

Article

# Multi-Horizon Forecasting of Global Horizontal Irradiance Using Online Gaussian Process Regression: A Kernel Study

Hanany Tolba <sup>1</sup>, Nouha Dkhili <sup>1,2</sup>, Julien Nou <sup>1</sup>, Julien Eynard <sup>1,2</sup> , Stéphane Thil <sup>1,2</sup>   
and Stéphane Grieu <sup>1,2,\*</sup>

<sup>1</sup> PROMES-CNRS Laboratory (UPR 8521), Rambla de la Thermodynamique, Tecnosud, 66100 Perpignan, France; hanany.tolba@promes.cnrs.fr (H.T.); nouha.dkhili@promes.cnrs.fr (N.D.); julien.nou@promes.cnrs.fr (J.N.); julien.eynard@univ-perp.fr (J.E.); stephane.thil@univ-perp.fr (S.T.)

<sup>2</sup> University of Perpignan Via Domitia, 52 avenue Paul Alduy, 66860 Perpignan, France

\* Correspondence: grieu@univ-perp.fr

Received: 17 June 2020; Accepted: 12 August 2020; Published: 13 August 2020



**Abstract:** In the present paper, global horizontal irradiance (GHI) is modelled and forecasted at time horizons ranging from 30 min to 48 h, thus covering intrahour, intraday and intraweek cases, using online Gaussian process regression (OGPR) and online sparse Gaussian process regression (OSGPR). The covariance function, also known as the kernel, is a key element that deeply influences forecasting accuracy. As a consequence, a comparative study of OGPR and OSGPR models based on simple kernels or combined kernels defined as sums or products of simple kernels has been carried out. The classic persistence model is included in the comparative study. Thanks to two datasets composed of GHI measurements (45 days), we have been able to show that OGPR models based on quasiperiodic kernels outperform the persistence model as well as OGPR models based on simple kernels, including the squared exponential kernel, which is widely used for GHI forecasting. Indeed, although all OGPR models give good results when the forecast horizon is short-term, when the horizon increases, the superiority of quasiperiodic kernels becomes apparent. A simple online sparse GPR (OSGPR) approach has also been assessed. This approach gives less precise results than standard GPR, but the training computation time is decreased to a great extent. Even though the lack of data hinders the training process, the results still show the superiority of GPR models based on quasiperiodic kernels for GHI forecasting.

**Keywords:** solar resource; global horizontal irradiance; time series forecasting; machine learning; online Gaussian process regression; online sparse Gaussian process regression

## 1. Introduction

Over the past few decades, the penetration of distributed generation in the power grid has been gaining momentum due to environmental concerns and increases in global power demand. Therefore, being able to handle an increased share of fluctuating power generation from intermittent renewable energy sources—in particular, solar photovoltaics—has become a critical challenge for power grid operators. In fact, the decentralisation of power generation has made compliance with voltage constraints, current levels and voltage drop gradients increasingly difficult and has spawned a multitude of stability, quality and safety issues. The intermittency of solar power is poorly accounted for by conventional planning strategies for the daily operation of power grids and increases the complication of voltage regulation [1]. Therefore, grid stability has become dependent upon the successful compensation of supply/demand variation. As a result, perceiving changes in the state of

the power grid in real time is necessary to maintain the continuity and quality of service. Thus, smart management strategies—especially predictive management strategies—enabling the autonomous and real-time monitoring and optimisation of power grid operation are required. Predictive management strategies aggregate in situ measurements and multi-horizon forecasts—in particular, those of the grid's load and photovoltaic power generation—in order to maintain the balance between supply and demand while ensuring grid stability at all times. Using forecasts of global horizontal irradiance (GHI), which is the total amount of shortwave radiation received from above by a surface horizontal to the ground, photovoltaic power generation can be anticipated. To this end, this paper focuses on forecasting GHI over different time horizons (i.e., from 30 min to 48 h).

GHI is a non-stationary signal, essentially because of periodic patterns (daily and also seasonal if longer time-series are considered) arising from solar geometry and absorption and scattering in Earth's atmosphere. Depending on how this non-stationarity is handled, two approaches can be defined. Since most classical time-series models require stationarity (or weak stationarity), a popular approach is to suppress or strongly reduce non-stationarity before modelling and forecasting the residuals. Several possibilities exist to achieve this, which are listed below.

- Using the clearness index, defined as the ratio of irradiance at ground level to extraterrestrial irradiance on a horizontal plane (see [2,3], as well as references therein).
- Using the clear-sky index, defined as the ratio of irradiance at ground level to clear-sky irradiance on a horizontal plane, for which a clear-sky model is needed (see [4,5], as well as references therein).
- Fitting and then suppressing seasonal and daily patterns, without using solar geometry or radiative transfer models. Various methods can be considered, such as polynomial [6], modified gaussian [7], trigonometric [8], Fourier [9], exponential smoothing [10] or dynamic harmonic regression methods [11].

Another approach is to model and forecast raw GHI time series, without any specific preprocessing. Indeed, the transformations necessary to achieve stationarity may hinder the performance of forecasting algorithms, since the resulting time series exhibit very low correlation [12]. Incidentally, it is shown in [13] that artificial intelligence-based models using raw GHI data outperform conventional approaches based on forecasting the clearness index and are able to capture the periodic component of this time series as well as its stochastic variation. Not having to rely on clear-sky or radiative transfer models also implies that all errors come from the forecasting method itself. This is especially relevant when using clear-sky models based on mean values, as those can lead to significant errors [14].

In the last few years, GHI and photovoltaic power generation forecasting with machine learning techniques have gained in popularity. Artificial neural networks are the most commonly used tools; for further information, see the works presented in [15–17] or the review in [18] for a summary of existing approaches. Next are support vector machines (SVMs), which are called support vector regression (SVR) in the context of regression; more information can be found in [19] or the review in [20]. Other machine learning techniques, such as nearest neighbor, regression tree, random forest or gradient boosting, have also been tested and compared to classical time-series forecasting [21–23].

The use of Gaussian process regression (GPR) dedicated to GHI forecasting, however, has remained limited. In [21,23], GPR was used to obtain short-term forecasts of the clear-sky index. A more recent study has been carried out but for daily and monthly forecasting of GHI in [24]. In a spatio-temporal context, some studies based on GPR have been carried out to obtain short-term forecasts of the clear-sky index [25–27] (it should be noted that in these papers, the spatio-temporal covariance has been obtained by fitting empirical variograms instead of using spatio-temporal kernels). It must be stressed that, in all of these studies, in addition to forecasting the clear-sky index, no optimal kernel has been selected: as a default choice, the authors used only the square exponential kernel. In this work, GPR is used to model and forecast raw GHI data at forecast horizons ranging from 30 min to 48 h, thus covering intrahour, intraday and intraweek cases. A comparison between simple kernels and combined kernels is made, demonstrating that an appropriate choice is key to obtaining

the best forecasting performance, especially at higher forecast horizons. It will be shown that, when an appropriate kernel is chosen, GPR models are sufficiently flexible to model and forecast GHI by jointly optimizing short-term variability (mainly due to atmospheric disturbances) and long-term trends (mainly due to solar geometry). As a consequence, in addition to the aforementioned benefits of not using preprocessing, the proposed models can be applied without modification in different locations, since any difference will directly be learned from the data.

The paper is organized as follows: Gaussian processes as well as commonly used covariance functions and ways of combining these functions are defined in Section 2. In Section 3, the principles of Gaussian process regression are introduced, including both online and online sparse models, as well as the estimation of their parameters and hyperparameters. The datasets used to train and validate the models included in the comparative study, derived from GHI measurements taken at the laboratory PROMES-CNRS in Perpignan (southern France), are described in Section 4. In Section 5, the forecasting accuracy of the developed OGPR and OSGPR models is evaluated. The paper ends with a conclusion. The hyperparameters' estimated values and detailed numerical results can be found in Appendices A and B, respectively.

## 2. Gaussian Processes

In this section, Gaussian processes are defined. In addition, commonly used covariance functions and ways of combining these functions are presented.

### 2.1. Definition

A Gaussian process (GP) can be defined as a collection of random variables, any finite number of which have a joint Gaussian distribution (see Definition 2.1 in [28]). A GP defines a prior over functions, which can be converted into a posterior over functions once some data have been observed. Let us use  $f(x) \sim \mathcal{GP}(\mu(x), k(x, x'))$  to indicate that a random function  $f(x)$  follows a Gaussian process.  $\mu(x) = \mathbb{E}[f(x)]$  is the mean function (usually assumed to be zero),  $k(x, x') = \mathbb{E}[(f(x) - \mu(x))(f(x') - \mu(x'))^\top]$  is the covariance function and  $x$  and  $x'$  are arbitrary input variables. In mathematics, GPs have been studied extensively, and their existence was first proven by the Daniell–Kolmogorov theorem (more familiar as Kolmogorov's existence/extension/consistency theorem; see [29]). In the field of machine learning, GPs first appeared in [30], as the limiting case of Bayesian inference performed on artificial neural networks with an infinite number of hidden neurons. However, their appearance in the geostatistics research community, where they are known as kriging methods, was less recent [31,32].

### 2.2. Covariance Functions

A covariance function, also known as kernel, evaluates the similarity between two points and encodes the assumptions about the function to be learned. This initial belief could concern the smoothness of the function or whether the function is periodic. Any function can be used as a kernel as long as the resulting covariance matrix is positive and semi-definite. For scalar-valued inputs ( $x$  and  $x'$ ), the commonly used kernels ( $k_{\text{WN}}$ ,  $k_{\text{SE}}$ ,  $k_{\text{RQ}}$ ,  $k_{\text{M}_v}$  and  $k_{\text{Per}}$ ) are briefly described below. An exhaustive list of kernels can be found in [28].

- The white noise kernel  $k_{\text{WN}}$  is given by

$$k_{\text{WN}}(x, x') = \sigma^2 \delta(x - x') \quad (1)$$

where  $\sigma^2$  is the variance. In all the considered kernels, it plays the role of a scale factor.

- The squared exponential kernel  $k_{\text{SE}}$  is given by

$$k_{\text{SE}}(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\ell^2}\right) \quad (2)$$

where  $\sigma^2$  is the variance and  $\ell$  is the correlation length or characteristic length-scale, sometimes called “range” in geostatistics.  $\ell$  controls how quickly the functions are sampled from the GP oscillate. If  $\ell$  is large, even points that are far away have meaningful correlation; therefore, GP functions would oscillate slowly and vice versa.

- The rational quadratic kernel  $k_{\text{RQ}}$  is given by

$$k_{\text{RQ}}(x, x') = \sigma^2 \left( 1 + \frac{(x - x')^2}{2\alpha\ell^2} \right)^{-\alpha} \quad (3)$$

where  $\sigma^2$  is the variance and  $\ell$  is the correlation length. In this kernel,  $\alpha$  defines the relative weighting of large-scale and small-scale variations.  $k_{\text{RQ}}$  can be seen as a scale mixture of squared exponential kernels with different correlation lengths distributed according to a Beta distribution [28].

