



**HAL**  
open science

## GHI forecasting using Gaussian process regression

Hanany Tolba, Nouha Dkhili, Julien Nou, Julien Eynard, Stéphane Thil,  
Stéphane Grieu

► **To cite this version:**

Hanany Tolba, Nouha Dkhili, Julien Nou, Julien Eynard, Stéphane Thil, et al.. GHI forecasting using Gaussian process regression. IFAC Workshop on Control of Smart Grid and Renewable Energy Systems, Jun 2019, Jeju, South Korea. hal-02051993

**HAL Id: hal-02051993**

**<https://hal.science/hal-02051993>**

Submitted on 28 Feb 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# GHI forecasting using Gaussian process regression

Hanany Tolba\* Nouha Dkhili\*\*\* Julien Nou\*  
Julien Eynard\*\*\* Stéphane Thil\*\*\* Stéphane Grieu\*\*\*

\* PROMES-CNRS, Rambla de la thermodynamique, Tecnosud, 66100  
Perpignan, France (e-mail: [firstname.lastname@promes.cnrs.fr](mailto:firstname.lastname@promes.cnrs.fr))

\*\* University of Perpignan Via Domitia, 52 Avenue Paul Alduy, 66860  
Perpignan, France (e-mail: [firstname.lastname@promes.cnrs.fr](mailto:firstname.lastname@promes.cnrs.fr))

---

**Abstract:** In this paper, online Gaussian process regression (GPR) is used to model and forecast Global Horizontal Irradiance, at forecast horizons ranging from 30 min to 5 h. It is shown that the covariance function (or kernel) is a key element, deeply influencing forecast results. As a consequence, Gaussian processes with simple kernels and with more complex kernels have been tested and compared to the classic persistence model. Using two datasets of 45 days, it is shown that online GPR models based on quasiperiodic kernels outperform both the persistence model and GPR models based on simple kernels, including the widely used squared exponential kernel.

Keywords: Global horizontal irradiance, forecasts, Gaussian processes, non-parametric regression, machine learning.

---

## 1. INTRODUCTION

Integration of fluctuating power generation (from renewable energy sources) in the electricity grid has always represented a major challenge. A grid operator has to ensure balance between electricity supply and demand. Unfortunately, many difficulties oppose the fulfillment of this goal, even more so in the context of smart grids due to the deployment of distributed power generators. The intermittency of solar power aggravates the problem of voltage regulation in distribution grids (Moreno-Munoz et al., 2008). Therefore, the stability of the network becomes dependent upon its successful compensation of supply-demand variations. As a result, maintaining the continuity and quality of service requires the ability to account for changes in the state of the network in real-time. To answer this question, statistical models, have been extensively used in recent years to forecast Global Horizontal Irradiance (GHI). These predictive models include the persistence model and auto-regressive models (AR) (Boland, 2008). Different models based on Artificial Neural Networks (ANN) can be found in the literature to forecast solar irradiance from meteorological and geographical data (Mellit and Pavan, 2010; Marquez and Coimbra, 2011). A comparative analysis of these methods can be found in (Lauret et al., 2015; Reikard, 2009), among others. However, the use of sophisticated methods remains scarce in the literature of solar irradiance. Especially, the relatively recent development in non-linear modeling, such as Gaussian Process Regression (GPR), remains very limited. In Lauret et al. (2015) and Voyant et al. (2017), a GPR model is used to forecast GHI and is compared with other models. The conclusion is that using GPR models lead most of the time to the best results. However, in these studies, no procedure has been conducted to select an optimal kernel for the GPR model. In fact, only the Squared Exponential (SE) kernel, i.e. the default

choice, has been considered. The present paper will show that this approach can be drastically improved by choosing an adapted kernel.

