Probabilistic Forecasting of Electricity Demand Incorporating Mobility Data

Israt Fatema * , **Gang Lei** **and Xiaoying Kong**

School of Electrical and Data Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia;
gang.lei@uts.edu.au (G.L.); xkong@mit.edu.au (X.K.)
* Correspondence: israt.fatema@student.uts.edu.au

Abstract: Due to extreme weather conditions and anomalous events such as the COVID-19 pandemic, utilities and grid operators worldwide face unprecedented challenges. These unanticipated changes in trends introduce new uncertainties in conventional short-term electricity demand forecasting (EDF) since its result depends on recent usage as an input variable. In order to quantify the uncertainty of EDF effectively, this paper proposes a comprehensive probabilistic EFD method based on Gaussian process regression (GPR) and kernel density estimation (KDE). GPR is a non-parametric method based on Bayesian theory, which can handle the uncertainties in EDF using limited data. Mobility data is incorporated to manage uncertainty and pattern changes and increase forecasting model scalability. This study first performs a correlation study for feature selection that comprises weather, renewable and non-renewable energy, and mobility data. Then, different kernel functions of GPR are compared, and the optimal function is recommended for real applications. Finally, real data are used to validate the effectiveness of the proposed model and are elaborated with three scenarios. Comparison results with other conventional adopted methods show that the proposed method can achieve high forecasting accuracy with a minimum quantity of data while addressing forecasting uncertainty, thus improving decision-making.

Keywords: probabilistic forecasting; Gaussian process regression; kernel density estimation; electricity demand forecasting; uncertainties

1. Introduction

Electricity consumption profiles worldwide have shifted in magnitude and daily patterns due to uncertain events, such as the COVID-19 pandemic, changing climate, and adverse effects of more frequent extreme conditions. These changes have caused significant difficulties and declined accuracy for traditional short-term electricity demand forecasting (EDF) methods. Accurate forecasting simultaneously becomes more important and challenging as the penetration of renewable energy increases the bi-directional communication between the supplier and the end-users. Optimal decisions can be obtained by improving forecast accuracy and quantifying uncertainty. Any uncertain events, such as the current COVID-19 pandemic, have brought much more effective practices for utility demand forecasting [1].

The socioeconomic severances affected total levels of electricity consumption, demand, price, and usage trends worldwide during the pandemic [1]. In Australia, the nationwide restrictions, which started in mid-March 2020 to control the spread of COVID-19, also decreased electricity demand. Figure 1 shows the monthly electricity time of use before and during the COVID-19 pandemic and the same period in 2021 and 2022, ranging from February to March, in NSW State, Australia [2]. It suggests that the electricity demand experienced significant drops during the pandemic compared with the same period in 2019. Moreover, the accuracy of the conventional models is not robust since time-varying electricity demand is more likely a non-stationary stochastic process, showing the challenges in forecasting caused by unprecedented events [3].

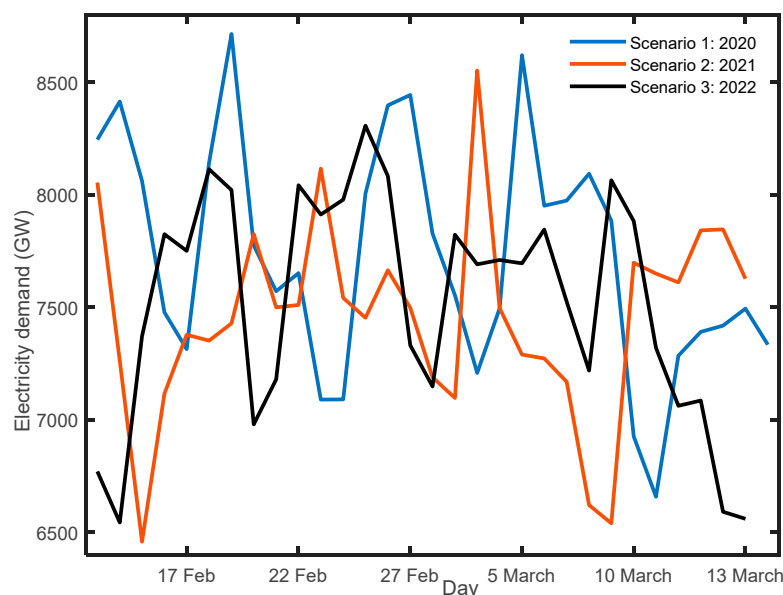



Figure 1. Comparison of electricity demand patterns from February to March 2020–2022.

Table 1 lists the annual consumption forecast accuracy (performance) in five regions of Australia between 2019 and 2022. It shows the percentage differences (errors) between actual and forecast values of the published energy forecasts (-ve error implies the actual is lower than forecast by %) adopted by the Australian Energy Market Operator (AEMO 2022) [4]. As can be seen from the table, forecast accuracy declined mostly over the pandemic period 2020–2022.

Table 1. Energy forecast accuracy (percentage error) in Australia (AEMO report 2022).

One-Year-Ahead Annual Operational Consumption Accuracy (%)	2018–2019	2019–2020	2020–2021	2021–2022
New South Wales	-2.0%	-0.26%	-1.1%	-3.9%
South Australia	-1.5%	2.6%	-0.3%	-0.8%
Tasmania	1.2%	2.2%	2.4%	-1.3%
Queensland	-3.9%	0.0%	-2.4%	-5.2%
Victoria	3.0%	1.3%	-1.7%	-8.4%

A few major factors contribute to the steep decline in forecasting accuracy [1,5]. (1) Limited data—small amounts of historical data generally decrease model accuracy. There are no historical events similar to the COVID-19 pandemic that can follow the consumption pattern since only months of data are available for model training and testing processes. (2) Any unknown underlying feature that affects usual usage patterns—conventional short-term forecasting models rely on long-term patterns and are not adaptive enough to learn about unprecedented events.

Both the spatial and temporal electricity demand patterns have changed in comparison with the non-pandemic period once the majority of people started to work from home, and the forced closure of industries slowed down other commercial activities [1,2]. These changing working conditions were eventually reflected in electricity grid planning, demand scheduling, renewable source integration, and spot pricing. EDF has an important role in the economy and is frequently used in business planning, policymaking, and market setting [6]. Moreover, it is crucial for the smart grid operation and may pose a technical risk during unusual situations when forecast accuracy significantly declines. Consequently, the aforementioned uncertainties affect the short-term forecasting algorithm's performance. Therefore, it is imperative to improve forecasting accuracy in terms of possible error reduction and uncertainty depiction.

The conventional deterministic forecasting (point prediction) methods provide a single expected value, which cannot incorporate the uncertainty information of the forecasting results [7]. In this context, an emerging method known as probabilistic forecasting can make effective inductive reasoning and therefore is more efficient for decision-making under dynamic scenarios [8]. EDF has been extensively studied, and numerous methods have been developed. However, earlier studies of EDF are mostly conducted during regular operations utilizing point forecasts and rarely address the uncertainties caused by any unprecedented events. EDFs are generally classified into two main categories: the statistical approaches and the artificial intelligence (AI) (machine learning) approaches [7]. Since electricity demand is affected by different independent variables such as weather/meteorological factors, statistical approaches which are based on linear models, such as exponential smoothing [7,8], multiple linear regression models (MLR), generalized autoregressive conditional heteroskedasticity (GARCH) [9], and autoregressive moving average (ARMA), are unable to account for the nonlinearity, non-stationarity, and randomness of time series data [9,10]. As shown in [9], GARCH models may capture a few aspects of uncertainty but not the volatile properties of time series.

Meanwhile, AI-based approaches, such as support vector regression (SVR) [11,12] and neural network (NN) methods [13,14], have been used in data analysis, pattern recognition, and EDF with accurate values. SVR is a complex computation involving a large dataset. In contrast, NN models are popular since they can simulate complicated nonlinear relationships. In addition, hybrid models [15–17] that combine the benefits of single models are gaining popularity for increasing forecast accuracy and resilience. The above-mentioned NN-based parametric methods often require a considerable quantity of data to discover relative patterns from samples due to their various parameters [13]. Given the limited amount of data during anomalous events, direct NN application is not recommended.

The above-listed time series approaches are all part of the point forecasting methods that predict a single expected value in look-ahead times to guide decision-making. However, earlier studies of EDF [5–17] were mostly conducted during regular operations utilizing point forecasts and rarely addressed the inherent uncertainties caused by unprecedented events such as pandemics. In contrast, a probabilistic forecast estimates the respective probabilities for all the possible future outcomes of a random variable concerning energy uncertainty, which is essential for making better decisions [8,18].

Three broad types of probabilistic forecasting methods have been developed: (1) input scenario based with simulated predictors [19], (2) interval construction and probabilistic forecasting models [20–23], and (3) post-processing through residual simulation [24,25]. Each of these can give a probabilistic result in the form of quantiles, intervals, or density functions [7].

Nevertheless, probabilistic forecasting approaches generally require a lot of data for training [8], making them inappropriate for EDF during unusual events when only a small sample of data is available. Gaussian process regression (GPR), a non-parametric method based on Bayesian theory, is preferable to parameterized methods for probabilistic regression analysis with a small number of training samples. [26]. In addition, GPR makes it possible to quantify the uncertainty of the forecasted values, which is extremely useful for security planning and operation of the system in safety-critical fields, such as energy systems [27].

GPR has been used for electricity demand prediction in several studies, and it can generate better estimates than benchmark methods for time series data, such as NN and SVR [16,28]. Moreover, when the Gaussian process (GP) is used to solve the regression issue, the accuracy of the forecast is sensitive to the choice of the covariance function. The kernel function solves the high-dimensional prediction distribution covariance matrix without extensive computing [29]. For the prediction of future electricity demand, a covariance kernel is built that incorporates daily/weekly trends and meteorological variables [27]. The effectiveness of three distinct kernels has been assessed in [16]. Decision-makers can pick kernel functions by comparing their performance across categories in time series

data, which improves probability prediction accuracy [30]. The deep Gaussian process (DGP) was used for EDF in [31]. The authors of [31] used sparse GP (SGP) and double stochastic variational inference DGP with mobility data to improve computational efficiency. However, SGP captures only higher-level latent space uncertainty [32]. DGP learns deep features from variable-sized data. Hence, they scale poorly with data size [33].

In [34], Gaussian process quantile regression (GPQR) was used to quantify the uncertainty in electricity demand forecasting since GPR is inherently capable of handling the complex interactions seen in time series.

In [35], applied weighted GPR was used to forecast the uncertainty of solar power, while [36] used GPR to predict the residential electricity demand rather than the system-level EDF. The researchers in [25] integrated both forecast combination and residual simulation approaches to model the forecast uncertainty. In [37], a hybrid model was proposed, which consists of kernel density estimation and Quantile regression (QR). QR has certain drawbacks; however, parameter estimation is more difficult than in Gaussian or generalized regression, which is a significant limitation [38]. In addition to that, the derived quantile curves may cross each other, resulting in an invalid response distribution [38]. This issue is mostly driven by the fact that these strategies estimate models individually for each quantile.

Interval prediction approaches provide lower and upper limits of the future forecast based on a confidence level yet cannot represent the full probability distribution of electricity demand [37]. By creating a probability density function (PDF) of forecasting results, probability density forecasting helps quantify uncertainty. Density estimation methods should be applied to enable GPR to obtain the PDF of the predicted electricity demand. This study takes advantage of the Gaussian kernel density estimation (KDE) approach to construct a probability density forecasting method due to its excellent generalization capabilities that greatly affect the distribution of response variables [39].