- The Matérn class of kernels  $k_{\text{M}_\nu}$  is given by

$$k_{\text{M}_\nu}(x, x') = \sigma^2 \frac{1}{2^{\nu-1}\Gamma(\nu)} \left( \sqrt{2\nu} \frac{|x - x'|}{\ell} \right)^\nu \mathcal{B}_\nu \left( \sqrt{2\nu} \frac{|x - x'|}{\ell} \right) \quad (4)$$

where  $\sigma^2$  is the variance,  $\Gamma$  is the standard Gamma function,  $\ell$  is the correlation length and  $\mathcal{B}_\nu$  is the modified Bessel function of second kind of order  $\nu$ .  $\nu$  controls the degree of regularity (differentiability) of the resultant GP. Kernels of the Matérn class are continuous but not differentiable. If  $\nu = 1/2$ ,  $k_{\text{M}_{1/2}} = k_{\text{E}}$ , the exponential kernel is obtained:

$$k_{\text{E}}(x, x') = \sigma^2 \exp \left( -\frac{|x - x'|}{\ell} \right) \quad (5)$$

This kernel corresponds to the continuous version of a classical discrete autoregressive model AR(1), also known as the Ornstein–Uhlenbeck process. As discussed in [28], when  $\nu = p - 1/2$ , with  $p \in \mathbb{N}$ , the continuous version of classical AR( $p$ ) is obtained. Two other kernels of the Matérn class  $k_{\text{M}_{3/2}}$  and  $k_{\text{M}_{5/2}}$  are widely exploited in the scientific literature. These kernels correspond to the cases  $\nu = 3/2$  and  $\nu = 5/2$ , respectively. They are given by

$$k_{\text{M}_{3/2}}(x, x') = \sigma^2 \left( 1 + \sqrt{3} \frac{|x - x'|}{\ell} \right) \exp \left( -\sqrt{3} \frac{|x - x'|}{\ell} \right) \quad (6)$$

and:

$$k_{\text{M}_{5/2}}(x, x') = \sigma^2 \left( 1 + \sqrt{5} \frac{|x - x'|}{\ell} + \sqrt{5} \frac{(x - x')^2}{3\ell^2} \right) \exp \left( -\sqrt{5} \frac{|x - x'|}{\ell} \right) \quad (7)$$

- The periodic kernel  $k_{\text{Per}}$  is given by

$$k_{\text{Per}}(x, x') = \sigma^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell^2} \right) \quad (8)$$

where  $\sigma^2$  is the variance and  $\ell$  is the correlation length.  $k_{\text{Per}}$  assumes a globally periodic structure (of period  $P$ ) in the function to be learned. A larger  $P$  causes a slower oscillation, while a smaller  $P$  causes a faster oscillation.

### 2.3. Kernel Composition

Several kernels can be combined to define efficient GPs. The only thing to keep in mind is that the resulting covariance matrix must be a positive semi-definite matrix. Two ways of combining kernels while keeping the positive semi-definite property are addition and multiplication. As a result, a quasiperiodic GP can be modelled by multiplying a periodic kernel by a non periodic kernel. This gives us a way to transform a global periodic structure into a local periodic structure.

For example, in [28], the product of  $k_{SE}$  and  $k_{Per}$  was used to model monthly average atmospheric CO<sub>2</sub> concentrations.

In this paper, we were inspired by [33] to construct a wide variety of kernels to model global horizontal irradiance by adding or multiplying two kernels from the kernel families presented in Section 2.2.

### 3. Gaussian Process Regression

In this section, standard Gaussian process regression (GPR), online Gaussian process regression (OGPR) and online sparse Gaussian process regression (OSGPR) are presented. The section ends with the training of a GPR model.

#### 3.1. Standard Gaussian Process Regression

Consider the standard regression model with additive noise, formulated as follows:

$$y = f(x) + \varepsilon \quad (9)$$

where  $y$  is the observed value,  $f$  is the regression function,  $x \in \mathbb{R}^{D \times 1}$  is the input vector (in most cases,  $1 \leq D \leq 3$ ; in this paper,  $x$  is the time, with a dimension of 1) and  $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$  is independent and identically distributed Gaussian noise. GPR can be defined as a Bayesian nonparametric regression which assumes a GP prior over the regression functions [28]. Basically, this consists of approximating  $f(x) \sim \mathcal{GP}(\mu(x), k(x, x'))$  with a training set of  $n$  observations  $\mathcal{D} = \{(x_i, y_i), 1 \leq i \leq n\}$ . In the following, all the input vectors  $x_i$  are merged into a matrix  $X \in \mathbb{R}^{n \times D}$ , and all corresponding outputs  $y_i$  are merged into a vector  $\mathbf{y} \in \mathbb{R}^{n \times 1}$ , so that the training set can be written as  $(X, \mathbf{y})$ . From Equation (9), one can see that  $\mathbf{y} \sim \mathcal{N}(\mu(X), K + \sigma_\varepsilon^2 \mathbf{I})$ , where  $K = k(X, X) \in \mathbb{R}^{n \times n}$ . The joint distribution of the observed data  $\mathbf{y}$  and the latent noise-free function on the test points  $f_* = f(X_*)$  is given in this setting by

$$\begin{bmatrix} \mathbf{y} \\ f_* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mu(X) \\ \mu(X_*) \end{bmatrix}, \begin{bmatrix} K + \sigma_\varepsilon^2 \mathbf{I} & K_* \\ K_*^\top & K_{**} \end{bmatrix} \right) \quad (10)$$

where  $K_* = k(X, X_*) \in \mathbb{R}^{n \times n_*}$  and  $K_{**} = k(X_*, X_*) \in \mathbb{R}^{n_* \times n_*}$ .

The posterior predictive density is also Gaussian:

$$f_* | X, \mathbf{y}, X_* \sim \mathcal{N}(\boldsymbol{\mu}_*, \sigma_*^2) \quad (11)$$

where:

$$\boldsymbol{\mu}_* = \mu(X_*) + K_*^\top (K + \sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{y} - \mu(X)) \quad (12)$$

and:

$$\sigma_*^2 = K_{**} - K_*^\top (K + \sigma_\varepsilon^2 \mathbf{I})^{-1} K_* \quad (13)$$

Let us examine what happens for a single test point  $x_*$  and let  $\mathbf{k}_*$  be the vector of covariances between this test point and the  $n$  training points:

$$\mathbf{k}_* = [k(x_*, x_1) \dots k(x_*, x_n)]^\top \quad (14)$$

Equation (12) becomes

$$\mu_* = \mu(x_*) + \mathbf{k}_*^\top (K + \sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{y} - \mu(x_*)) \quad (15)$$

Thus,  $\mu_*$  can be formulated as a linear combination of kernels, each one centered on a training point:

$$\mu_* = \mu(x_*) + \mathbf{k}_*^T \boldsymbol{\alpha} = \mu(x_*) + \sum_{i=1}^n \alpha_i k(x_i, x_*) \quad (16)$$

where  $\boldsymbol{\alpha} = (K + \sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{y} - \mu(x_*))$ .

Each time a new observation is added to the observation set, the coefficients  $\alpha_i$ —referred to as *parameters* in the following—are updated. By contrast, the parameters of the kernel, referred to as *hyperparameters*, are not updated once the training is completed (see Section 3.4 and Appendix A).

### 3.2. Online Gaussian Process Regression

When using standard GPR, it is difficult to incorporate a new training point or a new set of training points. In case a forecasting algorithm is run in situ, there is no fixed data set  $\mathcal{D} = \{(x_i, y_i), 1 \leq i \leq n\}$ ; rather, one or a few observations are collected at a time and, as a result, the total amount of information available grows perpetually and incrementally. Besides, updates are often required every hour or minute, or even every second. In this case, using these methods would incur prohibitive computational overheads. Each time we wish to update our beliefs in the light of a new observation, we would be forced to invert the matrix  $K + \sigma_\varepsilon^2 \mathbf{I}$  with a complexity  $\mathcal{O}(n^3)$ .

A solution to this is to use online Gaussian process regression (OGPR) [34–38]. Let us suppose that the distribution of the GP given the first  $n$  training points  $(X, \mathbf{y})$  is known. For simplicity, let us consider  $\mu = 0$  (without loss of generality). Suppose that  $n_+$  new observations are regrouped in the matrix  $X_+$ ; then,

$$\left( \begin{bmatrix} K \\ K_+ \end{bmatrix} + \sigma_\varepsilon^2 \mathbf{I}_{n+n_+} \right)^{-1} = \begin{pmatrix} k(X, X) + \sigma_\varepsilon^2 \mathbf{I}_n & k(X, X_+) \\ k(X_+, X) & k(X_+, X_+) + \sigma_\varepsilon^2 \mathbf{I}_{n_+} \end{pmatrix}^{-1} \quad (17)$$

This results in the following:

$$\left( \begin{bmatrix} K \\ K_+ \end{bmatrix} + \sigma_\varepsilon^2 \mathbf{I}_{n+n_+} \right)^{-1} = \begin{pmatrix} \underbrace{K + \sigma_\varepsilon^2 \mathbf{I}_n}_A & \underbrace{K_+}_B \\ \underbrace{K_+^T}_{B^T} & \underbrace{K_{++} + \sigma_\varepsilon^2 \mathbf{I}_{n_+}}_D \end{pmatrix}^{-1} \quad (18)$$

Using the block matrix inversion theorem, we obtain

$$\left( \begin{bmatrix} K \\ K_+ \end{bmatrix} + \sigma_\varepsilon^2 \mathbf{I}_{n+n_+} \right)^{-1} = \begin{pmatrix} A^{-1} + A^{-1} B \Delta^{-1} B^T A^{-1} & -A^{-1} B \Delta^{-1} \\ \Delta^{-1} B^T A^{-1} & \Delta^{-1} \end{pmatrix} \quad (19)$$

where  $\Delta = D - B^T A^{-1} B \in \mathbb{R}^{n_+ \times n_+}$ .

Now, inverting the  $(n + n_+) \times (n + n_+)$  matrix only requires the inverse of  $A$ , which is known and the inversion of a  $n_+ \times n_+$  matrix. Thus, the computational cost is  $\mathcal{O}(n_+^3 n^2)$  rather than  $\mathcal{O}((n + n_+)^3)$  when performing direct matrix inversion.

### 3.3. Online Sparse Gaussian Process Regression

As presented in Section 3.1, a challenge of using standard GPR is its computational complexity of  $\mathcal{O}(n^3)$  [28], within the context of training, with  $n$  being the number of training points. This principally comes from the need to invert the matrix  $K + \sigma_\varepsilon^2 \mathbf{I}$  in Equations (12) and (13). The simplest and most obvious approach to reducing computational complexity in Gaussian process regression is to ignore a part of the available training data: a forecast is then generated using only a subset of the available data during training, also known as the active set or inducing set. This leads to another version of

OGPR: the so-called online sparse Gaussian process regression (OSGPR). OSGPR includes both the deterministic training conditional (DTC) [39] and the fully independent training conditional (FITC) [40] approximation methods. Here, computational complexity is reduced to  $\mathcal{O}(n_s^2 n)$ , where  $n_s$  is the length of the active-set (usually  $n_s \ll n$ ). More information can be found in [38–42] and in the reviews presented in [34,43].

The choice of a selection procedure of the active set from the total  $n$  training points constitutes its own research field [38–41]. Sophisticated choices are based on various information criteria that quantify the pertinence of adding each point to the active-set. However, a simple way is to choose the  $n_s$  points randomly—a technique known as the subset of data—which, in some situations, may be as efficient as more complex choices [44].

