourly residual wind \( Y_t(s_i), \ i = 1, \ldots, n \), is bound to show a highly nonlinear behaviour that traditional time series approaches, such as ARIMA or more general linear space-time models, may fail to capture. The inability of traditional methods to capture the nonlinearity of spatio-temporal dynamics (McDermott & Wikle, 2017) necessarily translates into sub-optimal forecasts, as will be illustrated in Section 5. We propose an ESN model with nonlinear dynamics generated by a state space with neural network dynamics. We will introduce the model with the standard statistical terminology, but we also put in parenthesis the equivalent machine learning terms and use them henceforth.

Quadratic interactions have been shown to be effective in characterizing the nonlinear dynamical evolution in many physical processes (Gladish & Wikle, 2014; Wikle & Hooten, 2010). In this work, and consistently with McDermott and Wikle (2017), we propose to model the relationship between the data (output) of the wind speed residual at time \( t \) and all locations,
\[ y_t := \left( Y_t(s_1), \ldots, Y_t(s_n) \right)^T, \] and the \( n_h \)-dimensional latent variables (hidden states or reservoir states), \( h_t \), by a quadratic ESN:

\[ y_t = V_1 h_t + V_2 (h_t \odot h_t) + \epsilon_t, \]

where \( \odot \) is the Hadamard (element-wise) product, \( V_1 \) and \( V_2 \) are weight matrices whose entries are unknown, and \( \epsilon_t \) is the error at time \( t \), assumed to follow a zero-mean multivariate normal distribution. Here, \( V_1 \) and \( V_2 \) are estimated by least squares, as will be shown later in Section 4.2.1, and are independent from the spatial covariance matrix of \( \epsilon_t \). The uncertainty of the estimated \( y_t \) relies on the covariance of \( \epsilon_t \), but we will show in Section 4.1 that the prediction uncertainty can be calibrated with the ESN mean, by developing an approach similar to Bonas and Castruccio (2021).

The dynamics of the reservoir state vector \( h_t \) is formulated as

\[ h_t = \phi f \left( \frac{\delta}{\lambda_W} W h_{t-1} + U x_t \right) + (1 - \phi) h_{t-1}. \]

The vector \( x_t \) represents all the covariates (input), and here \( x_t := (1, y_{t-1}^T, \ldots, y_{t-m}^T)^T \), that is, we consider the intercept and the lagged wind speed residuals up to \( m \) time steps before; \( f \) is the element-wise activation function (here we choose the commonly used hyperbolic tangent function, but other choices are possible such as the rectified linear unit (Goodfellow et al., 2016)); \( \lambda_W \) is the maximal absolute eigenvalue of \( W \); \( \delta \) is a tuning parameter to scale \( W \); and \( \phi \) is the leaking rate to control the speed of the dynamics. In summary, the temporal evolution of \( h_t \) is controlled by two parts. The first one is a nonlinear evaluation of a linear combination of \( h_{t-1} \) and \( x_t \), and the second is a memory term that allows for long-range dependence.

The \( n_h \times n_h \)-dimensional weight matrix \( W = [w_{ij}] \) controls the interaction among reservoir states with time evolution, and the \( n_h \times (mn + 1) \)-dimensional weight matrix \( U = [u_{ij}] \) determines the mapping from the inputs to the reservoir states. The key difference between the ESN and the standard RNN lies in the structure of \( W \) and \( U \): an ESN assumes that these matrices are random with a controlled level of sparsity, while an RNN assumes full matrices with unknown fixed entries. Therefore, in the ESN, the interaction of each pair of reservoir states, as well as each pair of inputs and reservoir states, is assumed to exist with a given probability. The sparsity in both matrices prevents overfitting and dramatically reduces the model dimensionality. Specifically, the entries of \( W \) and \( U \) are generated as follows:

\[ w_{ij} | \gamma_w^w \sim \Gamma_w \text{Unif}(-a_w, a_w), \gamma_w^w \sim \text{Bern}(\pi_w), \]
\[ u_{ij} | \gamma_u^u \sim \Gamma_u \text{Unif}(-a_u, a_u), \gamma_u^u \sim \text{Bern}(\pi_u), \]

where \( a_w \) and \( a_u \) control the magnitude of nonzero entries in \( W \) and \( U \), respectively, and \( \pi_w \) and \( \pi_u \) determine the expected sparsity.

An extensive literature in neural networks has proposed stochastic optimization methods, the most common being the random initialization of the weight matrices in the learning algorithm (Yehudai & Shamir, 2019). ESNs are fundamentally different, as the weight matrices \( W \) and \( U \) are instead randomly generated from a given distribution; once drawn, they are fixed and do not need to be estimated. The randomness of \( W \) and \( U \) results in an ensemble of predictions, and its mean can be used as point prediction given its reduced variance with respect to a single ensemble member.
3.3 Spatial modelling

The proposed ESN model could in principle be used to directly produce forecasts of $Y_t$; however, this would imply the use of input comprising a spatially referenced vector of $n = 53,333$ points at each time $t$, and consequently a high-dimensional vector of reservoir states $h_t$, matrices $V_1$, $V_2$ in Equation (2) and matrices $W$, $U$ in Equation (3) to capture the dynamics. A brute-force approach disregarding the spatial information would therefore be computationally intractable.