## 2. GAUSSIAN PROCESSES

### 2.1 Definition

A Gaussian Process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution (Rasmussen and Williams, 2006). A Gaussian process defines a prior over functions, which can be converted into a posterior over functions once some data has been observed. To indicate that a random function  $f(\mathbf{x})$  follows a Gaussian process, we write  $f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ , where  $\mathbf{x}$  and  $\mathbf{x}'$  are arbitrary input variables,  $\mu(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$  is the mean function (usually assumed to be zero) and  $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))^T]$  is the covariance function or kernel.

### 2.2 Covariance function (kernel)

A covariance function encodes our assumptions about the function which we wish to learn. This initial belief could be how smooth the function is or whether the function is periodic. The covariance function is also known as the kernel. Any function could be a covariance function as long as the resulting covariance matrix is positive semi-definite. For scalar-valued inputs, we briefly describe commonly used kernels (see (Rasmussen and Williams, 2006) for an exhaustive list of kernels).

The Squared Exponential (SE) kernel is given by:

$$k_{\text{SE}}(x, x') = \sigma^2 \exp\left(-\frac{(x-x')^2}{2\ell^2}\right) \quad (1)$$

where  $\sigma > 0$  is the amplitude and  $\ell > 0$  is the correlation length parameter or characteristic length-scale. Intuitively,

$\ell$  controls how fast the functions sampled from your GP oscillate.

The Rational Quadratic (RQ) kernel is given by:

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

with  $\alpha > 0$  and  $\ell > 0$ . This kernel is equivalent to a scale mixture of SE kernels with different correlation length distributed according to a Beta distribution with parameters  $\sigma$ ,  $\alpha$  and  $\ell$  (Rasmussen and Williams, 2006).

The Matérn class of kernels is defined by:

$$k_{\text{M}\nu}(x, x') = \sigma^2 \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\sqrt{2\nu}\frac{|x-x'|}{\ell}\right)^{\nu} \cdot \mathcal{B}_{\nu}\left(\sqrt{2\nu}\frac{|x-x'|}{\ell}\right) \quad (3)$$

where  $\sigma$  is the amplitude,  $\ell$  is the correlation length parameter,  $\Gamma$  is the standard Gamma function and  $\mathcal{B}_{\nu}$  is the modified Bessel function of second kind of order  $\nu$ . The parameter  $\nu$  controls the degree of regularity (differentiability) of the resultant GP. If  $\nu = 1/2$ , the exponential kernel is obtained:

$$k_{\text{E}}(x, x') = \sigma^2 \exp(-|x-x'|/\ell) \quad (4)$$

This kernel corresponds to the continuous version of a classical discrete autoregressive model AR(1) (also known as Ornstein-Uhlenbeck process). Two other kernels of Matérn class are widely exploited in the literature. They correspond to the cases  $\nu = 3/2$  and  $\nu = 5/2$ , respectively, and are defined by:

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

$$k_{\text{M}_{5/2}}(x, x') = \sigma^2 \left(1 + \sqrt{5}\frac{|x-x'|}{\ell} + \sqrt{5}\frac{(x-x')^2}{3\ell^2}\right) \cdot \exp\left(-\sqrt{5}\frac{|x-x'|}{\ell}\right) \quad (6)$$

The periodic kernel is given by:

$$k_{\text{Per}}(x, x') = \sigma^2 \exp\left(-\frac{2\sin^2(\pi(x-x')/P)}{\ell^2}\right) \quad (7)$$

A periodic kernel assumes a globally periodic structure of period  $P$  in the function we wish to learn. The parameters  $\sigma$  and  $\ell$  have the same interpretation found in the kernels previously defined.

### 2.3 Kernel composition

It is possible to combine several kernel functions to obtain a more complex kernel function: the only constraint is that the resulting covariance matrix must be a positive semi-definite matrix. Addition and multiplication are two ways of combining covariance functions while keeping the positive semi-definite property:

$$k(x, x') = k_1(x, x') + k_2(x, x') \quad (8)$$

$$k(x, x') = k_1(x, x') \times k_2(x, x') \quad (9)$$

For example, we can model a quasiperiodic GP by multiplying a periodic kernel by a non periodic kernel: this gives a way to transform a global periodic structure into a local periodic structure. In this paper, a wide variety

of kernel structures are built to model GHI by adding or multiplying two kernels from the kernel families discussed in subsection 2.2.

