



# Methodological framework for short-and medium-term energy, solar and wind power forecasting with stochastic-based machine learning approach to monetary and energy policy applications

Tanveer Ahmad <sup>a, b,\*</sup>, Dongdong Zhang <sup>c</sup>, Chao Huang <sup>a, d, e</sup>

<sup>a</sup> State Key Laboratory of Internet of Things for Smart City and Department of Electrical and Computer Engineering, University of Macao, Macao, 999078, China

<sup>b</sup> Energy and Electricity Research Center, International Energy College, Jinan University, Zhuhai, Guangdong Province, 519070, China

<sup>c</sup> School of Electrical Engineering, Guangxi University, Nanning, 530004, China

<sup>d</sup> School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, 10083, China

<sup>e</sup> Shunde Graduate School, University of Science and Technology Beijing, Guangdong, 528399, China

## ARTICLE INFO

### Article history:

Received 30 January 2021

Received in revised form

3 April 2021

Accepted 7 May 2021

Available online 21 May 2021

### Keywords:

Renewable energy forecasting

Stochastic Gaussian process model

Machine learning

Artificial neural networks

Objective functions

Multi-objective optimization

## ABSTRACT

Anomalous seasons such as low-wind summers and extremely cold winters can seriously disrupt energy reliability and productivity. Better short/medium-term forecasts that provide reliable and strategic planning insights will allow the energy industry to plan for these extremes. In order to efficiently quantify uncertainty, this study proposes a Gaussian stochastic-based machine learning process model (GPR) for short/medium-term energy, solar, and wind (ESW) power forecasts using two different temporal resolutions of data. Four experimental steps (EXMS) were designed. Each EXMS is designed with four distinct fitting and predicting methods, and the GPR model uses seven kernel covariance functions for hyperparameter optimization. Real-time data is used for the forecasting analysis at three different locations. The forecasting results are validated using three existing models. The percent coefficient of variation of  $CV_{GPR1}$  and  $CV_{GPR2}$  of EXMS-1 and EXMS-3 for ESW power forecasts is 0.017%, 0.057%, 0.025%, and 0.223%, 0.225%, 0.170%, respectively. Accuracy has shown that the proposed model can predict ESW power simultaneously at two different temporal resolution data. The GPR accuracy with four EXMS methodologies is promising by addressing ESW power forecasts under the GPR framework of significant utilities, independent power producers, and public interest.

© 2021 Elsevier Ltd. All rights reserved.

## 1. Introduction

In view of the constant increase in energy demand as well as environmental challenges, renewable energy sources have been used extensively for generating electricity. RESs will account for two-thirds of the global energy supply by 2050, according to an International Renewable Energy Agency statement [1]. The energy industry and markets are volatile and sensitive to new technological innovations and breakthroughs and, changes in fiscal as well as monetary policies, consumer trend changes, and major global

events [2]. Many commercial and political organizations, think tanks, governmental agencies, and academics and researchers around the world consider numerous new energy policies as well as their impacts when addressing growing energy sustainability concerns, energy scarcity, pollution, and energy independence, all of which are majorly caused by the use of energy [3]. Energy has an impact on both individuals and institutions worldwide [4]. Therefore, it's essential to precisely quantify energy [5,6], solar [7], and wind power [8] to understand how important developments, energy, and monetary policies impact it [9].

Power/smart grids can decrease variable renewable generation operating effects using short/medium-term ESW forecasts [10]. Smart/intelligent wind turbines may have millisecond wind forecasts to maximize power production by modifying the turbine blade's pitch dynamically. Short/medium-term solar forecasts include ground-based sky imagery for cloud speed estimation and

\* Corresponding author. State Key Laboratory of Internet of Things for Smart City and Department of Electrical and Computer Engineering, University of Macao, Macao, 999078, China.

E-mail addresses: [tanveer.ahmad.pk11@gmail.com](mailto:tanveer.ahmad.pk11@gmail.com), [\(T. Ahmad\).](mailto:tahmad@um.edu.mo)

## Nomenclature

### Abbreviations

|          |  |
|----------|--|
| BR-BPNN  | Bayesian regularized backpropagation neural networks     |
| CV       | Coefficient of variation                                 |
| CI       | Computational intelligence                               |
| SCG-BPNN | Conjugate scale gradient backpropagation neural networks |
| DL       | Deep learning  |
| OFME     | Energy model objective function                          |
| ESW      | Energy, solar and wind                                   |
| EOMT     | Estimated objective minimum trace                        |
| EGPR     | Exact gaussian process regression                        |
| EXMS     | Experimental steps                                       |
| GPR      | Gaussian stochastic-based machine learning process model |
| M-Mean   | GPR mean values  |
| ITTF     | Iteration time trace function                            |
| LM-BPNN  | Levenberg–Marquardt backpropagation neural networks      |
| MAE      | Mean absolute error                                      |
| MWh      | Megawatt hour  |
| MMF      | Model minimum feasible point                             |
| OETT     | Objective evaluation time trace                          |
| EB-FV    | Objective function estimation                            |
| RTR      | Real-time reading  |
| SSE      | Residual sum of squares                                  |
| RMSE     | Root mean square error                                   |
| OFMS     | Solar model objective function                           |
| FICAM    | The fully independent conditional approximation method   |
| N-Point  | The next point   |
| SDPAM    | The subset of data points approximation method           |
| SRAM     | The subset of regressor approximation method             |
| OFMW     | The wind model objective function                        |

### Acronyms

|   |   |
|---|---|
| $\mathcal{A} \subset \mathcal{N} = \{1, 2, 3, \dots, o\}$ | Active set  |
| $z_{actual(m)}$   | Actual ESW forecasting                                      |
| $A_{density}(\text{kg/m}^3)$                              | Air density   |
| $T_{amb}(\text{°C})$                                      | Ambient temperature   |
| $f_{BEFP}$  | Best estimated feasible point                               |
| $P_{bcd}$   | Block coordinate descent                                    |
| $\sigma_m$  | Characteristics length scale of ESW                         |
| $C_{cov.}(\text{fraction})$                               | Cloud cover   |
| $C_{cov.}(\text{fraction})$                               | Cloud cover   |
| $CV_{GPR1}$   | Coefficient of variation of Gaussian regression process one |
| $CV_{GPR2}$   | Coefficient of variation of Gaussian regression process two |
| $Q(z_{new} z, Y, z_{new})$                                | Conditional probability function                            |
| $I(y, y' \theta)$   | Covariance function   |
| $o$   | Data samples  |
| $Q(z_j g(y_j), y_j)$                                      | Density function  |

|                                       |  |
|---------------------------------------|--|
| $T_{dpt}(\text{°C})$                  | Dew point temperature  |
| $S_{idiff}(\text{kW/m}^2)$            | Diffuse irradiance   |
| $S_{id}(\text{kW/m}^2)$               | Direct irradiance  |
| $T_{dbt}(\text{°C})$                  | Dry-bulb temperature   |
| $f_{EOFV}$                            | Estimated objective function value                                   |
| $s$                                   | Euclidean distance between $y_j$ and $y_k$                           |
| $E_{exact}$                           | Exact Gaussian process regression                                    |
| $z$                                   | Explicates the target values   |
| $f_m$                                 | Fit Method   |
| $z_{forecasting(m)}$                  | Forecasted ESW values  |
| $F_{fic}$                             | Fully independent conditional approximation                          |
| $\mathcal{A}$                         | Hilbert Spaces Kernel replication space                              |
| $H_r$                                 | Hours  |
| $H_r(\%)$                             | Humidity ratio   |
| $k_{kf}$                              | Kernel function  |
| $\beta$                               | Knowledge of coefficients of fixed basis function                    |
| $g_{new}$ and $g$ to $z_{new}$ to $z$ | Latent variables   |
| $\sigma_g^2$                          | Length scale characteristics   |
| $d_k s \in \mathbb{R}$                | Linear combination coefficients for approximation                    |
| $I(y, y_s \theta)$                    |  |
| $M_{52}$                              | Matern kernel with parameter 5/2                                     |
| $ARD_{32}$                            | Matern kernel with variable 3/2                                      |
| $M_{32}$                              | Matern Kernel with variable 3/2 function                             |
| $D$                                   | Matrix which comprises all $d_{ks}$ coefficients                     |
| $\mu$                                 | Mean value   |
| $f_{OBFV}$                            | Observed objective function value                                    |
| $P_{prec}(\text{mm/h})$               | Precipitation  |
| $P_m$                                 | Predict method   |
| $P_{24-h-l}(\text{MWh})$              | Previous 24-h energy consumption                                     |
| $P_{dl}(\text{MWh})$                  | Previous day energy consumption                                      |
| $P_{dwp}(\text{kW})$                  | Previous day wind speed  |
| $P_{wl}(\text{MWh})$                  | Previous week energy consumption                                     |
| $P_{week,wk}(\text{kW})$              | Previous week wind speed   |
| $R_{toe}(\text{kW/m}^2)$              | Radiation toa  |
| $AQ_{quadratic}$                      | Rational quadratic kernel with a separate length scale per predictor |
| $\mathcal{N} = \{1, 2, 3, \dots, o\}$ | Set of all energy, solar and wind observation                        |
| $B_{size}$                            | Size of the block used as a comma-separated pairs                    |
| $S_{mass}(\text{kg/m}^2)$             | Snow mass  |
| $S_{fall}(\text{mm/h})$               | Snowfall   |
| $S_{sr}$                              | Subset of regressors approximation                                   |
| $\mu_{TS}, \sum TS$                   | Subset of regressors to $\mu$ and $\sum$                             |
| $R_{surf.}(\text{kW/m}^2)$            | Surface radiation  |
| $f_{TET}$                             | Total elapsed time   |
| $E_{total}(\text{MWh})$               | Total energy consumption   |
| $f_{TFE/I}$                           | Total function evaluations/iterations                                |
| $f_{TOFET}$                           | Total objective function evaluation time                             |
| $S_{power}(\text{kW})$                | Total solar power production   |
| $W_{power}(\text{kW})$                | Total wind power production  |
| $Y$                                   | Training values of input feature variables                           |
| $\sigma^2$                            | Variance noise   |
| $T_{wbt}(\text{°C})$                  | Wet bulb temperature   |
| $W_s(\text{m/s})$                     | Wind speed   |

performance [11,12]. In general, increasing the grid's ability to forecast will lead to more flexible energy markets and versatile day-to-day dispatch in real time [13].

The challenge is evident in the recent and ongoing rise in the energy and renewable energy industry [14]: how can this intermittent power source be embedded in the power/smart grid [15]?

Better forecasts will position ESW on the global energy balance for penetration and sustainable growth [16]. The implications of the value of the forecast are: i) decreased penalties and charges; ii) competitive real-time benefit of information and resources trading on the “day ahead”; iii) more efficient project operations, construction, and maintenance planning [17].

Gabriel et al. [18] conducted research which provides further data to assist strategic decision by discussing the use of ML algorithms to create more reliable predictions in the international energy market. Tanveer et al. [19] proposed four ML models for energy demand forecasting for climate-sensitive and non-climate sensitive conditions. Another study on ML models for Summer, Autumn, Winter, and Spring season load demand forecasting has been conducted which gives the least mean absolute percentage error of 0.061%, 0.010%, 0.005% and 0.049%, respectively [20]. The key concern about the growth of wind power generated in electrical systems is to preserve the electrical system stability, as wind power output will swing from 20% to 100% of the wind farm capacity [21,22]. In contrast, the economical approach depends on a more reliable wind energy prediction to help schedule the short run, leading to the optimization of storage decisions when the country has the largest hydro power plants. Photovoltaic power generation highly accurate forecasts are the basis for the energy generation, and transmission, maintaining continuity and sustainability of electricity systems [23].

The key goal is to deliver a better forecast with better accuracy based on recent technologies [9]. The technical reliability of ESW prevision is evolving dynamically. The key drivers behind innovation are market forecast service providers trying to constantly develop their forecasting methodologies because of their highly competitive situation. The criteria for successful ESW forecasting are: i) reliability (e.g., use of real-time climate data) [24]; ii) usefulness (e.g., model should be used successfully) [25]; iii) simplicity (e.g., simple system) [26]; iv) cost-effectiveness (e.g., affordable) [27]; v) importance (e.g., highly reliable for the desired tasks) [28]; and vi) multiple purpose applicability (e.g., decision-making and monitoring tool for several applications) [29]. Existing forecasting models are most useful for specific tasks, for illustrative purposes, for temporal resolution of data (e.g., 5 min, 15 min, 30 min, and 1 h), for specific locations, and for specific applications (e.g., only for energy, solar, and wind forecasting). Only very few models were built to forecast ESW at the same time. But their accuracy in forecasting these tasks is not up to date. This study proposed a GPR-based machine learning model with a different combination of "Fit Methods" and "Predict Methods" for ESW power forecasting at the same time. The study differs from existing studies in terms of accuracy, reliability, forecasting tasks, and time-series temporal resolution of ESW power data.

### 1.1. Background and related literature

Two key methods have appeared in the short history of the ESW power forecasting: *i) physical approach*, and *(ii) statistical approach* [30]. Physical-based approaches used irradiance transfer and meteorology schemes to measure the meteorological inputs, i.e., speed of wind at the wind hub height, and then use power curves to transfer it into power [31]. Physical models have been largely used for global solar irradiance forecasting [23]. More precise predictive accuracy can be achieved with the use of physical models if weather conditions are constant, while accuracy can not be assured or accurately forecasted when weather conditions change significantly. The statistical approaches are mathematical relation modeling between numerical forecasting as measured power output and input is then "learned" or "trained" based on available data [32]. These methods constitute the basis for both predictions, and interference [33,34]. Modern forecasting systems have recently been introduced, which illustrate that the two approaches converge to integrate physical and statistical methods as necessary to obtain greater accuracy [30].

ML models, which are developed based on statistical methods, have been well recognized in the area of engineering and science. It

follows a method of data preparation, algorithm formation, ML model generation and prediction development and further refining. Largely applied ML models for energy, solar, and wind power forecasting include support vector machine, extreme learning machine, and artificial neural networks. Examples of hybrid models used in energy, wind, and solar power forecasting include analog ensemble, wavelet transform, particle swarm optimization, and generative adversarial networks, and convolutional neural networks methods. Each method of ML has particular benefits and requirements for implementation [35,36].

The popular forecasting approaches are: autoregressive integrated moving average [37], support vector machine [38], genetic algorithm [39], fuzzy logic [40], particle swarm optimization [41], linear regression [42], grey prediction, and autoregressive moving average [43]. Concerning precision, computational intelligence methods are better than statistical methods, particularly for the parameters with higher variability in the source data [44]. Hybrid approaches have greater precision than stand-alone methods [45]. Statistical techniques are only applied for the short to medium range, while computational intelligence methods for all time scales are preferable (short-medium-long term).