One commonly used approach to reduce the size of spatially referenced data in geosciences is using the EOF (or equivalently PCA) method. This approach allows to capture the main modes of spatial variability; for example, Gladish and Wikle (2014) used the first 10 EOFs to represent the sea surface temperature on a region in the Pacific Ocean comprised of more than two thousand locations. In this work, we propose a stochastic approach with a Gaussian random field to model the spatial data. After fitting the Gaussian random field with all the available data, we consider a set of knots and use them to represent the entire process. Results in Section 5.1 show that this modelling choice results in better predictive performances than EOFs.

We assume a zero-mean Gaussian random field to model $Y_t(s)$ at each time $t$, that is, $Y_t(s) \sim \text{GP}(0, C(\cdot, \cdot))$. To reflect the spatially varying dependence structure dictated, among others, by the different topographical features in Saudi Arabia, we use a nonstationary Matérn covariance function (Paciorek & Schervish, 2006) for $Y_t(s)$. More specifically, for any pair of locations $(s, s')$, the covariance is

$$C(s, s') = \sigma(s)\sigma(s')R^{NS}(s, s'; \Sigma(\cdot), \nu(\cdot)) + \tau^2(s)I_{\{s=s'\}},$$

(5)

where $\sigma^2(s)$, $\nu(s)$, and $\tau^2(s)$ are the spatially varying partial sill, smoothness, and nugget parameters, respectively, $\Sigma(s)$ is the spatially varying covariance matrix of a Gaussian kernel centred at $s$,

$$R^{NS}(s, s'; \Sigma(\cdot), \nu(\cdot)) = |\Sigma(s)|^{1/2}|\Sigma(s')|^{1/2}|\{\Sigma(s) + \Sigma(s')\}|/2^{\nu/2} \times R^{S}(s - s'; \{\Sigma(s) + \Sigma(s')\}/2, \{\nu(s) + \nu(s')\}/2)$$

is the nonstationary Matérn correlation function, $R^{S}(h; \Sigma, \nu) = 2^{1-\nu}\Gamma(\nu)^{-1}(h^T\Sigma^{-1}h)^{\nu/2}/K_\nu((h^T\Sigma^{-1}h)^{1/2})$ is the stationary Matérn correlation function (Stein, 1999), $K_\nu(\cdot)$ is the modified Bessel function of the second kind with order $\nu$, and $I_{\{\cdot\}}$ is the indicator function. We use the R (R Core Team, 2019) package convoSPAT (Risser & Calder, 2017) to fit the nonstationary random field with $Y_t(s_i)$ $i = 1, \ldots, n$, in 2015.

The package convoSPAT uses a mixture component model for $\sigma^2(\cdot)$, $\nu(\cdot)$, $\Sigma(\cdot)$ and $\tau^2(\cdot)$ and locally estimates the mixture parameters. We place 42 mixture components spanning the entire spatial domain as illustrated in Figure 2a. Our results suggest that the nugget estimate $\hat{\tau}^2(\cdot)$ is negligible, and the spatially varying partial sill estimate $\hat{\sigma}^2(\cdot)$ and smoothness estimate $\hat{\nu}(\cdot)$ are shown in Figure 2b and c. The covariance of the Gaussian kernel characterizes anisotropy and the dependence scale, and we show in Figure 2a the isoline of the estimate $\hat{\Sigma}(\cdot)$ at the mixture components.

We adopt a moderately large set of knots across the spatial domain and then model their space-time dynamics with the ESN. Once forecasts on the knots have been provided, spatial interpolation (kriging) is used to produce spatial predictions. A formal comparison between the approximate process using knots and the true process has been studied, for example, in Titsias (2009) and Matthews et al. (2016) using the Kullback–Leibler divergence.
**FIGURE 2** (a): Ellipses of estimate $\hat{\Sigma}(\cdot)$ at 42 mixture components (locations indicated as the red dot). (b) and (c): Spatially varying partial sill estimate $\hat{\sigma}^2(\cdot)$ and smoothness estimate $\hat{\nu}(\cdot)$ in the mixture component model [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE 3** Overview of the $n^* = 3173$ chosen knots: 2,756 knots closest to a 0.25°-resolution regular grid are shown in light grey, and 417 knots with an average wind speed from 2013 to 2015 that is greater than 6 m/s are shown in red [Colour figure can be viewed at wileyonlinelibrary.com]