## 3. GAUSSIAN PROCESS REGRESSION (GPR)

### 3.1 Standard Gaussian process regression

Consider the standard regression model:

$$y = f(\mathbf{x}) + \varepsilon \quad (10)$$

where  $\mathbf{x} \in \mathbb{R}^{D \times 1}$  is the input vector,  $f$  is the regression function,  $y$  is the observed value and  $\varepsilon \sim \mathcal{N}(0, \sigma_{\varepsilon}^2)$  is an independent, identically distributed Gaussian noise. Gaussian process regression is a Bayesian nonparametric regression which assumes a GP prior over the regression functions (Rasmussen and Williams, 2006): it consists in approximating  $f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$  using a training set of  $n$  observations  $\mathcal{D} = \{(\mathbf{x}_i, y_i), 1 \leq i \leq n\}$ . As shorthand notation, we merge all the input vectors  $\mathbf{x}_i$  into a matrix  $X \in \mathbb{R}^{n \times D}$  and all corresponding outputs  $y_i$  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 (10), it can be seen that:

$$\mathbf{y} \sim \mathcal{N}(\mu(X), K + \sigma_{\varepsilon}^2 \mathbf{I}) \quad (11)$$

where  $K = k(X, X) \in \mathbb{R}^{n \times n}$ . In this setting, the joint distribution of the observed data  $\mathbf{y}$  and the latent noise-free function on the test points  $\mathbf{f}_* = f(X_*)$  is given by:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(X) \\ \mu(X_*) \end{bmatrix}, \begin{bmatrix} K + \sigma_{\varepsilon}^2 \mathbf{I} & K_* \\ K_*^{\top} & K_{**} \end{bmatrix}\right) \quad (12)$$

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

It can be shown that the posterior predictive density is then also Gaussian (see Rasmussen and Williams (2006)):

$$\mathbf{f}_* | X, \mathbf{y}, X_* \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\sigma}_*^2) \quad (13)$$

where:

$$\boldsymbol{\mu}_* = \mu(X_*) + K_*^{\top} (K + \sigma_{\varepsilon}^2 \mathbf{I})^{-1} (\mathbf{y} - \mu(X)) \quad (14)$$

$$\boldsymbol{\sigma}_*^2 = K_{**} - K_*^{\top} (K + \sigma_{\varepsilon}^2 \mathbf{I})^{-1} K_* \quad (15)$$

Let us examine what happens in the case of a single test point  $\mathbf{x}_*$ . Let  $\mathbf{k}_*$  be the vector of covariances between the test point and the  $n$  training points:

$$\mathbf{k}_* = [k(\mathbf{x}_*, \mathbf{x}_1) \dots k(\mathbf{x}_*, \mathbf{x}_n)]^{\top} \quad (16)$$

Equation (14) becomes:

$$\boldsymbol{\mu}_* = \mu(\mathbf{x}_*) + \mathbf{k}_*^{\top} (K + \sigma_{\varepsilon}^2 \mathbf{I})^{-1} (\mathbf{y} - \mu(\mathbf{x}_*)) \quad (17)$$

Thus  $\boldsymbol{\mu}_*$ , the mean prediction for  $f(\mathbf{x}_*)$ , can be written as a linear combination of kernel functions, each one centered on a training point:

$$\boldsymbol{\mu}_* = \mu(\mathbf{x}_*) + \mathbf{k}_*^{\top} \boldsymbol{\alpha} = \mu(\mathbf{x}_*) + \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, \mathbf{x}_*) \quad (18)$$

where  $\boldsymbol{\alpha} = (K + \sigma_{\varepsilon}^2 \mathbf{I})^{-1} (\mathbf{y} - \mu(\mathbf{x}_*))$ . In the sequel, these coefficients  $\alpha_i$  will be referred to as *parameters*; they are updated each time a new observation is made (as opposed to the parameters of the kernel, referred to as *hyperparameters*, which are not updated once training is over (see subsection 3.3).