The artificial neural network is already used in 40% of the ESW power forecasting tasks [43]. The expert systems are mostly used for short-term forecasting [46,47]. Mathematical programming gives the best solution for ESW power forecasting from a set of available data under some circumstances. Nonlinear programming is the best example of mathematical modeling for forecasting. The disadvantages of mathematical modeling are hard to implement in real-time applications. The hybrid models include: i) statistical-statistical methods, ii) statistical-computational intelligence (CI) methods, ii) CI–CI methods, and iv) statistical-MP methods. Each model has further sub-classes of forecasting tasks, and further details can be found in Ref. [43]. These methods were used to develop the parameters and assumptions in different forecasting models.

Deep learning (DL) is a promising name for forecasting tasks [48,49]. The basic structure of deep learning includes stacked auto-encoder [50], deep recurrent neural network [51,52], deep belief network [53], deep convolutional neural network [54], generative adversarial networks [55], stacked extreme learning machine [56], etc. The DL forecasting includes data preprocessing methods, probabilistic forecasting methods, deterministic forecasting methods, feature extractor based on DL, error post-processing methods, and parameters tuning and network structure optimization [43]. The DL models have two major challenges, including modeling problems and theoretical issues [57]. However, the main challenge is how to construct a hierarchical model of strong functional learning. One of the difficulties to face is setting the most fitting DL predictive models for a particular forecast data set.

With the popularity of ESW in the world's electricity network, increased precision of renewable power operation, planning, and management are key to energy system planning. This is a challenging task because renewable energy is unreliable and chaotic [57]. Different methodologies have been proposed to improve renewable forecasts' accuracy, amongst others, physical models, computational methods, and artificial intelligence techniques. The GPR model is a promising form of machine learning capable of finding the underlying nonlinear features and high-level invariant ESW data structures. Our proposed ESW forecasting GPR model differs from the existing studies listed above in terms of hyper-parameter optimization, used temporal resolution of data, different forecasting tasks, accuracy, kernel covariance function of the network fitting approximation for four proposed EXMS. The major contribution of this study is demonstrated in section 1.2.

## 1.2. Contribution of this study

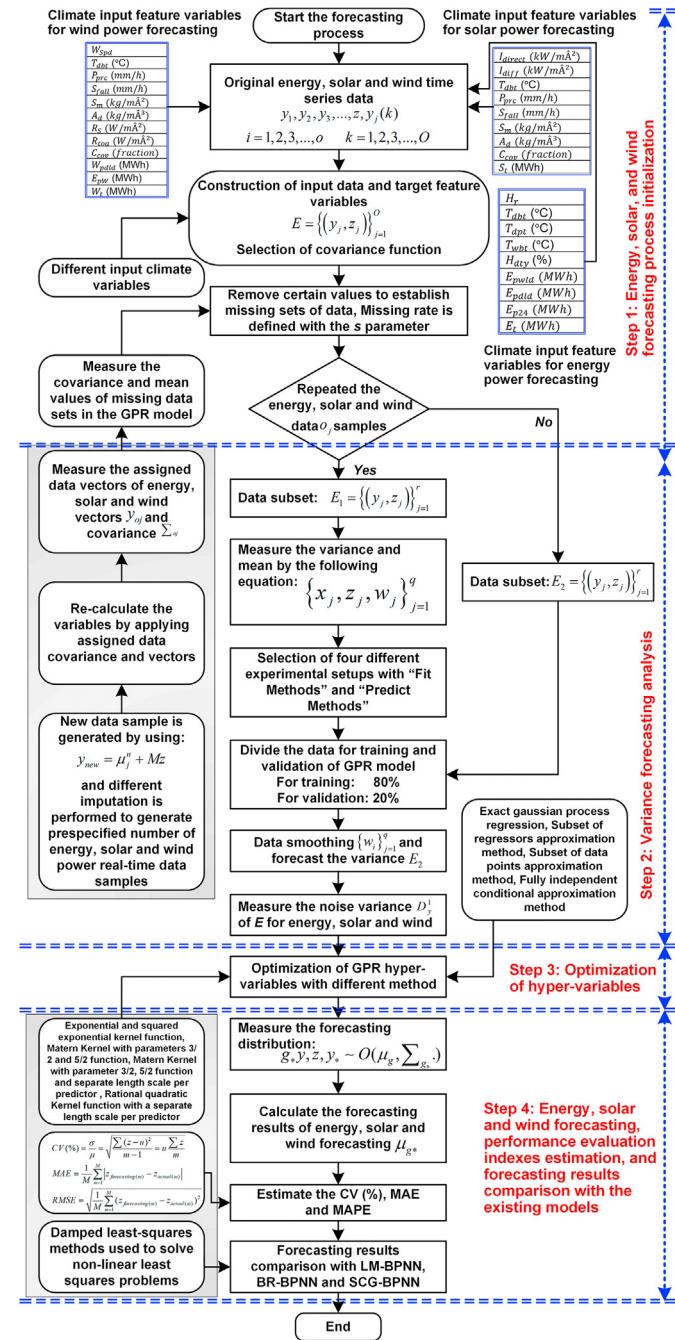
The contribution of this study is described in the following points:

- 1) This paper proposed a stochastic-based ML approach for forecasting short/mid-term energy, solar, and wind, that can be used to help decision-making in control programs and energy flexibility initiatives approach to monetary and energy policy applications. Real-time data is taken from a closed source at three different locations. The results demonstrate that the proposed model is feasible and renders an accurate short/mid term energy, solar, and wind power forecasting for three different regions.
- 2) The GPR model is divided into four EXMS for the ESW forecast. Each EXMS consists of both the “Fit Method” and the “Predict Method.” In addition, the “Fit Method” and “Predict Method” used four GPR forecasting methods (e.g., “exact”, “sr”, “sd”, and “fic”) and seven kernel covariance methods. Seven distinct kernel functions are compared, and we choose the perfect kernel for the GPR method. The squared-exponential kernel function is used with a separate-scale-length for the predictors of each method. Hyperparameters are calculated to reduce the five-fold cross-validation losses by applying automatic-hyperparameter optimization.
- 3) The performance of four EXMS is calculated using two temporal resolutions of forecasting intervals, namely 30 min and 1 h. The energy forecasting analysis is performed for short/medium-term forecasting of EWS power. The proposed GPR and four EXMS give advantages in terms of low forecasting error for EWS forecasting at the same time as compared to the commonly used models in a different number of timescales.
- 4) The accuracy of GPR and four EXMS are compared to three existing artificial neural networks and four performance assessment indexes. The results obtained from the ESW forecast explain that the quality of the GPR model with the proposed four EXMS models is significantly better than the existing models.

The rest of this study is organized into the following sections. Section 2 provides an overview of the construction of the machine learning model and the proposed methods. It is further extended with the sections “Fit Methods” and “Predict Methods” with different kernel covariance functions. Four experimental setups, performance evaluation indexes, and a detailed description of the data source are also provided in section 2. Section 3 provides a complete and thorough analysis of the GPR forecast for ESW. Forecasting results validation, discussion, benefits, challenges, and real-time applications are part of section 3. Section 4 concludes this study.

## 2. Construction of the machine learning model and the methods proposed

The construction of the proposed ML model and the proposed ESW forecasting methods are briefly discussed in this section. These methods are divided into four basic components/steps, including: i) Fit methods for estimating GPR model parameters and initialization of the ESW forecasting process; and ii) Kernel covariance function and variance of the forecasting analysis; iii) Experimental setup of ESW demand forecasting with the use of hyper-variable optimization; and iv) Performance assessment and comparison of results with existing methods of artificial neural networks. Fig. 1 visualized the structure of the proposed ESW forecasting method. The parameters  $y_1, y_2, y_3, \dots, o$  explicate the



**Fig. 1.** Structure of the proposed ESW forecasting method.

input feature variables, and  $z$  explicate the target variables. The parameter  $i$  represents the total number of data observations or data samples.  $E_1$  and  $E_2$  are a subset of data sets. A detailed analysis of each part is given below.

### 2.1. Fits and predicts methods for estimating GPR model parameters

The fitness and predictability methods for estimating the GPR model parameter and making forecasts are discussed in this section. The fit methods are used to calculate the weights/parameters of the training data, e.g., the parameters restored by the GPR model coefficient and saved as an internal state object. Forecast methods are used to measure the weights of the test data sets in order to

make accurate predictions. More details and analysis are provided below.

### 2.1.1. Exact Gaussian process regression

It is essential to overcome the space and time requirement of solving linear systems to execute the exact gaussian process regression (EGPR) interface on a large number of ESW datasets. In order to solve linear structures for the inference method, most EGPR deployments use the Cholesky decomposition [58]. An instance of target value  $z$  from EGPR can be modeled as

$$Q(z_j|g(y_j), y_j) \sim O\left(z_j|z(y_j)^U \beta + g(z_j), \sigma^2\right) \quad (1)$$

However, making forecasting for ESW, the EGPR needs:

- The value of  $\beta$  which represents the knowledge of coefficients of fixed basis function.
- The evaluation of covariance function  $l(y, y'|\theta)$  for arbitrary  $y$  and  $y'$ , rendering the EGPR kernel hyperparameters  $\theta$ .
- The information of variance noise  $\sigma^2$  which appears in the density function  $Q(z_j|g(y_j), y_j)$ .

First, it is important to calculate the values of  $\beta, \theta$ , and  $\sigma^2$  from the input feature variables  $Y, z$ .  $Y$  represents the training values of input feature variables (in case, ESW) and  $z$  explicates the target values. Making ESW forecasting from EGPR with known input feature variables needs density  $Q(z_{new}|z, Y, z_{new})$ . The values of  $Q(z_{new}|z, Y, z_{new})$  is taken from the conditional probability function:

$$Q(z_{new}|z, Y, z_{new}) = \frac{Q(z_{new}|z, Y, z_{new})}{Q(z|Y, y_{new})} \quad (2)$$

The density of the numerator is taken from the latent variables  $g_{new}$  and  $g$  to  $z_{new}$  to  $z$ , respectively. Based on these values, it is possible to measure the joint distribution for  $g_{new}$  and  $g$  to  $z_{new}$  to  $z$  to measure the  $Q(z_{new}|z, Y, z_{new})$ :

$$Q(z_{new}|z, Y, z_{new}) = \int \int Q(z_{new}, z|g_{new}, g) Q(z_{new}|z, Y, z_{new}) \quad (3)$$

$$Q(z_{new}|z, Y, z_{new}) = \int \int Q(z_{new}, z|g_{new}, g, Y, z_{new}) \\ Q(g_{new}, g|Y, y_{new}) dfdf_{new} \quad (4)$$

Hence, the large number of estimation parameters (for illustrate,  $o$  data samples), computing forecasting can be very expensive. The advantages of EGPR normally rearrange the computation to neglect the computing cost.

### 2.1.2. Subset of the regressors approximation method

The subset of regressors approximation method (SRAM) comprised of replacing the network kernel function  $l(y, y_s|\theta)$  in the EGPR method by using it is an approximation  $\hat{l}_{TS}(y, y_s|\theta, A)$ , the active set which is given  $\mathcal{A} \subset \mathcal{N} = \{1, 2, 3, \dots, o\}$ . For the EGPR, the expected forecasting in GPR relays on the number of  $\mathcal{N}$  functions  $\zeta_{\mathcal{N}} = \{l(y, y_j|\theta), j = 1, 2, 3, \dots, o\}$ , where  $\mathcal{N} = \{1, 2, 3, \dots, o\}$  is the indices of set of all ESW observation, or data samples. The basic objective is to span approximation of these function from a smaller function set  $\mathcal{S}_{\mathcal{A}}$ , where  $\mathcal{A} \subset \mathcal{N} = \{1, 2, 3, \dots, o\}$  indices subset of all data points to be used in the active set considering  $\zeta_{\mathcal{A}} = \{l(y, y_j|\theta, j = \mathcal{A}\}$ . The basic objective behind this was to approximate

the  $\zeta_{\mathcal{N}}$  as a linear elements combination of  $\mathcal{S}_{\mathcal{A}}$ . The approximation supposes  $l(y, y_s|\theta)$  taking the function  $\zeta_{\mathcal{A}}$  is as follows:

$$\hat{l}(y, y_r|\theta) = \sum_{k \in \mathcal{A}} d_{ks} l(y, y_k|\theta) \quad (5)$$

$d_{ks} \in \mathbb{R}$  represents the linear combination coefficients for approximation  $l(y, y_s|\theta)$ . For example,  $D$  represents the matrix which comprises all  $d_{ks}$  coefficients. We consider  $D$  is a  $|\mathcal{A}| \times o$  matrix like that  $D(l, s) = d_{ks}$ . The MATLAB is found the best fit for elements  $\zeta_{\mathcal{N}}$  by applying the function of active set  $\mathcal{A} \subset \mathcal{N} = \{1, 2, 3, \dots, o\}$  and decreasing the error function of the network.

$$F(\mathcal{A}, D) = \sum_{s=1}^o l(y, y_s|\theta) - \hat{l}(y, y_s|\theta)_{\mathcal{A}}^2 \quad (6)$$

$\mathcal{A}$  is the Hilbert Spaces kernel replication space connected with kernel function  $l$  [58].

The following equation is the coefficient of the matrix which reduces the  $F(\mathcal{A}, D)$  is:

$$\hat{D}_{\mathcal{A}} = L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta)^{-1} L(Y_{\mathcal{A}}, Y|\theta) \quad (7)$$

The kernel function approximation applying the active elements set  $\mathcal{A} \subset \mathcal{N} = \{1, 2, 3, \dots, o\}$  is:

$$\hat{l}(y, y_r|\theta) = \sum_{k \in \mathcal{A}} d_{ks} l(y, y_k|\theta) = L(y^U, Y_{\mathcal{A}}|\theta) D(:, s) \quad (8)$$

The subset of regressors approximation to the SRAM based kernel approach applying the active data set of ESW is  $\mathcal{A} \subset \mathcal{N} = \{1, 2, 3, \dots, o\}$  explicated as:

$$\hat{l}_{TS}(y, y_s|\theta, \mathcal{A}) = L(y^U, Y_{\mathcal{A}}|\theta) \hat{D}_{\mathcal{A}}(:, s) = L(y^U, Y_{\mathcal{A}}|\theta) L(Y_A, Y_A|\theta)^{-1} \\ L(Y_{\mathcal{A}}, y|\theta) L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta) \quad (9)$$

The regressor approximation to  $L(Y, Y|\theta)$  is defined as

$$\hat{L}_{ST}(Y, Y_s|\theta, A) = L(Y, Y_{\mathcal{A}}|\theta) L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta)^{-1} L(Y_{\mathcal{A}}, Y|\theta) \quad (10)$$

The disadvantages of SRAM are that it gives a small predictive variance resulting from an error in large-scale forecasting datasets. This limitation is overcome by using prior distribution.