Moreover, the literature reports that numerous articles incorrectly assumed the GPR approach was deterministic and failed to fully use its benefits [40–45]. Using deterministic error metrics with the probabilistic EDF methods is one reason why they have not been developed enough. When using deterministic error metrics, the probabilistic EDF methods may have been underestimated because they did not work as well as their deterministic counterparts [7].

This paper aims to develop a comprehensive short-term probabilistic electricity demand forecasting method based on GPR-KDE that can deal with anomalous events. The contributions of this paper are summarized as follows:

- Considering the EDF uncertainty during unprecedented events, GPR is effectively applied to deal with the regression problem with limited historical data to perform successful inductive reasoning. Incorporating mobility data as an important feature that better represents the underlying shifts in practice theory [1,46] further improves its performance.
- Identifying highly correlated features with electricity demand output through a pre-processing stage makes the model simpler, demanding less data while prediction accuracy is still high.
- The performance and robustness of various GPR kernel and KDE models in point, interval, and density predictions of short-term electricity demand are compared comprehensively. Three different kernel functions are compared, and the optimal one for real applications is recommended.
- Comprehensive comparisons with other benchmark machine learning methods have been carried out using three scenario datasets. To validate the proposed model, a 5-fold cross-validation technique is used. Moreover, in order to test the accuracy of the model, statistical and probabilistic forecasting evaluation metrics are used. It shows that the proposed approach can better deal with anomalous events.

The remainder of the paper is organized as follows. In Section 2, the proposed method is described. Section 3 explains the whole procedure for implementing the EDF in detail.

Evaluation and different scenarios with real data are also demonstrated in this section. Section 4 provides an analysis, a discussion of the results, and conclusions.

2. Materials and Methods

This section presents a detailed description of how probabilistic ED forecasting handles uncertainty. It is challenging to produce a suitable adequate dataset for the parametric algorithm's training under anomalous events such as the COVID-19 pandemic because of the unprecedented changes in consumption patterns and the quantity of data. As parametric approaches optimize multiple parameters, they need a lot of data to detect relative trends. Therefore, it is not suitable in this scenario because only limited ED and mobility observations have been available from the beginning of the pandemic. Parametric distributions for the random variables are generally expected [47]. However, conditional predictive densities are not guaranteed to follow the same distribution, even when observations form a known and well-behaved marginal distribution [47]. Incorrect distributional assumptions potentially affect analyses and interpretation of the findings [47]. This work trains a model using a non-parametric GPR-KDE method, which does not require prior knowledge or assumptions about data distribution [39].

The forecasting set-up for electricity demand time series data is represented as follows: at time t, electricity demand at time $t + k$, k is referred to as the lead time, and y_{t+k} is the random variable which is electricity demand at time $t + k$. The historical electricity demand data $[y_{t-1}, \ldots y_t]$ and the training set (x_t) include different selected features, such as mobility data and temperatures. The objective of EDF is to develop a model $f(.)$ for y_{t+k} based on the gradient gained from optimizing the loss functions.

Probabilistic EDF typically involves three phases. Initially, the influential factors which are relevant to electricity demand variability should be identified from the historical dataset through a feature selection process. Next, the GPR algorithm is utilized to predict the electricity demand on different kernels. Finally, the probability density function can be obtained by KDE. The flowchart of the proposed GPR-KDE-based probability density forecasting method is shown in Figure 2. The evaluation metrics are also discussed in this section.

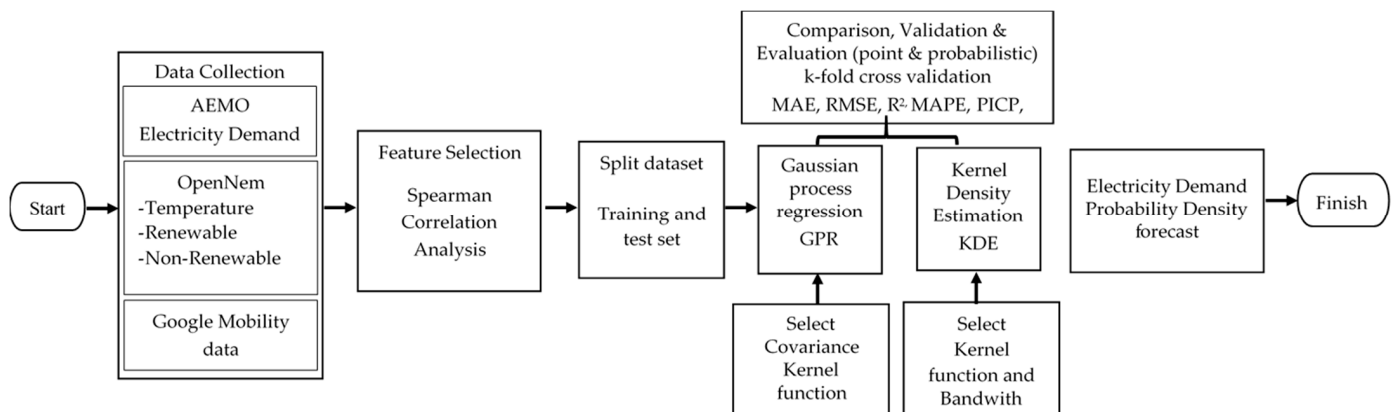



Figure 2. The flowchart of the GPR-KDE probability density forecasting model.

2.1. Gaussian Process Regression for Probabilistic Forecast

Gibbs and Mark initially presented the GPR model [48], which was later extended by Kersting et al. [49] and Tolvanen et al. [50]. GPR is a kernel-based nonlinear non-parametric regression technique based on Bayes' theorem. The covariance function is crucial in defining the relation between input data and output. GPR implies each sample follows the Gaussian distribution, and every linear combination of samples follows the joint Gaussian distribution [34].

Let the training data set $D_{tr} = \{(x_n, y_n)\}_{n=1}^N$, where x_n is the input of n number and y_n is the output, p stands for the number of selected features, i.e., $p = 5$; $X = [x_1, x_2, \ldots, x_n]^T$ is the input set containing all selected features and historical demand data, and $Y =$

$[y_1, y_2, \ldots, y_n]^T$ is the output set at time T. Assuming the input variables $f(X) = f(x_i)$ follow the n vector joint Gaussian distribution, then f is thus a *GP*.

It defines a probability distribution over functions and can be described as:

$$f(X) \sim GP\big(m(X), K(X, X')\big) \tag{1}$$

where $m(X)$ and $K(X, X')$ are the mean and covariance functions, respectively. The GPR model takes the mapping from X to Y as a *GP* and may relate the input with the output terms as follows:

$$Y = f(X) + h(X)^T \beta + \varepsilon \tag{2}$$

where Y is an observation, $f(X)$ is the mapping function, $h(X)$ is a set of basis functions that transform the original feature vector X into a new feature vector $h(X)$ which is in $p \times 1$ vector, β is a p-by-1 vector of basis function coefficients, and $\varepsilon \sim N(0, \sigma_n^2 I)$ is noise; σ_n^2 is the standard deviation; I is an identity matrix with appropriate dimensions. Since the noise is independent, Y is also a *GP*:

$$Y \sim GP\Big(m(X), K(X, X') + \sigma_n^2 I\Big) \tag{3}$$

The prior distribution of Y can be learned from the training data using Bayes' theory:

$$Y^p \sim N\Big(0, K(X, X') + \sigma_n^2 I\Big) \tag{4}$$

The prior joint distribution of training set output Y and the test output f_* are:

$$\begin{bmatrix} Y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X, X) + \sigma_n^2 I_n & K(X, x_*) \\ K(x_*, X) & K(x_*, x_*) \end{bmatrix}\right) \tag{5}$$

where $K(X, X) = \big(K(x_i, x_j)\big)$ represents the $N \times N$ covariance matrix on inputs of the training set D_{tr}; $\big(K(x_i, x_j)\big)$ is the kernel function; $K(X, x_*) = (k(x_t, x_*))$ represents the $N \times 1$ vector of covariance between the test point x_* and the training inputs in D_{tr} $K(x_*, x_*)$ is the covariance of the test points. The posterior distribution of test set prediction f_* can be obtained as follows:

$$p(y|x_*, D) \sim N(y \big| h\beta + \overline{f}_*, cov(f_*)) \tag{6}$$

$$\overline{f}_* = k(x_*, x)^T (k + \sigma^2 I)^{-1} y \tag{7}$$

$$cov(f_*) = k(x_*, x_*) - k(x_*, x)^T (k + \sigma^2 I)^{-1} k(x_*, x_*) \tag{8}$$

where $\overline{\mu} = \overline{f}_*$ and $\overline{\sigma}_{f_*}^2 = cov(f_*)$. There are many different covariance kernel functions for the GPR model to select:

1. Squared exponential kernel function:

$$K(x_i, x_j) = \sigma_f^2 \exp[-\frac{1}{2}\frac{r^2}{\sigma_l^2}] \tag{9}$$

where σ_l is the characteristic length scale and σ_f^2 is the variance of time series. Both σ_f^2 and σ_l are parameters to be optimized during the training. $r = \sqrt{(x_i - x_j)^T (x_i - x_j)}$ is the Euclidean distance between x_i and x_j.

2. Exponential Kernel:

$$K(x_i, x_j) = \sigma_f^2 \exp(-\frac{r}{\sigma_l}) \tag{10}$$

3. Matern 3/2 Kernel:

$$K(x_i, x_j) = \sigma_f^2 \exp\left(1 + \frac{\sqrt{3}\,r}{\sigma_l}\right) \exp\left(-\frac{\sqrt{3}\,r}{\sigma_l}\right) \tag{11}$$

In addition to the covariance matrix K, the parameters of $\sigma^2 I$ can be learned together. The optimization parameters can be specified as $\theta = \left\{\sigma_f^2, \sigma_l^2, \sigma^2\right\}$. To estimate the parameters, we maximize the following marginal likelihood function:

$$\hat{\theta} = \mathrm{argmax}\, p(D|\beta, \theta) \tag{12}$$

The mean $\overline{\mu}$ and variance $\overline{\sigma}_{f_*}{}^2$ of the test point x_* can be calculated according to Equations (7) and (8) by identifying the optimal parameters. The variance enables quantification of the EDF uncertainties. To quantify the uncertainty, the interval predictions corresponding to certain confidence levels (such as 95%) for n samples are: ($l = \overline{f}_* - \frac{t\left(\frac{\alpha}{2}, n-1\right)*\overline{\sigma}_{f_*}{}^2}{\sqrt{n}}$, $\mathrm{u} = \overline{f}_* + \frac{t\left(\frac{\alpha}{2}, n-1\right)*\overline{\sigma}_{f_*}{}^2}{\sqrt{n}}$), where l and u represent the lower and upper bounds of the confidence interval, respectively; α is the degree of confidence. The value of the t distribution (denoted as $t(.)$) may be found by referencing the t table.

In addition, to improve prediction accuracy, this research explores the effects of different scenarios of electricity demand datasets on different kernel functions and selects the best kernel function.