### 3.2 Online Gaussian Process Regression (OGPR)

An obstacle that could be met when using the standard GPR approach is the difficulty with which to incorporate

a new training point or set of points. When a forecasting algorithm is run *in situ*, we do not have a fixed data set  $\mathcal{D} = \{(\mathbf{x}_i, y_i), 1 \leq i \leq n\}$ . Rather, data is received one or a few observations at a time, and the total amount of information keeps growing. Moreover, updates are often required every hour, minute or even second. Using the above-mentioned methods in this case would rapidly incur prohibitive computational overheads, since the matrix  $(K + \sigma_\varepsilon^2 \mathbf{I})$  has to be inverted, with a complexity  $\mathcal{O}(n^3)$ .

A solution is to use the so-called Online Gaussian Process Regression (OGPR) approaches (Csato, 2002; Huber, 2014; Ranganathan et al., 2011; Kou et al., 2013; Bijl et al., 2015). Formally, let us suppose that we know the distribution of the GP given the first  $n$  training points  $(X, \mathbf{y})$  and for simplicity, let us consider  $\mu = 0$  (without loss of generality). Suppose that we make  $n_+$  new observations regrouped in the matrix  $X_+$ . Then:

$$\begin{aligned} & \left( \begin{bmatrix} K \\ K_+ \end{bmatrix} + \sigma_\varepsilon^2 \mathbf{I}_{n+n_+} \right)^{-1} \\ &= \begin{pmatrix} k(X, X) + \sigma_\varepsilon^2 \mathbf{I}_n & k(X, X_+) \\ k(X_+, X) & k(X_+, X_+) + \sigma_\varepsilon^2 \mathbf{I}_{n_+} \end{pmatrix}^{-1} \end{aligned} \quad (19)$$

$$= \begin{pmatrix} \underbrace{K + \sigma_\varepsilon^2 \mathbf{I}_n}_A & \underbrace{K_+}_B \\ \underbrace{K_+^\top}_B^\top & \underbrace{K_{++} + \sigma_\varepsilon^2 \mathbf{I}_{n_+}}_D \end{pmatrix}^{-1} \quad (20)$$

Using the block matrix inversion theorem, we get:

$$\begin{aligned} & \left( \begin{bmatrix} K \\ K_+ \end{bmatrix} + \sigma_\varepsilon^2 \mathbf{I}_{n+n_+} \right)^{-1} \\ &= \begin{pmatrix} A^{-1} + A^{-1} B \Delta^{-1} B^\top A^{-1} & -A^{-1} B \Delta^{-1} \\ \Delta^{-1} B^\top A^{-1} & \Delta^{-1} \end{pmatrix} \end{aligned} \quad (21)$$

where  $\Delta = D - B^\top A^{-1} B \in \mathbb{R}^{n_+ \times n_+}$ . It can be seen that the inversion of the  $(n + n_+) \times (n + n_+)$  matrix now only requires the inverse of  $A$ , which is already known, and the inversion of a  $n_+ \times n_+$  matrix: the computational cost is thus  $\mathcal{O}(n_+^3 + n_+^2)$  rather than  $\mathcal{O}((n + n_+)^3)$  when performing direct matrix inversion.

### 3.3 Training a GPR model

As mentioned in subsection 2.2, there is freedom in choosing the kernel function of a GP. Common choices, such as the kernel previously covered, include hyperparameters. The hyperparameters of the GPR model (10), denoted  $\theta$ , group those of its kernel and the variance of the noise. For instance, the squared exponential  $k_{SE}$  kernel defined by 1 contains two hyperparameters: the amplitude  $\sigma$  and the correlation length  $\ell$ , to which is added the noise variance  $\sigma_\varepsilon^2$ : thus, in this case  $\theta = [\sigma \ell \sigma_\varepsilon^2]^\top$ .