### 2.1.3. Datapoint approximation method subset

The subset of data points approximation method (SDPAM) is used as the subset of regressors to the ESW data distribution  $z_{new}$  given  $z, Y, z_{new}$  is:

$$Q(z_{new}|z, Y, z_{new}) = \mathcal{N}(z_{new}|y(z_{new})^U \beta + \mu_{TS}, \delta^2_{new} + \sum TS) \quad (11)$$

From equation (11),  $\mu_{TS}$  and  $\sum TS$  are a subset of regressors to  $\mu$  and  $\sum$ . The values of  $\mu_{TS}$  and  $\sum TS$  are taken by replacing  $l(y, y_s|\theta)$  by it is a subset of regressors approximation  $\hat{l}_{TS}(y, y_s|\theta, \mathcal{A})$  in  $\mu$  and  $\sum$ , respectively, as given in Section 2.1.2. The  $\mu_{TS}$  is measured by the following equation:

$$\mu_{TS} = \underbrace{\hat{L}_{TS}(y_{new}^U, Y|\theta, \mathcal{A})}_{(1)} \underbrace{\left( \hat{L}_{TS}(Y, Y|\theta, \mathcal{A}) + \sigma^2 J_0 \right)^{-1}}_{(2)} (z - I\beta) \quad (12)$$

Since the first part (1) of equation (12) can be written as:

$$(1) = L_{TS}(y_{new}^U, Y|\theta, \mathcal{A}) L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta)^{-1} L(Y_{\mathcal{A}}, Y|\theta) \quad (13)$$

The second part (2) of equation (12) can be written as:

$$(2) = \frac{J_0}{\sigma^2} - \frac{L(Y_{\mathcal{A}}, Y|\theta)}{\sigma^2} \left[ L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta) + \frac{L(Y_{\mathcal{A}}, Y|\theta)L(Y, Y_{\mathcal{A}}|\theta)}{\sigma^2} \right]^{-1} \frac{L(Y_{\mathcal{A}}, Y|\theta)}{\sigma^2} \quad (14)$$

It is a fact, that  $J_0 - C(B+C)^{-1} = B(B+C)^{-1}$  based on it, equation (12),  $\mu_{TS}$  value is defined as:

$$\mu_{TS} = L_{new}(y_{new}^U, Y_{\mathcal{A}}|\theta) \left[ L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta) + \frac{L(Y_{\mathcal{A}}, Y|\theta)L(Y, Y_{\mathcal{A}}|\theta)}{\sigma^2} \right]^{-1} \frac{L(Y_{\mathcal{A}}, Y|\theta)}{\sigma^2} (z - I\beta) \quad (15)$$

Similarly, the value of  $\sum TS$  is defined as follows:

$$\sum TS = \underbrace{\hat{L}_{TS}(y_{new}, y_{new}|\theta, \mathcal{A})}_i - \underbrace{\hat{L}_{TS}(y_{new}^U, Y|\theta, \mathcal{A})}_{ii} \underbrace{\hat{L}_{TS}(Z, Z|\theta, \mathcal{A} + \sigma^2 J_0)^{-1}}_{iii} \underbrace{\hat{L}_{TS}(Y, y_{new}^U|\theta, \mathcal{A})}_{iv} \quad (16)$$

Due to

$$ii = L(y_{new}^U, Y_{\mathcal{A}}|\theta) L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta)^{-1} L(Y_{\mathcal{A}}, Y|\theta) \quad (17)$$

iii = (2) in equation  $\mu_{TS}$

$$iv = L(Y, Y_{\mathcal{A}}|\theta) L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta)^{-1} L(Y_{\mathcal{A}}, y_{new}^U|\theta)$$

The value of the  $\sum TS$  is measured from the following equation:

$$\sum TS = L(y_{new}^U, Y_{\mathcal{A}}|\theta) \left[ L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta) + \frac{L(Y_{\mathcal{A}}, Y|\theta)L(Y, Y_{\mathcal{A}}|\theta)}{\sigma^2} \right]^{-1} L(Y_{\mathcal{A}}, y_{new}^U|\theta) \quad (18)$$

The SDPAM method is simple, flexible to implement, as well as suitable for a wide range of issues in classification and regression. The proof structure given above and applied to the SDPAM makes it easy to apprehend the kernel function's ESW parameters.

#### 2.1.4. Fully independent conditional approximation method

The fully independent conditional approximation method (FICAM) approach is a comprehensive technique for approximating true GPR kernel capacity in such a way as to minimize predictive-variance problems of subset of regressors data approximation while still providing a correct Gaussian method. The FICAM to  $I(y_j, y_k|\theta)$  for a network active set  $\mathcal{A} \subset \mathcal{N} = \{1, 2, 3, \dots, o\}$  is a rendering of the following equation:

$$\begin{aligned} \hat{l}_{GJD}(y_j, y_k|\theta, \mathcal{A}) &= \hat{l}_{TS}(y_j, y_k|\theta, \mathcal{A}) \\ &+ \mathcal{I}_{jk}(y_j, y_k|\theta) - \hat{l}_{TS}(y_j, y_k|\theta, \mathcal{A}) \end{aligned} \quad (19)$$

$$\mathcal{I}_{jk} \begin{cases} 1 & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \quad (20)$$

The approximation of FICAM is almost equal to the subset of regressors if the values of  $j \neq k$ . If  $j = k$ , the MATLAB used the fit kernel value rather than an approximation. The diagonal matrix  $\mathcal{Q}(Y|\theta, \mathcal{A})$  defined the number is defined as:

$$[\mathcal{Q}(Y|\theta, \mathcal{A})]_{jk} = \mathcal{I}_{jk} \left( (y_j, y_k|\theta) - \hat{l}_{TS}(y_j, y_k|\theta, \mathcal{A}) \right) \quad (21)$$

$$[\mathcal{Q}(Y|\theta, \mathcal{A})]_{jk} = \begin{cases} \mathcal{I}_{jk} \left( (y_j, y_k|\theta) - \hat{l}_{TS}(y_j, y_k|\theta, \mathcal{A}) \right) & \text{if } j = k \\ 0 & \text{if } j \neq k \end{cases} \quad (22)$$

The FICAM approximation to  $L(Y, Y|\theta)$  is taken by the following equation:

$$\hat{l}_{GJD}(y_j, y_k|\theta, \mathcal{A}) = \hat{l}_{TS}(Y, Y|\theta, \mathcal{A}) + \mathcal{Q}(Y|\theta, \mathcal{A}) \quad (23)$$

$$\begin{aligned} \hat{l}_{GJD}(y_j, y_k|\theta, \mathcal{A}) &= L(Y, Y_{\mathcal{A}}|\vartheta) L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta)^{-1} L(Y_{\mathcal{A}}, Y|\vartheta) \\ &+ \mathcal{Q}(Y|\theta, \mathcal{A}) \end{aligned} \quad (24)$$

For FICAM approximation distribution of  $z_{new}$  given by  $z$ ,  $Y$ , and  $z_{new}$  is

$$Q(z_{new}|z, Y, y_{new}) = \mathcal{A} \left( z_{new} | i(y_{new})^T \beta + \mu_{GJD}, \sigma_{new}^2 + \sum_{GJD} * \right) \quad (25)$$

$\mu_{GJD}$  and  $\sum_{GJD} *$  are the FICAM approximation to  $\mu$  and  $\sum$  which is rendered in forecasting using the EGPR approach. In a subset of regressors case,  $\mu_{GJD}$  and  $\sum_{GJD} *$  are taken by replacing and using all occurrences of right (true value) kernel with it is FICAM data approximation. The overall final conditions or forms of  $\mu_{GJD}$  and  $\sum_{GJD} *$  are acquired from the following equation:

$$\mu_{GJD} = L(y_{new}^U, Y_{\mathcal{A}}|\theta) \beta_{\mathcal{A}}^{-1} L(Y_{\mathcal{A}}, Y|\theta) A(\theta, \sigma^2, \mathcal{A})^{-1} (z - I\beta) \quad (26)$$

$$\begin{aligned} \sum_{GJD} * &= l(y_{new}, y_{new}|\theta) - L(y_{new}^U, Y_{\mathcal{A}}|\theta) L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta)^{-1} L \\ &\quad \left( Y_{\mathcal{A}}, y_{new}^U \middle| \theta + L(y_{new}^U, Y_{\mathcal{A}}|\theta) \beta_{\mathcal{A}}^{-1} L(Y_{\mathcal{A}}, y_{new}^U|\theta) \right) \end{aligned} \quad (27)$$

Where

$$\beta_{\mathcal{A}} = L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta) + L(Y_{\mathcal{A}}, Y|\theta) A(\theta, \sigma^2, \mathcal{A})^{-1} L(Y_{\mathcal{A}}, Y_{\mathcal{A}}|\theta) \quad (28)$$

$$A(\theta, \sigma^2, \mathcal{A}) = Q(Y|\theta, \mathcal{A}) + \sigma^2 J_0 \quad (29)$$

The FICAM is flexible, simple to implement, and fully probabilistic. This is a powerful approach applied in many applications [59]. Our proposed FICAM is a simple but powerful algorithm for ESW forecasting tasks.

## 2.2. Kernel covariance function

In statistics and mathematical theory, covariance calculates the combining of two parameters (input and target parameters) and how often two of the parameters mutate in a random parameter is represented by a kernel function. In the GPR model, the kernel covariance function is used to express the similarity between input and target variables [58]. For illustrate, in the GPR model, the input predictor values  $y_j$  normally have a close response to the target values  $z_j$ . It is further used to specify the covariance between to lantern parameters  $g(y_j)$  and  $g(y_k)$  and  $y_j, y_k$  are  $e - by - 1$  vectors. It is used to determine if the response values at a single point  $y_j$  are changes/affected by other targets/response points, then  $y_k, k \neq j, j = 1, 2, 3, \dots, e$ . The kernel covariance function  $I(y_k, y_j)$  can be defined by various kernel covariance functions. We used seven different kernel covariance functions in this study. These kernel covariance functions include: i) Exponential covariance function ii) Squared-exponential kernel function; iii) Matern-kernel with parameters 3/2; iv) Matern-kernel with parameters 5/2; v) Matern-kernel with parameters 3/2 and separate length per predictor; vi) Matern-kernel with parameters 5/2 and separate length per predictor; and vii) Rational-quadratic-kernel function with a separate length scale per predictor [60], and detail of each covariance function is described below.

### 2.2.1. Exponential and squared exponential kernel functions

The exponential function of the kernel is defined in the following equation:

$$I(y_j, y_k | \theta) \sigma_g^2 \exp\left(-\frac{s}{\sigma_m}\right) \quad (30)$$

Where  $\sigma_m$  explicates the characteristics length scale of ESW data samples, and

$$s = \sqrt{(y_j - y_k)^U (y_j - y_k)} \quad (31)$$

Parameter  $s$  represents the Euclidean distance between  $y_j$  and  $y_k$ .

The squared-exponential-kernel function is the most widely applied kernel covariance function for the GPR method. The following equation defines it:

$$I(y_j, y_k | \theta) = \sigma_g^2 \exp\left[-\frac{1}{2} \frac{(y_j - y_k)^U (y_j - y_k)}{\sigma_m^2}\right] \quad (32)$$

In equation (32),  $\sigma_g^2$  explicates the length scale characteristics.

### 2.2.2. Matern kernel with 3/2 and 5/2 parameters

Maternal kernel covariance with 3/2 parameters is defined by the following equation:

$$I(y_j, y_k | \theta) = \sigma_g^2 \exp\left(1 + \frac{\sqrt{3}s}{\sigma_m}\right) \exp\left(-\frac{\sqrt{3}s}{\sigma_m}\right) \quad (33)$$

The Matern kernel covariance with parameters 5/2 is defined by the following equation:

$$I(y_j, y_k | \theta) = \sigma_g^2 \exp\left(1 + \frac{\sqrt{5}s}{\sigma_m} + \frac{5r^2}{3\sigma_m^2}\right) \exp\left(-\frac{\sqrt{5}s}{\sigma_m}\right) \quad (34)$$

Euclidean distance between  $y_j$  and  $y_k$  is taken  $s = \sqrt{(y_j - y_k)^U (y_j - y_k)}$  as given in equation (31).

### 2.2.3. Matern kernel with 3/2, 5/2 parameter function and separate-length-scale for each predictor

Matern-kernel with variables 3/2, 5/2, and separate-length-scales per predictor is defined in the following equations:

$$I(y_j, y_k | \theta) = \sigma_g^2 (-s) \quad (35)$$

The equation (35) shows the Matern-kernel with variable 3/2, and 5/2 is defined as the following equation:

$$I(y_j, y_k | \theta) = \sigma_g^2 (1 + \sqrt{3} s) \exp(-\sqrt{3}s) \quad (36)$$

In this scenario,  $s$  is defined from the following equation:

$$s = \sqrt{\sum_{n=1}^e \frac{(y_{kn} - y_{jn})^2}{\sigma_n^2}} \quad (37)$$

### 2.2.4. Rational-quadratic kernel function with a separate-length-scale for each predictor

Rational-quadratic kernel function with a separate-length-scale for each predictor is defined in the following equation:

$$I(y_j, y_k | \theta) = \sigma_g^2 \left(1 + \frac{1}{2\alpha} \frac{(y_{jn} - y_{jn})^2}{\sigma_n^2}\right)^{-\alpha} \quad (38)$$

The rational quadratic kernel function allows different length scales to be used  $\sigma_n^2$  for every forecasted  $n, n = 1, 2, 3, \dots, e$ , the more detail is given in Ref. [61]. The value of unconstrained parameterization in  $\theta$  case is considered as:

$$\theta = \begin{cases} \theta_n = \log \theta_n, & \text{for } n = 1, 2, 3, \dots, e \\ \theta_{e+1} = \log \sigma_g \end{cases} \quad (39)$$

## 2.3. Experimental setup for energy, solar, and wind forecasting

This section of the study is divided into two parts. The first shows the four experimental setups used for different ESW forecast input feature variables. The second part explains the detailed description of the ESW data. Details of each part are briefly discussed below:

### 2.3.1. Four setups for experimental design

The ESW forecasting optimization process was divided into four experimental design setups. Each experimental setup shows the different "Fit Method" and "Predict Method." The difference between the "Fit Method" and the "Predict Method" is already indicated in Section 2.1 and Section 2.2. The four combinations of these algorithms will help to improve the ESW forecasting tasks. The details of each experimental setup are shown step-by-step below:

**Setup-1:** in this setup, two GPR model equations using "Fit

Method" and "Predict Method" are applied.

$$\text{EXMS-1} = \begin{cases} GPR_1 = y_{\text{input}}, z_{\text{target}}, k_{kf}, E_{\text{exp}}, f_m, E_{\text{exact}}, P_m, E_{\text{exact}} \\ GPR_2 = y_{\text{input}}, z_{\text{target}}, k_{kf}, M_{32}, f_m, E_{\text{exact}}, P_m, P_{bcd}, B_{\text{size}} \end{cases} \quad (40)$$

In equation (40),  $y_{\text{input}}$  and  $z_{\text{target}}$  describes the input feature variables and target variables,  $f_m$  is the Fit Method,  $E_{\text{exact}}$  is the exact Gaussian process regression,  $P_m$  is the predict method,  $M_{32}$  is the Matern kernel with variable 3/2 function,  $P_{bcd}$  is the block coordinate descent is used to solve the ESW forecasting optimization problem, and  $B_{\text{size}}$  represents the block's size as a comma-separated pair, which consists of an integer value in a range of one to  $o$  data samples, where  $o$  demonstrates is the length of input feature variables.