Our chosen knots consist of two sets. The first set comprises 2756 data locations closest to a 0.25°-resolution regular grid across the entire domain, while the second set is chosen at locations with a high wind power potential. We calculate the average wind speed from 2013 to 2015 and focus on the locations with values greater than 6 m/s, which is approximately the 98% quantile. A set of 417 locations are then selected, such that no pairs have differences less than 0.005° in longitude or latitude. Figure 3 shows a total number of $n^* = 3173$ chosen knots, which we denote by $\{s^*_1, \ldots, s^*_n\}$. The output $y_t$ is then replaced by $y^*_t := (Y_t(s^*_1), \ldots, Y_t(s^*_n))^\top$. The input $x_t$ is substituted accordingly with $x^*_t := (1, y^*_{t-1}, \ldots, y^*_{t-m})^\top$, and the reservoir states $h_t$ in Equation (3) are updated by plugging in $x^*_t$ as $h_t = \phi f(\frac{\delta}{|\lambda|} Wh_{t-1} + Ux^*_t) + (1 - \phi) h_{t-1}$.

**4 | INFEERENCE AND FORECASTING**

This section discusses both the inference and the forecasting approach. In Section 4.1 we show how forecasting is performed given all parameter estimates, and then in Section 4.2 we elaborate on the inference procedure, where the tuning parameters are estimated via cross-validation based on the forecasting performance.
## 4.1 Forecasting

Let the training period be \(\{0, \ldots, T\}\) and the forecasting period be \(\{T+1, \ldots, T_{\text{max}}\}\). In this work, we are not only interested in forecasts for 1 h ahead, but also for 2 and 3 h ahead to make our predictions conform with the majority of wind energy markets (Hering & Genton, 2010). When we forecast \(\mathbf{y}_{t+2}^*\) at time \(t\), the observation \(\mathbf{y}_{t+1}^*\) is required but unavailable, so we substitute the one time step forecast \(\hat{\mathbf{y}}_{t+1}^*\) for \(\mathbf{y}_{t+1}^*\) to forecast \(\mathbf{y}_{t+2}^*\). Section S1 in the Supplementary Material formalizes the forecasts for multiple time steps ahead.

At time \(t\), once the forecast \(\hat{\mathbf{y}}_{t+h}^* = (\hat{\mathbf{y}}_{t+h}(s_1^*), \ldots, \hat{\mathbf{y}}_{t+h}(s_{n^*}^*))^T\), \(h = 1, 2, 3\), at the knots is obtained, the wind speed residuals at all locations \(\mathbf{y}_{t+h} = (\mathbf{y}_{t+h}(s_1), \ldots, \mathbf{y}_{t+h}(s_{n}))^T\) can be obtained via spatial interpolation, that is, \(\hat{\mathbf{y}}_{t+h} = \mathbf{K}^f\mathbf{K}^{-1}\hat{\mathbf{y}}_{t+h}^*\), where \(\mathbf{K}^f = [k_{ij}^f]\) is an \(n \times n^*\)-dimensional covariance matrix with entries \(k_{ij}^f = \text{cov}(Y_{t+h}(s_i), Y_{t+h}(s_j^*))\) and \(\mathbf{K} = [k_{ij}]\) is an \(n^* \times n^*\)-dimensional covariance matrix with entries \(k_{ij} = \text{cov}(Y_{t+h}(s_i^*), Y_{t+h}(s_j^*))\).

As pointed out by Efron and Hastie (2016), averaging a set of unbiased estimators is an effective way to reduce estimator variance. Rougier (2016) also investigated why the ensemble mean often yields a smaller mean-squared error (MSE) than the majority of the ensemble members. Throughout this work, we use 100 ensemble members to forecast, each consisting of an independent random realization of \(\mathbf{W}\) and \(\mathbf{U}\). The mean of the 100 ensemble forecasts is then used as the final forecast. In order to quantify the uncertainty, an empirical calibration similar to that proposed in Bonas and Castruccio (2021) is used here. An ESN is trained with the wind speed residuals \(Y_t(s)\) from 2013 to 2014, and the forecast for 2015 (calibration period) is produced. At each location \(s\), the quantiles of \(Y_t(s) - \hat{Y}_t(s)\) in 2015 are estimated. Then, these quantiles will be used to build the forecast distribution for \(\hat{Y}_t(s)\) in 2016.