### 3.4. Training a GPR Model

As previously mentioned in the paper, there is freedom when choosing the kernel of a GPR model. The model hyperparameters, denoted as  $\theta = [\theta_1 \dots \theta_m]^T$  below, group those of the chosen kernel and the noise variance. As an example, the squared exponential kernel  $k_{SE}$  (this kernel is widely used for GHI forecasting), defined by Equation (2), has two hyperparameters— $\sigma$  and  $\ell$  (see Section 2.2)—to which the noise variance  $\sigma_\varepsilon^2$  is added. Thus,  $\theta = [\sigma \ell \sigma_\varepsilon^2]^T$ . The hyperparameters of all the kernels included in this comparative study are given in Appendix A (see Equations (A1)–(A16)). These hyperparameters are estimated from the data. To do so, the probability of the data given the aforesaid hyperparameters is computed. Assuming a GPR model with Gaussian noise (9), the log marginal likelihood is given by [28]

$$\mathcal{L}(\theta) = \log [P(\mathbf{y}|X, \theta)] \quad (20)$$

Thus,

$$\mathcal{L}(\theta) = -\frac{1}{2} \left( \mathbf{y}^T (K + \sigma_\varepsilon^2 \mathbf{I})^{-1} \mathbf{y} + \log \left( \det (K + \sigma_\varepsilon^2 \mathbf{I}) \right) + n \log(2\pi) \right) \quad (21)$$

Usually, the estimation of  $\theta$  is achieved by maximising the log marginal likelihood (21). If its gradient is known and a gradient-based algorithm, such as a conjugate gradient method, can be used (as proposed in [45]), the maximisation process may be accelerated. More information about using the gradient and Hessian of Equation (21) to speed the learning phase in the GPR algorithm can be found in [28,46,47].

## 4. Data Description

The data used in this comparative study originate from a rotating shadowband irradiometer (RSI), whose uncertainty range is about  $\pm 5\%$ , housed at the laboratory PROMES-CNRS in Perpignan, roughly 20 km west of the Mediterranean Sea. This device consists of two horizontally leveled silicon photodiode radiation detectors, situated in the center of a spherically curved shadowband. The region is characterised by strong winds resulting in an often clear sky, in addition to mild winters and hot and dry summers.

The comparative study enables the evaluation of GPR models through two GHI datasets, each sampled at a 30 min rate. The first one, hereafter referred to as the “summer” dataset, covers a period of 45 days, from 5 June to 18 July 2015. The second one, hereafter referred to as the “winter” dataset, also covers a period of 45 days, from 1 November to 15 December 2015.

The motivation for using two datasets instead of one larger dataset is twofold. On one hand, clear-sky days are frequent during summertime, while in winter, more complicated situations occur (overcast, broken clouds, etc.); as will be seen in the forecasting results (see Section 5.3), GHI is thus easier to predict in summer, while the winter dataset allows the highlighting of the models’ robustness in complex atmospheric conditions. On the other hand, this method allows us to show that it is

not necessary to gather a huge dataset to obtain good forecasting results. If not specified otherwise, the first 30 days and the last 15 days of each GHI dataset are used for training (see Section 5.1) and to evaluate the models' generalisation ability, respectively. Finally, the forecast horizons considered in this comparative study range from 30 min to 48 h, thus covering intrahour, intraday and intraweek cases.

## 5. Modeling and Forecasting Results

In this section, the modeling and forecasting results obtained using the OGPR and OSGPR models included in the comparative study as well as the evaluation metrics used are presented. The methods of initialising and estimating the hyperparameters are also explained.

### 5.1. Hyperparameter Initialisation and Estimation

To model GHI using Gaussian process regression, we first have to identify the most adapted kernel. In this comparative study, kernel identification is inspired by the procedure described in [33]. The main idea is to construct a wide variety of kernel structures by adding or multiplying simple kernels from the kernel families discussed in Section 2.2. While all the possible combinations of simple kernels have been evaluated, only the following kernels and combinations have been included in the comparative study.

- Simple kernels:  $k_E$ ,  $k_{SE}$ ,  $k_{RQ}$ ,  $k_{M_{3/2}}$ ,  $k_{M_{5/2}}$ , and  $k_{Per}$ .
- Quasiperiodic kernels, formulated as
  - products of simple kernels—i.e.,  $k_{Per \times E}$ ,  $k_{Per \times SE}$ ,  $k_{Per \times RQ}$ ,  $k_{Per \times M_{3/2}}$ , and  $k_{Per \times M_{5/2}}$ ;
  - sums of simple kernels—i.e.,  $k_{Per + E}$ ,  $k_{Per + SE}$ ,  $k_{Per + RQ}$ ,  $k_{Per + M_{3/2}}$ , and  $k_{Per + M_{5/2}}$ .

As a matter of fact, results emanating from other combinations of non-periodic kernels are not presented here because they demonstrate a behaviour similar to that of simple kernels.

The maximisation of the log marginal likelihood (21) allows the estimation of kernel hyperparameters from GHI training data. Convergence to the global maximum cannot be guaranteed since  $\mathcal{L}(\theta)$  is non-convex with respect to  $\theta$ . Several techniques can be used to remedy this issue, the most classical of which is to use multiple starting points which are randomly selected from a specific prior distribution; for instance,  $\theta_i \sim \text{Uniform}(0, 1)$ . In [48], the authors incorporate various prior types for the hyperparameters' initial values, then examine the impact that priors have on the GPR models' predictability. Results show that hyperparameters' estimates, when using  $k_{SE}$ , are indifferent to the prior distributions, as opposed to  $k_{Per}$ , whose hyperparameters' estimates differ along with the priors. This leads to the conclusion that prior distributions have a great influence on the hyperparameters' estimates in the case of a periodic kernel.

As mentioned in Section 4, the two GHI datasets (i.e., the summer and winter datasets) are split into a training subset and a testing subset that cover periods of 30 days and 15 days, respectively. As for the initial values of the hyperparameters  $\theta$ , we have made the following choices.

- The correlation length  $\ell$  has been chosen to be equal to the training data's standard deviation.
- In case a periodic kernel is involved in the regression process, an initial value equal to one day has been chosen for the period  $P$  in order to surpass issues that arise while estimating it.
- The initial values of remaining hyperparameters, if any, are randomly drawn from a uniform distribution.

### 5.2. Evaluation of Models' Accuracy

Two evaluation metrics are considered in this comparative study. The first one is the normalized root mean square error (nRMSE):

$$\text{nRMSE} = \frac{\sqrt{\frac{1}{n_*} \sum_{i=1}^{n_*} (y_{\text{test}}(i) - y_{\text{forecast}}(i))^2}}{\frac{1}{n_*} \sum_{i=1}^{n_*} y_{\text{test}}(i)} \quad (22)$$

The second evaluation metric is the correlation coefficient (R):

$$R = \frac{\sum_{i=1}^{n_*} (y_{\text{test}}(i) - \bar{y}_{\text{test}}) (y_{\text{forecast}}(i) - \bar{y}_{\text{forecast}})}{\sqrt{\sum_{i=1}^{n_*} (y_{\text{test}}(i) - \bar{y}_{\text{test}})^2} \sqrt{\sum_{i=1}^{n_*} (y_{\text{forecast}}(i) - \bar{y}_{\text{forecast}})^2}} \quad (23)$$

where  $\bar{\cdot}$  is the arithmetic mean and  $\mathbf{y}_{\text{test}} \in \mathbb{R}^{n_* \times 1}$  and  $\mathbf{y}_{\text{forecast}} \in \mathbb{R}^{n_* \times 1}$  are the test data and forecasts given by the GPR models, respectively.

These evaluation metrics enable performance (generalisation) assessment during testing, which is a challenging task when time horizons and time scales vary, as highlighted in [21].

### 5.3. Forecasting Results Using OGPR

Forecasts of GHI obtained at different time horizons are compared to enable the assessment of the models' global performance. The persistence model is used as a reference. For all the OGPR models catalogued in Section 5.1, the nRMSE vs. the forecast horizon is displayed in Figure 1. Further numerical results for the summer dataset (Tables A2 and A3) and the winter dataset (Tables A4 and A5) are presented in Appendix B. Broadly speaking, there are three classes of models, each possessing a different performance level. The worst performance is exhibited by the persistence model, and improved performance is witnessed when using OGPR models based on simple kernels; lastly, considerably better performance is exhibited by OGPR models based on quasiperiodic kernels, particularly considering higher forecast horizons.

For both datasets, even at the lowest forecast horizon (30 min), OGPR models based on simple kernels give forecasts comparable to those given by the persistence model (nRMSE  $\simeq$  0.22 in summer and nRMSE  $\simeq$  0.30 in winter), while OGPR models based on quasiperiodic kernels already give better forecasts (nRMSE  $\simeq$  0.18 in summer and nRMSE  $\simeq$  0.21 in winter). As the forecast horizon increases, the persistence model's performance degradation is more rapid than that of OGPR models. At the highest forecast horizon (5 h), the persistence model gives nRMSE  $\simeq$  1.11 in summer and nRMSE  $\simeq$  1.50 in winter; for OGPR models based on simple kernels, nRMSE  $\simeq$  0.55 in summer and nRMSE  $\simeq$  0.70 in winter; for OGPR models based on quasiperiodic kernels, nRMSE  $\simeq$  0.30 in summer and nRMSE  $\simeq$  0.40 in winter.

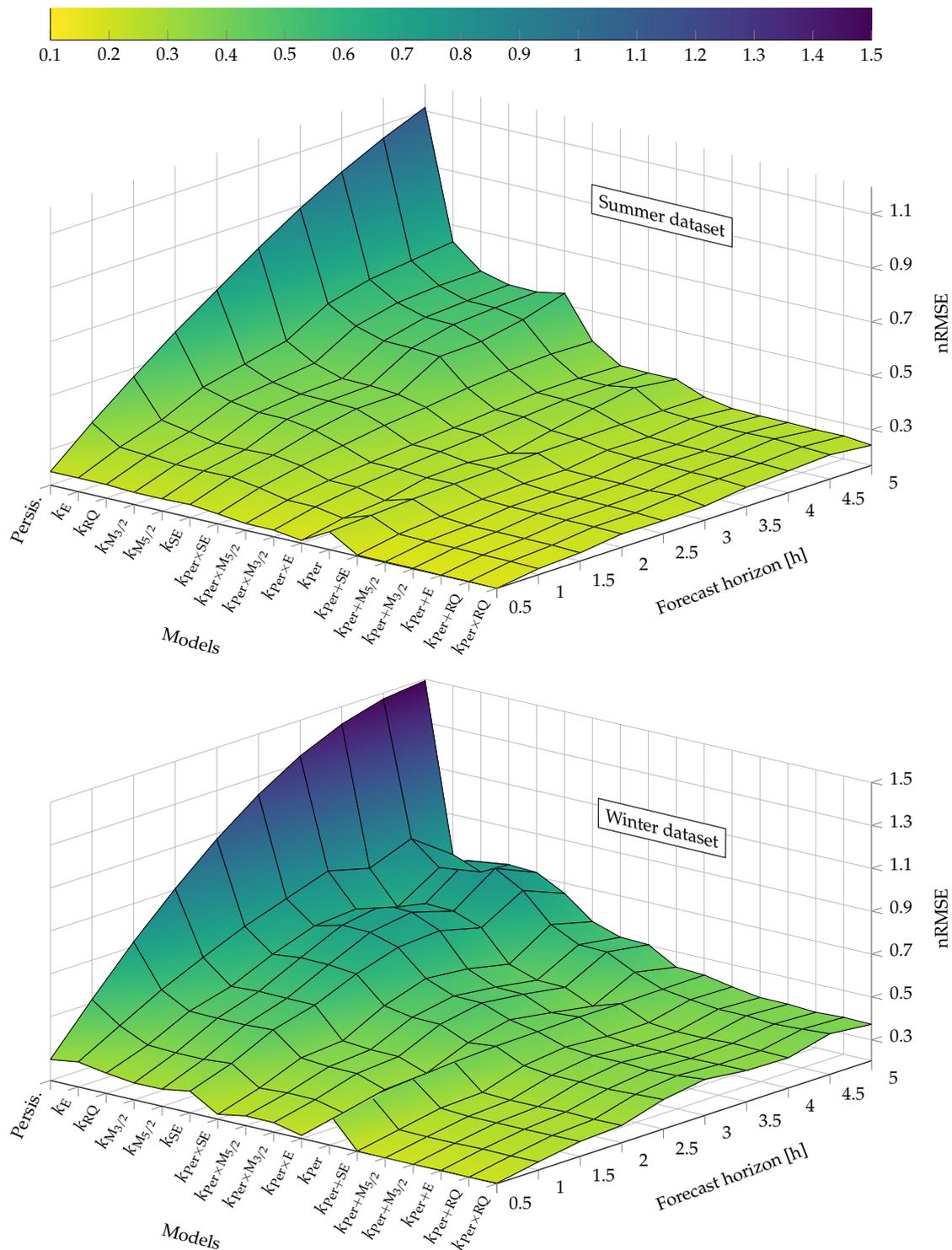
Regarding OGPR models based on simple kernels, no best-performing models have been found. Depending on the forecast horizon,  $k_{M_{3/2}}$ ,  $k_{M_{5/2}}$ ,  $k_{SE}$  and  $k_{RQ}$  all alternatively give the best forecasts, while  $k_E$  does not manage to perform competitively. An interesting observation is that  $k_{SE}$ —this kernel is often used to forecast GHI; see [21,23]—does not ensure the best results among the simple kernels.