Setup 2 is given by the following equation:

$$\text{EXMS-2} = \begin{cases} GPR_1 = y_{\text{input}}, z_{\text{target}}, k_{kf}, SE_{\text{exp}}, f_m, S_{sd}, P_m, S_{sd} \\ GPR_2 = y_{\text{input}}, z_{\text{target}}, k_{kf}, M_{52}, f_m, E_{\text{exact}}, P_m, P_{bcd}, B_{\text{size}} \end{cases} \quad (41)$$

Where  $M_{52}$  is represents the Matern kernel with parameter 5/2, and the subset of data points approximation.

Setup 3 is given by the following equation:

$$\text{EXMS-3} = \begin{cases} GPR_1 = y_{\text{input}}, z_{\text{target}}, k_{kf}, ARD_{32}, f_m, S_{sr}, P_m, S_{sr} \\ GPR_2 = y_{\text{input}}, z_{\text{target}}, k_{kf}, ARD_{32}, f_m, E_{\text{exact}}, P_m, P_{bcd}, B_{\text{size}} \end{cases} \quad (42)$$

Where  $ARD_{32}$  is Matern kernel with variable 3/2 as well as a separate length scale per predictor,  $S_{sr}$  and is the subset of regressors approximation.

Setup 4 is given by the following equation:

$$\text{EXMS-4} = \begin{cases} GPR_1 = y_{\text{input}}, z_{\text{target}}, k_{kf}, ARD_{32}, f_m, F_{fic}, P_m, F_{fic} \\ GPR_2 = y_{\text{input}}, z_{\text{target}}, k_{kf}, AQ_{\text{quadratic}}, f_m, E_{\text{exact}}, P_m, P_{bcd}, B_{\text{size}} \end{cases} \quad (43)$$

Where  $F_{fic}$  represents the fully independent conditional approximation, and  $AQ_{\text{quadratic}}$  explicates the rational quadratic kernel with a separate length scale per predictor. The basic difference in the four different experimental setups given above is the replacement of kernel function ( $k_{kf}$ ), "Fit Method ( $f_m$ )" and the "Predict Method ( $P_m$ )". The combination of different methods/algorithms gives us insight into the choice of the most appropriate experimental setup for the different nature of the load curve (e.g., in our case, ESW forecasting). Table 1 visualized the estimation of the fitted value of the approximation between the two methods of  $GPR_1$  and  $GPR_2$ . Fitted values for solar and wind power forecasting are more similar.

### 2.3.2. Performance evaluation index

In this study, three performance evaluation indicators are used

**Table 1**

Estimation of the value of the approximation between the two methods  $GPR_1$  and  $GPR_2$ .

| Forecasting source | Experimental setups |          |        |        |
|--------------------|---------------------|----------|--------|--------|
|                    | EXMS-1              | EXMS-2   | EXMS-3 | EXMS-4 |
| Energy             | 318.783             | 143.669  | 23.014 | 21.155 |
| Solar power        | 1.0005              | 3.853    | 0.146  | 0.421  |
| Wind power         | 85.799              | 262.1497 | 0.2114 | 1.236  |

to evaluate the accuracy of ESW power forecasting. The mathematical formulation of each indicator and the potential purpose of the indicator used is briefly defined in Table 2. In Table 2,  $\sigma$  represents the standard deviation,  $\mu$  shows the mean value,  $m$  is the total number of ESW power data samples,  $z_{\text{actual}(m)}$  is the actual ESW forecasting values and  $z_{\text{forecasting}(m)}$  explicates the forecasted ESW values.

### 2.3.3. Data source and description of the data

Real-time data is taken from a closed source at three different locations. The data is comprised of three distinct parts. The first part includes real-time energy consumption and weather data is taken from the ISO New England website. Solar power production data is taken from "Huanghe Hydropower Golmud Solar Park." The production of wind power is taken from "Longyuan Rudong Intertidal Wind Farm." Descriptive statistics of ESW data in real-time, including minimum, median, maximum, mean, and  $\sigma$  are shown in Table 3. The temporal resolution of energy consumption data is provided in 30-min intervals. The temporal resolution of solar and wind data consists of 1-h intervals, respectively. The start and end dates for energy consumption, solar, and wind power are March 1, 2019, to March 15, 2019, January 1, 2019, to June 30, 2019, and January 1, 2019, to June 30, 2019, respectively. Total input/target data samples of energy, solar, and wind are taken at 720, 4336, and 4336, respectively. The ESW data is divided into training and validation/target parts. The first 80% of data is used for training, and the remaining 20% of data is used for validation/target purposes.

Table 4, Table 5, and Table 6 demonstrate the correlation coefficients between input and output variables for real-time energy consumption, solar and wind power data. Input and target variables of energy consumption forecasting include hours ( $H_r$ ), dry-bulb temperature ( $T_{dbt}$  (°C)), dew point temperature ( $T_{dpt}$  (°C)), wet bulb temperature ( $T_{wbt}$  (°C)), humidity ratio ( $H_r$  (%)), previous week's energy consumption ( $P_{wl}$  (MWh)), previous day's energy consumption ( $P_{dl}$  (MWh)), previous 24-h energy consumption ( $P_{24-h-l}$  (MWh)) and total energy consumption ( $E_{total}$  (MWh)). Total energy consumption is the aggregate of different types of domestic, commercial, and industrial loads. In Table 4, it can be seen that the  $E_{total}$  and  $P_{wl}$ ,  $P_{dl}$  shows strong correlation coefficients.

## 3. Forecasting results and discussion

This section discusses the detailed analysis and discussion of the proposed GPR model with four fit and predictive methods. The validation of ESW power forecasting results compared to existing studies is briefly discussed. A detailed description of the forecast results for each section is provided below.

### 3.1. Optimization results of GPR model

As discussed in section 2.3.1, the setup of the GPR model is divided into four parts. The forecasting results of the ESW forecast for each experimental setup is shown in Fig. 2. Thomson's multitaper spectral density estimates a clear view of the temporal resolution of the forecasting results. The legends of Fig. 2 shows (RTR-E), (FM-EXACT-E), (FM-BCD-E), (RTR-S), (FM-EXACT-S), (FM-BCD-S), (RTR-W), (FM-EXACT-W), and (FM-EXACT-W) (FM-BCD-W). FM is the forecasting method, and RT is the ESW real-time power source, respectively. EXMS represents all four experimental setups. It can be seen that ESW forecasting in EXMS-1, EXMS-3, and EXMS-4 is very accurate. There is a slight dispersion in energy and wind forecast in EXMS-2. The basic reason for this is that the combination of the GPR model with the FM-SD-W and the FM-BCD-W method in EXMS-2 is not appropriate for forecasting wind power.

**Table 2**

Performance evaluation statistics.

| Sr. # | Performance evaluation indicators | Mathematical formulation   | Purpose  |
|-------|-----------------------------------|--|--|
| 1     | CV (%)                            | $\frac{\sigma}{\mu} = \sqrt{\frac{(z-u)^2}{m-1}} = \mu \sum_m z$         | The CV demonstrates the extent of dispersion in the data samples about the average values of ESW forecasting.                          |
| 2     | MAE                               | $\frac{1}{M} = \sum_{m=1}^M  z_{forecasting(m)} - z_{actual(m)} $        | The MAE is the forecasting measure between paired observations expressing a similar phenomenon.  |
| 3     | RMSE                              | $\sqrt{\frac{1}{M} \sum_{m=1}^M (z_{forecasting(m)} - z_{actual(m)})^2}$ | The RMSE is the widely used identification of variations between the actual values displayed in each design and the forecasted values. |

**Table 3**

Descriptive statistics of real-time ESW data.

| Data sources                          | Minimum  | Median   | Maximum   | Mean     | $\sigma$ |
|---------------------------------------|----------|----------|-----------|----------|----------|
| Real-time energy consumption data     | 6049.370 | 8710.030 | 10826.280 | 8577.227 | 1262.801 |
| Real-time solar power generation data | 0.000    | 0.238    | 75.797    | 14.737   | 20.340   |
| Real-time wind power generation data  | 0.000    | 48.149   | 477.494   | 87.059   | 100.806  |

**Table 4**

Correlation coefficients between input and output feature real-time energy consumption variables.

| Variables          | Input and output feature variables |           |           |           |             |             |             |              |             |
|--------------------|------------------------------------|-----------|-----------|-----------|-------------|-------------|-------------|--------------|-------------|
|                    | $H_r$                              | $T_{dbt}$ | $T_{dpt}$ | $T_{wbt}$ | $H_r$       | $P_{wl}$    | $P_{dl}$    | $P_{24-h-l}$ | $E_{total}$ |
| $H_r$              | 1.00                               | 0.30      | -0.04     | 0.08      | -0.23       | 0.48        | 0.48        | -0.03        | 0.48        |
| $T_{dbt}$ (°C)     | 0.30                               | 1.00      | 0.31      | 0.63      | -0.31       | 0.54        | 0.58        | 0.16         | 0.59        |
| $T_{dpt}$ (°C)     | -0.04                              | 0.31      | 1.00      | 0.93      | <b>0.80</b> | -0.14       | -0.15       | -0.26        | -0.07       |
| $T_{wbt}$ (°C)     | 0.08                               | 0.63      | 0.93      | 1.00      | 0.54        | 0.07        | 0.09        | -0.16        | 0.15        |
| $H_r$ (%)          | -0.23                              | -0.31     | 0.80      | 0.54      | 1.00        | -0.49       | -0.52       | -0.37        | -0.46       |
| $P_{wl}$ (MWh)     | 0.48                               | 0.54      | -0.14     | 0.07      | -0.49       | 1.00        | 0.86        | 0.27         | 0.97        |
| $P_{dl}$ (MWh)     | 0.48                               | 0.58      | -0.15     | 0.09      | -0.52       | 0.86        | 1.00        | 0.31         | 0.88        |
| $P_{24-h-l}$ (MWh) | -0.03                              | 0.16      | -0.26     | -0.16     | -0.37       | 0.27        | 0.31        | 1.00         | 0.31        |
| $E_{total}$ (MWh)  | 0.48                               | 0.59      | -0.07     | 0.15      | -0.46       | <b>0.97</b> | <b>0.88</b> | 0.31         | 1.00        |

Input and target variables of solar power forecasting include direct irradiance ( $S_{id}$ (kW/m $\hat{A}^2$ )), diffuse irradiance ( $S_{idiff}$ (kW/m $\hat{A}^2$ )), ambient temperature ( $T_{amb}$ (°C)), precipitation ( $P_{prec}$ (mm/h)), snowfall ( $S_{fall}$ (mm/h)), snow mass ( $S_{mass}$ (kg/m $\hat{A}^2$ )), air density ( $A_{density}$ (kg/m $\hat{A}^3$ )), cloud cover ( $C_{cov}$ (fraction)), and total solar power production ( $S_{power}$ (kW)). The correlation of each input feature variable can be seen in Table 5.

**Table 5**

Correlation coefficients between input and output feature real-time solar data variables.

| Variables                          | Input and output feature variables |             |           |            |            |            |               |           |             |
|------------------------------------|------------------------------------|-------------|-----------|------------|------------|------------|---------------|-----------|-------------|
| Solar data                         | $S_{id}$                           | $S_{idiff}$ | $T_{amb}$ | $P_{prec}$ | $S_{fall}$ | $S_{mass}$ | $A_{density}$ | $C_{cov}$ | $S_{power}$ |
| Input and output feature variables | $S_{id}$                           | $S_{idiff}$ | $T_{amb}$ | $P_{prec}$ | $S_{fall}$ | $S_{mass}$ | $A_{density}$ | $C_{cov}$ | $S_{power}$ |
| $S_{id}$ (kW/m $\hat{A}^2$ )       | 1.00                               | 0.58        | 0.22      | 0.03       | 0.00       | 0.04       | 0.06          | -0.01     | 0.87        |
| $S_{idiff}$ (kW/m $\hat{A}^2$ )    | 0.58                               | 1.00        | 0.47      | 0.06       | -0.04      | -0.04      | -0.21         | 0.06      | 0.90        |
| $T_{amb}$ (°C)                     | 0.22                               | 0.47        | 1.00      | 0.14       | -0.10      | -0.27      | -0.78         | 0.16      | 0.38        |
| $P_{prec}$ (mm/h)                  | 0.03                               | 0.06        | 0.14      | 1.00       | 0.28       | 0.05       | -0.09         | 0.24      | 0.05        |
| $S_{fall}$ (mm/h)                  | 0.00                               | -0.04       | -0.10     | 0.28       | 1.00       | 0.31       | 0.20          | 0.14      | -0.02       |
| $S_{mass}$ (kg/m $\hat{A}^2$ )     | 0.04                               | -0.04       | -0.27     | 0.05       | 0.31       | 1.00       | 0.53          | 0.16      | 0.00        |
| $A_{density}$ (kg/m $\hat{A}^3$ )  | 0.06                               | -0.21       | -0.78     | -0.09      | 0.20       | 0.53       | 1.00          | 0.00      | -0.08       |
| $C_{cov}$ (fraction)               | -0.01                              | 0.06        | 0.16      | 0.24       | 0.14       | 0.16       | 0.00          | 1.00      | 0.02        |
| $S_{power}$ (kW)                   | <b>0.87</b>                        | 0.90        | 0.38      | 0.05       | -0.02      | 0.00       | -0.08         | 0.02      | 1.00        |

Input and target variables for wind power forecasting include wind speed ( $W_s$ (m/s)), dry-bulb temperature ( $T_{amb}$ (°C)), perception ( $P_{prec}$ (mm/h)), snowfall ( $S_{fall}$ (mm/h)), snow mass ( $S_{mass}$ (kg/m $\hat{A}^2$ )), air density ( $A_{density}$ (kg/m $\hat{A}^3$ )), surface radiation ( $R_{surf}$ (kW/m $\hat{A}^2$ )), radiation toa ( $R_{toe}$ (kW/m $\hat{A}^2$ )), cloud cover ( $C_{cov}$ (fraction)), previous day wind speed ( $P_{dwp}$ (kW)), previous week wind speed ( $P_{week,wk}$ (kW)), and total wind power production ( $W_{power}$ (kW)). The  $W_{power}$  with  $W_s$  and  $W_{power}$  shows a strong correlation. The detailed analysis of correlation coefficients listed above with each input feature variables is given in Table 6.