2.2. Probability Density Prediction Based on Kernel Density Estimation

The kernel density estimator estimates a smooth density from data samples by assigning each sample point a density function [51]. All of these contributions are added together to determine the distribution. The normal distribution has been widely employed to estimate the density function in research. However, it only works effectively for data that follows a bell-shaped distribution. This method utilizes the kernel density estimate (KDE) method, a non-parametric representation of the probability density function, to avoid making assumptions about the distribution of the data. Unlike interval estimation, KDE is based on the observed data to construct the underlying PDF without distributional assumptions for the shape of the density, allowing the estimated density to be, for example, multi-modal, fat-tailed, or skewed. The selection of the kernel function and the determination of bandwidth are crucial to KDE [37]. Applying KDE to the predicted distribution from GPR gives a smooth curve estimation of the underlying PDF. It is helpful for creating a more accessible and understandable representation of the distribution, including by plotting a histogram with a fitted density curve [8]. The KDE smooths the distribution and estimates the PDF continuously, eliminating the separate-out impact of the histogram bins [51].

If sequence Y consists of $n \times 1$ dimensional observations, the drawing sample is $Y = [y_1, y_2, \ldots, y_n]$, and then they can be organized into a histogram's bins. Depending on the distance between the sample drawings, the histogram contains a variety of bins with higher heights than others. For instance, if the values of two drawings are similar and the size is small, these drawings will be placed in the same bin. Therefore, the kernel density estimator attempts to average away the effect of each data point y_i, $(where\ i = [1 \ldots .n])$ in a non-smooth histogram by providing each point a kernel function with a specific width. The KDE method estimates the actual probability density function f through the following function:

$$\hat{f}_n(y) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{y - y_i}{h}\right) \tag{13}$$

where h is bandwidth, $k(.)$ is the kernel function, which is a symmetric function that is integrated into one and has a mean of zero. These kernel functions usually represent some kind of similarity between two points in a space [34]. Several kernel density functions can estimate PDF, such as uniform, triangular, triweight, Epanechnikov, Gaussian, etc.,

empirically; however, it makes little difference which one is used [52]. To implement the KDE in this study, a Gaussian kernel is utilized. It substitutes each sample point with a Gaussian-shaped kernel and adds these Gaussians to estimate the density.

$$k\left(\frac{y - y_i}{h}\right) = \left(\frac{1}{\sqrt{2\pi}}\right) \exp\left(\frac{(y - y_i)^2}{2h^2}\right) \tag{14}$$

The smoothing parameter $h > 0$ is the bandwidth, which modifies the distribution's overall appearance. The use of Gaussian KDE has certain benefits over other bandwidth selection methods since it can automatically determine the bandwidth using a rule of thumb [53,54].

2.3. Evaluation Metric

The most important issue with a probabilistic forecast is that the actual distribution of the underlying method is unknown [55]. The predicted and real distributions of EDF cannot be compared using only previous demand. There are several approaches to assess the effectiveness of probabilistic forecasts, with the approach selected based on the intended objective. We may use tests and parameters to validate the model and choose the best model. The model's accuracy may be verified, and selection criteria can be established with the use of tests and parameters. To test the accuracy of the proposed method for EDF, different tests and validation methods are utilized. Both deterministic and probabilistic error metrics are evaluated in this study.

2.3.1. Metrics for Point Forecasting

Several methods have been used to assess the efficacy of prediction models in the literature. Commonly used metrics to evaluate deterministic/point forecast accuracy are the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and correlation coefficient value (R^2) [56]. A lower RMSE and MAE imply a more accurate forecast, which evaluates the difference between actual and predicted values. The R^2 value determines the correlation between actual and predicted values, which is between 0 and 1 (where 0 means no correlation and 1 means the model has no error). The three error measures are defined as follows:

$$\text{RMSE}(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left(h(x^{(i)}) - y^{(i)}\right)^2} \tag{15}$$

$$\text{MAE}(X, h) = \frac{1}{m} \sum_{i=1}^{m} |h(x^{(i)}) - y^{(i)}| \tag{16}$$

$$R = \frac{\sum_{i=1}^{m} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{m}(x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{m}\left(y_i - \overline{y}\right)^2}} \tag{17}$$

where x is considered to be the actual values, y defines the predicted values, $\overline{x}$ defines the mean of x, $\overline{y}$ defines the mean of y, X is a matrix containing all the features value, h is the prediction function, and m is the total number of instances in the test set.

2.3.2. k-Fold Cross-Validation

The most popular validation strategies in recent research are k-fold cross-validation and hold-out validation [8,57,58]. Hold-out validation does not perform well with small datasets due to increased volatility in performance estimation and a larger risk of overfitting with limited training data [57]. The k-fold cross-validation provides insight into GPR model performance by partitioning the dataset into k-folds and training the model on k-1 folds while testing on the remaining fold. The process is then repeated k times, with each fold being used exactly once as the test data. This study has used 5-fold cross-validation, and the model's performance is estimated by averaging each evaluation, such as MAE, RMSE,

and R^2. Furthermore, it provides a more accurate estimation of the model's generalization error performance than a single train–test split, which is used to identify overfitting or underfitting [57,58].

k-Fold cross-validation compares training and test error to assess GPR overfitting. The model's training error is made on the complete dataset, while the model makes a test error on the test set during cross-validation. The model is overfitting the data if the test error is considerably larger than the training error. However, large training and test errors indicate that the model is underfitting the data and not reflecting the underlying pattern [57].

2.3.3. Metrics for Probabilistic Forecasting

Probabilistic forecasting is typically evaluated in terms of its reliability, sharpness, and competence (skill). The prediction interval coverage probability (PICP) measures the reliability of the predictions via coverage rate [59]. It shows the percentage of the actual values that will be covered within certain prediction interval limits [59]. The considerably larger the PICP, the more likely it is that the real values will be within the prediction interval (PI). PICP is defined as follows:

$$PICP = \frac{1}{N} \sum_{i=1}^{N} C_i,$$ (18)

where N is the number of samples, and C_i is a Boolean variable defined as follows:

$$C_i = \begin{cases} 1, & y_i \in [L_i, U_i]; \\ 0, & y_i \notin [L_i, U_i], \end{cases}$$ (19)

where L_i, and U_i are the lower and upper PI boundaries of the target y_i, respectively. PICP ranges between 0–100%. The PI is accepted as valid if the PICP value is greater than the prediction interval nominal confidence $PINC = 100(1 - \alpha)\%$; here, α indicated the probability of error. In this study, the PI limits are supposed to cover 95%, 90%, 80%, and 50% of the PDF of the forecast. The evaluation of the PICP alone is misleading since high PICP values can be easily reached when the width of the PIs is large. In contrast, a forecast with a narrow PI and a large PICP is more reliable. The effectiveness of PIs is determined by their widths. Thus, larger PIs are less informative for making decisions since it presents increased uncertainty [59]. Therefore, to evaluate the widths of the PI, a supplementary metric is required, and lower MPIW and high PICP values help make better decisions [59]. The metric is the Mean Prediction Interval Width (MPIW), which is defined as follows:

$$MPIW = \frac{1}{N} \sum_{i=1}^{N} (U_i - L_i)$$ (20)

To evaluate the calibration and sharpness of the forecasted PDF simultaneously, the continuous ranked probability score (CRPS) metric is used and can be expressed as follows [60]:

$$CRPS = \int_{-\infty}^{\infty} (F(x) - 1\{x \geq y\})^2 dx.$$ (21)

where $F(x)$ represents the predicted cumulative distribution function (CDF) of the variable of interest x, and y is the verifying observation. If the forecast variable x equals the observation y, then the value jumps from zero to one. The squared difference between the two CDFs is averaged over the number of observation pairs. The smaller the CRPS metric, the better the accuracy of the PDF.

2.4. Dataset and Data Pre-Processing

Several real-world data sources are used to evaluate and demonstrate the effectiveness of the proposed method. This study focuses on data from three years: 2020, 2021, and 2022, onset and during the COVID-19 pandemic as three different scenarios to capture and illustrate

the forecasting uncertainty. The data are from New South Wales (NSW), Australia's most populous state; a similar range of data is used, dating from 12 February–13 March.

To achieve optimal data quality, the entire dataset is examined, and missing data points are populated by neighborhood values. The data sources of this study are as follows:

- Electricity demand data: AEMOs' open dataset is used, which contains accumulated daily electricity demand sampling rate (30 min) [3].
- Energy and weather data: OpenNEM, an open platform for National Electricity Market (NEM) data, is used, which contains time series hourly weather (temperature), renewable and non-renewable energy generation, and consumption data [61].
- Mobility data: The mobility data is obtained through the Google COVID-19 Community Mobility Reports and have six location-specific metrics (parks, workplaces, residential, retail and recreation, and grocery and pharmacy) [62].

Feature Selection

To improve the forecasting accuracy, a correlation analysis is performed to determine which features are most strongly related to electricity demand. Due to the influence of numerous factors, such as temperature, day of the week, and volatility of renewable sources, it is difficult and complex to establish the true impact of the pandemic on electricity demand. Studies have found that temperature significantly affects energy use compared to other parameters [17].

Feature selection reduces change sensitivity and overfitting in the proposed model. Therefore, this study uses the correlation coefficient between temperature indicators and electricity demand to determine whether incorporating mobility data have the same or greater importance in highlighting their qualities. There are different methods for choosing features, such as Pearson and Spearman correlations. This study adopts Spearman's rank correlation coefficient to test the electricity demand–mobility correlation since it is a non-parametric test and defines the strength and direction of the monotonic relationship between two variables [63]. The Spearman's rank correlation coefficient statistic can be defined as follows:

$$s = \frac{\sum_{i=1}^{n}(R_i - \overline{R})(S_i - \overline{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \overline{R})^2 \sum_{i=1}^{n}(S_i - \overline{S})^2}} = 1 - 6\sum_{i=1}^{n}\frac{d_i^2}{n(n^2 - 1)} \tag{22}$$

where R_i and S_i are the i_{th} rank of the first and second data inputs, respectively, and $d_i^2 = (R_i - S_i)^2$ represents the distance between two data inputs.

The S value of Spearman correlation coefficients of different features during the pandemic period (February–March 2020–2022) is shown in Table 2 and Figure 3, which reveal that temperature and non-renewable sources of energy have a strong correlation with the electricity demand. Selected features for this study are temperature, renewable and non-renewable energy, and workplace and residential mobility data.

Table 2. The Spearman Correlation Coefficient (*S*) results from electricity demand and mobility indicators, temperature, and energy sources.