## 4.2 Inference

### 4.2.1 Estimating the output matrices

Estimating \(\mathbf{V}_1\) and \(\mathbf{V}_2\) can be performed via least square regression. The large size of the output vector would potentially result in a high dimension of the reservoir state space. Therefore, it is necessary to penalize the coefficient estimation to reduce its variance and prevent overfitting. If we denote

\[
\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_T^T \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T, (\mathbf{h}_1 \odot \mathbf{h}_1)^T \\ \vdots \\ \mathbf{h}_T^T, (\mathbf{h}_T \odot \mathbf{h}_T)^T \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} \mathbf{e}_1^T \\ \vdots \\ \mathbf{e}_T^T \end{bmatrix},
\]

then, we can express Equation (2) in a matrix linear regression form as \(\mathbf{Y} = \mathbf{HB} + \mathbf{E}\). We use ridge regression or Tikhonov regularization to estimate \(\mathbf{B}\) such that

\[
\hat{\mathbf{B}} = \arg \min_{\mathbf{B}} (\|\mathbf{Y} - \mathbf{HB}\|_F^2 + \lambda \|\mathbf{B}\|_F^2),
\]

where \(\|\cdot\|_F\) is the Frobenius norm and \(\lambda\) is the ridge penalty. The ridge estimator \(\hat{\mathbf{B}}\) has the following closed form:

\[
\hat{\mathbf{B}} = (\mathbf{H}^T\mathbf{H} + \lambda \mathbf{I})^{-1}\mathbf{H}^T\mathbf{Y}.
\]
### Table 1: Parameter estimates in the ESN model via cross-validation

| Parameter explanation                  | Notation | Estimate |
|----------------------------------------|----------|----------|
| Number of reservoir states             | $n_h$    | 2500     |
| Input lag                              | $m$      | 1        |
| Leaking rate                           | $\phi$   | 1        |
| Scaling matrix parameter for $W$       | $\delta$ | 0.9      |
| Ridge penalty                          | $\lambda$| 0.15     |
| Magnitude of entries in $W$            | $a_w$    | 0.05     |
| Magnitude of entries in $U$            | $a_u$    | 0.01     |
| Sparsity of $W$                        | $\pi_w$  | 0.1      |
| Sparsity of $U$                        | $\pi_u$  | 0.01     |

*Note: $y^*_t$ from 2013 to 2014 is used as the training data set, and $y^*_t$ in 2015 the validation data set.*

Once $\hat{B}$ is obtained, the forecast for one time step ahead (and iteratively for any future time step) on the knots can be computed as follows:

$$
\hat{y}^*_t = \hat{B}^T \begin{bmatrix} h_t \\ h_t \odot h_t \end{bmatrix},
$$

for $t \in \{T + 1, \ldots, T_{\text{max}}\}$.

#### 4.2.2 Cross-validation

Except for $V_1$ and $V_2$, several other parameters need to be estimated via cross-validation: the dimension of the reservoir states $n_h$, the number of lags $m$ in $x_t$, the leaking rate $\phi$ and the scaling parameter $\delta$ in Equation (3), the magnitude and sparsity parameters $a_w$, $\pi_w$, $a_u$, and $\pi_u$ in Equation (4), and the ridge penalty $\lambda$ in Equation (6). We consider $y^*_t$ from 2013 to 2014, for a total of 17,520 hourly observations, as the training data to obtain the ridge estimator $\hat{B}$ in Equation (6). Subsequently, $y^*_t$ in 2015, for a total of 8760 hourly observations, is used as the validation data. More specifically, at any time $t$ in 2015, $y^*_0$, $\ldots$, $y^*_t$ are available; we forecast $y^*_{t+1}$, we compare it with the true validation data, and we then calculate the MSE at all knots and time points for the mean of 100 ensemble forecasts (see Section S2 in the Supplementary Material for more details). We search over a large grid (shown in Table S1) to obtain the optimal values of the parameters, see Table 1.

The large number of reservoir states indicates that a complex model is necessary to capture the nonlinearity of the multivariate time series with $n^* = 3173$ dimensions. The estimated ridge penalty avoids overfitting using the large reservoir space. The obtained optimal leaking rate is one (i.e. the reservoir states are fully updated by new activations at each time). Since we removed the low-frequency periodicities to the original wind speed (see Equation (1)), the optimal leaking rate of one suggests a short-term dynamic in the residual field $Y_t(s)$. In our ESN, we use the lagged wind speed residuals as the input, and we observe that the optimal model only requires one lagged observation as the input at each time. The parameter $\delta$ determines the spectral radius of the matrix $W$ after scaling and influences the fading speed of the input in the hidden neurons.
In practice, $\delta < 1$ always guarantees the echo state property (i.e. the reservoir states are independent on the initial conditions after a sufficiently long time). Our estimated value of $\delta$ indicates that even though only one lagged observation is necessary at each time, the hidden neurons preserve a long memory of the input (lagged wind speed residuals) by employing a moderately large spectral radius of $W$, 0.9. The values of $a_w$, $\pi_w$, $a_u$, and $\pi_u$ are all comparatively small, implying a sparse collection of weak connections among reservoir states and between the reservoir states and the input.