These hyperparameters have to be 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 (10), the log marginal likelihood is given by (Rasmussen and Williams, 2006):

$$\mathcal{L}(\theta) = \log [\mathbb{P}(\mathbf{y}|X, \theta)] \quad (22)$$

$$= -\frac{1}{2} \left( \mathbf{y}^\top (K + \sigma_\varepsilon^2 \mathbf{I})^{-1} \mathbf{y} + \log (\det (K + \sigma_\varepsilon^2 \mathbf{I})) + n \log(2\pi) \right) \quad (23)$$

Estimation of the hyperparameters  $\theta$  is usually achieved by maximizing the log marginal likelihood (22). The maximization process may be accelerated if the gradient of the log likelihood is known and a gradient-based algorithm, such as a conjugate gradient method, can be used (Blum and Riedmiller, 2013). The interested reader can turn to (Moore et al., 2016), (Ambikasaran et al., 2016) or (Rasmussen and Williams, 2006) for more information.

## 4. DATA DESCRIPTION

The dataset is derived from measurements taken in Perpignan (southern France), at the PROMES-CNRS laboratory, using a Rotating Shadowband Irradiometer (RSI). Typical uncertainties of the RSI are about  $\pm 5\%$ . The station is located approximately 20 km west of the Mediterranean Sea. Winter is mild and summer is hot and dry. In addition, there is a lot of wind, often resulting in a cloudless sky.

Evaluation of the GPR models included in the comparative study has been made using a dataset covering a period of 45 days, from June 5<sup>th</sup>, 2015 to July 18<sup>th</sup>, 2015, containing GHI data with a 30 min time step. If not specified otherwise, the first 30 days of the dataset are used for training (see subsection 5.1), whereas the last 15 days are used to evaluate models' forecasting skill.

## 5. MODELING AND FORECASTING RESULTS

### 5.1 Initialization and estimation of hyperparameters

For modelling GHI using a GPR model, we first need to identify the kernel function most adapted to this type of time series. The procedure of kernel identification adopted in this study is inspired by the work developed in (Duvenaud et al., 2013). The idea is to construct a wide variety of kernel structures by adding or multiplying kernels. In particular, the kernel families discussed in subsection 2.2, as well as their sum and product are considered. While all the possible combinations have been evaluated, only the following kernels have been included in this study:

- simple kernels:  $k_E, k_{SE}, k_{RQ}, k_{M_{3/2}}, k_{M_{5/2}}, k_{Periodic}$ ;
- quasiperiodic kernels; products:  $k_{Per \times E}, k_{Per \times SE}, k_{Per \times RQ}, k_{Per \times M_{3/2}}, k_{Per \times M_{5/2}}$ ; sums:  $k_{Per+E}, k_{Per+SE}, k_{Per+RQ}, k_{Per+M_{3/2}}, k_{Per+M_{5/2}}$ .

Indeed, the other combinations of non-periodic kernels exhibited the same behavior as simple kernels. For clarity's sake, their results have not been displayed.

The hyperparameters of each kernel have been estimated from GHI training data via the minimization of the log marginal likelihood (22). Due to the non-convexity of  $\mathcal{L}(\theta)$ , the optimization may not converge to the global maximum. The classical approach to tackle this issue is to use multiple starting points randomly selected from a specific prior distribution. In (Chen and Wang, 2018), the authors consider different types of priors for the initial values of hyperparameters and investigate the influence of the priors on the predictability of GPR models. The results reveal that the estimates for  $k_{SE}$  are robust regardless of the prior distributions, whilst they are very different using different priors for  $k_{Per}$  which implies that the prior