Spearman Coefficient S	Weather	Energy Source		Mobility Feature					
	Temperature	Renewable	Non-Renewable	Workplace	Residential	Retail & Recreation	Grocery & Pharmacy	Parks	Transit Stations
2020	0.956	0.077	0.681	0.427	-0.436	-0.419	-0.483	-0.303	0.3116
2021	0.810	0.023	0.577	-0.063	0.225	-0.394	-0.236	-0.252	-0.119
2022	0.828	0.261	0.606	-0.399	0.447	-0.488	0.116	-0.363	-0.047

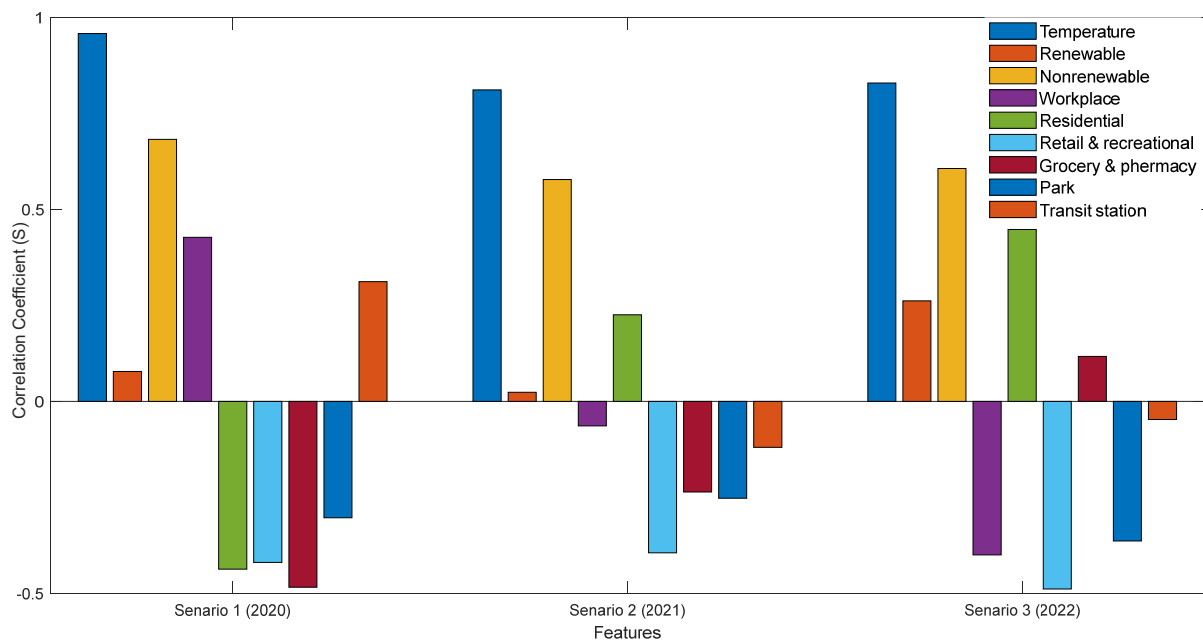



Figure 3. The correlation coefficient *(S)* of different features with the electricity demand.

In addition, all of the 6 mobility indicators have a stronger correlation with electricity demand. This indicates that mobility indicators can reflect electricity demand during an anomalous situation such as a pandemic. The *S* value of the workplace was high just before the start of the pandemic, and it became negative during the lockdown period of 2021–2022 and vice versa in the case of other mobility indicators, especially the residential feature, which showed a negative value just before the pandemic and became positive during the year 2021–2022. It shows that staying longer at home for a longer length of time reduces electricity demand, which corresponds to the actual scenario throughout the pandemic. Hence, incorporating mobility features in electricity demand forecasting could increase model performance.

3. Results

This section first executed several tests on different real-world data sets to demonstrate the effectiveness of the proposed forecasting model. Next, test evaluations between a few existing point forecasting models are demonstrated. These simulation results explain the performance of the proposed model in addition to the error metrics. Then, comparative results for probabilistic forecasting are presented and discussed. Finally, it discusses the ablation study.

Three common covariance functions are tested to choose an appropriate one for the proposed GPR model, namely the following: squared-exponential, exponential, and Matern 3/2. The three scenarios select 95%, 90%, 80%, and 50% confidence levels of the PI for the probabilistic GPR model training. Finally, the best-performing PI result is used as input to derive PDFs of electricity demand forecasting over time, which are then checked against observed values using the KDE method.

In this study, the offline method is used to make training as straightforward as possible. All the forecasting processes are implemented with Matlab R2022b and tested using a personal computer with Intel(R) Core(TM) i7-8650U CPU @ 1.90GHz 2.11 GHz, 16 GB RAM.

To determine the optimal kernel function, this paper first does comparative studies of three common kernels. Table 3 presents the forecasting errors of the proposed model. Four forecasting error measurements, MAE, RMSE, and R^2 are utilized to evaluate the model. To summarize the results of Table 3:

Table 3. Forecast test set error on three scenarios.

	Scenario 1 (2020)			Scenario 2 (2021)			Scenario 3 (2022)		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
GPR (Exponential)	35.76	48.21	0.99	49.85	61.86	0.98	38.97	48.36	0.99
GPR (Squared exp.)	55.92	71.21	0.98	49.95	60.18	0.98	87.35	107.37	0.95
GPR (Matern 3/2)	39.65	53.97	0.98	40.74	50.43	0.98	76.99	94.77	0.96

Scenario 1: the year 2020 gives the best prediction result for EDF using the Exponential covariance function, the MAE 35.76, RMSE 48.21, and RMSE 0.99.

Scenario 2: the year 2021 gives the best prediction result for EDF using the Matern 3/2 covariance function, the MAE 49.95, RMSE 60.18, and RMSE 0.98.

Scenario 3: the year 2022 gives the best prediction result for EDF using the Exponential covariance function, the MAE 38.97, RMSE 48.36, and RMSE 0.99.

Therefore, the test results observe that the Exponential covariance function adequately captures the electricity demand pattern in two scenarios (1 and 3) and performs well.

Table 4 presents the training set error statistics. The testing data error measurements between actual and predicted values are based on the average error of all 5-folds for the training set. To interpret these values, notice that the difference between test and training set error are very small.

Table 4. Training set errors based on three scenarios.

	Scenario 1 (2020)			Scenario 2 (2021)			Scenario 3 (2022)		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
GPR (Exponential)	35.01	48.11	0.99	48.02	60.47	0.98	38.40	48.01	0.99
GPR (Squared exp.)	54.80	70.34	0.98	48.27	61.32	0.98	86.57	106.98	0.95
GPR (Matern 3/2)	38.43	53.01	0.98	41.44	50.10	0.98	77.09	93.98	0.96

3.1. Point Forecasting

Given that the proposed comprehensive GPR model also provides point forecasts, its performance is compared to four baselines models: (1) Regression tree (RT), (2) Support vector regression (SVR), (3) Back-propagation NN (BPNN), and (4) QR.

The purpose of this evaluation is to demonstrate that the proposed method can improve forecasting models with limited data better than the baseline models. The test data set is used to fine-tune the hyper-parameters of each approach, and the best-performing model is then selected. The comparison of the forecasting results of various methods on test data is summarized, and Table 5 outlines the best possible results for each method. It is apparent from Table 5 that the GPR-based approach outperforms the four baselines based on the results in three different scenarios (the year 2020–2022).

Table 5. Forecasting errors comparison of SVR, RT, BPNN, and GPR in three scenarios.

	Scenario 1 (2020)			Scenario 2 (2021)			Scenario 3 (2022)		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
GPR	35.76	48.21	0.99	40.74	50.43	0.98	38.97	48.36	0.99
BPNN	90.11	114.29	0.94	90.85	105.77	0.93	110.66	136.74	0.92
RT	120.3	181.9	0.86	250.02	300.67	0.29	232.58	237.69	0.69
SVR	70.21	73.248	0.96	81.771	91.00	0.94	100.48	125.59	0.92
QR	54.18	69.42	0.98	47.31	57.50	0.98	86.53	106.30	0.95

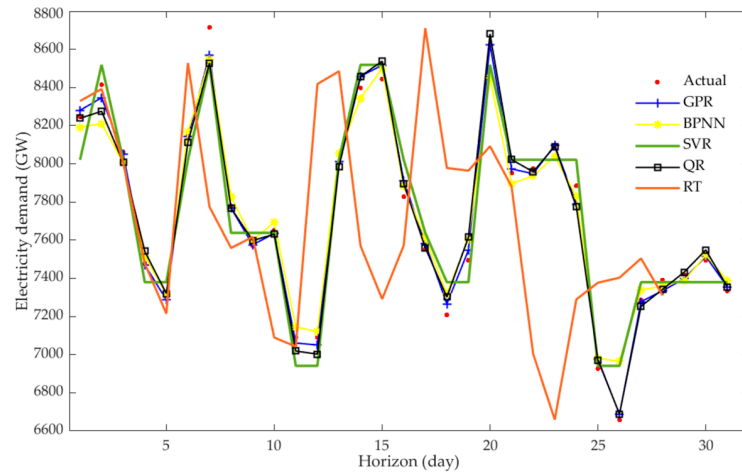
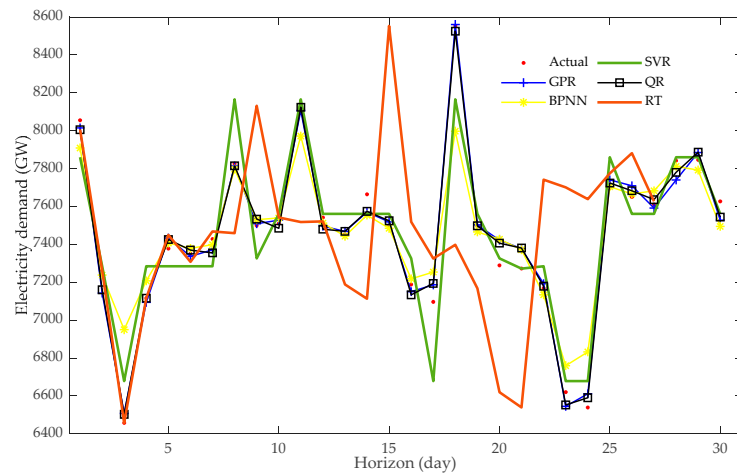
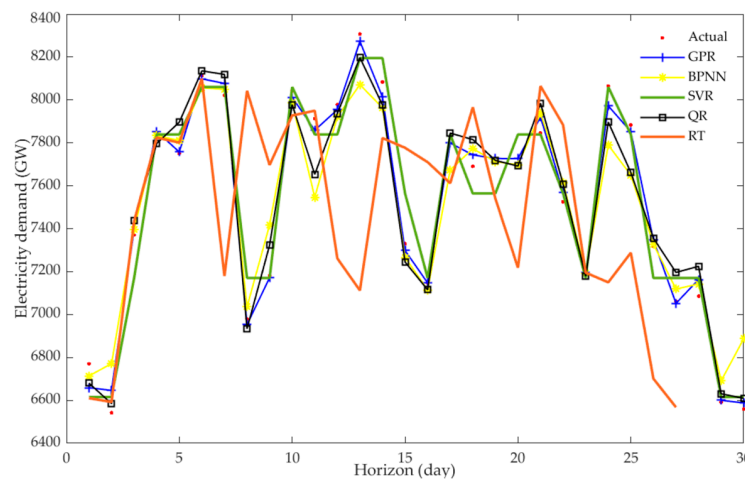
The time series forecasting results (of three scenarios) of electricity demand on various methods are shown in Figure 4. It can be observed that the RT suffers from larger deviations from the actual demand compared with GPR-based methods. The proposed approach gets closer to the actual data than the QR, SVR, and BPNN methods. The results are consistent with those in Table 5.

(**a**) Scenario 1

(**b**) Scenario 2

(**c**) Scenario 3

Figure 4. Day-ahead time series forecasting results of different methods for three scenarios.

QR and the proposed model have the highest R^2 values of 0.98 and 0.99, respectively. This displays a significant correlation between actual and predicted values. As compared to other approaches, RT's performance was subpar. Hence, regression tree methods overfit and have large variances due to poor performance on unseen test data [64]. Regarding BPNN, this parametric algorithm has several parameters to optimize. Hence, samples from small training data cannot reveal all hidden patterns. For QR, parameter estimation is difficult with limited data [38]. After QR, Kernel function-based SVR has performed well compared to RT and BPNN with an average R^2 of 0.94. As a result, GPR, QR, and SVR are preferred models for estimating the impact of uncertainty.