Because OGPR models based on a periodic kernel produce a periodic signal, one can say that such models are similar to clear-sky GHI models whose parameters have been fitted to the data. As a consequence, these models simply recurrently reproduce the same bell-shaped curve and produce practically the same nRMSE with respect to increasing forecast horizons. Although good forecasts can be produced by these OGPR models on clear-sky days, they are unable to predict atmospheric-disturbance-induced variability in GHI data.

However, OGPR models based on quasiperiodic kernels—these kernels combine a periodic kernel with a non-periodic kernel—possess the advantage of the periodic kernel, while still managing to predict rapid changes in GHI during the day. Among quasiperiodic kernels,  $k_{\text{Per} \times \text{RQ}}$  surpasses other kernels for the summer dataset, with  $k_{\text{Per} + M_{3/2}}$ ,  $k_{\text{Per} + M_{5/2}}$  and  $k_{\text{Per} + E}$  all coming in a close second (see Table A2); for the winter dataset, however, there is no clear best-performing kernel, as those four kernels alternatively take the first place (see Table A4).

An in-depth analysis of the temporal evolution of GHI during the models' training and testing phases sheds more light on their performance. Once more, the persistence model serves as a reference. Three OGPR models are selected: the classic  $k_{SE}$ -based OGPR model and two of the best-performing models based on quasiperiodic kernels; i.e., the  $k_{\text{Per} \times \text{RQ}}$ -based OGPR model and the  $k_{\text{Per} + SE}$ -based OGPR model. Here, a dataset of nine days (selected from the summer dataset) is used, split into a

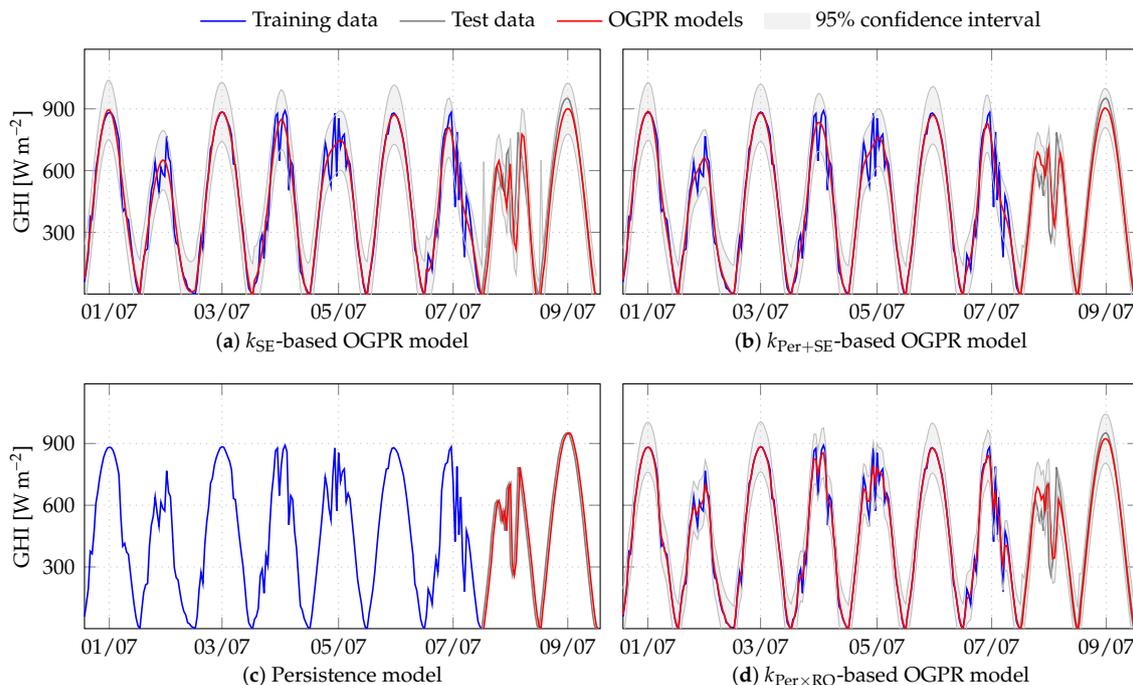
seven-day training subset and a two-day testing subset. In Figures 2–4, 30 min, 4 h, and 48 h forecasts are shown.



**Figure 1.** Normalized root mean square error (nRMSE) vs. forecast horizon for the persistence model (the reference model) and all online Gaussian progress regression (OGPR) models included in the comparative study, for both the summer dataset (top plot) and the winter dataset (bottom plot). The forecast horizon ranges between 30 min and 5 h.

Recall that, while each data sample is used during training, during testing, a new observation is added to the observation set only each whole multiple of the forecast horizon. This means that, first, for all three figures, the training results are identical; second, the coefficients  $\alpha_i$  in Equation (16) are updated every 30 min in Figure 2 and every 4 h in Figure 3, while for the 48-h forecast horizon, no update occurs.

An inspection of the seven-day training phase reveals that the data are well-fitted by every OGPR model. Signals generated by both  $k_{SE}$  and  $k_{Per+SE}$  are quite smooth and show few differences, as opposed to  $k_{Per \times RQ}$ , whose capability for generating more irregular signals allows it to follow the temporal evolution of GHI more closely in the case of atmospheric disturbances.

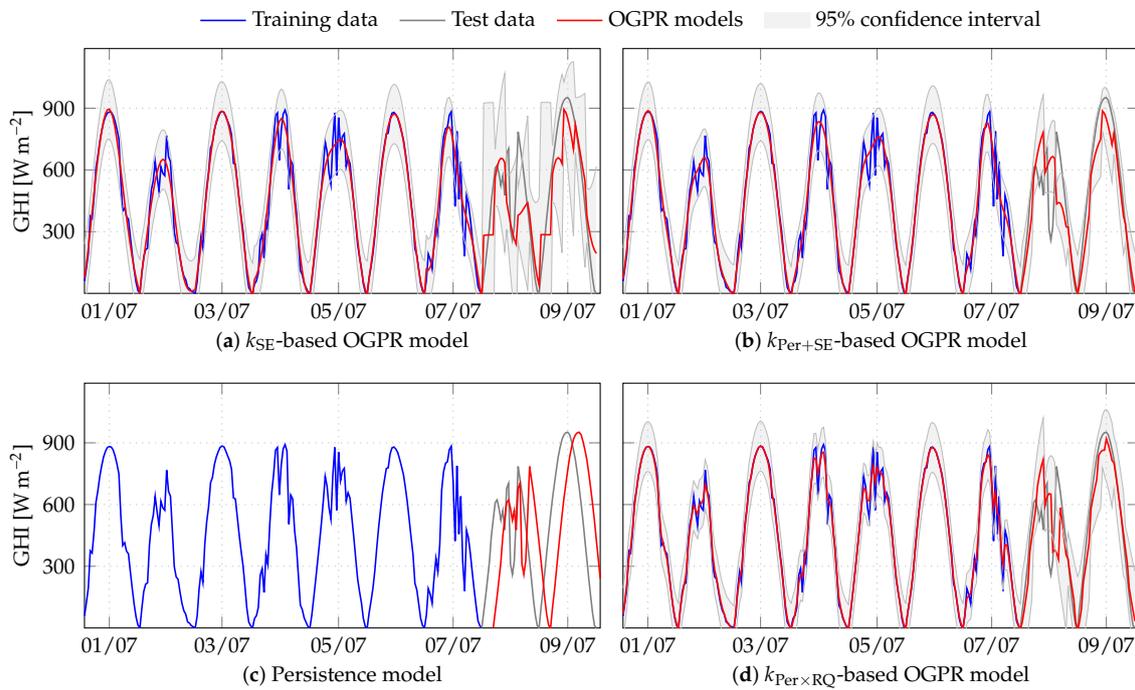


**Figure 2.** Thirty minute forecasts of GHI given by the persistence model and OGPR models based on  $k_{SE}$ ,  $k_{Per+SE}$  and  $k_{Per \times RQ}$  (nine-day dataset, with seven days used for training and two days used for testing).