5 | FORECASTING EVALUATION

5.1 | Forecasting results for wind speed

Once the parameters have been estimated in the training step from 2013 to 2015, we proceed to forecast $y^{*}_t$ in 2016 and compare it with other common strategies for short-term wind forecasts. The calibration algorithm explained in Section 4.1 is used to obtain the probabilistic forecast distribution in 2016. Table 2 shows the proportion of the true wind speed residuals $y^{*}_t$ in 2016 falling into the associated prediction intervals, and we observe that overall they match the nominal levels closely.

We compare the ESNs with persistence forecasting and the ARIMA model. The persistence forecasting method uses the last observed values as estimates for the near future (i.e. $\hat{y}_{t+3} = \hat{y}_{t+2} = \hat{y}_{t+1} = y_t$). This method is very simple, but frequently used in practice because of its good performance for very short-term forecasting. Additionally, we propose an ARIMA approach, which assumes that at each knot $s^*_i$, $i = 1, \ldots, n^*$, the time series of wind residual $Y_t(s^*_i)$ has the following model:

$$
\left(1 - \sum_{k=1}^{p(s^*_i)} \psi_k(s^*_i) \mathcal{L}^k\right) (1 - \mathcal{L})^{d(s^*_i)} Y_t(s^*_i) = c(s^*_i) + \left(1 + \sum_{k=1}^{q(s^*_i)} \theta_k(s^*_i) \mathcal{L}^k\right) \epsilon_t(s^*_i),
$$

where $p(s^*_i)$, $d(s^*_i)$, and $q(s^*_i)$ are the order of the autoregressive, differencing and moving-average terms respectively. Here, $\mathcal{L}$ is the lag operator (e.g. $\mathcal{L} Y_t(s^*_i) = Y_{t-1}(s^*_i)$), $\epsilon_t(s^*_i)$ is the error term, and $c(s^*_i)$, $\psi_1(s^*_i)$, $\ldots$, $\psi_{p(s^*_i)}(s^*_i)$ and $\theta_1(s^*_i)$, $\ldots$, $\theta_{q(s^*_i)}(s^*_i)$ are unknown parameters at $s^*_i$. The optimal orders for the ARIMA model at each knot are independently optimized through cross-validation using the data from 2013 to 2015. For each time in 2016, the ARIMA model with the optimal

| Prediction interval coverage | 1 h ahead | 2 h ahead | 3 h ahead |
|------------------------------|-----------|-----------|-----------|
| 95%                          | 94.8% (0.5%) | 94.6% (0.6%) | 94.6% (0.7%) |
| 80%                          | 80.7% (1.0%) | 80.6% (1.1%) | 80.6% (1.2%) |
| 60%                          | 60.7% (1.3%) | 60.3% (1.3%) | 60.1% (1.5%) |
Table 3 summarizes the MSE results for $y^*_t$ at all the $n^*$ knots and time points in 2016. The ESN outperforms the persistence and ARIMA in forecasts, with better results as the lead time increases from 1 to 3 h ahead.

After forecasting $Y_t(s)$ at the $n^*$ knots, we produce a map at the $n$ locations by spatial interpolation from the inferred spatial Gaussian random field described in Section 3.3, and we denote this approach by S-ESN. We compare this method with ARIMA or persistence approaches trained on all the $n$ locations, a choice which puts the S-ESN at a competitive disadvantage, as it can learn the dynamics only on the knots. Table 4 shows the outperformance of ESN over the ARIMA and persistence in terms of the average MSE, and in Figure 4 we show the maps of the MSE difference between S-ESN and persistence for 1 to 3 h ahead. We observe smaller errors for our forecasts than the persistence forecasts at the majority of locations. We also provide probabilistic forecasts of $Y_t(s)$ at all locations in 2016 based on the calibrated quantiles from 2015. Table S2 in the Supplementary Material shows the prediction interval coverage, and we observe that the uncertainty is well captured.
As mentioned in Section 3.3, the EOF approach is also commonly used to reduce dimensions (McDermott & Wikle, 2017). We denote this method by EOF-ESN, and we compare it with our method. We use $\mathbf{y}_t$ at all locations from 2013 to 2015 to calculate its covariance matrix and extract the first $n_{\text{EOF}} = 3173$ EOFs (principal components). We then reduce the dimension of $\mathbf{y}_t$ to $n_{\text{EOF}}$, apply the ESN to the projected vector, compute forecasts up to 3 h ahead, and recover the forecasted map at all locations by projecting the EOFs back into the original space. Results in Table 4 indicate better forecast skills from the S-ESN, which relies on spatial information to reduce dimensionality, as opposed to EOF-ESN, a data-driven, non-spatially informed method for dimension reduction. We reiterate that in Table 4, the persistence forecasting method relies on past values at all locations, whereas S-ESN and EOF-ESN only model the data in a reduced space. Despite the use of a more limited set of data, even the EOF-ESN is competitive against persistence: for 2 and 3 h ahead, it produces better forecasts.