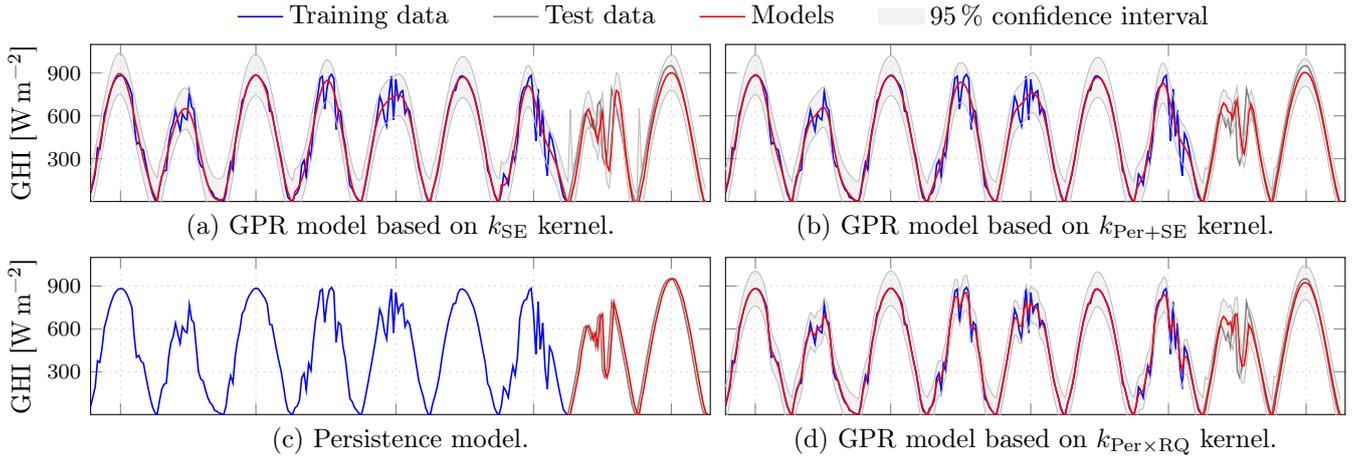


Fig. 2. 30 min forecasts, for the persistence model and OGPR models based on three different kernels:  $k_{SE}$ ,  $k_{Per \times RQ}$  and  $k_{Per+SE}$  (dataset of nine days, with a training dataset of seven days and a testing dataset of two days).