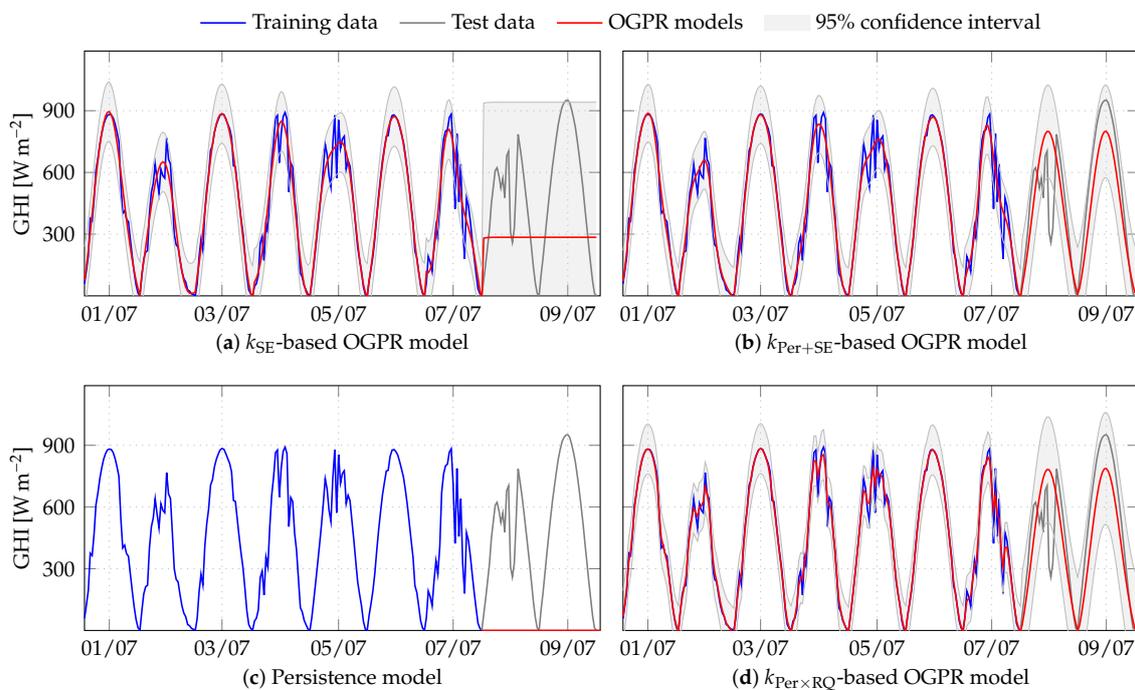
A study of the two-day testing phase reveals the following: all models perform very well when a new observation is added to the observation set every 30 min (Figure 2), although OGPR models based on quasiperiodic kernels, especially  $k_{Per \times RQ}$ , perform slightly better. The performance gap becomes more apparent when the forecast horizon increases. Thus, when a new observation is added to the observation set every 4 h, the  $k_{SE}$ -based OGPR model struggles to predict changes in GHI accurately, as conveyed by the substantial confidence interval between observations (Figure 3a). As soon as a new observation is made, the model fits to the data, but in the absence of an update, it converges to the constant mean value learned during training (around  $280 \text{ W m}^{-2}$ ). In Figure 4a, this behaviour is more obvious: no update is made throughout the entire two-day testing period, and without additional information, the OGPR model simply makes an “educated guess”, consisting of this constant mean value associated with a large confidence interval. Quasiperiodic kernels, however, do possess additional information on GHI, showing that it has a daily pattern. As with OGPR models based on simple kernels, when a new observation is added to the observation set, they fit to the data (see Figure 3b,d).

In the opposite case, OGPR models based on quasiperiodic kernels reproduce the periodic value learned during training (see Figure 4b,d), giving a result that is distinctly more faithful to the desired behaviour than a constant mean value. This explains why the performance of OGPR models based on

quasiperiodic kernels degrades more slowly when the forecast horizon increases, as seen in Figure 1. Based on the results shown in Figure 3b,d,  $k_{Per \times RQ}$  is the best choice among quasiperiodic kernels as it permits sharper and more drastic changes in the modelled signal.



**Figure 3.** Four hour forecasts of GHI given by the persistence model and OGPR models based on  $k_{SE}$ ,  $k_{Per+SE}$  and  $k_{Per \times RQ}$  (nine-day dataset, with seven days used for training and two days used for testing).



**Figure 4.** Forty-eight hour forecasts of GHI given by the persistence model and OGPR models based on  $k_{SE}$ ,  $k_{Per+SE}$  and  $k_{Per \times RQ}$  (nine-day dataset, with seven days used for training and two days used for testing).

The nRMSE gain relative to the persistence model is presented in Table 1, for OGPR models based on the classical  $k_{SE}$  and the three best-performing quasiperiodic kernels (winter dataset).

**Table 1.** Performance comparison, in terms of nRMSE, between the persistence model and OGPR models based on different kernels: the classical  $k_{SE}$  and the three best-performing quasiperiodic kernels (winter dataset). The forecast horizon ranges between 30 min and 5 h. The best results are highlighted in green.

Forecast Horizon		30 min	1 h	2 h	3 h	4 h	5 h
Persistence	nRMSE	0.3009	0.5167	0.9062	1.2168	1.4182	1.4934
$k_{SE}$	nRMSE	0.3012	0.3681	0.5163	0.6692	0.6118	0.6676
	Gain w.r.t. persistence	0%	28.8%	43.0%	45.0%	56.9%	55.3%
$k_{Per+M_{3/2}}$	nRMSE	0.2028	0.2388	0.2810	0.3704	0.3446	0.3839
	Gain w.r.t. persistence	32.6%	53.8%	69.0%	69.6%	75.7%	74.3%
	Gain w.r.t. $k_{SE}$	32.6%	35.1%	45.6%	44.6%	43.7%	42.5%
$k_{Per+E}$	nRMSE	0.2044	0.2410	0.2864	0.3591	0.3517	0.3728
	Gain w.r.t. persistence	32.1%	53.4%	68.4%	70.5%	75.2%	75.0%
	Gain w.r.t. $k_{SE}$	32.1%	34.5%	44.5%	46.3%	42.5%	44.2%
$k_{Per \times RQ}$	nRMSE	0.2054	0.2422	0.2833	0.3819	0.3420	0.3801
	Gain w.r.t. persistence	31.7%	51.3%	68.7%	68.6%	75.9%	74.5%
	Gain w.r.t. $k_{SE}$	31.8%	34.2%	45.1%	42.9%	44.1%	43.1%

#### 5.4. Forecasting Results Using OSGPR

The impact of training data sparsity on forecasting accuracy (generalisation) is assessed in this section of the paper. The subset of data technique is used; this technique simply consists of ignoring a part of the data available for training. Lowering the quantity of data used to train the models reduces computation time during training and also during testing, since the number of parameters is reduced (see Equation (16)). In addition, this allows us to evaluate the models' ability to handle missing data in the training sequence (i.e., their generalisation ability in the case of missing data in that sequence).

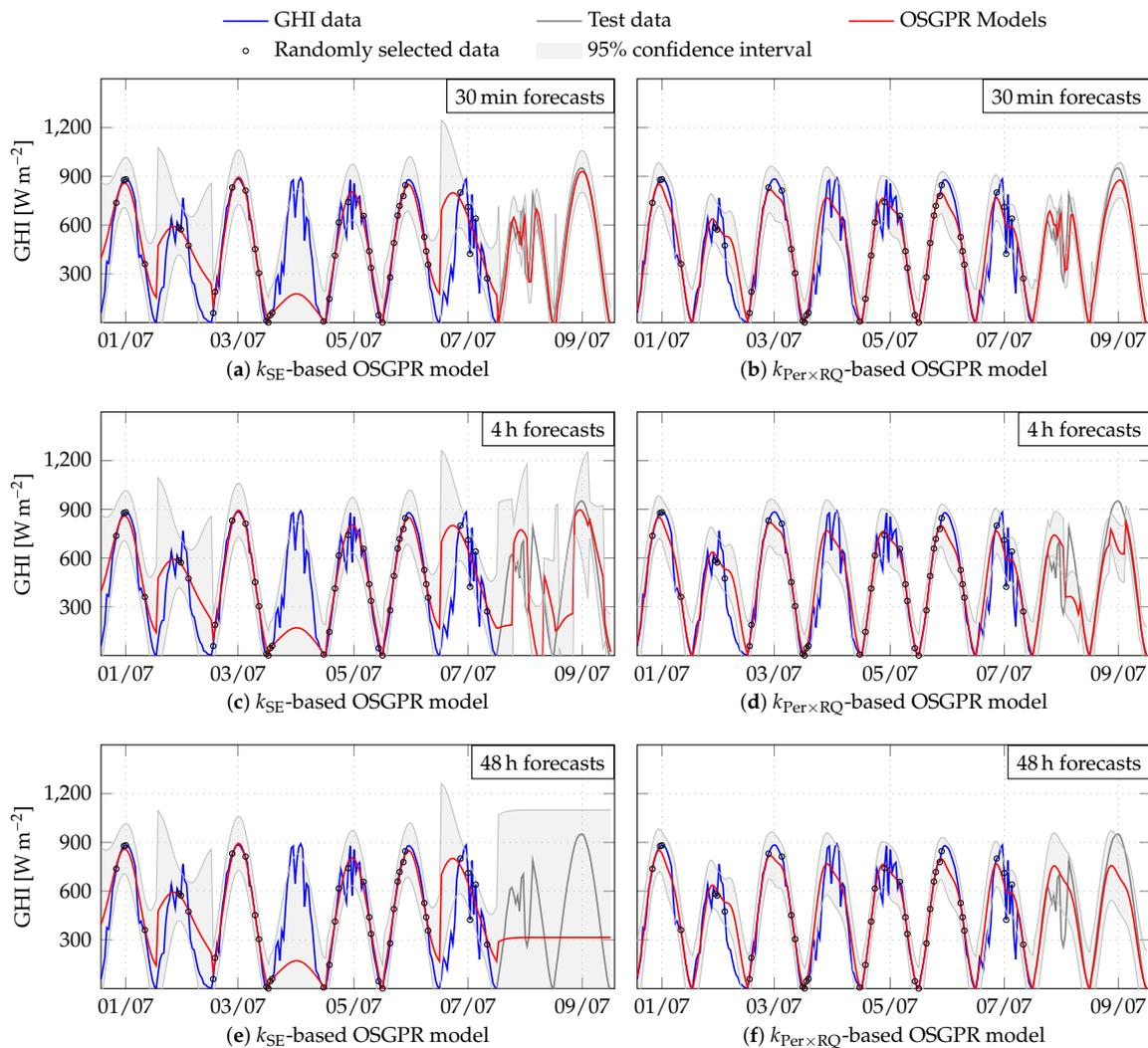
Thirty minute, 4 h and 48 h forecasts given by both the  $k_{SE}$ -based OSGPR model and the  $k_{Per \times RQ}$ -based OSGPR model are displayed in Figure 5. The dataset constitutes the same measurements of GHI (nine days) that were considered with OGPR models (see Section 5.3). Here, however, only 20% of the available data—randomly selected from the summer dataset—have been used during training.

With this low number of training points, the OSGPR model based on  $k_{SE}$  does not provide acceptable results. Compare, for example, the third and fourth days of training: when given sufficient data, a good fit is obtained, but if the only training points are at the beginning and the end of the day (as during the fourth day), the OSGPR model based on  $k_{SE}$  simply makes the best inference, which, in this case, is not sufficient. In contrast, the OSGPR model based on  $k_{Per \times RQ}$  still manages to give an acceptable fit. The key here is that even a low number of training points is enough for the models based on quasiperiodic kernels to learn the daily periodic behavior of GHI.