Traditional spatio-temporal random process models can also be used to provide wind forecasts. However, given the high dimension in space and time, approximation techniques need to be applied for inference and prediction. We use the fixed rank kriging (FRK; Cressie & Johannesson, 2008) approach as one example for comparison. FRK proposes a spatio-temporal random effect model, where the spatio-temporal random process is decomposed by a series of basis functions, thus achieving dimensionality reduction and naturally accounting for nonstationarity. We use the recently developed R package FRK (Zammit-Mangion & Cressie, 2021) to model $\mathbf{y}_t$, and the results are shown in Table 4. FRK yields more accurate forecasts than persistence and EOF-ESN for 1 and 2 h ahead. We notice that the outperformance of FRK over EOF-ESN reduces as the lead time increases (FRK becomes worse than EOF-ESN for 3 h ahead), indicating that the temporal dynamic may not be well characterized by the spatio-temporal process. For all three forecast horizons, S-ESN and ARIMA have consistently more accurate results than FRK. One possible reason for the outperformance of ARIMA over FRK is that ARIMA computes the location-wise forecast using all the available data directly but FRK uses basis functions for the spatio-temporal data, which implies some level of approximation.

In conclusion, we stress the remarkable result of our proposed S-ESN approach: by only relying on data at the knots (about 6% of all the locations), S-ESN is able to outperform alternative methods, even for very short-term forecasts.

Once the wind speed forecasts are produced, we proceed to evaluate the associated predictions for wind power. While it may be interesting to perform a comparison over the entire spatial domain, the practical interest lies in the locations suitable for wind farm construction. In Section 6, we investigate the generated wind power for locations highlighted as suitable for wind farms by Giani et al. (2020). It is also noteworthy that MSE may not be a perfect metric to assess wind speed or associated wind power due to the possibly asymmetric economic cost between under- and over-forecasts. The evaluation can be performed differently by replacing the MSE with a case-specific commercial loss function.

5.2 Simulation study on a modified Lorenz 96 model

To better understand the ability of ESN to capture nonlinear dynamics, we also perform a simulation study from a model with a controlled level of nonlinearity in the temporal dynamics. We use one of the most popular models in numerical weather prediction, the Lorenz 96 model (Lorenz, 1996), and we modify it by adding a factor to control the nonlinearity in the dynamics. We have observed that ESN leads to the most accurate forecasts compared to the ARIMA, vector
autoregression and persistence methods. When the nonlinearity of the dynamics is stronger, the
outperformance of the ESN is more significant. Details of the simulation study for the modified
Lorenz 96 model are given in Section S3 in the Supplementary Material.

6 | WIND POWER FORECASTS IN SAUDI ARABIA

Once our forecasting method has been validated against existing alternatives, we proceed to pro-
duce the forecasts of wind power. From an operational perspective, the electricity produced by
wind power cannot be stored and must be consumed once it is injected into the power grid. An
accurate forecast of wind power helps plan the correct amount of additional electricity to be dis-
patched from other sources. Penalties or fines are applied in utility markets in case of unfulfilled
commitment to provide an agreed amount of power. The typical forecast horizon for schedul-
ing electricity transmission, allocating resources and dispatching generated power is 2 to 4 h
(Gneiting et al., 2006; Hering & Genton, 2010). Herein, we focus on 2-h-ahead wind forecasts.
We study the wind power forecasts based on the optimal wind farm build-out in Saudi Arabia as
shown in Giani et al. (2020) and assess how much power would be saved with our forecasts if all
the identified wind farms were installed and in operation.

The 2-h-ahead forecast of \( Y_t(s) \) is transformed into near-surface wind speed \( Z_t(s) \) according
to Equation (1). Since the wind speed is different at different heights, we convert the wind speed
near the surface (10 m above the ground) to the wind turbine hub height. The wind power law is
commonly used to perform this conversion:

\[
Z_t^{(h)}(s) = Z_t(s)(h/10)^{\alpha(s,t)},
\]

where \( h \) is the hub height and \( \alpha(s,t) \) is the location- and time-specific factor. Traditionally, the
factor \( \alpha(s,t) \) is fixed as a constant for all locations and times, and the value 1/7 is usually used
over open land surfaces (Pryor & Barthelmie, 2011) and specifically over Saudi Arabia (Rehman
et al., 2007; Tagle et al., 2019). However, Crippa et al. (2021) showed that if the location- and
time-specific factor \( \alpha(s,t) \) is independently estimated at each location from different profile
heights, the recovered wind speed at the hub height is on average 36% more accurate, so we use
their model to convert the wind speed vertically. It is noteworthy that this vertical extrapolation
model by the wind power law would lead to some errors. In this study, we do not take into account
the uncertainty of \( \alpha(s,t) \) for the probabilistic forecasts of wind power.