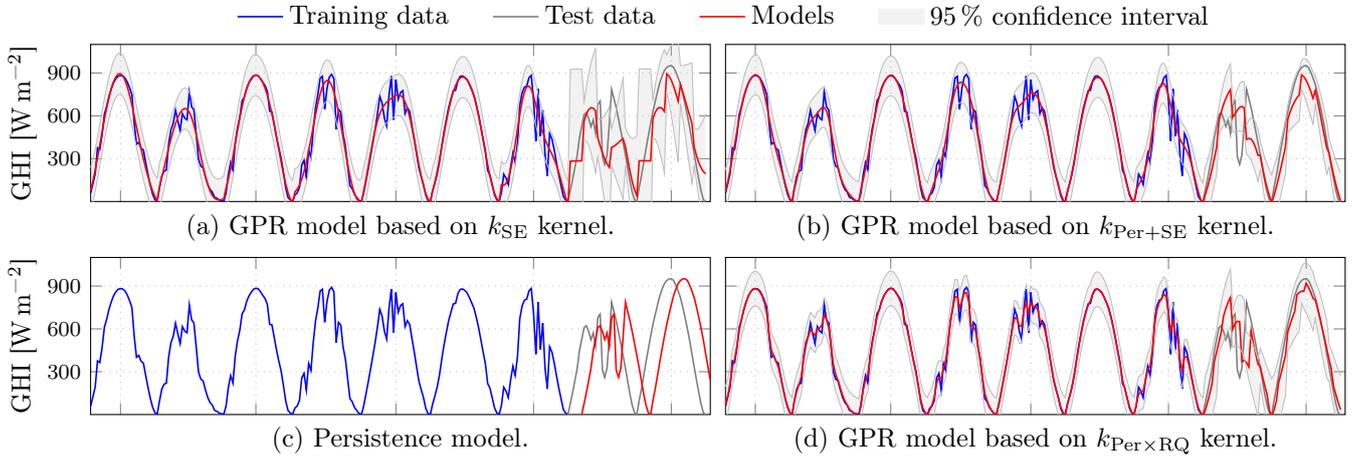


Fig. 3. 4 h forecasts, for the persistence model and OGPR models based on three different kernels:  $k_{SE}$ ,  $k_{Per \times RQ}$  and  $k_{Per+SE}$  (dataset of nine days, with a training dataset of seven days and a testing dataset of two days).

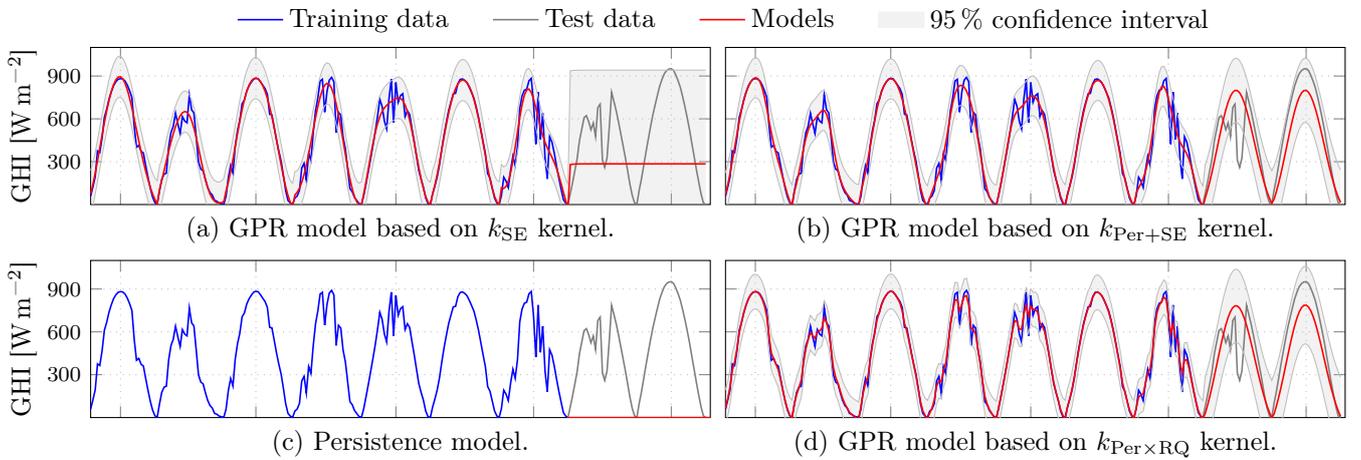


Fig. 4. 48 h forecasts, for the persistence model and OGPR models based on three different kernels:  $k_{SE}$ ,  $k_{Per \times RQ}$  and  $k_{Per+SE}$  (dataset of nine days, with a training dataset of seven days and a testing dataset of two days).

taken into account every 30 min in Figure 2, every 4 h in Figure 3 and every 48 h in Figure 4 (which results in no update at all, since the testing dataset lasts two days). It also implies that training is identical in these three figures; only testing differs.

When looking at the training stage, it can be seen that each GPR model fits the data quite well. There are very few differences between  $k_{SE}$  and  $k_{Per+SE}$ , because they both generate rather smooth signals; the  $k_{Per \times RQ}$  kernel, however, allows for more irregular signals and thus follows GHI more closely during atmospheric disturbances.