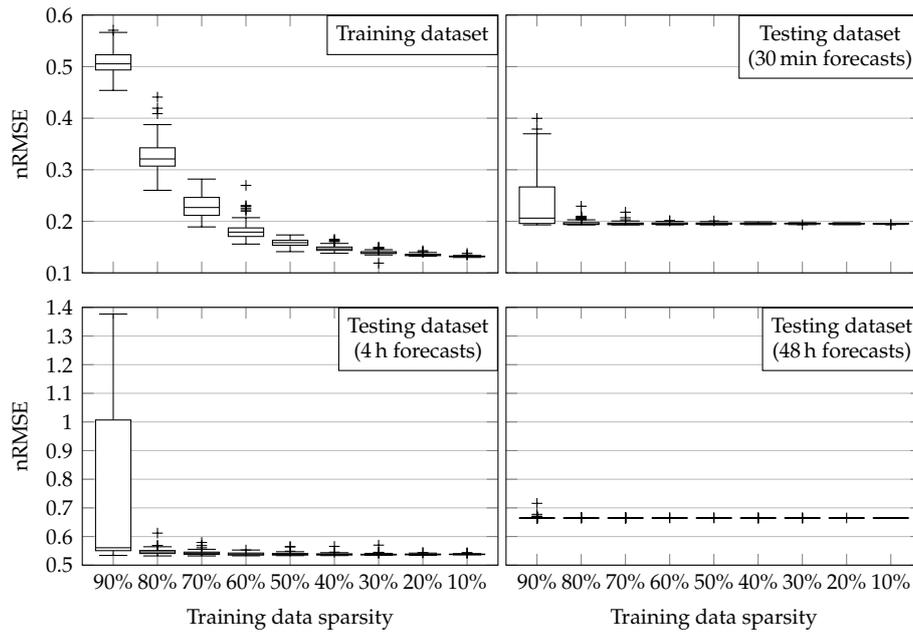
As expected, the forecast results are good when considering the 30 min horizon, with an advantage shown for the  $k_{SE}$ -based OSGPR model. However, as for OGPR models, when the forecast horizon increases, the superiority of the quasiperiodic kernel is shown again. It should be noted that, with a low number of data, periodic behaviour is not learned as well as before: compare Figure 5f, where the two testing days are not bell-shaped but rather resemble the last four training days, to Figure 4d). Nonetheless, this example demonstrates the usefulness of Gaussian process regression models, even in the case of (severely) missing data.

Figures 6 and 7 show nRMSE vs. training data sparsity for the  $k_{SE}$ -based OSGPR model and the  $k_{Per \times RQ}$ -based OSGPR model, respectively. Figure 8 shows the computation time vs. training data sparsity for the latter model. To avoid making conclusions based on a particular realisation (keep in mind that training data are randomly selected), 100 Monte Carlo runs have been conducted; thus, the use of box plots (the Tukey box plots display the median, 0.25-quartile and 0.75-quartile,

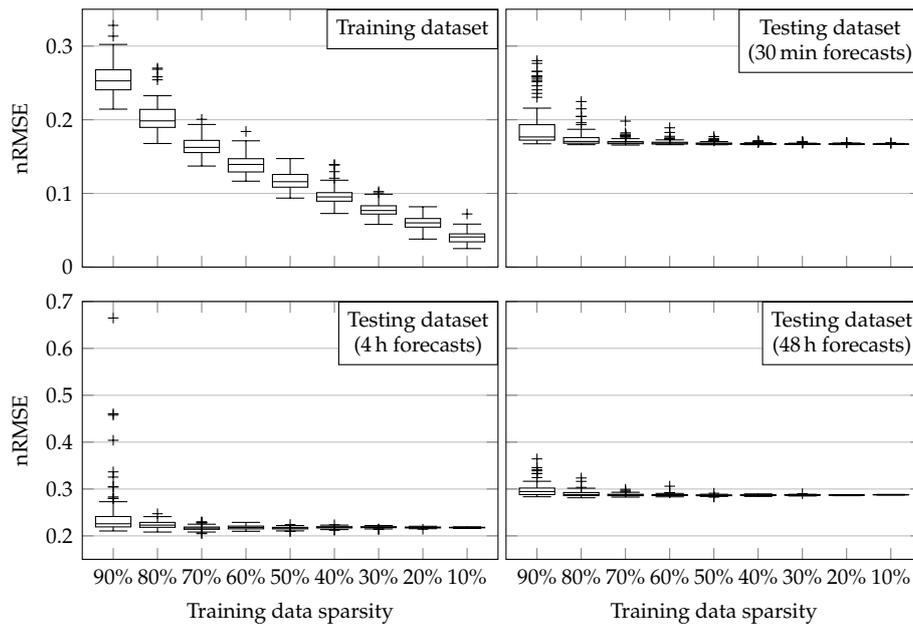
and whiskers corresponding to  $\pm 1.5$  times the interquartile range) in Figures 6–8. As can be seen, if during training, nRMSE decreases steadily with data sparsity, during testing, it quickly reaches a threshold level, and dispersion also quickly vanishes. In this study, this threshold seems to be around 70%. As a consequence, using only 30% of the available training data appears to be sufficient to achieve the same nRMSE results to those obtained when using the whole training dataset. The gain in computation time is significant, as it falls from a median of 175 s at a sparsity level of 0% to a median of 17 s at a sparsity level of 70%, which amounts to around a 90% reduction.



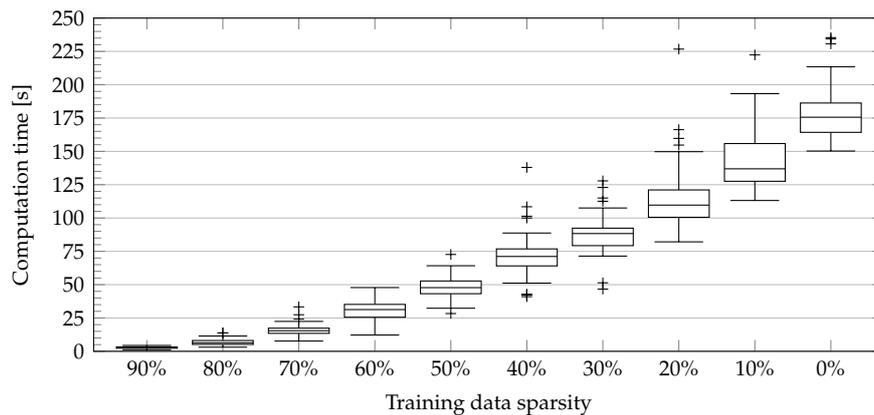
**Figure 5.** Thirty minute, 4 h and 48 h forecasts of GHI given by OSGPR models based on  $k_{SE}$  and  $k_{Per \times RQ}$  (nine-day dataset, with seven days used for training and two days used for testing). Training data are randomly selected to obtain a data sparsity of 20%.



**Figure 6.** Box plots depicting nRMSE vs. training data sparsity for 30 min, 4 h and 48 h forecasts of GHI given by the  $k_{SE}$ -based OSGPR model (summer dataset). Number of Monte Carlo runs: 100.



**Figure 7.** Box plots depicting nRMSE vs. training data sparsity for 30 min, 4 h and 48 h forecasts of GHI given by the  $k_{Per \times RQ}$ -based OSGPR model (summer dataset). Number of Monte Carlo runs: 100.



**Figure 8.** Computation time vs. training data sparsity for the  $k_{\text{Per}\times\text{RQ}}$ -based OSGPR model. 100 Monte Carlo runs have been conducted. Data are randomly selected in each run to obtain the specified sparsity level. The computer used has 32 GB of RAM and an Intel Xeon CPU E3-1220 V2 @3.10 GHz.

## 6. Conclusions

The present paper aimed to design a Gaussian process regression model in the service of GHI forecasting at various time horizons (intrahour, intraday and intraweek). In existing research, only the squared exponential kernel  $k_{\text{SE}}$  was used as the default choice (see [21,23]); no due attention was paid to the kernel selection and the ensuing impact of the model's performance, which was thoroughly investigated and proven to be pertinent in this paper. A comparative study was carried out on several GPR models, ranging from rather simple models to more complex versions, such as quasiperiodic kernels that are the sum or product of a simple kernel and a periodic kernel. Two separate 45-day datasets have been used. The conclusions which arise are that GPR models predictably outperform the persistence model, and that quasiperiodic-kernel-based GPR models are particularly apt to model and forecast GHI. In the case of low forecast horizons, all OGPR models perform satisfactorily. The case of higher forecast horizons, however, is an entirely different matter, as quasiperiodic kernels prove themselves to be much more suitable for the application at hand. The proposed interpretation is that the structure of GHI is better modelled through an omnipresent periodic component representing its global structure and a random latent component explaining rapid variations due to atmospheric disturbances. The periodic component in the proposed quasiperiodic kernels opens the door to GHI forecasting at higher forecast horizons (several hours to a few days) by minimising performance degradation. Accurate GHI forecasting from intrahour to intraweek horizons thus becomes a real possibility, provided that the proper kernel is fashioned.

A simple sparse GPR approach, the subset of data technique, has also been assessed; this consists of using only a part of the available data for training. This approach gives less precise results than standard GPR, but the training computation time is decreased to a great extent: this study indicates that comparable performance (in terms of nRMSE) is obtained when using only 30% of the available training data, which leads to a 90% decrease of computation time. Additionally, even though the lack of data hinders the training process, the results still show the superiority of GPR models based on quasiperiodic kernels.

Future work will focus on some of the best-performing quasiperiodic kernels that have been put forward in this paper and assess their performance using larger GHI datasets.

**Author Contributions:** All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research (project Smart Occitania) was funded by the French Environment and Energy Management Agency (ADEME).

**Acknowledgments:** The authors would like to thank ADEME for its financial support. The authors also thank all the Smart Occitania consortium members—in particular, ENEDIS, the French distribution system operator—for their contribution to this work.

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

## Abbreviations

The following abbreviations are used in this manuscript:

GHI	Global horizontal irradiance
GP	Gaussian process
GPR	Gaussian process regression
OGPR	Online Gaussian process regression
OSGPR	Online sparse Gaussian process regression
nRMSE	Normalized root mean square error

## Appendix A. Kernel Hyperparameters

In this appendix, the mathematical expression of the simple and quasiperiodic kernels considered in Section 5 are given. One can also find in Table A1 the estimated values of the model hyperparameters for all OGPR models, using both summer and winter datasets. As stated in Section 3.4, these hyperparameters are denoted as  $\theta = [\theta_1 \dots \theta_m]^T$ .

**Table A1.** Estimated hyperparameters of all the OGPR models included in the comparative study for both summer and winter datasets.

Kernel	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$
	sum./win.	sum./win.	sum./win.	sum./win.	sum./win.	sum./win.
$k_E$	342.3/175.6	0.510/0.290	—/—	—/—	—/—	—/—
$k_{M_{3/2}}$	354.1/202.2	0.169/0.160	—/—	—/—	—/—	—/—
$k_{M_{5/2}}$	355.0/199.5	0.186/0.170	—/—	—/—	—/—	—/—
$k_{SE}$	351.0/168.6	0.140/0.080	—/—	—/—	—/—	—/—
$k_{Per}$	420.6/172.2	0.998/1.000	1.627/0.767	—/—	—/—	—/—
$k_{RQ}$	348.7/176.6	0.110/0.089	0.664/4.507	—/—	—/—	—/—
$k_{Per \times E}$	3027.7/201.0	0.999/1.001	6.279/0.717	84.22/4.127	—/—	—/—
$k_{Per \times M_{3/2}}$	270.1/157.0	0.972/1.019	0.600/0.513	1.664/1.283	—/—	—/—
$k_{Per \times M_{5/2}}$	271.1/157.3	0.971/1.020	0.599/0.512	1.801/1.405	—/—	—/—
$k_{Per \times SE}$	271.0/154.2	0.950/1.000	0.570/0.524	0.884/0.810	—/—	—/—
$k_{Per \times RQ}$	349.9/252.6	0.998/1.000	1.210/0.889	0.065/0.226	0.016/0.016	—/—
$k_{Per+E}$	374.3/173.1	0.998/1.000	1.497/0.693	146.2/86.76	0.136/0.176	—/—
$k_{Per+M_{3/2}}$	430.2/153.5	0.998/1.000	1.600/0.635	137.4/85.02	0.071/0.112	—/—
$k_{Per+M_{5/2}}$	371.7/151.2	0.998/1.000	1.455/0.629	137.5/85.01	0.079/0.123	—/—
$k_{Per+SE}$	355.1/153.9	0.998/1.000	1.410/0.609	128.4/81.39	0.060/0.070	—/—
$k_{Per+RQ}$	385.9/153.6	0.999/1.000	1.494/0.643	154.9/84.68	0.038/0.074	0.170/0.736