The rest of the EXMS are given higher forecast accuracy. The full view with all temporal resolution data samples intervals of Fig. 2 is given in supplementary material (A).

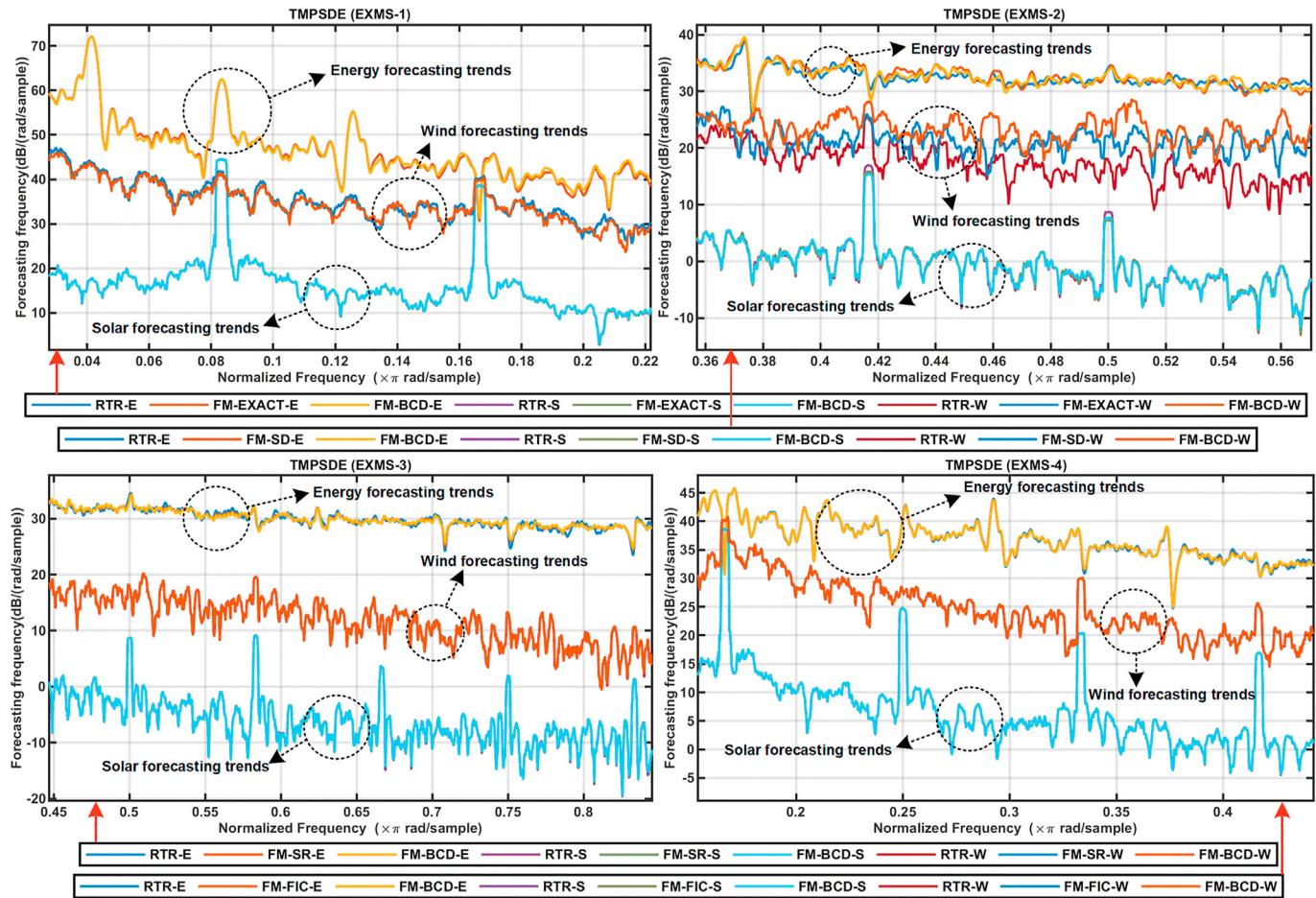
The descriptive statistics (minimum, median, maximum, mean and  $\sigma$ ) of the ESW power forecast are shown in Table 7. Each data

source represents a unique experimental setup as well as a unique combination of GPR fit and predict methods. The existing models used to compare and validate GPR results are the Levenberg–Marquardt backpropagation neural networks (LM-BPNN), the Bayesian regularized backpropagation neural networks

**Table 6**

Correlation coefficients between input and output feature variables of real-time wind consumption data.

| Variables              | Input and output feature variables |           |            |            |            |               |             |            |            |           |               |             |
|------------------------|------------------------------------|-----------|------------|------------|------------|---------------|-------------|------------|------------|-----------|---------------|-------------|
| Wind data              | $W_s$                              | $T_{amb}$ | $P_{prec}$ | $S_{fall}$ | $S_{mass}$ | $A_{density}$ | $R_{surf.}$ | $R_{toe.}$ | $C_{cov.}$ | $P_{dwp}$ | $P_{week.wk}$ | $W_{power}$ |
| $W_s(m/s)$             | 1.00                               | 0.03      | 0.10       | 0.02       | -0.02      | -0.02         | 0.04        | 0.03       | 0.04       | -0.04     | 0.41          | <b>0.96</b> |
| $T_{amb}(^{\circ}C)$   | 0.03                               | 1.00      | 0.18       | -0.22      | -0.46      | -0.99         | 0.39        | 0.35       | 0.01       | 0.03      | 0.00          | 0.08        |
| $P_{prec}(mm/h)$       | 0.10                               | 0.18      | 1.00       | 0.28       | 0.03       | -0.19         | -0.04       | 0.06       | 0.22       | -0.04     | 0.06          | 0.09        |
| $S_{fall}(mm/h)$       | 0.02                               | -0.22     | 0.28       | 1.00       | 0.40       | 0.21          | -0.08       | -0.07      | 0.07       | -0.06     | 0.01          | 0.00        |
| $S_{mass}(kg/mA^2)$    | -0.02                              | -0.46     | 0.03       | 0.40       | 1.00       | 0.47          | -0.06       | -0.05      | 0.08       | -0.16     | -0.04         | -0.06       |
| $A_{density}(kg/mA^2)$ | -0.02                              | -0.99     | -0.19      | 0.21       | 0.47       | 1.00          | -0.26       | -0.22      | -0.02      | -0.04     | -0.01         | -0.07       |
| $R_{surf.}(kW/mA^2)$   | 0.04                               | 0.39      | -0.04      | -0.08      | -0.06      | -0.26         | 1.00        | 0.91       | -0.21      | -0.04     | -0.08         | 0.08        |
| $R_{toe.}(kW/mA^2)$    | 0.03                               | 0.35      | 0.06       | -0.07      | -0.05      | -0.22         | 0.91        | 1.00       | -0.03      | -0.06     | -0.07         | 0.07        |
| $C_{cov.}(fraction)$   | 0.04                               | 0.01      | 0.22       | 0.07       | 0.08       | -0.02         | -0.21       | -0.03      | 1.00       | -0.10     | -0.02         | 0.01        |
| $P_{dwp}(kW)$          | -0.04                              | 0.03      | -0.04      | -0.06      | -0.16      | -0.04         | -0.04       | -0.06      | -0.10      | 1.00      | -0.04         | -0.03       |
| $P_{week.wk}(kW)$      | 0.41                               | 0.00      | 0.06       | 0.01       | -0.04      | -0.01         | -0.08       | -0.07      | -0.02      | -0.04     | 1.00          | 0.39        |
| $W_{power}(kW)$        | <b>0.96</b>                        | 0.08      | 0.09       | 0.00       | -0.06      | -0.07         | 0.08        | 0.07       | 0.01       | -0.03     | 0.39          | 1.00        |

**Fig. 2.** Thomson multi-taper power spectral density estimates predict ESW forecast trends.

(BR-BPNN), and the conjugate scale gradient backpropagation neural networks (SCG-BPNN). For example, the minimum values of FM-EXACT-EXMS-1, FM-BCD-EXMS-1, FM-SD-EXMS-2, FM-BCD-EXMS-2, FM-SR-EXMS-3, FM-BCD-EXMS-3, FM-FIC-EXMS-4, FM-BCD-EXMS-4, LM-BPNN, BR-BPNN, SCG-BPNN are noted as 6049.279 MWh, 6036.848 MWh, 6074.086 MWh, 6046.884 MWh, 6048.213 MWh, 6055.356 MWh, 6048.246 MWh, 6054.949 MWh, 6122.214 MWh, 6028.333 MWh, 6169.530 MWh, respectively. The

forecast reading provides accurate forecasting accuracy of the proposed four experimental GPR model configurations as well as validation of existing models. The median, maximum, mean and  $\sigma$  values give higher accuracy.

The minimum RTR value for solar and wind is zero. The forecasted solar and wind minimum values in Table 7 show some negative values that demonstrate the forecast error. The error of FM-BCD-EXMS-1, FM-SD-EXMS-2 and FM-BCD-EXMS-2 in the

**Table 7**

Descriptive statistics on ESW power forecasting.

| Forecasting source                  | Data sources    | Minimum  | Median   | Maximum   | Mean     | $\sigma$ |
|-------------------------------------|-----------------|----------|----------|-----------|----------|----------|
| Energy forecasting indicators (MWh) | FM-EXACT-EXMS-1 | 6049.279 | 8710.613 | 10825.683 | 8577.244 | 1262.676 |
|                                     | FM-BCD-EXMS-1   | 6036.848 | 8705.873 | 10812.723 | 8577.267 | 1259.112 |
|                                     | FM-SD-EXMS-2    | 6074.086 | 8702.220 | 10819.363 | 8577.692 | 1258.578 |
|                                     | FM-BCD-EXMS-2   | 6046.884 | 8705.531 | 10824.319 | 8577.606 | 1258.725 |
|                                     | FM-SR-EXMS-3    | 6048.213 | 8705.090 | 10825.958 | 8577.239 | 1262.434 |
|                                     | FM-BCD-EXMS-3   | 6055.356 | 8702.300 | 10829.795 | 8577.231 | 1262.362 |
|                                     | FM-FIC-EXMS-4   | 6048.246 | 8704.798 | 10825.699 | 8577.247 | 1262.431 |
|                                     | FM-BCD-EXMS-4   | 6054.949 | 8703.423 | 10825.420 | 8577.245 | 1262.376 |
|                                     | LM-BPNN         | 6122.214 | 8676.141 | 10833.123 | 8581.351 | 1246.516 |
|                                     | BR-BPNN         | 6028.333 | 8713.166 | 10837.261 | 8580.418 | 1258.948 |
|                                     | SCG-BPNN        | 6169.530 | 8666.920 | 10835.250 | 8590.625 | 1244.501 |
| Solar forecasting indicators (kW)   | FM-EXACT-EXMS-1 | -0.032   | 0.250    | 75.796    | 14.737   | 20.338   |
|                                     | FM-BCD-EXMS-1   | -0.988   | 0.637    | 75.901    | 14.746   | 20.350   |
|                                     | FM-SD-EXMS-2    | -0.940   | 0.712    | 75.927    | 14.731   | 20.339   |
|                                     | FM-BCD-EXMS-2   | -0.828   | 0.742    | 75.811    | 14.758   | 20.326   |
|                                     | FM-SR-EXMS-3    | -0.009   | 0.340    | 75.715    | 14.737   | 20.340   |
|                                     | FM-BCD-EXMS-3   | -0.018   | 0.355    | 75.817    | 14.736   | 20.341   |
|                                     | FM-FIC-EXMS-4   | -0.010   | 0.389    | 75.677    | 14.737   | 20.340   |
|                                     | FM-BCD-EXMS-4   | -0.009   | 0.363    | 75.748    | 14.736   | 20.340   |
|                                     | LM-BPNN         | -0.030   | 0.296    | 75.699    | 14.737   | 20.340   |
|                                     | BR-BPNN         | -0.019   | 0.256    | 75.757    | 14.737   | 20.339   |
|                                     | SCG-BPNN        | -5.718   | 1.093    | 72.225    | 14.664   | 20.271   |
| Wind forecasting indicators (kW)    | FM-EXACT-EXMS-1 | -0.113   | 48.159   | 477.480   | 87.059   | 100.796  |
|                                     | FM-BCD-EXMS-1   | -85.802  | 66.842   | 499.936   | 87.066   | 97.789   |
|                                     | FM-SD-EXMS-2    | -62.419  | 63.208   | 615.324   | 91.749   | 99.655   |
|                                     | FM-BCD-EXMS-2   | -216.071 | 60.239   | 637.488   | 85.989   | 101.184  |
|                                     | FM-SR-EXMS-3    | -0.003   | 48.245   | 477.707   | 87.059   | 100.806  |
|                                     | FM-BCD-EXMS-3   | 0.003    | 48.293   | 477.495   | 87.065   | 100.803  |
|                                     | FM-FIC-EXMS-4   | -0.003   | 48.244   | 478.753   | 87.060   | 100.809  |
|                                     | FM-BCD-EXMS-4   | -0.534   | 48.275   | 477.517   | 87.062   | 100.808  |
|                                     | LM-BPNN         | -0.371   | 48.214   | 477.276   | 87.059   | 100.807  |
|                                     | BR-BPNN         | -0.086   | 48.235   | 477.615   | 87.059   | 100.807  |
|                                     | SCG-BPNN        | -74.221  | 48.386   | 466.162   | 87.036   | 100.773  |

forecasting of wind power is moderate compared to others. FM-BCD-EXMS-3 ( $-0.018$ ) and FM-FIC-EXMS-4 ( $-0.010$ ) for solar forecasting show the least forecast error. More detailed descriptive statistics on ESW forecasting can be briefly visualized in Table 7.

Fig. 3 visualized the predicted intervals and responses of the GPR model. Only EXMS-1 responses of the GPR model for ESW forecasting are shown to avoid multiple numbers of plots. The objective is to predict median values and 99% forecast ESW intervals based on 8, 8, 11 predictors or input variables. The yellow area showed the 99% forecast intervals. Forecasting intervals give users details about the probability distribution of the target variable, and we match the range of these intervals with the notion of reliability: small intervals are more accurate than longer intervals. The forecasting intervals for Fig. 3 gives a small error.