Compared to QR, BPNN, RT, and SVR, the RMSE index of the proposed model has been averagely improved in the case of electricity demand by 36%, 58%, 79%, and 49%, respectively. Similarly, the performance error MAE shows an improvement of 38%, 60%, 80%, and 54%, respectively. The findings indicate that for point forecasting, the proposed GPR model has better performance with an acceptably smaller error than the other four baseline models despite trend change phenomena triggered by unprecedented events such as pandemics and extreme weather conditions.

3.2. Probabilistic Forecasting—Prediction Interval

Predictions Interval (PI) under four confidence levels (CI) (95%, 90%, 80%, and 50%) are carried out by various kernel functions of the GPR and QR model in support of verifying the effectiveness of the proposed model. Both the PICP and MPIW of the PI for three scenarios are then calculated, and the performance comparisons are displayed in Table 6. The actual electricity demand is represented in Figure 5a–c by the black line and GPR prediction with a black dotted line. The 95%, 90%, 80%, and 50% values of PI are depicted in Figure 5. The results for RT, SVR, and BPNN are not shown since these are only point forecasting methods.

Table 6. Prediction Interval evaluation achieved by various kernel methods in three scenarios.

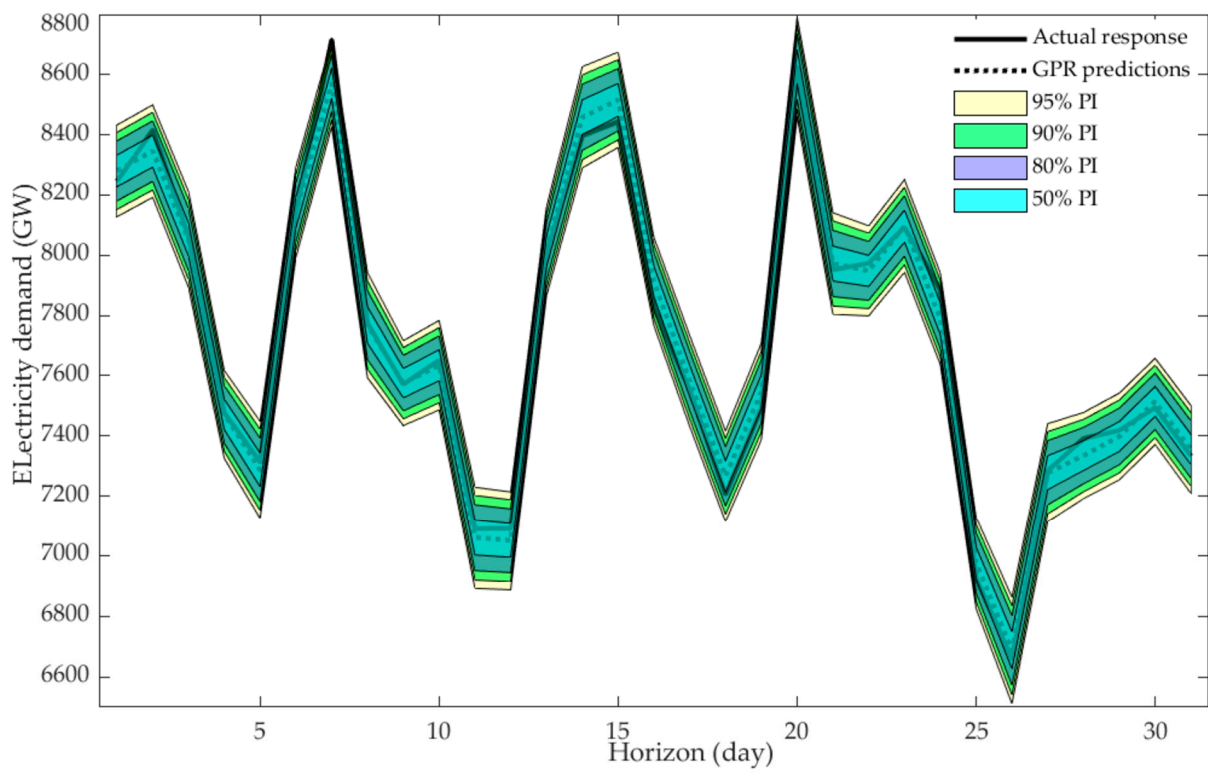
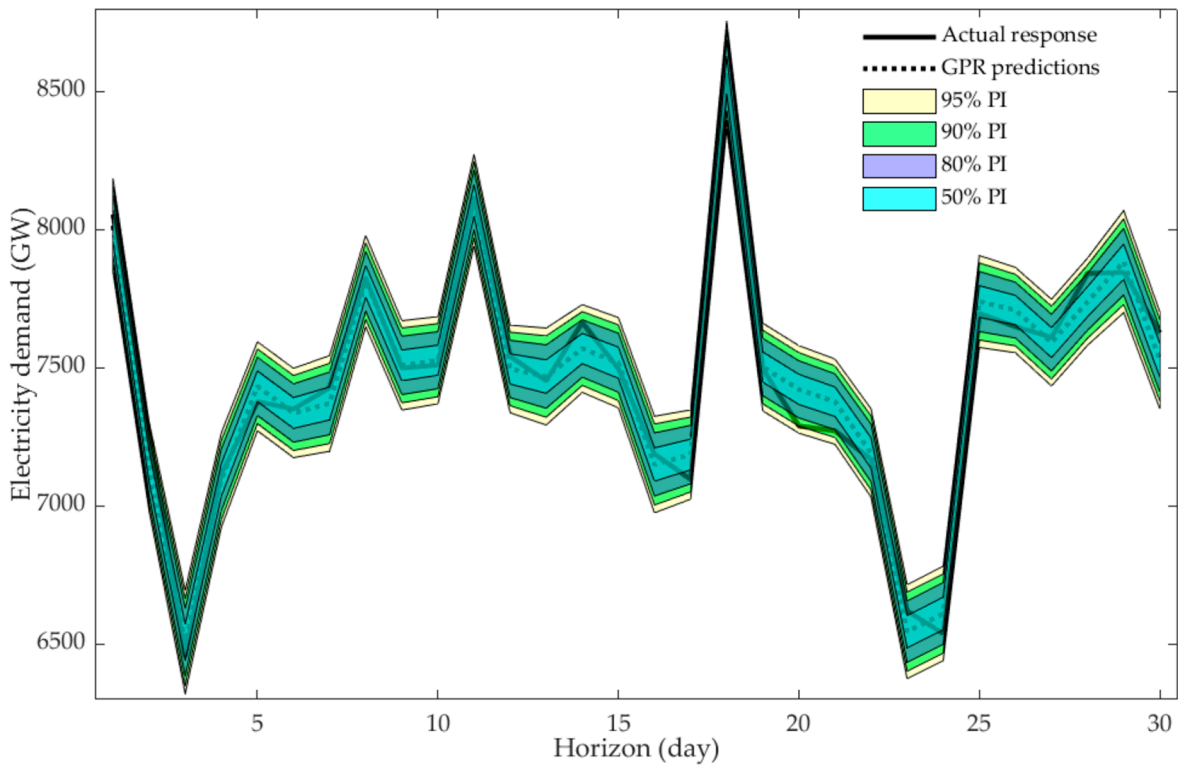
Scenario	Covariance Function	PICP (%)				MPIW (%)			
		95%	90%	80%	50%	95%	90%	80%	50%
Scenario 1 2020	GPR (Exponential)	98%	96%	96%	77%	30%	26%	20%	11%
	GPR (Squared exp.)	96%	96%	93%	64%	38%	32%	25%	13%
	GPR (Matern 3/20)	97%	94%	93%	74%	35%	30%	23%	12%
	QR	95%	95%	92%	63%	39%	33%	26%	14%
Scenario 2 2021	GPR (Exponential)	97%	96%	87%	61%	33%	27%	21%	11%
	GPR (Squared exp.)	96%	96%	90%	64%	35%	29%	23%	12%
	GPR (Matern 3/20)	96%	96%	93%	74%	34%	28%	22%	11%
	QR	95%	95%	89%	63%	36%	30%	24%	13%
Scenario 3 2022	GPR (Exponential)	96%	96%	96%	36$%	37%	31%	24%	12%
	GPR (Squared exp.)	96%	93%	90%	58%	57%	37%	48%	19%
	GPR (Matern 3/20)	96%	96%	90%	64%	56%	47%	37%	19%
	QR	95%	92%	89%	57%	58%	48%	38%	20%

(**a**) Scenario 1

(**b**) Scenario 2

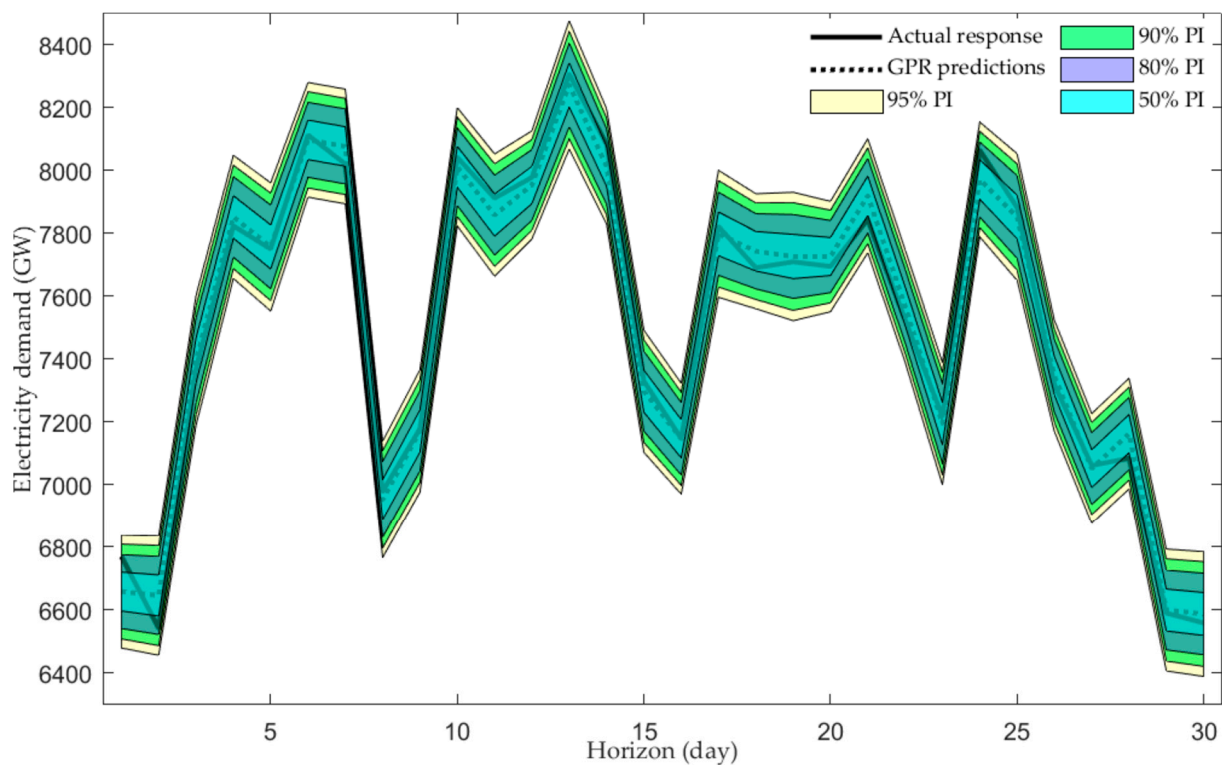
Figure 5. *Cont.*

(**c**) Scenario 3

Figure 5. Day-ahead prediction interval results of three scenarios (using GPR—Exponential).