- Simple kernels:

$$k_{SE}(x, x') = \theta_1^2 \exp\left(-\frac{(x-x')^2}{2\theta_2^2}\right) \quad (\text{A1})$$

$$k_{RQ}(x, x') = \theta_1^2 \left(1 + \frac{(x-x')^2}{2\theta_3\theta_2^2}\right)^{-\theta_3} \quad (\text{A2})$$

$$k_E(x, x') = \theta_1^2 \exp\left(-\frac{|x-x'|}{\theta_2}\right) \quad (\text{A3})$$

$$k_{M_{3/2}}(x, x') = \theta_1^2 \left(1 + \sqrt{3}\frac{|x-x'|}{\theta_2}\right) \exp\left(-\sqrt{3}\frac{|x-x'|}{\theta_2}\right) \quad (\text{A4})$$

$$k_{M_{5/2}}(x, x') = \theta_1^2 \left(1 + \sqrt{5}\frac{|x-x'|}{\theta_2} + \sqrt{5}\frac{(x-x')^2}{3\theta_2^2}\right) \exp\left(-\sqrt{5}\frac{|x-x'|}{\theta_2}\right) \quad (\text{A5})$$

$$k_{Per}(x, x') = \theta_1^2 \exp\left(-\frac{2\sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2}\right) \quad (\text{A6})$$

- Products of simple kernels:

$$k_{Per \times SE}(x, x') = \theta_1^2 \exp\left(-\frac{2\sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2} - \frac{(x-x')^2}{2\theta_4^2}\right) \quad (\text{A7})$$

$$k_{Per \times E}(x, x') = \theta_1^2 \exp\left(-\frac{2\sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2} - \frac{|x-x'|}{\theta_4}\right) \quad (\text{A8})$$

$$k_{Per \times M_{3/2}}(x, x') = \theta_1^2 \exp\left(-\frac{2\sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2} - \sqrt{3}\frac{|x-x'|}{\theta_4}\right) \left(1 + \sqrt{3}\frac{|x-x'|}{\theta_4}\right) \quad (\text{A9})$$

$$k_{Per \times M_{5/2}}(x, x') = \theta_1^2 \exp\left(-\frac{2\sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2} - \sqrt{5}\frac{|x-x'|}{\theta_4}\right) \cdot \left(1 + \sqrt{5}\frac{|x-x'|}{\theta_4} + \sqrt{5}\frac{(x-x')^2}{3\theta_4^2}\right) \quad (\text{A10})$$

$$k_{Per \times RQ}(x, x') = \theta_1^2 \exp\left(-\frac{2\sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2}\right) \left(1 + \frac{(x-x')^2}{2\theta_5\theta_4^2}\right)^{-\theta_5} \quad (\text{A11})$$

- Sums of simple kernels:

$$k_{\text{Per+SE}}(x, x') = \theta_1^2 \exp\left(-\frac{2 \sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2}\right) + \theta_4^2 \exp\left(-\frac{(x-x')^2}{2\theta_5^2}\right) \quad (\text{A12})$$

$$k_{\text{Per+E}}(x, x') = \theta_1^2 \exp\left(-\frac{2 \sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2}\right) + \theta_4^2 \exp\left(-\frac{|x-x'|}{\theta_5}\right) \quad (\text{A13})$$

$$k_{\text{Per+M}_{3/2}}(x, x') = \theta_1^2 \exp\left(-\frac{2 \sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2}\right) + \theta_4^2 \left(1 + \sqrt{3} \frac{|x-x'|}{\theta_5}\right) \exp\left(-\sqrt{3} \frac{|x-x'|}{\theta_5}\right) \quad (\text{A14})$$

$$k_{\text{Per+M}_{5/2}}(x, x') = \theta_1^2 \exp\left(-\frac{2 \sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2}\right) + \theta_4^2 \left(1 + \sqrt{5} \frac{|x-x'|}{\theta_5} + \sqrt{5} \frac{(x-x')^2}{3\theta_5^2}\right) \exp\left(-\sqrt{5} \frac{|x-x'|}{\theta_5}\right) \quad (\text{A15})$$

$$k_{\text{Per+RQ}}(x, x') = \theta_1^2 \exp\left(-\frac{2 \sin^2\left(\frac{\pi(x-x')}{\theta_2}\right)}{\theta_3^2}\right) + \theta_4^2 \left(1 + \frac{(x-x')^2}{2\theta_6\theta_5^2}\right)^{-\theta_6} \quad (\text{A16})$$

## Appendix B. Detailed Numerical Results

In this appendix, detailed numerical results are presented. The values of nRMSE and the correlation coefficient R can be found in Tables A2 and A3 for the summer dataset and in Tables A4 and A5 for the winter dataset.

**Table A2.** Numerical values of nRMSE obtained when performing forecasts with persistence and OGPR models (summer dataset). The forecast horizon ranges between 30 min and 5 h. The best results are highlighted in green.

Forecast Horizon	30 min	1 h	2 h	3 h	4 h	5 h
Persistence	0.2177	0.3485	0.5833	0.7935	0.9749	1.1130
$k_E$	0.2183	0.2818	0.4028	0.4737	0.5998	0.6382
$k_{M_{3/2}}$	0.2125	0.2652	0.3574	0.3815	0.5075	0.5263
$k_{M_{5/2}}$	0.2126	0.2651	0.3558	0.3792	0.5042	0.5235
$k_{\text{RQ}}$	0.2179	0.2743	0.3739	0.4144	0.5377	0.5532
$k_{\text{SE}}$	0.2153	0.2666	0.3387	0.3707	0.4660	0.5427
$k_{\text{Per}}$	0.2340	0.2441	0.2596	0.2720	0.2772	0.2779
$k_{\text{Per+E}}$	0.1689	0.1931	0.2209	0.2292	0.2538	0.2518
$k_{\text{Per+M}_{3/2}}$	0.1703	0.1954	0.2223	0.2308	0.2508	0.2548
$k_{\text{Per+M}_{5/2}}$	0.1705	0.1956	0.2226	0.2312	0.2510	0.2553
$k_{\text{Per+RQ}}$	0.1692	0.1929	0.2200	0.2355	0.2526	0.2529
$k_{\text{Per+SE}}$	0.1714	0.1986	0.2268	0.2345	0.2548	0.2588
$k_{\text{Per}\times E}$	0.1779	0.2098	0.2585	0.2480	0.3118	0.3192
$k_{\text{Per}\times M_{3/2}}$	0.1922	0.2280	0.2660	0.2920	0.3211	0.3188
$k_{\text{Per}\times M_{5/2}}$	0.1927	0.2289	0.2677	0.2947	0.3237	0.3220
$k_{\text{Per}\times \text{RQ}}$	0.1675	0.1899	0.2176	0.2197	0.2484	0.2424
$k_{\text{Per}\times \text{SE}}$	0.2045	0.2506	0.3061	0.3462	0.3825	0.3886

**Table A3.** Numerical values of the correlation coefficient R obtained when performing forecasts with persistence and OGPR models (summer dataset). The forecast horizon ranges between 30 min and 5 h. The best results are highlighted in green.

Forecast Horizon	30 min	1 h	2 h	3 h	4 h	5 h
Persistence	0.9462	0.8618	0.6119	0.2835	−0.0773	−0.3992
$k_E$	0.9445	0.9057	0.7973	0.7029	0.5052	0.4196
$k_{M_{3/2}}$	0.9468	0.9140	0.8356	0.8282	0.6392	0.5999
$k_{M_{5/2}}$	0.9468	0.9143	0.8375	0.8300	0.6448	0.6058
$k_{RQ}$	0.9448	0.9096	0.8211	0.7914	0.5893	0.5528
$k_{SE}$	0.9468	0.9165	0.8619	0.8268	0.7165	0.6108
$k_{Per}$	0.9478	0.9361	0.9230	0.9134	0.9091	0.9085
$k_{Per+E}$	0.9673	0.9574	0.9439	0.9413	0.9254	0.9259
$k_{Per+M_{3/2}}$	0.9649	0.9531	0.9396	0.9390	0.9269	0.9222
$k_{Per+M_{5/2}}$	0.9649	0.9530	0.9394	0.9388	0.9268	0.9220
$k_{Per+RQ}$	0.9657	0.9564	0.9435	0.9376	0.9268	0.9240
$k_{Per+SE}$	0.9657	0.9530	0.9372	0.9368	0.9230	0.9202
$k_{Per \times E}$	0.9640	0.9491	0.9218	0.9281	0.8837	0.8807
$k_{Per \times M_{3/2}}$	0.9585	0.9407	0.9158	0.9079	0.8798	0.8807
$k_{Per \times M_{5/2}}$	0.9583	0.9401	0.9147	0.9061	0.8776	0.8781
$k_{Per \times RQ}$	0.9677	0.9584	0.9450	0.9451	0.9278	0.9311
$k_{Per \times SE}$	0.9525	0.9274	0.8856	0.8688	0.8224	0.8201

**Table A4.** Numerical values of nRMSE obtained when performing forecasts with persistence and OGPR models (winter dataset). The forecast horizon ranges between 30 min and 5 h. The best results are highlighted in green.

Forecast Horizon	30 min	1 h	2 h	3 h	4 h	5 h
Persistence	0.3009	0.5167	0.9062	1.2168	1.4182	1.4934
$k_E$	0.3193	0.4250	0.6154	0.7622	0.7765	0.7128
$k_{M_{3/2}}$	0.2765	0.3542	0.5130	0.7086	0.6589	0.7448
$k_{M_{5/2}}$	0.2776	0.3542	0.5112	0.7049	0.6535	0.7383
$k_{RQ}$	0.2947	0.3633	0.5119	0.6776	0.6165	0.6793
$k_{SE}$	0.3012	0.3682	0.5163	0.6692	0.6118	0.6676
$k_{Per}$	0.3479	0.3662	0.3988	0.4129	0.4171	0.4218
$k_{Per+E}$	0.2044	0.2410	0.2864	0.3591	0.3517	0.3729
$k_{Per+M_{3/2}}$	0.2028	0.2388	0.2810	0.3704	0.3446	0.3839
$k_{Per+M_{5/2}}$	0.2028	0.2389	0.2810	0.3710	0.3447	0.3848
$k_{Per+RQ}$	0.2032	0.2399	0.2815	0.3678	0.3466	0.3817
$k_{Per+SE}$	0.2042	0.2421	0.2873	0.3813	0.3575	0.4043
$k_{Per \times E}$	0.2166	0.2608	0.3180	0.4533	0.3714	0.4408
$k_{Per \times M_{3/2}}$	0.2434	0.2896	0.3632	0.4977	0.4264	0.5138
$k_{Per \times M_{5/2}}$	0.2445	0.2909	0.3651	0.5000	0.4294	0.5186
$k_{Per \times RQ}$	0.2054	0.2422	0.2833	0.3819	0.3420	0.3801
$k_{Per \times SE}$	0.2256	0.2847	0.3776	0.5404	0.4800	0.5505

**Table A5.** Numerical values of the correlation coefficient R obtained when performing forecasts with persistence and OGPR models (winter dataset). The forecast horizon ranges between 30 min and 5 h. The best results are highlighted in green.