There are two turbine models chosen for the 75 identified optimal wind farms (Giani et al.,
2020): Nordex N131/3300 and GE Energy 2.75-100, with hub heights 84 and 75 m above the
ground respectively. We compute the wind speed for each wind farm at the turbine hub heights
and then use the power curve provided by the wind turbine manufacturers to convert the wind
speed to electric power (see Figure S4). Each curve has three critical points that divide the power
curve into four zones. When the wind speed is less than the cut-in speed, the turbine motor does
not rotate. When the wind speed is greater than the cut-in but less than the rated speed, the tur-
bine motor produces more power with faster wind. Once the wind speed reaches the rated speed
but not greater than the cut-out speed, the turbine keeps producing the maximal power output.
If the wind speed goes beyond the cut-out speed, then the turbine motor stops rotating to avoid
damage and produces no power. Therefore, better wind speed forecasts do not necessarily yield
better wind power forecasts, as their relationship is not strictly monotonic.
We also convert the forecasted near-surface wind speed by the ARIMA and persistence methods as well as the true wind speed at the 75 wind farms, and we compare their error in wind power forecasts against our approach. Note that the economic cost may be asymmetric for the under- and over-estimation of wind power. Under-forecast occurrence means that the power grid has ordered more electricity than needed, which results in a power surplus. Then, the power generation must be reduced, and it is usually more expensive than the opposite case. This asymmetry is case specific and can vary in different countries. For simplicity, we do not distinguish the direction of the wind power forecast error in this work. We calculate the absolute difference between the S-ESN (similarly, the ARIMA or persistence) wind power forecasts and the true power for each hour in 2016. The annual sum of the absolute differences in wind energy for all the 75 wind turbines is $2.77 \times 10^8$ kW·h for S-ESN, $3.05 \times 10^8$ kW·h for ARIMA and $3.12 \times 10^8$ kW·h for persistence. Thus, we obtain a 9% improvement against the ARIMA and an 11% improvement against the persistence forecasts. The bid price in the electricity market varies based on many factors (demand and supply, providers, etc.). Modelling the imbalance price is generally difficult and out of the scope of our work; see Zhu and Genton (2012) and Pinson (2013) for some discussion on the operational management in an electricity market. To gain some insights, using a medium bid price of 0.025 US dollars per kW·h nowadays in Saudi Arabia as a reference, the forecasts from our approach could yield a saving of nearly one million US dollars over a year against the persistence forecasts. When probabilistic forecasts are considered, Figure 5 shows the whole-year difference between the S-ESN forecast quantiles and the truth in wind energy. The difference curve around centre quantiles is quite flat. Specifically, the differences from 40% to 70% quantiles vary from $2.77 \times 10^8$ kW·h to $3 \times 10^8$ kW·h, and the maximum difference from 20% to 80% quantiles reaches $3.95 \times 10^8$ kW·h. It is also observed that the difference is inflated drastically for extreme quantiles. Finally, we notice that the wind energy forecast error is asymmetric in quantiles, owing to the characteristics of the power curve illustrated in Figure S4.

7 | DISCUSSION AND CONCLUSION

In this work, we have introduced a novel ESN framework for spatio-temporal wind data. The approach is shown to have superior predictive performances than the operational standards in
the wind energy market, contributes to the spatio-temporal model literature, and has practical implications for the wind energy sector in Saudi Arabia.

The proposed model uses the high dimensionality of neural networks to capture the non-linear dynamics of hourly wind and provides two solutions to ease the computational burden of the model. The paradigm of reservoir computing allows achieving sparsity in the RNN matrices, thereby dramatically reducing the parameter space; the use of spatial (problem-informed) knots allows reducing the spatial dimensionality while preserving a larger amount of information than previous approaches. Although our method shows high predictive power, it represents only one of the many possibilities available to bridge the gap between machine learning methods and spatio-temporal statistics. Many possible alternatives linking state-of-the-art geospatial models and neural network forecasting techniques are still unexplored.

This work presents a fast and efficient approach to hourly wind forecasting over large areas. More detailed quantification of the economic gains from an improved forecast is reliant upon a market model for trading and bidding energy. The energy sector of Saudi Arabia is currently under full governmental control; hence, the integration of our forecasting approach would require considerable speculation on the market model that will be employed in the future. Therefore, we choose not to pursue this route but rather opt for a simpler comparison against operational standards, such as persistence, under the current market prices in Saudi Arabia. Some of our current studies focus on engaging with local leaders in the energy sectors to detail the range of possible future market models that could be integrated with our approach. In more immediate terms, this work will provide additional information on where to build turbines aside from the wind abundance and construction and operation costs in Giani et al. (2020). Finally, while this work is reliant upon (validated) high-resolution WRF simulations, it represents a template to assess wind predictions over a large spatial area for developing countries in the Arab Gulf, and more broadly worldwide, with an emerging wind energy portfolio.