It can be seen that the PI generated with the proposed method covers the actual value most of the time. For example, Figure 5c shows that the proposed method has a 95% confidence interval that can cover most of the actual electricity demands. Table 6 results show a good performance and indicates that the PICP and MPIW are exceeding the CI levels in most of the cases. For example, the PICP of scenario 1 (before the pandemic), scenario 2 and 3 (during the pandemic) are 98%, 97%, and 96%, respectively, exceeding the CI level of 95% with an MPIW of 30%, 33%, and 37%, respectively, using the exponential kernel function. In comparison, the 95% confidence level by the other two covariance functions (Squared exponential and Matern 3/2) does not cover many actual electricity demand points. However, the MPIW values by the Matern 3/2 covariance function of the GPR model is wider than that of the other two kernels in scenario 1 and 3. In addition, the PICP values of the QR model is marginally covering the preassigned PICP, and the MPIW values are also wider than that of the other GPR kernel models.

The results show that the created PIs of the proposed model have a high probability of covering the target values and better captures uncertainty with sudden peak and descend values during the pandemic (scenario 2 and 3) that is hardly following any pattern. As shown in Figure 5, the actual values are mostly located in narrow PIs with 98%, 97%, and 96% PICP for scenarios 1, 2, and 3, respectively.

As shown in Figure 5b, the difference between the two peaks of some daily electricity demand curves is relatively small. For example, there is not much variance between days 15 and 16. Meanwhile, the peaks of some daily EDF curves have a large disparity; for example, day 17, 18, and 19 has a sudden peak and drop in electricity demand. The fluctuation range of the electricity demand varies daily, reflecting the uncertainty. Concurrently, scenario 1, Figure 5a, has comparatively less disparity between different days of EDF. Given that the short-term EDF is very unpredictable, the probabilistic performance of the proposed method is adequate.

3.3. Probabilistic Density Prediction

From the analysis of the previous sections, Predictive intervals provide a lower and upper bound between which the actual future value will fall within a certain probability (confidence interval). Finally, probabilistic forecasts can be provided as predictive probability density functions, which provide a full estimation of the probability distribution around a point forecast. The lower and upper bounds of the PI in Section 2.1 are used to estimate the KDE as a range of possible values and create a histogram with the values falling within this range. Using KDE, the underlying density function can be estimated after the histogram is obtained [8]. To estimate the underlying density, KDE includes integrating the histogram with the Gaussian kernel function. This study uses the CRPS to quantify the calibration and sharpness for density forecast evaluation. Table 7 compares the CRPS values with different kernel functions of GPR, and QR probabilistic forecasting techniques are also tested for comparison.

Table 7. This Probabilistic density prediction results from three scenarios.

Scenario	Covariance Function	CRPS
Scenario 1 2020	GPR (Exponential)	21.35
	GPR (Squared expo.)	43.23
	GPR (Matern 3/2)	27.48
	QR	44.74
Scenario 2 2021	GPR (Exponential)	36.80
	GPR (Squared expo.)	38.76
	GPR (Matern 3/2)	30.26
	QR	39.85
Scenario 3 2022	GPR (Exponential)	22.60
	GPR (Squared expo.)	69.55
	GPR (Matern 3/2)	53.21
	QR	70.78

Figure 6 provides the complete electricity demand probability density curves. To make the comparisons, the actual electricity demand values are also presented with the point forecasting value of different baseline methods of three scenarios. To provide the proposed method's potential to incorporate uncertainty, the prediction results from Figure 5a–c are illustrated in Figure 6 (density curve). From the prediction interval (Figure 5), the highest actual electricity demand values from three scenarios are day 7 in 2020, day 18 in 2021, and day 13 in 2022.

Based on Figure 6, most of the actual values distribute in the middle of probability density curves. This indicates these values are anticipated in the forecasting distributions with high probability. Moreover, Figure 6 shows that the baseline point forecasting approaches do not yield reliable results and cannot capture the uncertainty caused by unprecedented events, such as the COVID-19 pandemic. It is also evident that the observation lines in scenario 1 are closer to the center of the PDF while others (scenarios 2 and 3) are slightly distant from the center, which indicates that the probabilistic forecasting is reliable. Concurrently, the other possible explanation is that many variables influence consumption patterns and become critical to EDF with precision. Hence, the suggested technique adopts a conservative method to reach conservative decisions under varying impacts.

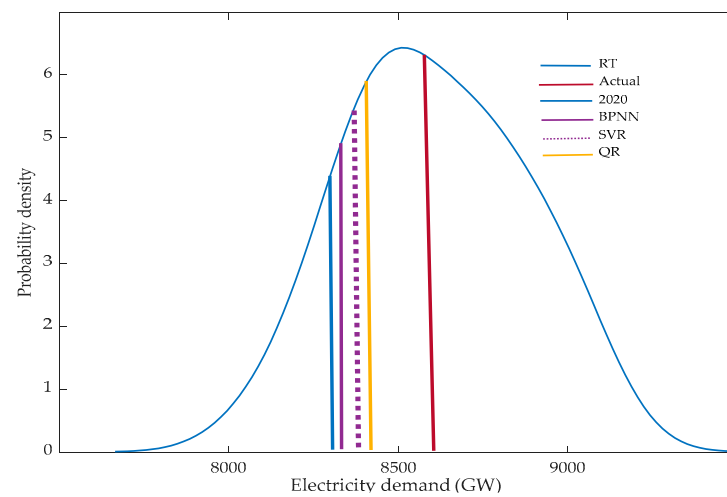
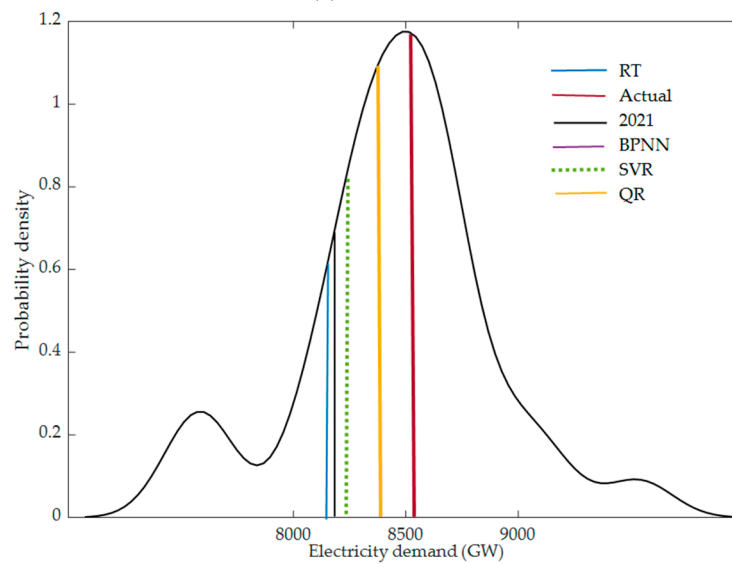
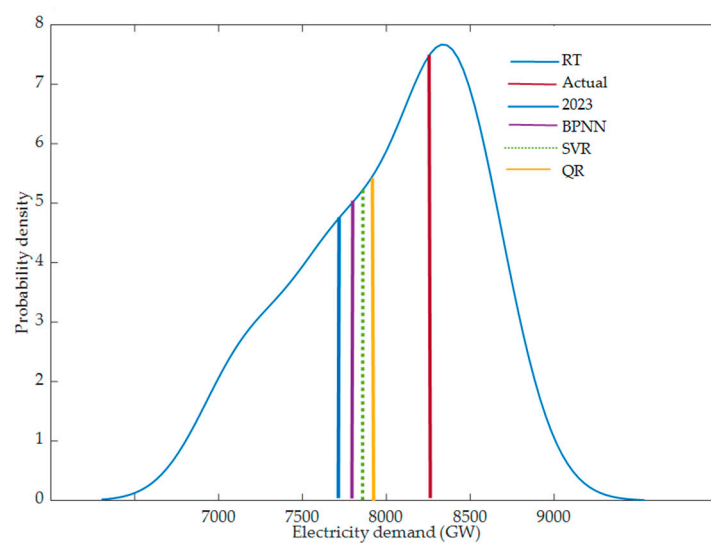
Figure 7 shows the histograms and the PDFs based on the Gaussian kernel density estimation (KDE) (red line) functions of residuals between the actual and the predicted electricity demand under three scenarios. It can be seen that the estimated KDE functions well matched the histograms of the EDF residuals in three scenarios, approximately obeying normal distribution. As shown in Figure 7, the forecast error values (residuals) for all testing sets are normally distributed around the zero value. The forecast error histogram plots showed a high number of cases centered on zero error, indicating that the suggested model achieved excellent performance with less bias.

(**a**) Scenario 1

(**b**) Scenario 2

(**c**) Scenario 3

Figure 6. Probabilistic density curves based on GPR-KDE and point forecasting results of different methods for three scenarios.

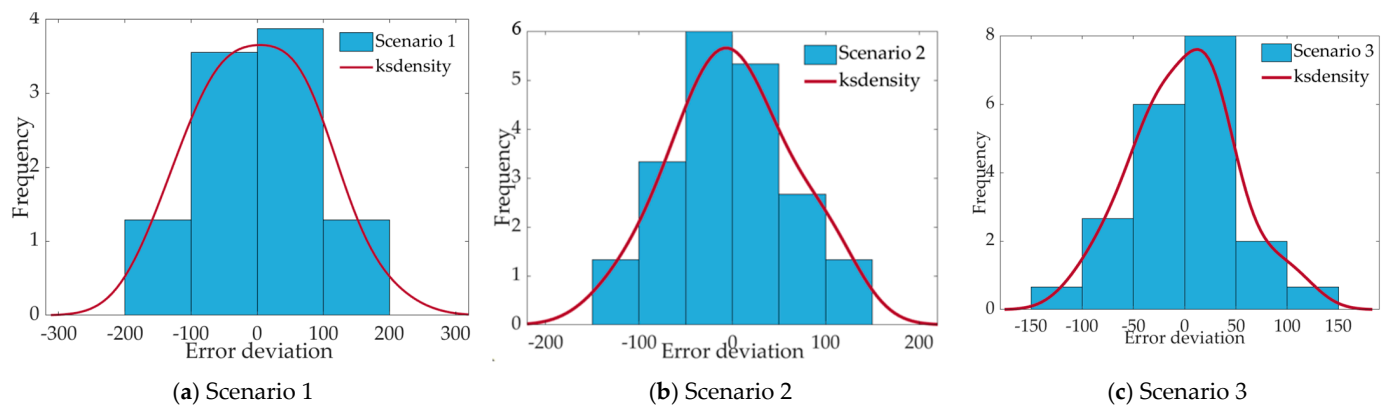



Figure 7. Probability histogram and the Gaussian kernel density estimation (red line) of electricity demand forecasting residuals of three scenarios.