Forecast Horizon	30 min	1 h	2 h	3 h	4 h	5 h
Persistence	0.9299	0.7932	0.3666	−0.1357	−0.5436	−0.7250
$k_E$	0.9108	0.8494	0.6579	0.4591	0.3958	0.5041
$k_{M_{3/2}}$	0.9397	0.8999	0.7761	0.5770	0.5969	0.5018
$k_{M_{5/2}}$	0.9392	0.8998	0.7774	0.5796	0.6029	0.5071
$k_{RQ}$	0.9306	0.8928	0.7724	0.5846	0.6426	0.5557
$k_{SE}$	0.9273	0.8893	0.7672	0.5861	0.6499	0.5666
$k_{Per}$	0.9147	0.8944	0.8687	0.8583	0.8555	0.8522
$k_{Per+E}$	0.9675	0.9547	0.9354	0.8999	0.9035	0.8880
$k_{Per+M_{3/2}}$	0.9678	0.9554	0.9375	0.8928	0.9067	0.8809
$k_{Per+M_{5/2}}$	0.9678	0.9553	0.9375	0.8924	0.9067	0.8803
$k_{Per+RQ}$	0.9677	0.9550	0.9373	0.8945	0.9059	0.8825
$k_{Per+SE}$	0.9673	0.9539	0.9343	0.8851	0.8989	0.8678
$k_{Per \times E}$	0.9631	0.9461	0.9186	0.8315	0.8871	0.8371
$k_{Per \times M_{3/2}}$	0.9532	0.9330	0.8949	0.7871	0.8525	0.7696
$k_{Per \times M_{5/2}}$	0.9528	0.9324	0.8936	0.7849	0.8501	0.7647
$k_{Per \times RQ}$	0.9671	0.9539	0.9362	0.8846	0.9068	0.8825
$k_{Per \times SE}$	0.9599	0.9354	0.8840	0.7456	0.8035	0.7334

## References

1. Dkhili, N.; Eynard, J.; Thil, S.; Grieu, S. A survey of modelling and smart management tools for power grids with prolific distributed generation. *Sustain. Energy Grids Netw.* **2020**, *21*, 100284. [\[CrossRef\]](#)
2. Lorenz, E.; Heinemann, D. Prediction of Solar Irradiance and Photovoltaic Power. In *Comprehensive Renewable Energy*; Sayigh, A., Ed.; Elsevier: Amsterdam, The Netherlands, 2012; pp. 239–292.
3. Inman, R.H.; Pedro, H.T.; Coimbra, C.F. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* **2013**, *39*, 535–576. [\[CrossRef\]](#)
4. Ineichen, P. Comparison of eight clear sky broadband models against 16 independent data banks. *Sol. Energy* **2006**, *80*, 468–478. [\[CrossRef\]](#)
5. Nou, J.; Chauvin, R.; Thil, S.; Grieu, S. A new approach to the real-time assessment of the clear-sky direct normal irradiance. *Appl. Math. Model.* **2016**, *40*, 7245–7264. [\[CrossRef\]](#)
6. Al-Sadah, F.H.; Ragab, F.M.; Arshad, M.K. Hourly solar radiation over Bahrain. *Energy* **1990**, *15*, 395–402. [\[CrossRef\]](#)
7. Baig, A.; Akhter, P.; Mufti, A. A novel approach to estimate the clear day global radiation. *Renew. Energy* **1991**, *1*, 119–123. [\[CrossRef\]](#)
8. Kaplanis, S.N. New methodologies to estimate the hourly global solar radiation; Comparisons with existing models. *Renew. Energy* **2006**, *31*, 781–790. [\[CrossRef\]](#)
9. Dong, Z.; Yang, D.; Reindl, T.; Walsh, W.M. Short-term solar irradiance forecasting using exponential smoothing state space model. *Energy* **2013**, *55*, 1104–1113. [\[CrossRef\]](#)
10. Yang, D.; Sharma, V.; Ye, Z.; Lim, L.I.; Zhao, L.; Aryaputera, A.W. Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. *Energy* **2015**, *81*, 111–119. [\[CrossRef\]](#)
11. Trapero, J.R.; Kourentzes, N.; Martin, A. Short-term solar irradiation forecasting based on Dynamic Harmonic Regression. *Energy* **2015**, *84*, 289–295. [\[CrossRef\]](#)
12. Hocaoglu, F.O.; Gerek, Ö.N.; Kurban, M. Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks. *Sol. Energy* **2008**, *82*, 714–726. [\[CrossRef\]](#)
13. Sfetos, A.; Coonick, A. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Sol. Energy* **2000**, *68*, 169–178. [\[CrossRef\]](#)
14. Chauvin, R.; Nou, J.; Eynard, J.; Thil, S.; Grieu, S. A new approach to the real-time assessment and intraday forecasting of clear-sky direct normal irradiance. *Sol. Energy* **2018**, *167*, 35–51. [\[CrossRef\]](#)

15. Gutierrez-Corea, F.V.; Manso-Callejo, M.A.; Moreno-Regidor, M.P.; Manrique-Sancho, M.T. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. *Sol. Energy* **2016**, *134*, 119–131. [[CrossRef](#)]
16. Rana, M.; Koprinska, I.; Agelidis, V.G. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Convers. Manag.* **2016**, *121*, 380–390. [[CrossRef](#)]
17. Alzahrani, A.; Shamsi, P.; Dagli, C.; Ferdowsi, M. Solar Irradiance Forecasting Using Deep Neural Networks. *Procedia Comput. Sci.* **2017**, *114*, 304–313. [[CrossRef](#)]
18. Yadav, A.K.; Chandel, S.S. Solar radiation prediction using Artificial Neural Network techniques: A review. *Renew. Sustain. Energy Rev.* **2014**, *33*, 772–781. [[CrossRef](#)]
19. Jiang, H.; Dong, Y. Global horizontal radiation forecast using forward regression on a quadratic kernel support vector machine: Case study of the Tibet Autonomous Region in China. *Energy* **2017**, *133*, 270–283. [[CrossRef](#)]
20. Zendehboudi, A.; Baseer, M.; Saidur, R. Application of support vector machine models for forecasting solar and wind energy resources: A review. *J. Clean. Prod.* **2018**, *199*, 272–285. [[CrossRef](#)]
21. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Foulloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582. [[CrossRef](#)]
22. Li, J.; Ward, J.K.; Tong, J.; Collins, L.; Platt, G. Machine learning for solar irradiance forecasting of photovoltaic system. *Renew. Energy* **2016**, *90*, 542–553. [[CrossRef](#)]
23. Lauret, P.; Voyant, C.; Soubdhan, T.; David, M.; Poggi, P. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Sol. Energy* **2015**, *112*, 446–457. [[CrossRef](#)]
24. 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](#)]
25. Yang, D.; Gu, C.; Dong, Z.; Jirutitijaroen, P.; Chen, N.; Walsh, W.M. Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging. *Renew. Energy* **2013**, *60*, 235–245. [[CrossRef](#)]
26. Bilonis, I.; Constantinescu, E.M.; Anitescu, M. Data-driven model for solar irradiation based on satellite observations. *Sol. Energy* **2014**, *110*, 22–38. [[CrossRef](#)]
27. Aryaputera, A.W.; Yang, D.; Zhao, L.; Walsh, W.M. Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging. *Sol. Energy* **2015**, *122*, 1266–1278. [[CrossRef](#)]
28. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006.
29. Dudley, R.M. *Real Analysis and Probability*, 2nd ed.; Cambridge Studies in Advanced Mathematics; Cambridge University Press: Cambridge, UK, 2002.
30. Neal, R.M. *Bayesian Learning for Neural Networks*; Lecture Notes in Statistics; Springer Science + Business Media: Berlin/Heidelberg, Germany, 1996.
31. Matheron, G. Principles of geostatistics. *Econ. Geol.* **1963**, *58*, 1246–1266. [[CrossRef](#)]
32. Cressie, N.A. *Statistics for Spatial Data*; Wiley Series in Probability and Statistics; Wiley-Interscience: Hoboken, NJ, USA, 2015.
33. Duvenaud, D.; Lloyd, J.; Grosse, R.; Tenenbaum, J.; Zoubin, G. Structure Discovery in Nonparametric Regression through Compositional Kernel Search. In Proceedings of the 30th International Conference on Machine Learning, PMLR, Atlanta, GA, USA, 16–21 June 2013; Volume 28, pp. 1166–1174.
34. Csató, L. Gaussian Processes–Iterative Sparse Approximations. Ph.D. Thesis, Aston University, Birmingham, UK, 2002.
35. Huber, M.F. Recursive Gaussian process: On-line regression and learning. *Pattern Recognit. Lett.* **2014**, *45*, 85–91. [[CrossRef](#)]
36. Ranganathan, A.; Yang, M.H.; Ho, J. Online sparse Gaussian process regression and its applications. *IEEE Trans. Image Process.* **2011**, *20*, 391–404. [[CrossRef](#)]
37. Kou, P.; Gao, F.; Guan, X. Sparse online warped Gaussian process for wind power probabilistic forecasting. *Appl. Energy* **2013**, *108*, 410–428. [[CrossRef](#)]
38. Bijl, H.; van Wingerden, J.W.; Schön, T.B.; Verhaegen, M. Online sparse Gaussian process regression using FITC and PITC approximations. *IFAC-PapersOnLine* **2015**, *48*, 703–708. [[CrossRef](#)]

39. Seeger, M.W.; Williams, C.K.I.; Lawrence, N.D. Fast Forward Selection to Speed Up Sparse Gaussian Process Regression. In Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West, FL, USA, 3–6 January 2003.
40. Snelson, E.L.; Ghahramani, Z. Sparse Gaussian processes using pseudo-inputs. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2005; pp. 1257–1264.
41. Csató, L.; Opper, M. Sparse on-line Gaussian processes. *Neural Comput.* **2002**, *14*, 641–668. [[CrossRef](#)] [[PubMed](#)]
42. Keerthi, S.; Chu, W. A matching pursuit approach to sparse Gaussian process regression. In Proceedings of the Advances in Neural Information Processing Systems 18 (NIPS 2005), Vancouver, BC, Canada, 5–8 December 2005.
43. Quiñonero Candela, J.; Rasmussen, C.E. A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.* **2005**, *6*, 1939–1959.
44. Snelson, E.L. Flexible and Efficient Gaussian Process Models for Machine Learning. Ph.D. Thesis, University of Cambridge, Cambridge, UK, 2007.
45. Blum, M.; Riedmiller, M.A. Optimization of Gaussian process hyperparameters using RPROP. In Proceedings of the 21st European Symposium on Artificial Neural Networks (ESANN'13), Bruges, Belgium, 24–26 April 2013.
46. Moore, C.J.; Chua, A.J.K.; Berry, C.P.L.; Gair, J.R. Fast methods for training Gaussian processes on large datasets. *R. Soc. Open Sci.* **2016**, *3*, 160125. [[CrossRef](#)] [[PubMed](#)]
47. Ambikasaran, S.; Foreman-Mackey, D.; Greengard, L.; Hogg, D.W.; O’Neil, M. Fast direct methods for Gaussian processes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 252–265. [[CrossRef](#)]
48. Chen, Z.; Wang, B. How priors of initial hyperparameters affect Gaussian process regression models. *Neurocomputing* **2018**, *275*, 1702–1710. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).