### 3.2. Analysis of hyperparameter optimization

Hyperparameter optimization is a variable of GPR methods used to control the learning process of the GPR model. For hyperparameter optimization, the ESW data has one forecast parameter and continuous response in this study. The squared exponential kernel function is used to adjust the hyperparameter. Hyperparameters are detected (see Table 8), which reduces the five-fold cross-validation loss by using the automatic hyperparameter optimization function. In some cases, the forecasting accuracy is not as intended (e.g., low forecasting error), and in this case, the random seed function is further used to improve the learning performance of GPR methods for different experimental setups.

Fig. 4 visualized the hyperparameter optimization of the objective function model. The horizontal axis describes the sigma's values, and the vertical axis represents the values of the objective

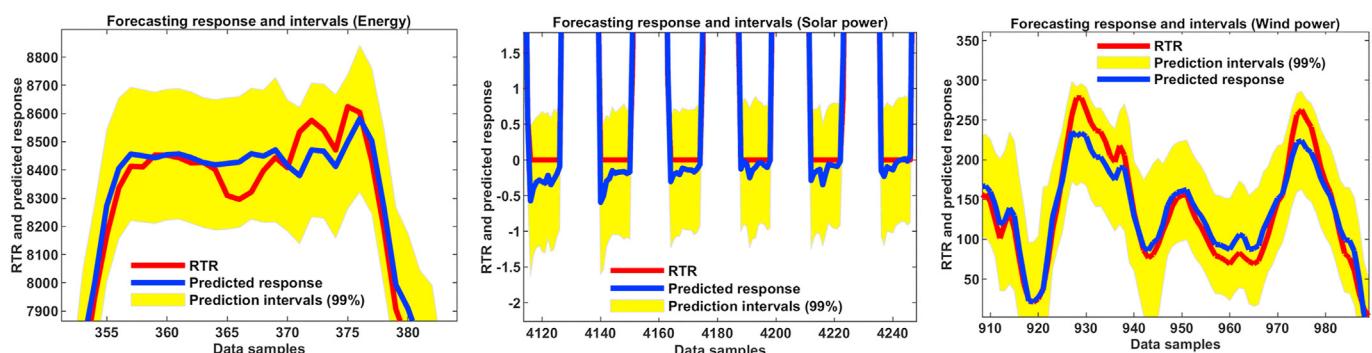
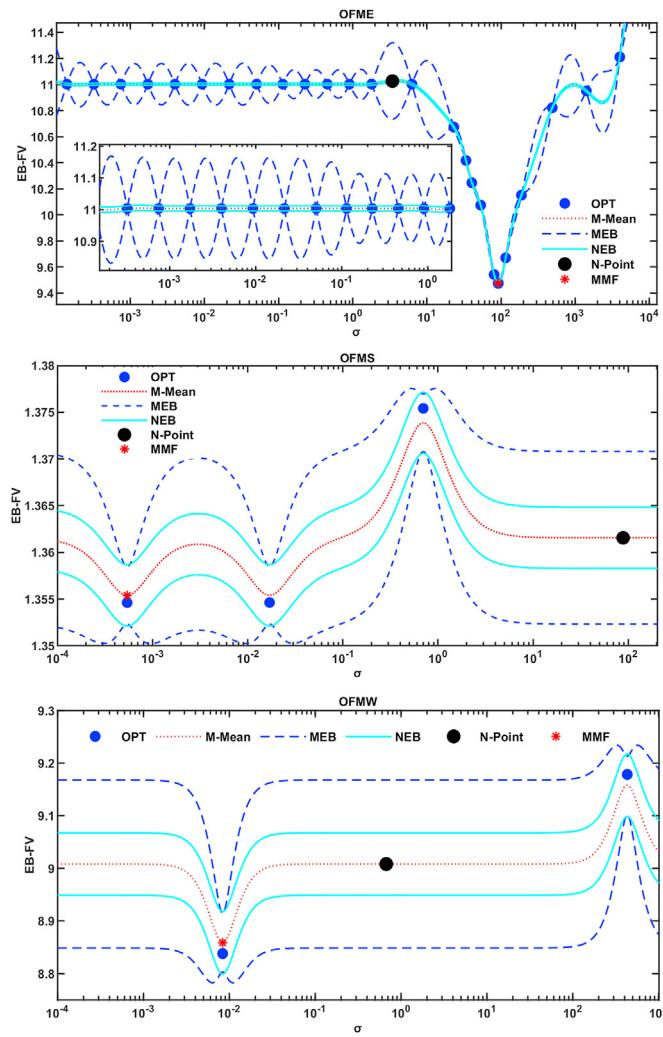


Fig. 3. Predicted intervals and response of the ESW forecast GPR model.

**Table 8**

Optimized GPR model parameters for ESW forecasting.

| Optimized function performed | Energy forecasting | Solar power forecasting | Wind power forecasting |
|------------------------------|--------------------|-------------------------|------------------------|
| $f_{OBFV}$                   | 9.474              | 1.354                   | 6.172                  |
| $f_{EOFV}$                   | 9.476              | 1.354                   | 6.299                  |
| $f_{BEPF}$                   | 91.45              | 0.009                   | 26.48                  |
| $f_{TOFET}$                  | 40.977             | 1077.911                | 1989.444               |
| $f_{TET}$                    | 56.998             | 1093.460                | 2025.126               |
| $f_{ITFE/I}$                 | 30                 | 30                      | 30                     |

**Fig. 4.** Hypermeter optimization analysis of GPR model for ESW forecasting.

function estimation (EB-FV). These functions include the values of the objective optimization parameter, e.g., the observed points (OPT), the GPR mean values (M-Mean), the error bar between actual and forecast load (MEB), the network noise error (NEB), the next point (N-Point) and the model minimum feasible point (MMF). The detailed analysis of each of these parameters given above is provided in the Supplementary Material (B). The NEB gives the least boundary between the MEB. This means that the energy model objective function (OFME), the solar model objective function (OFMS), and the wind model objective function (OFMW) make the least forecast error. OFMS and OFMW were stopped at the iteration to acquire Fig. 3 to visualize all error bars better. The MMF is the point where the network makes the least forecasting error possible.

The M-Mean, NEB, MEB, N-Point and MMF curves of OFME, OFMS and OFMW are different because of the nature of the input variables' temporal resolution. Fig. 5 visualized the optimized estimated objective minimum trace (EOMT), iteration time trace function (ITTF), and objective evaluation time trace (OETT) function of optimized GPR parameters. More details of Fig. 5 and Table 8 are included in the Supplementary Material (B).

Table 9 shows the hyperparameter optimization coefficients with 95% confidence bounds of Fig. 5. Parameters  $a$ ,  $b$ ,  $c$ ,  $m$ ,  $w$ , SSE and RMSE represent the coefficient of ITTF, OETT, and EOMT, respectively. SSE is the residual sum of the squares, and  $w$  has visualized the weight of the coefficients. The following general model/custom equation

$$f(x,y) = a + b * \sin(m * \pi * x * y) + c * \exp(-w * y)^2$$
 is used to calculate coefficients of ESW hypermeter parameters. The RMSE demonstrates the goodness of fit. There are small yellow dots in Fig. 5 that have visualized an SSE error. The RMSE indicates the least forecast error.

### 3.3. Analysis of length scale log and predictor variables

There are 8, 8, 11 continuous energy, solar, and wind power forecast parameters are used. The GPR model is equipped with an exponential kernel function with a different length-scale for each predictor applying equation (44). This covariance function is taken as:

$$l(y_j, y_k | \theta) = \sigma_g^2 \exp \left[ -\frac{1}{2} \sum_{n=1}^e \frac{(y_{jn} - y_{kn})^2}{\sigma_n^2} \right] \quad (44)$$

Where  $\sigma_n^2$  explains the scale length for forecaster  $n$ ,  $n = 1, 2, 3, \dots, e$  and  $\sigma_g^2$  is the feature variables  $\sigma$ . The parametrization of unconstrained function  $\theta$  is taken as:

$$\theta = \begin{cases} \theta_n = \log \theta_n, & \text{for } n = 1, 2, 3, \dots, e \\ \theta_{e+1} = \log \sigma_g \end{cases} \quad (45)$$

To analyze the log predictor variables' length, first set the kernel function length scales to 10, and the noise and signal standard deviations to the standard response deviation. The predictor is standardized in the training data and uses the exact fitting method. It's shown in Fig. 6, that the scale length of the 3rd predictor parameter for energy forecasting is relatively higher than that of others. For solar power forecasting, the length of predictor 1st, 2nd and 3rd predictor parameters are different from those of other predictors. These parameters do not appear to influence the predictive response in the same way as the other parameters.

Fig. 7 depicted the loss of the GPR method when compared to the separated test data. The loss function is used to determine how accurately the proposed method forecasts future demand for energy, solar, and wind power. To calculate the loss of the GPR method, the total loss of energy, solar, and wind power forecasting was calculated using the 20% separate test data. It was noted that

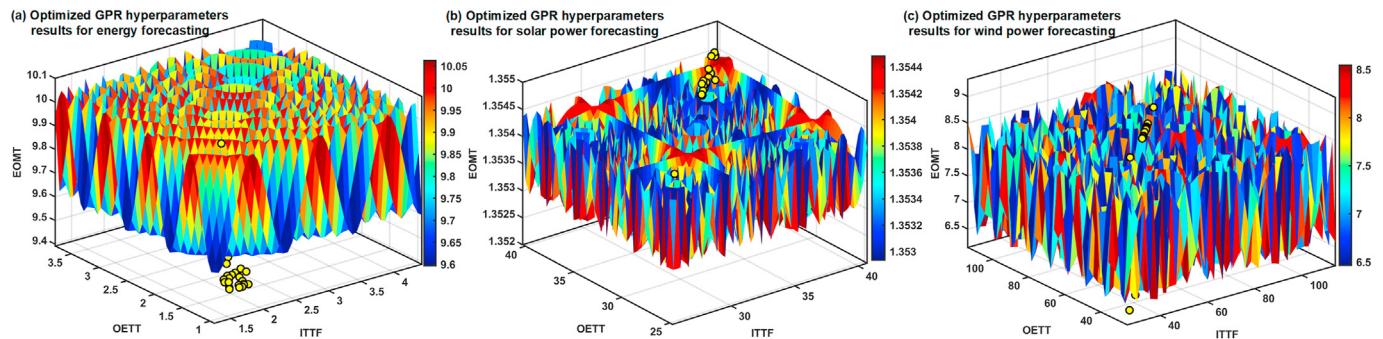


Fig. 5. Optimized EOMT, ITTF, OETT model functions for ESW forecasting.

**Table 9**  
Hyperparameter optimization coefficient analysis with 95% confidence bounds.

| Coefficients | Energy coefficients | Solar coefficients | Wind coefficients |
|--------------|---------------------|--------------------|-------------------|
| <i>a</i>     | 34.58               | 1.354              | 7.506             |
| <i>b</i>     | -0.2325             | -0.0007734         | -1.041            |
| <i>c</i>     | -24.75              | 0.266              | 0.817             |
| <i>m</i>     | 2.08                | 0.6468             | 0.748             |
| <i>w</i>     | -6.637e-05          | 0.2913             | 0.902             |
| SSE          | 7.83                | 0.002076           | 44.670            |
| RMSE         | 0.5596              | 0.009112           | 1.330             |

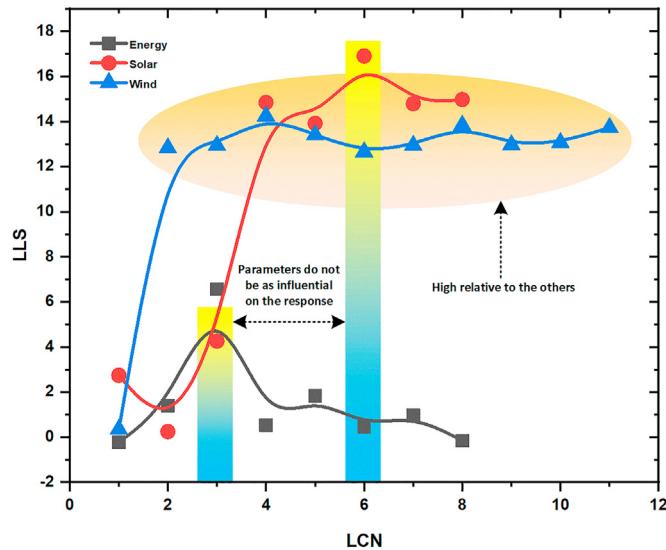


Fig. 6. For predictor variables, different length scales have been used for response predictor parameters.

the loss of solar and wind forecasting is minimal. The loss of solar and wind power, for example, was 0.01, and 0.03, respectively. When compared to solar and wind power forecasting, the measured loss for energy demand forecasting was considerably higher. However, the overall difference in loss measured with the separate test data test is very small.

#### 3.4. Validation of the model GPR

Three performance evaluation indexes and three existing models are used to verify the GPR model's effectiveness and accuracy. Existing models used for validation of results include the LM-BPNN [62], BR-BPNN [63], and SCG-BPNN [64]. The reason for using these three models is the accuracy, efficiency, low cost, high

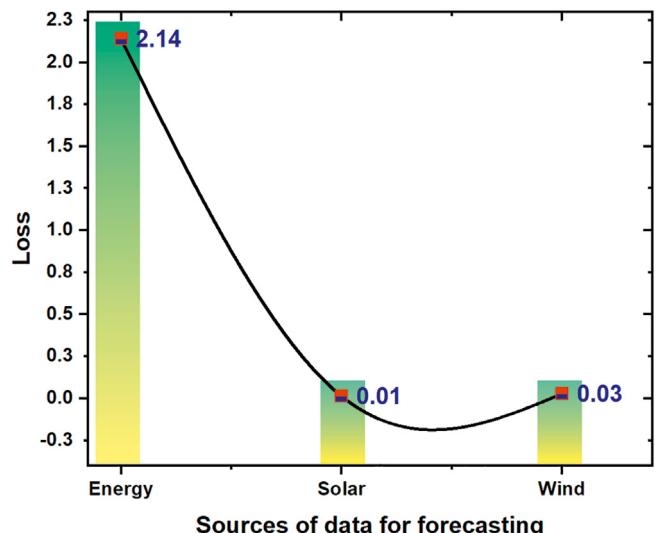


Fig. 7. The loss of the GPR method compared with the separated test data.

computational time, and a large number of applications for different ESW real-time forecasting applications. Researchers and the scientific community have a strong attachment to these models. Various types of input variables, different locations, and different data sets (e.g., weather and ESW) are used in these existing studies. Our study used the same network structure with varying network parameters (as used in existing studies) for ESW power forecasting. The differences in the results of existing models and our GPR model can be seen in Fig. 8. The forecast accuracy of the GPR model is promising. The accuracy of SCG-BPNN for energy forecasting is low compared to the proposed four EXMS for the GPR model. But this is not true of the forecasting of solar power. The accuracy of four EXMS and existing LM-BPNN, BR-BPNN, and SCG-BPNN models is the same for forecasting solar power. The accuracy of the FM-SD-EXMS-2 forecast for wind power is low compared to other methods. The box plot of Fig. 8 provides an accurate comparison of all observations/data samples. The box plot shows the exact fit between the RTR and the ESW power forecasting demand. Some outliers for FM-SD-EXMS-2 and FM-BCD-EXMS-3 can be noted. This means that the fit and predict methods BCD and SD are not a promising combination for wind power forecasting.