From the results in Table 7, the CRPS values show the superiority of the exponential kernel function in predicting the full distribution. In scenarios 1 and 3, the CRPSs of the exponential kernel function are 21.35 and 22.60, respectively, which are the smallest metrics, indicating that it is the best on these two datasets. The kernel functions with the best probabilistic prediction performance on dataset 2 is Matern 3/2 with CRPS 30.26. Compared to QR, the CRPS index of the proposed model GPR (Exponential, Squared Expo. and Matern 3/2) has been averagely improved in the case of electricity demand by 48%, 2%, and 28%, respectively.

CRPS and MAE (for deterministic forecasts) can be directly compared, which makes probabilistic and point-forecast comparisons easy [65]. Therefore, CRPS values in this section for the probabilistic forecast are better than the MAE value and averagely improved by 30%, as shown in Table 5 in Section 3.1.

3.4. Ablation Study

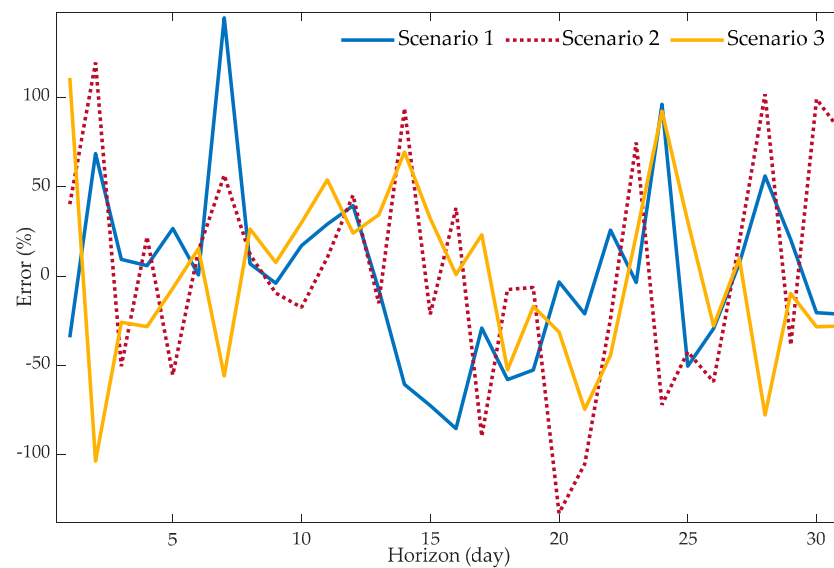
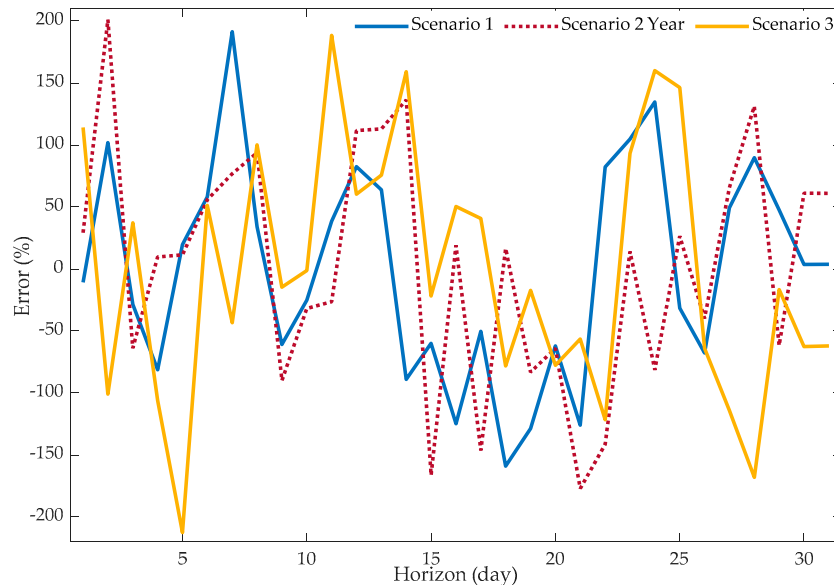
To investigate the behavior of the proposed model, this paper has conducted an ablation study. The same GPR-KDE model was trained without mobility data, which have drastically different test errors as seen by the two separate Figure 8a,b. This shows that mobility data has a strong regularizing effect and gives a huge improvement across all scenarios of forecasting since Figure 8b shows test errors are more dispersed than Figure 8a.

(**a**) Error plot with mobility data

(**b**) Error plot without mobility data

Figure 8. Test error plot for GPR (Exponential kernel) model of three scenarios.

4. Discussion and Conclusions

EDF is critical in the energy sector, and uncertainties that affect energy demand predictions must be considered. Conventional point EDF cannot precisely capture this uncertainty. This paper utilizes GPR, which can deal with anomalous events with minimal data. GPR has recently acquired attention in the literature as a non-parametric probabilistic forecasting method. Most articles incorrectly assumed the GPR approach was deterministic and failed to fully use its benefits. To bridge this gap, this research aimed to model EDF as a regression issue and developed a comprehensive short-term probabilistic EDF based on GPR-KDE methods to depict unprecedented events. This study also proposed using mobility data to include social and economic behaviors in forecasting algorithms to demonstrate the human behavior pattern (practice theory [46]) as further characteristics to deal with the uncertainty caused by anomalous events such as floods, bushfires, and pandemics. Therefore, firstly, a correlation study is carried out to identify the features highly related to electricity demand. Four factors were identified as significantly impacting the EDF: temperature, renewable and non-renewable energy sources, and human mobility (specifically workplace and res-

idential) data. After identifying the highly correlated features, three widely-used kernel functions were examined to find the optimal one for modelling complex interactions. It identified the Exponential covariance function as a great kernel function. Finally, GPR is applied for EDF and compared with different models and kernel functions to provide point forecasting along with comprehensive probability estimation for prediction intervals. Then, the model forecasting results under different CI are input into the KDE function to estimate the probability distribution.

Several testing and validation techniques were used to examine the performance of the proposed method for predicting EDF using real data; 5-fold cross-validation was used for model validation in addition to the other statistical technique. Results showed that the Exponential covariance function performed well among other kernels used in this study to evaluate the EDF.

In addition, it is crucial to compare with other recent studies to further demonstrate the efficacy of the proposed method. It is imperative to utilize the same data set to make the comparison significant. However, to the best of our knowledge, no study in the literature has investigated the EDF using a similar data set and timeframe. Therefore, this study demonstrated the proposed method's implementation in three different scenarios in Australia to establish the model's accuracy. The results show:

- GPR has been averagely improved compared to BPNN, RT, and SVR in the case of point forecasting by 59%, 79%, and 51%, respectively.
- The system was validated for one-day-ahead weather forecasts and the most typical time horizon using the PICP and MPIW indicators. The PI results of three scenarios with different confidence levels (95%, 90%, 80%, and 50%) indicate that most of the real electricity demand falls into the PIs with 50–80% CI, which demonstrates the accuracy of the proposed method. Under the same PICP condition, the MPIW of the proposed method GPR (exponential kernel) is much smaller than that of other compared kernels.
- Meanwhile, a large number of real electricity demand values are located near the peak of probability density curves. The CRPS for the probabilistic forecast is better than the MAE value of the point forecasting.

This study's results show that GPR-KDE is a reliable method and can still produce a very accurate forecast of complex electricity demand patterns with limited historical data. Different restriction measures in both scenarios 2 and 3 during the pandemic and their impact on people's activities have considerably changed the electricity demand profile distinctively. In contrast, heavy rainfall and associated flooding in NSW, which started in February 2021 and 2022, declared a disaster and forced people to evacuate overnight as floodwaters engulfed houses [66]. These two opposite scenarios made EDF challenging during anomalous events. Hence, according to the practice theory [1,46], people's activities at residence would show up as routine repeating trends in terms of electricity usage, despite the irregularity. Therefore, incorporating mobility data greatly enhances the EDF model's capacity to detect significant shifts in electricity demand behavior.

The ablation study shows that the proposed method may significantly cut down the error between predicted and real EDF by incorporating mobility patterns. In the future, the proposed method will be tested in other energy forecasting areas, such as electricity prices, with sparse data. More relevant features, such as mobile traffic, mobile phone, and Electric vehicle (EV) data, will be tested to examine correlation with EDF.

Author Contributions: I.F. developed the theory, performed the computations, and participated in original draft preparation under the supervision of G.L. and X.K.; I.F. and X.K. proposed the scheme; G.L. reviewed, edited, and formally analysed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: To support the findings of this study, an open dataset has been used and can be accessed in Electricity demand data—AEMO https://aemo.com.au/ (accessed on 20 August 2022). Weather and energy source data—OpenNem National Electricity https://opennem.org.au/energy/nem (accessed on 26 September 2022). Google Mobility data https://www.google.com/COVID19/mobility/ (accessed on 15 October 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fatema, I.; Kong, X.; Fang, G. Analyzing and forecasting electricity demand and price using deep learning model during the COVID-19 pandemic. In Proceedings of the Parallel Architectures, Algorithms and Programming: 11th International Symposium, PAAP, Proceedings 11, Shenzhen, China, 28–30 December 2020. [CrossRef]
2. Liu, J.; Zhang, Z.; Fan, X.; Zhang, Y.; Wang, J.; Zhou, K.; Liang, S.; Yu, X.; Zhang, W. Power system load forecasting using mobility optimization and multi-task learning in COVID-19. *Appl. Energy* **2022**, *310*, 118303. [CrossRef] [PubMed]
3. Aggregated Demand and Price Data. Available online: https://aemo.com.au/ (accessed on 20 August 2022).
4. Australian Energy Market Operator (AEMO 2022). Available online: https://aemo.com.au/ (accessed on 20 January 2023).
5. Chen, Y.; Yang, W.; Zhang, B. Using mobility for electrical load forecasting during the covid-19 pandemic. *arXiv* **2006**, arXiv:08826.20206. [CrossRef]
6. Fatema, I.; Kong, X.; Fang, G. Electricity demand and price forecasting model for sustainable smart grid using comprehensive long short term memory. *Int. J. Sustain. Eng.* **2021**, *14*, 1714–1732. [CrossRef]
7. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]
8. Van der Meer, D.; Widén, J.; Munkhammar, J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1484–1512. [CrossRef]
9. Hor, C.-L.; Watson, S.; Majithia, S. Daily load forecasting and maximum demand estimation using ARIMA and GARCH. In Proceedings of the 2006 International Conference on Probabilistic Methods Applied to Power Systems, Stockholm, Sweden, 11–15 June 2006; pp. 1–6.
10. Bozkurt, Ö.Ö.; Biricik, G.; Tayşi, Z. Artificial neural network and SARIMA based models for power load forecasting in Turkish electricity market. *PLoS ONE* **2017**, *12*, 75915. [CrossRef]
11. Chen, B.-J.; Chang, M.-W.; Lin, C.-J. Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001. *IEEE Trans. Power Syst.* **2004**, *19*, 1821–1830. [CrossRef]
12. Fiot, J.-B.; Dinuzzo, F. Electricity Demand Forecasting by Multi-Task Learning. *IEEE Trans. Smart Grid* **2016**, *9*, 544–551. [CrossRef]
13. Ding, N.; Benoit, C.; Foggia, G.; Besanger, Y.; Wurtz, F. Neural Network-Based Model Design for Short-Term Load Forecast in Distribution Systems. *IEEE Trans. Power Syst.* **2015**, *31*, 72–81. [CrossRef]
14. Xu, F.Y.; Cun, X.; Yan, M.; Yuan, H.; Wang, Y.; Lai, L.L. Power Market Load Forecasting on Neural Network With Beneficial Correlated Regularization. *IEEE Trans. Ind. Inform.* **2018**, *14*, 5050–5059. [CrossRef]
15. Amjady, N.; Keynia, F.; Zareipour, H. Wind Power Prediction by a New Forecast Engine Composed of Modified Hybrid Neural Network and Enhanced Particle Swarm Optimization. *IEEE Trans. Sustain. Energy* **2011**, *2*, 265–276. [CrossRef]
16. Lloyd, J.R. GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes. *Int. J. Forecast.* **2014**, *30*, 369–374. [CrossRef]
17. Song, K.-B.; Ha, S.-K.; Park, J.-W.; Kweon, D.-J.; Kim, K.-H. Hybrid Load Forecasting Method With Analysis of Temperature Sensitivities. *IEEE Trans. Power Syst.* **2006**, *21*, 869–876. [CrossRef]
18. Li, T.; Wang, Y.; Zhang, N. Combining Probability Density Forecasts for Power Electrical Loads. *IEEE Trans. Smart Grid* **2019**, *11*, 1679–1690. [CrossRef]
19. Hong, T.; Wilson, J.; Xie, J. Long Term Probabilistic Load Forecasting and Normalization With Hourly Information. *IEEE Trans. Smart Grid* **2013**, *5*, 456–462. [CrossRef]
20. Hyndman, R.; Koehler, A.B.; Ord, J.K.; Snyder, R.D. *Forecasting with Exponential Smoothing: The State Space Approach*; Springer Science & Business Media: Berlin, Germany, 2008.
21. Charytoniuk, W.; Chen, M.; Kotas, P.; Van Olinda, P. Demand forecasting in power distribution systems using nonparametric probability density estimation. *IEEE Trans. Power Syst.* **1999**, *14*, 1200–1206. [CrossRef]
22. Bracale, A.; Caramia, P.; Carpinelli, G.; Di Fazio, A.R.; Varilone, P. A Bayesian-Based Approach for a Short-Term Steady-State Forecast of a Smart Grid. *IEEE Trans. Smart Grid* **2013**, *4*, 1760–1771. [CrossRef]
23. Liu, B.; Nowotarski, J.; Hong, T.; Weron, R. Probabilistic Load Forecasting via Quantile Regression Averaging on Sister Forecasts. *IEEE Trans. Smart Grid* **2015**, *8*, 730–737. [CrossRef]
24. Xie, J.; Hong, T.; Laing, T.; Kang, C. On Normality Assumption in Residual Simulation for Probabilistic Load Forecasting. *IEEE Trans. Smart Grid* **2015**, *8*, 1046–1053. [CrossRef]
25. Xie, J.; Hong, T. GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation. *Int. J. Forecast.* **2016**, *32*, 1012–1016. [CrossRef]
26. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2008.