ACKNOWLEDGEMENTS
This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No: OSR-2018-CRG7-3742.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in KAUST Repository at https://urldefense.com/v3/__http://doi.org/10.25781/KAUST-SW314__!!N11eV2iwdfsls7vhACgBsn1amFuRrItB5EicSzWApyp9h86KEo7tiiK2yVDPumqSvzQYIkanc6xQKVJYRDUaeAlskkw$.

ORCID
Huang Huang https://orcid.org/0000-0002-5950-4698
Stefano Castruccio https://orcid.org/0000-0002-6728-965X
Marc G. Genton https://orcid.org/0000-0001-6467-2998

REFERENCES
Benjamin, S.G., Weygandt, S.S., Brown, J.M., Hu, M., Alexander, C.R., Smirnova, T.G. et al. (2016). A North American hourly assimilation and model forecast cycle: the rapid refresh. *Monthly Weather Review*, 144(4), 1669–1694.

Bonas, M. & Castruccio, S. (2021) Calibration of spatial forecasts from citizen science urban air pollution data with sparse recurrent neural networks. *arXiv preprint arXiv:2105.02971*.
McDermott, P.L. & Wikle, C.K. (2019b) Deep echo state networks with uncertainty quantification for spatio-temporal forecasting. *Environmetrics*, 30(3), e2553.

Mellor, G.L. & Yamada, T. (1982) Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, 20(4), 851–875.

Organisation for Economic Co-operation and Development (2020) The impact of coronavirus (COVID-19) and the global oil price shock on the fiscal position of oil-exporting developing countries. Available from: https://www.oecd.org/coronavirus/policy-responses/the-impact-of-coronavirus-covid-19-and-the-global-oil-price-shock-on-the-fiscal-position-of-oil-exporting-developing-countries-8ba9fb95

Paciorek, C. & Schervish, M. (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5), 483–506.

Pinson, P. (2013) Wind energy: forecasting challenges for its operational management. *Statistical Science*, 28(4), 564–585.

Pryor, S. & Barthelmie, R. (2011) Assessing climate change impacts on the near-term stability of the wind energy resource over the United States. *Proceedings of the National Academy of Sciences*, 108(20), 8167–8171.

R Core Team (2019) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/

Rehman, S., El-Amin, I., Ahmad, F., Shaahid, S., Al-Shehri, A. & Bakhashwain, J. (2007) Wind power resource assessment for Rafha, Saudi Arabia. *Renewable and Sustainable Energy Reviews*, 11(5), 937–950.

Risser, M.D. & Calder, C.A. (2017) Local likelihood estimation for covariance functions with spatially-varying parameters: the convoSPAT package for R. *Journal of Statistical Software*, 81(14), 1–32.

Rougier, J. (2016) Ensemble averaging and mean squared error. *Journal of Climate*, 29(24), 8865–8870.

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G. et al. (2008) A description of the advanced research WRF version 3. NCAR Technical Note TN-475+STR, National Center for Atmospheric Research.

Stein, M. (1999) *Interpolation of spatial data: some theory for Kriging*. New York: Springer.

Tagle, F., Castruccio, S., Crippa, P. & Genton, M.G. (2019) A non-Gaussian spatio-temporal model for daily wind speeds based on a multi-variate skew-t distribution. *Journal of Time Series Analysis*, 40(3), 312–326.

Taqqu, M.S., Teverovsky, V. & Willinger, W. (1995) Estimators for long-range dependence. *Fractals*, 3(4), 785–798. 

Titsias, M. (2009) Variational learning of inducing variables in sparse Gaussian processes. In: *Proceedings of the 12th international conference on artificial intelligence and statistics*, vol. 5, pp. 567–574.

Wikle, C.K. & Hooten, M.B. (2010) A general science-based framework for dynamical spatio-temporal models. *Test*, 19(3), 417–451.

Yehudai, G. & Shamir, O. (2019) On the power and limitations of random features for understanding neural networks. In: Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F., Fox, E. & Garnett, R. (Eds.) *Advances in neural information processing systems*, vol. 32, pp. 6598–6608.

Zammit-Mangion, A. & Cressie, N. (2021) FRK: an R package for spatial and spatio-temporal prediction with large datasets. *Journal of Statistical Software*, 98(4), 1–48.

Zhang, J., Crippa, P., Genton, M.G. & Castruccio, S. (2021) Assessing the reliability of wind power operations under a changing climate with a non-Gaussian bias correction. *Annals of Applied Statistics*, 15(4), 1831–1849.

Zhu, X. & Genton, M.G. (2012) Short-term wind speed forecasting for power system operations. *International Statistical Review*, 80(1), 2–23.

**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher's website.

---

**How to cite this article:** Huang, H., Castruccio, S. & Genton, M.G. (2022) Forecasting high-frequency spatio-temporal wind power with dimensionally reduced echo state networks. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 71(2), 449–466. Available from: https://doi.org/10.1111/rssc.12540