The performance evaluation indexes for ESW forecasting are given in Table 10. The first column of Table 10 represents both the four EXMS and the existing models. The values of  $CV_{GPR_1}$  and  $CV_{GPR_2}$  of EXMS-1 for energy forecasting are reported as 0.017% and 0.745%, respectively.

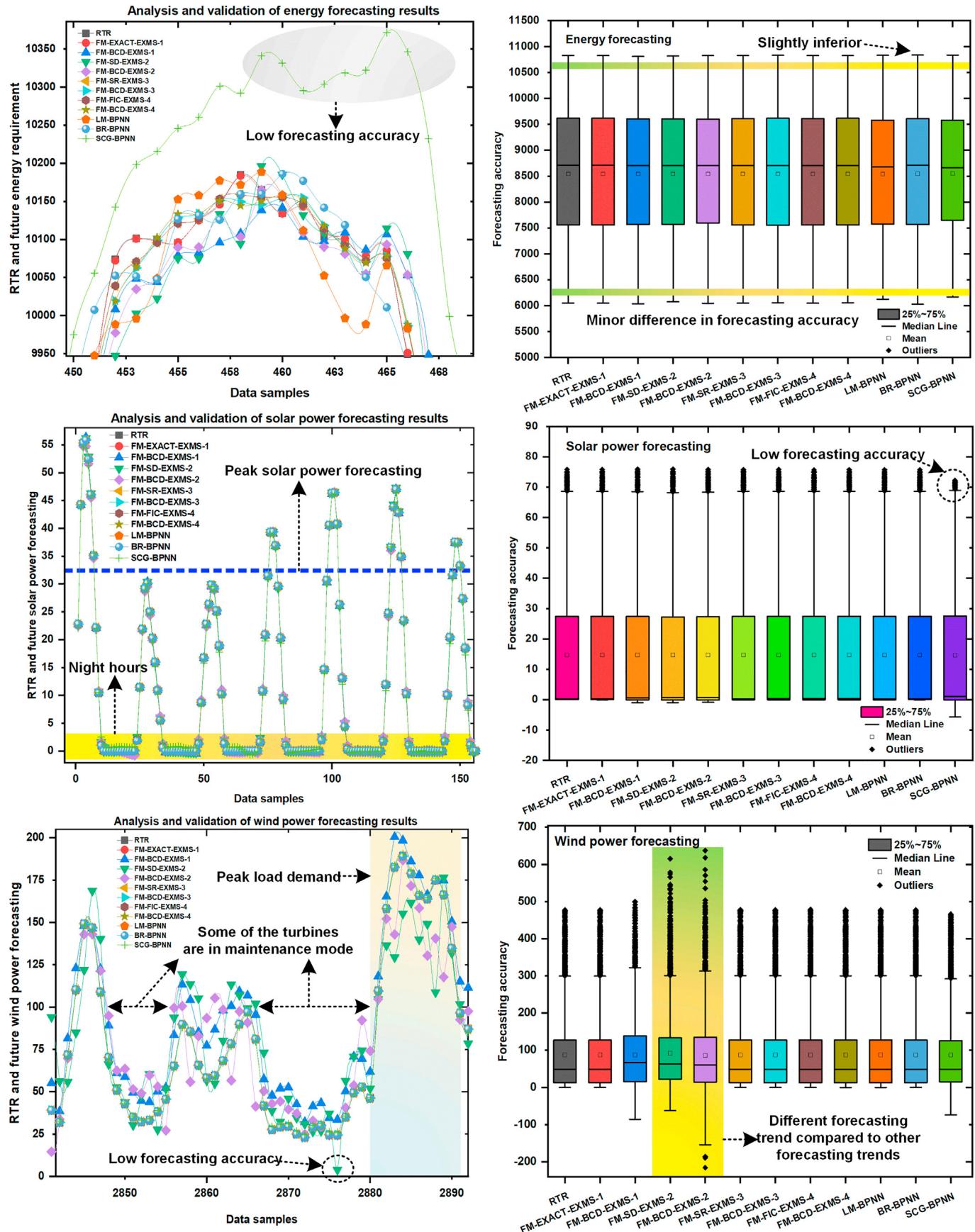


Fig. 8. Comparison and validation of results with three existing LM-BPNN, BR-BPNN, and SCG-BPNN models.

**Table 10**

Validation of performance evaluation indices and results with existing models for ESW forecasting.

| Experimental setups | Forecasting method | Performance evaluation indicators | Energy forecasting | Solar power forecasting | Wind power forecasting |
|---------------------|--------------------|-----------------------------------|--------------------|-------------------------|------------------------|
| EXMS-1              | GPR                | $CV_{GPR_1}$ (%)                  | 0.017              | 0.057                   | 0.025                  |
|                     |                    | $MAE_{GPR_1}$                     | 0.960              | 0.005                   | 0.014                  |
|                     |                    | $RMSE_{GPR_1}$                    | 1.463              | 0.008                   | 0.022                  |
|                     |                    | $CV_{GPR_2}$ (%)                  | 0.745              | 1.540                   | 25.084                 |
|                     |                    | $MAE_{GPR_2}$                     | 43.886             | 0.158                   | 17.254                 |
|                     |                    | $RMSE_{GPR_2}$                    | 63.715             | 0.226                   | 21.907                 |
| EXMS-2              | GPR                | $CV_{GPR_1}$ (%)                  | 0.970              | 1.796                   | 25.138                 |
|                     |                    | $MAE_{GPR_1}$                     | 59.407             | 0.192                   | 17.183                 |
|                     |                    | $RMSE_{GPR_1}$                    | 82.934             | 0.264                   | 21.954                 |
|                     |                    | $CV_{GPR_2}$ (%)                  | 0.898              | 2.329                   | 33.076                 |
|                     |                    | $MAE_{GPR_2}$                     | 53.554             | 0.235                   | 18.704                 |
|                     |                    | $RMSE_{GPR_2}$                    | 76.784             | 0.342                   | 28.888                 |
| EXMS-3              | GPR                | $CV_{GPR_1}$ (%)                  | 0.182              | 0.208                   | 0.166                  |
|                     |                    | $MAE_{GPR_1}$                     | 11.353             | 0.018                   | 0.110                  |
|                     |                    | $RMSE_{GPR_1}$                    | 15.573             | 0.030                   | 0.145                  |
|                     |                    | $CV_{GPR_2}$ (%)                  | 0.223              | 0.225                   | 0.170                  |
|                     |                    | $MAE_{GPR_2}$                     | 14.145             | 0.023                   | 0.115                  |
|                     |                    | $RMSE_{GPR_2}$                    | 19.056             | 0.033                   | 0.149                  |
| EXMS-4              | GPR                | $CV_{GPR_1}$ (%)                  | 0.182              | 0.252                   | 0.171                  |
|                     |                    | $MAE_{GPR_1}$                     | 11.408             | 0.022                   | 0.112                  |
|                     |                    | $RMSE_{GPR_1}$                    | 15.608             | 0.037                   | 0.149                  |
|                     |                    | $CV_{GPR_2}$ (%)                  | 0.215              | 0.227                   | 0.175                  |
|                     |                    | $MAE_{GPR_2}$                     | 13.581             | 0.021                   | 0.121                  |
|                     |                    | $RMSE_{GPR_2}$                    | 18.371             | 0.033                   | 0.153                  |
| Existing models     | LM-BPNN            | $CV_{LM-BPNN}$ (%)                | 1.375              | 0.205                   | 0.170                  |
|                     |                    | $MAE_{LM-BPNN}$                   | 89.178             | 0.020                   | 0.114                  |
|                     |                    | $RMSE_{LM-BPNN}$                  | 1.174              | 0.030                   | 0.148                  |
|                     | BR-BPNN            | $CV_{BR-BPNN}$ (%)                | 0.895              | 0.156                   | 0.186                  |
|                     |                    | $MAE_{LM-BPNN}$                   | 56.503             | 0.013                   | 0.111                  |
|                     |                    | $RMSE_{LM-BPNN}$                  | 76.484             | 0.023                   | 0.162                  |
|                     | SCG-BPNN           | $CV_{SCG-BPNN}$ (%)               | 2.311              | 5.235                   | 3.992                  |
|                     |                    | $MAE_{LM-BPNN}$                   | 1.571              | 0.538                   | 1.966                  |
|                     |                    | $RMSE_{LM-BPNN}$                  | 1.975              | 0.770                   | 3.486                  |

The  $CV_{GPR_1}$  values are low as compared to  $CV_{GPR_2}$ . It means  $CV_{GPR_1}$  combination of energy forecasting is more effective. If we compare the two values of EXMS-1 and EXMS-2, there is a slight difference as it shows the results of combinations of EXMS-1 are promising but  $MAE_{GPR_1}$  is quite more effective.  $RMSE_{GPR_1}$  for solar and wind power forecasting are similar. But this trend is not true for  $RMSE_{GPR_2}$ . The  $MAE_{GPR_2}$  values for ESW power forecast are 43.88 MWh, 0.158 kW, 17.254 kW, respectively. The  $RMSE_{GPR_2}$  for energy, the forecast is higher compared to solar and wind. A similar trend is noted for  $RMSE_{GPR_2}$  in EXMS-1. Validation of GPR EXMS-4 results with LM-BPNN, BR-BPNN, and SCG-BPNN provides higher accuracy. The GPR results are more accurate and effective compared to the different performance evaluation indexes. The brief discussion of performance evaluation indexes with existing models is further demonstrated in the Supplementary Materials section (C).

### 3.5. Discussion, advantages, challenges, and real-time applications

**Discussion:** In this study, we used three different forecasting tasks, e.g., energy, solar, and wind power forecasting. We have proposed a GPR model with four EXMS to perform this task. The basic objective behind this was to acquire accurate forecasting using different time-series data resolution, different time horizons, and different sets of input variables. Most studies have been conducted to perform only one task simultaneously, e.g., energy, solar power, or wind power forecasting. Only a few studies have been conducted in the recent past that can perform three different forecasts based on the different nature of the load curve and the different sets of input variables. Our study successfully performs this task with high accuracy and precision. The four GPR EXMS

models provide more options for choosing a better fit and predicting forecasting method. We applied different climate input variables without changing the new network function or parameters. The time horizon and time series length of the data samples were also different. Based on the different characteristics and nature of the ESW power load curve, the forecast accuracy is promising. For example, the accuracy of EXMS-1, EXMS-3, and EXMS-4 is highly reliable, with the least forecast error. And this accuracy trend is similar to the ESW power forecast. More precisely, the GPR method is like a Gaussian multivariate distribution with an infinity dimension, in which each array of data set labels is shared with Gaussian.

**Advantages:** GPR has many advantages, functions well on both small and large datasets, and can calculate uncertainty forecasts. The proposed EXMS of the GPR model can be accurately forecasted at the same time by the ESW power forecast without changing the parameters of the network structure. The GPR model can be applied efficiently through various libraries, such as MATLAB, GPytorch, scikit-learn, and GPy.

**Challenges:** As the number of input feature variables increases, so does the total time elapsed for the GPR model to learn, which is quite difficult. A large volume of data observation may increase the network's cost for forecasting tasks for the exact method. This can be a challenge for ultra-long forecasting tasks. Our study focused on short-and medium-term ESW forecasting tasks.

**Applications:** This study is applied simultaneously in three different applications, including ESW power forecasting. The previous reading of ESW power data with climate parameters is needed to implement GPR models and the proposed four EXMS models. Based on past trends, the GPR model makes it possible to perform past data readings using different fit and predict methods

and extract trends from these data sets to predict future demand for ESW power.

#### 4. Conclusion

This study provides insights for industry experts, academics, utilities, independent power producers and practitioners into plausible levels of forecast uncertainty and accuracy achieved in this context by using the stochastic GPR model. This study also provides a brief analysis of ESW forecasting based on different climate variables, two different temporal resolution data, two different observation/data samples, and four proposed experimental setups. The summary of results proposed that GPR achieves a satisfactory level of accuracy for ESW based on historical time-series climate and ESW data reading. In contrast to the existing models, the forecast accuracy of EXMS-1, EXMS-3, and EXMPS-4 is better than the existing models. The hyperparameter sensitivity test showed that the model to be used for improved performance should be carefully tuned or searched to obtain accurate hyperparameters. Most studies have been conducted to perform only one task simultaneously, e.g., energy, solar power, or wind power forecasting. Only a few studies have been conducted in the recent past that can perform three different forecasts based on the different nature of the load curve and the different sets of input variables. Our study performed accurately and presented at the same time the least for ESW forecasting based on four proposed experimental setups. The results are promising, and the errors are very low compared to existing common methods.

Future recommendations include incorporating various forecasts into decision-making in the light of risk-related volatility, developing federated learning GPR models that protect privacy, and merging automated weather forecasting into model architectures.

#### Credit author statement

**Tanveer Ahmad:** Conceptualization; Investigation; Methodology; Formal analysis; Software; Validation; Data curation; Writing - Review & Editing; Visualization, **Dongdong Zhang:** Methodology; Resources, Funding acquisition, **Chao Huang:** Conceptualization; Funding acquisition.

#### Declaration of competing interest

The authors declare that they have no knowledge of competing for financial interests or personal relationships that might have influenced the work reported in this paper.

#### Acknowledgements

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFE0118000), the Key Laboratory of Special Machine and High Voltage Apparatus (Shenyang University of Technology), Ministry of Education, Grant KFKT202006 and Guangxi Young and Middle-aged Scientific Research Basic Ability Promotion Project, 2020KY01009, National Natural Science Foundation of China under Grant 62002016, in part by Beijing Natural Science Foundation under Grant 9204028, in part by Guangdong Basic and Applied Research Foundation under Grant 2019A1515111165. The authors are grateful for the sponsorship of the UM-Macau Postdoc Talent Program under the cover of the Government of Macau, SAR.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at

<https://doi.org/10.1016/j.energy.2021.120911>.