27. Blum, M.; Riedmiller, M. Electricity demand forecasting using Gaussian processes. In Proceedings of the Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence, Bellevue, WA, USA, 14–18 July 2013.

28. Xie, G.; Chen, X.; Weng, Y. An Integrated Gaussian Process Modeling Framework for Residential Load Prediction. *IEEE Trans. Power Syst.* **2018**, *33*, 7238–7248. [CrossRef]

29. Zhu, S.; Yuan, X.; Xu, Z.; Luo, X.; Zhang, H. Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast. *Energy Convers. Manag.* **2019**, *198*, 111772. [CrossRef]

30. Zhang, Z.; Wang, C.; Peng, X.; Qin, H.; Lv, H.; Fu, J.; Wang, H. Solar radiation intensity probabilistic forecasting based on K-means time series clustering and Gaussian process regression. *IEEE Access* **2021**, *9*, 89079–89092. [CrossRef]

31. Cao, D.; Zhao, J.; Hu, W.; Zhang, Y.; Liao, Q.; Chen, Z.; Blaabjerg, F. Robust Deep Gaussian Process-Based Probabilistic Electrical Load Forecasting Against Anomalous Events. *IEEE Trans. Ind. Inform.* **2021**, *18*, 1142–1153. [CrossRef]

32. Li, Y.; Rao, S.; Hassaine, A.; Ramakrishnan, R.; Canoy, D.; Salimi-Khorshidi, G.; Mamouei, M.; Lukasiewicz, T.; Rahimi, K. Deep Bayesian Gaussian processes for uncertainty estimation in electronic health records. *Sci. Rep.* **2021**, *11*, 20685. [CrossRef]

33. Laradji, I.H.; Schmidt, M.; Pavlovic, V.; Kim, M. Efficient deep Gaussian process models for variable-sized inputs. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14 July 2019; pp. 1–7. [CrossRef]

34. Yang, Y.; Li, S.; Li, W.; Qu, M. Power load probability density forecasting using Gaussian process quantile regression. *Appl. Energy* **2018**, *213*, 499–509. [CrossRef]

35. Sheng, H.; Xiao, J.; Cheng, Y.; Ni, Q.; Wang, S. Short-Term Solar Power Forecasting Based on Weighted Gaussian Process Regression. *IEEE Trans. Ind. Electron.* **2017**, *65*, 300–308. [CrossRef]

36. Shepero, M.; van der Meer, D.; Munkhammar, J.; Widén, J. Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data. *Appl. Energy* **2018**, *218*, 159–172. [CrossRef]

37. Zhang, L.; Xie, L.; Han, Q.; Wang, Z.; Huang, C. Probability Density Forecasting of Wind Speed Based on Quantile Regression and Kernel Density Estimation. *Energies* **2020**, *13*, 6125. [CrossRef]

38. Waldmann, E. Quantile regression: A short story on how and why. *Stat. Model.* **2018**, *18*, 203–218. [CrossRef]

39. He, Y.; Zheng, Y. Short-term power load probability density forecasting based on Yeo-Johnson transformation quantile regression and Gaussian kernel function. *Energy* **2018**, *154*, 143–156. [CrossRef]

40. Alamaniotis, M.; Chatzidakis, S.; Tsoukalas, L. Monthly load forecasting using kernel based gaussian process regression. In Proceedings of the 9th Mediterranean Conference on Power Generation, Transmission, Distribution, and Energy Conversion: MEDPOWER, Athens, Greece, 2–5 November 2014. [CrossRef]

41. Leith, D.J.; Heidl, M.; Ringwood, J.V. Gaussian process prior models for electrical load forecasting. In Proceedings of the 2004 International Conference on Probabilistic Methods Applied to Power Systems, Ames, IA, USA, 12–16 September 2004; ISBN 0-9761319-1-9.

42. Mori, H.; Ohmi, M. Probabilistic short-term load forecasting with Gaussian processes. In Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems, Arlington, VA, USA, 6–10 November 2005; ISBN 1-59975-174-7. [CrossRef]

43. Alamaniotis, M.; Bargiotas, D.; Tsoukalas, L.H. Towards smart energy systems: Application of kernel machine regression for medium term electricity load forecasting. *Springerplus* **2016**, *5*, 58. [CrossRef] [PubMed]

44. Lourenço, J.; Santos, P. Short-term load forecasting using a Gaussian process model: The influence of a derivative term in the input regressor. *Intell. Decis. Technol.* **2012**, *6*, 273–281. [CrossRef]

45. Dong, B.; Li, Z.; Rahman, S.M.; Vega, R. A hybrid model approach for forecasting future residential electricity consumption. *Energy Build.* **2016**, *117*, 341–351. [CrossRef]

46. Stephen, B.; Tang, X.; Harvey, P.R.; Galloway, S.; Jennett, K.I. Incorporating Practice Theory in Sub-Profile Models for Short Term Aggregated Residential Load Forecasting. *IEEE Trans. Smart Grid* **2015**, *8*, 1591–1598. [CrossRef]

47. Golestaneh, F.; Pinson, P.; Gooi, H.B. Very Short-Term Nonparametric Probabilistic Forecasting of Renewable Energy Generation—With Application to Solar Energy. *IEEE Trans. Power Syst.* **2016**, *31*, 3850–3863. [CrossRef]

48. Gibbs, M.N. *Bayesian Gaussian Processes for Regression and Classification*; Citeseer: Princeton, NJ, USA, 1998.

49. Kersting, K.; Plagemann, C.; Pfaff, P.; Burgard, W. Most likely heteroscedastic Gaussian process regression. In Proceedings of the 24th International Conference on Machine Learning, Corvalis, OR, USA, 20 June 2007; pp. 393–400. [CrossRef]

50. Tolvanen, V.; Jylänki, P.; Vehtari, A. Expectation propagation for nonstationary heteroscedastic Gaussian process regression. In Proceedings of 2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Reims, France, 21–24 September 2014. [CrossRef]

51. Zhao, X.; Liu, J.; Yu, D.; Chang, J. One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data. *Energy Convers. Manag.* **2018**, *164*, 560–569. [CrossRef]

52. Thompson, J.R.; Tapia, R.A. *Non-Parametric Function Estimation, Modeling, and Simulation*; SIAM: Philadelphia, PA, USA, 1990.

53. Bashtannyk, D.M.; Hyndman, R.J. Bandwidth selection for kernel conditional density estimation. *Comput. Stat. Data Anal.* **2001**, *36*, 279–298. [CrossRef]

54. Scott, D.W. *Multivariate Density Estimation: Theory, Practice, and Visualization*; John Wiley & Sons: New York, NY, USA, 2015.

55. Viviani, E.; Di Persio, L.; Ehrhardt, M. Energy Markets Forecasting. From Inferential Statistics to Machine Learning: The German Case. *Energies* **2021**, *14*, 364. [CrossRef]

56. Kuo, P.-H.; Huang, C.-J. An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks. *Sustainability* **2018**, *10*, 1280. [CrossRef]

57. Rohani, A.; Taki, M.; Abdollahpour, M. A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). *Renew. Energy* **2018**, *115*, 411–422. [CrossRef]

58. Najibi, F.; Apostolopoulou, D.; Alonso, E. Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast. *Int. J. Electr. Power Energy Syst.* **2021**, *130*, 106916. [CrossRef]

59. Shrivastava, N.A.; Panigrahi, B.K. Point and prediction interval estimation for electricity markets with machine learning techniques and wavelet transforms. *Neurocomputing* **2013**, *118*, 301–310. [CrossRef]

60. Gneiting, T.; Balabdaoui, F.; Raftery, A. Probabilistic forecasts, calibration and sharpness. *J. R. Stat. Soc. Ser. B* **2007**, *69*, 243–268. [CrossRef]

61. National Electricity Market. Available online: https://opennem.org.au/energy/nem/ (accessed on 26 September 2022).

62. Google. COVID-19 Community Mobility Reports 2020. Available online: https://www.google.com/covid19/mobility/ (accessed on 15 October 2022).

63. Puth, M.-T.; Neuhäuser, M.; Ruxton, G. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Anim. Behav.* **2015**, *102*, 77–84. [CrossRef]

64. Tso, G.K.F.; Yau, K.W.K. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy* **2007**, *32*, 1761–1768. [CrossRef]

65. Hersbach, H. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast.* **2000**, *15*, 559–570. [CrossRef]

66. NSW State Emergency Service. Available online: https://www.ses.nsw.gov.au/ (accessed on 10 December 2022).