#### References

- [1] Zhao W, Zhang H, Zheng J, Dai Y, Huang L, Shang W, et al. A point prediction method based automatic machine learning for day-ahead power output of multi-region photovoltaic plants. Energy 2021;223:120026. <https://doi.org/10.1016/j.energy.2021.120026>.
- [2] Sherwin ED, Henrion M, Azevedo IML. Estimation of the year-on-year volatility and the unpredictability of the United States energy system. Nat Energy 2018;3:341–6. <https://doi.org/10.1038/s41560-018-0121-4>.
- [3] Chaturvedi V. Energy security and climate change: friends with asymmetric benefits. Nat Energy 2016;1:1–12. <https://doi.org/10.1038/nenergy.2016.75>.
- [4] Mertzanis C. Institutions, development and energy constraints. Energy 2018;142:962–82. <https://doi.org/10.1016/j.energy.2017.10.073>.
- [5] Zhang W, Quan H, Gandhi O, Rajagopal R, Tan CW, Srinivasan D. Improving probabilistic load forecasting using quantile regression NN with skip connections. IEEE Trans Smart Grid 2020;11:5442–50. <https://doi.org/10.1109/TSG.2020.2995777>.
- [6] Ahmad T, Chen H, Huang R, Yabin G, Wang J, Shair J, et al. Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment. Energy 2018;158:17–32. <https://doi.org/10.1016/j.energy.2018.05.169>.
- [7] Fouillot A, Voyant C, Notton G, Motte F, Paoli C, Nivet ML, et al. Solar irradiation prediction with machine learning: forecasting models selection method depending on weather variability. Energy 2018;165:620–9. <https://doi.org/10.1016/j.energy.2018.09.116>.
- [8] Wang J, Hu J. A robust combination approach for short-term wind speed forecasting and analysis – combination of the ARIMA (autoregressive integrated moving average), ELM (extreme learning machine), SVM (support vector machine) and LS-SVM (least square SVM) forecasts usi. Energy 2015;93: 41–56. <https://doi.org/10.1016/j.energy.2015.08.045>.
- [9] Sanjari MJ, Gooi HB, Nair NKC. Power generation forecast of hybrid PV-Wind system. IEEE Trans Sustain Energy 2020;11:703–12. <https://doi.org/10.1109/TSTE.2019.2903900>.
- [10] Howell S, Rezgui Y, Hippolyte JL, Jayan B, Li H. Towards the next generation of smart grids: semantic and holonic multi-agent management of distributed energy resources. Renew Sustain Energy Rev 2017;77:193–214. <https://doi.org/10.1016/j.rser.2017.03.107>.
- [11] Bessa RJ, Trindade A, Miranda V. Spatial-temporal solar power forecasting for smart grids. IEEE Trans Ind Inform 2015;11:232–41. <https://doi.org/10.1109/TII.2014.2365703>.
- [12] Sharadga H, Hajimirza S, Balog RS. Time series forecasting of solar power generation for large-scale photovoltaic plants. Renew Energy 2020;150: 797–807. <https://doi.org/10.1016/j.renene.2019.12.131>.
- [13] Kakran S, Chanaan S. Smart operations of smart grids integrated with distributed generation: a review. Renew Sustain Energy Rev 2018;81:524–35. <https://doi.org/10.1016/j.rser.2017.07.045>.
- [14] Zafar MW, Shahbaz M, Sinha A, Sengupta T, Qin Q. How renewable energy consumption contribute to environmental quality? The role of education in OECD countries. J Clean Prod 2020;268:122149. <https://doi.org/10.1016/j.jclepro.2020.122149>.
- [15] Das S, Basu M. Day-ahead optimal bidding strategy of microgrid with demand response program considering uncertainties and outages of renewable energy resources. Energy 2020;190:116441. <https://doi.org/10.1016/j.energy.2019.116441>.
- [16] Jurasz J. Modeling and forecasting energy flow between national power grid and a solar-wind-pumped-hydroelectricity (PV–WT–PSH) energy source. Energy Convers Manag 2017;136:382–94. <https://doi.org/10.1016/j.enconman.2017.01.032>.
- [17] Ahmad T, Zhang H, Yan B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustain Cities Soc 2020;55:102052. <https://doi.org/10.1016/j.scs.2020.102052>.
- [18] Herrera GP, Constantino M, Tabak BM, Pistori H, Su JJ, Naranpanawa A. Long-term forecast of energy commodities price using machine learning. Energy 2019;179:214–21. <https://doi.org/10.1016/j.energy.2019.04.077>.
- [19] Ahmad T, Huanxin C, Zhang D, Zhang H. Smart energy forecasting strategy with four machine learning models for climate-sensitive and non-climate sensitive conditions. Energy 2020;198:117283. <https://doi.org/10.1016/j.energy.2020.117283>.
- [20] Ahmad T, Chen H. Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment. Energy 2018;160:1008–20. <https://doi.org/10.1016/j.energy.2018.07.084>.
- [21] Vatanpour M, Sadeghi Yazdankhah A. The impact of energy storage modeling in coordination with wind farm and thermal units on security and reliability in a stochastic unit commitment. Energy 2018;162:476–90. <https://doi.org/10.1016/j.energy.2018.07.181>.
- [22] da Silva RG, Ribeiro MHDM, Moreno SR, Mariani VC, Coelho L dos S. A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting. Energy 2021;216:119174. <https://doi.org/10.1016/j.energy.2020.119174>.
- [23] Luo X, Zhang D, Zhu X. Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge. Energy 2021;225:120240.

- [https://doi.org/10.1016/j.energy.2021.120240.](https://doi.org/10.1016/j.energy.2021.120240)
- [24] Fathi S, Srinivasan R, Fenner A, Fathi S. Machine learning applications in urban building energy performance forecasting: a systematic review. *Renew Sustain Energy Rev* 2020;133:110287. <https://doi.org/10.1016/j.rser.2020.110287>.
- [25] Zhang L, Wen J, Li Y, Chen J, Ye Y, Fu Y, et al. A review of machine learning in building load prediction. *Appl Energy* 2021;285:116452. <https://doi.org/10.1016/j.apenergy.2021.116452>.
- [26] Demolli H, Dokuz AS, Ecemis A, Gokcek M. Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Convers Manag* 2019;198:112492. <https://doi.org/10.1016/j.enconman.2019.111823>.
- [27] Sharma A, Kakkar A. Forecasting daily global solar irradiance generation using machine learning. *Renew Sustain Energy Rev* 2018;82:2254–69. <https://doi.org/10.1016/j.rser.2017.08.066>.
- [28] Huertas-Tato J, Aler R, Galván IM, Rodríguez-Benítez FJ, Arbizu-Barrena C, Pozo-Vázquez D. A short-term solar radiation forecasting system for the Iberian Peninsula. Part 2: model blending approaches based on machine learning. *Sol Energy* 2020;195:685–96. <https://doi.org/10.1016/j.solener.2019.11.091>.
- [29] Liu H, Chen C. Data processing strategies in wind energy forecasting models and applications: a comprehensive review. *Appl Energy* 2019;249:392–408. <https://doi.org/10.1016/j.apenergy.2019.04.188>.
- [30] Zieher M, Lange M, Focken U. Variable renewable energy forecasting – integration into electricity grids and markets – a best practice guide. *VRE Discuss Ser – Pap # 06 2020;61–29*.
- [31] Mayer MJ, Gróf G. Extensive comparison of physical models for photovoltaic power forecasting. *Appl Energy* 2020;283:116239. <https://doi.org/10.1016/j.apenergy.2020.116239>.
- [32] Pearre NS, Swan LG. Statistical approach for improved wind speed forecasting for wind power production. *Sustain Energy Technol Assess* 2018;27:180–91. <https://doi.org/10.1016/j.seta.2018.04.010>.
- [33] Boland J, David M, Lauret P. Short term solar radiation forecasting: island versus continental sites. *Energy* 2016;113:186–92. <https://doi.org/10.1016/j.energy.2016.06.139>.
- [34] Trumper JR, Kouroutes N, Martin A. Short-term solar irradiation forecasting based on dynamic harmonic regression. *Energy* 2015;84:289–95. <https://doi.org/10.1016/j.energy.2015.02.100>.
- [35] Zhou Y, Zheng S, Zhang G. Machine learning-based optimal design of a phase change material integrated renewable system with on-site PV, radiative cooling and hybrid ventilations—study of modelling and application in five climatic regions. *Energy* 2020;192:116608. <https://doi.org/10.1016/j.energy.2019.116608>.
- [36] Alizamir M, Kim S, Kisi O, Zounemat-Kermani M. A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: case studies of the USA and Turkey regions. *Energy* 2020;197:117239. <https://doi.org/10.1016/j.energy.2020.117239>.
- [37] Kaytez F. A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption. *Energy* 2020;197:117200. <https://doi.org/10.1016/j.energy.2020.117200>.
- [38] Dai Y, Zhao P. A hybrid load forecasting model based on support vector machine with intelligent methods for feature selection and parameter optimization. *Appl Energy* 2020;279:115332. <https://doi.org/10.1016/j.apenergy.2020.115332>.
- [39] Prado F, Minutolo MC, Kristjanpoller W. Forecasting based on an ensemble autoregressive moving average - adaptive neuro - fuzzy inference system - neural network - genetic algorithm framework. *Energy* 2020;197:117159. <https://doi.org/10.1016/j.energy.2020.117159>.
- [40] Chen SX, Gooi HB, Wang MQ. Solar radiation forecast based on fuzzy logic and neural networks. *Renew Energy* 2013;60:195–201. <https://doi.org/10.1016/j.renene.2013.05.011>.
- [41] Song W, Cattani C, Chi CH. Multifractional Brownian motion and quantum-behaved particle swarm optimization for short term power load forecasting: an integrated approach. *Energy* 2020;194:116847. <https://doi.org/10.1016/j.energy.2019.116847>.
- [42] Lebotsa ME, Sigauke C, Bere A, Fildes R, Boylan JE. Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem. *Appl Energy* 2018;222:104–18. <https://doi.org/10.1016/j.apenergy.2018.03.155>.
- [43] Debnath KB, Moushamed M. Forecasting methods in energy planning models. *Renew Sustain Energy Rev* 2018;88:297–325. <https://doi.org/10.1016/j.rser.2018.02.002>.
- [44] Chahkoutahi F, Khashei M. A seasonal direct optimal hybrid model of computational intelligence and soft computing techniques for electricity load forecasting. *Energy* 2017;140:988–1004. <https://doi.org/10.1016/j.energy.2017.09.009>.
- [45] Kong W, Jia Y, Dong ZY, Meng K, Chai S. Hybrid approaches based on deep whole-sky-image learning to photovoltaic generation forecasting. *Appl Energy* 2020;280:115875. <https://doi.org/10.1016/j.apenergy.2020.115875>.
- [46] Rahman S, Hazim O. Load forecasting for multiple sites: development of an expert system-based technique. *Elec Power Syst Res* 1996;39:161–9. [https://doi.org/10.1016/S0378-7796\(96\)01114-5](https://doi.org/10.1016/S0378-7796(96)01114-5).
- [47] Short-term load forecasting of Taiwan power system using a knowledge-based expert system. *IEEE Transitions Power Syst* 1990;5:1214–21.
- [48] Shi H, Xu M, Li R. Deep learning for household load forecasting-A novel pooling deep RNN. *IEEE Trans Smart Grid* 2018;9:5271–80. <https://doi.org/10.1109/TSG.2017.2686012>.
- [49] Kong X, Li C, Zheng F, Wang C. Improved deep belief network for short-term load forecasting considering demand-side management. *IEEE Trans Power Syst* 2020;35:1531–8. <https://doi.org/10.1109/TPWRS.2019.2943972>.
- [50] Ye C, Ding Y, Wang P, Lin Z. A data-driven bottom-up approach for spatial and temporal electric load forecasting. *IEEE Trans Power Syst* 2019;34:1966–79. <https://doi.org/10.1109/TPWRS.2018.2889995>.
- [51] Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Trans Smart Grid* 2019;10:841–51. <https://doi.org/10.1109/TSG.2017.2753802>.
- [52] Ahmad T, Zhang H. Novel deep supervised ML models with feature selection approach for large-scale utilities and buildings short and medium-term load requirement forecasts. *Energy* 2020;209:118477. <https://doi.org/10.1016/j.energy.2020.118477>.
- [53] Chang GW, Lu HJ. Integrating gray data preprocessor and deep belief network for day-ahead PV power output forecast. *IEEE Trans Sustain Energy* 2020;11:185–94. <https://doi.org/10.1109/TSTE.2018.2888548>.
- [54] Khodayar M, Wang J. Spatio-temporal graph deep neural network for short-term wind speed forecasting. *IEEE Trans Sustain Energy* 2019;10:670–81. <https://doi.org/10.1109/TSTE.2018.2844102>.
- [55] Wang F, Zhang Z, Liu C, Yu Y, Pang S, Duić N, et al. Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. *Energy Convers Manag* 2019;181:443–62. <https://doi.org/10.1016/j.enconman.2018.11.074>.
- [56] Massaoudi M, Refaat SS, Chihi I, Trabelsi M, Oueslati FS, Abu-Rub H. A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting. *Energy* 2021;214:118874. <https://doi.org/10.1016/j.energy.2020.118874>.
- [57] Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting. *Energy Convers Manag* 2019;198:111799. <https://doi.org/10.1016/j.enconman.2019.111799>.
- [58] Rasmussen CE, Williams CKI. Gaussian processes for machine learning. MIT press; 2006. p. 1–13. <https://doi.org/10.7551/mitpress/3206.001.0001>.
- [59] Quiñonero-Candela J, Rasmussen CE. A unifying view of sparse approximate Gaussian process regression. *J Mach Learn Res* 2005;6:1939–59.
- [60] MathWorks. Kernel (covariance) function options. MATLAB Doc; 2019. <https://au.mathworks.com/help/stats/kernel-covariance-function-options.html>.
- [61] Neal R. Bayesian learning for neural networks. *Lect Notes Stat -New York-Springer Verlag*; 1996. p. 1.
- [62] Ahmad T, Chen H. Short and medium-term forecasting of cooling and heating load demand in building environment with data-mining based approaches. *Energy Build* 2018;166:460–76. <https://doi.org/10.1016/j.enbuild.2018.01.066>.
- [63] Saini LM. Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning based artificial neural networks. *Elec Power Syst Res* 2008;78:1302–10. <https://doi.org/10.1016/j.epsr.2007.11.003>.
- [64] Moreira MO, Balestrassi PP, Paiva AP, Ribeiro PF, Bonatto BD. Design of experiments using artificial neural network ensemble for photovoltaic generation forecasting. *Renew Sustain Energy Rev* 2021;135:110450. <https://doi.org/10.1016/j.rser.2020.110450>.