Now turn to the testing stage. When there is a new observation each 30 min (Figure 2), the persistence model is excellent and all GPR models give good results, with a slight advantage to quasiperiodic kernels, especially  $k_{Per \times RQ}$ . But when the horizon increases the difference between the models become apparent. When there is a new observation each 4 h, the simple kernel  $k_{SE}$  struggles to explain GHI variation, as demonstrated by the large confidence interval obtained (Figure 3a). When no update is available it simply converges to its mean value (around  $280 \text{ W m}^{-2}$ ). This phenomenon is accentuated in Figure 4a: no updates are made during the entire testing and since this GPR model does not have additional information on the GHI behavior, it simply gives a constant value equal to its mean. On the contrary, quasiperiodic kernels have an additional information on GHI: it has a daily pattern. Therefore, when no update is available they converge to a periodic value. This is the advantage of quasiperiodic kernels: when no additional information is available, they reproduce the periodic behavior learned during training (see Figures 4b and 4d). When comparing the quasiperiodic kernels,  $k_{Per \times RQ}$  is preferred because it allows for more drastic changes in the modeled signal, as can be seen in Figures 3b and 3d.

## 6. CONCLUSION

The main objective of this paper was the elaboration of a Gaussian process regression model adapted to multi-horizon GHI forecasting. In previous research on GHI forecasting using GPR models, the squared exponential kernel  $k_{SE}$  was used as the default choice (see e.g. (Lauret et al., 2015; Voyant et al., 2017)). However, as it has been explained throughout the present document, the kernel is a key element of GPR models. A comparison between several GPR models based on simple kernels and more complex kernels has shown that, while GPR models globally outperform the persistence model, GPR models based on quasiperiodic kernels are better suited to model and forecast GHI. The proposed interpretation is that the structure of GHI is better modeled through an omnipresent periodic component representing its global structure and a random latent component explaining rapid variations due to atmospheric disturbances.

## REFERENCES

- S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg, and M. O’Neil. Fast direct methods for gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):252–265, 2016.
- H. Bijl, J.-W. van Wingerden, T. B. Schön, and M. Verhaegen. Online sparse gaussian process regression using FITC and PITC approximations. *IFAC-PapersOnLine*, 48(28):703–708, 2015.
- M. Blum and M. A. Riedmiller. Optimization of gaussian process hyperparameters using Rprop. In *21st European Symposium on Artificial Neural Networks (ESANN’13)*, Bruges, Belgium, April 2013.
- J. Boland. Time series modelling of solar radiation. In Viorel Badescu, editor, *Modeling Solar Radiation at the Earth’s Surface: Recent Advances*, pages 283–312. Springer Berlin Heidelberg, 2008.
- Z. Chen and B. Wang. How priors of initial hyperparameters affect gaussian process regression models. *Neurocomputing*, 275:1702–1710, 2018.
- L. Csató. *Gaussian Processes - Iterative Sparse Approximations*. PhD thesis, Aston University, 2002.
- D. Duvenaud, J. Lloyd, R. Grosse, J. Tenenbaum, and G. Zoubin. Structure discovery in nonparametric regression through compositional kernel search. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1166–1174, June 2013.
- M. F. Huber. Recursive gaussian process: On-line regression and learning. *Pattern Recognition Letters*, 45:85–91, 2014.
- P. Kou, F. Gao, and X. Guan. Sparse online warped Gaussian process for wind power probabilistic forecasting. *Applied Energy*, 108:410–428, 2013.
- P. Lauret, C. Voyant, T. Soubdhan, M. David, and P. Poggi. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. *Solar Energy*, 112:446–457, 2015.
- R. Marquez and C. F. M. Coimbra. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Solar Energy*, 85(5):746–756, 2011.
- A. Mellit and A. M. Pavan. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Solar Energy*, 84(5):807–821, 2010.
- C. J. Moore, A. J. K. Chua, C. P. L. Berry, and J. R. Gair. Fast methods for training gaussian processes on large datasets. *Royal Society Open Science*, 3(5):160125, May 2016.
- A. Moreno-Munoz, J. J. G. de la Rosa, R. Posadillo, and F. Bellido. Very short term forecasting of solar radiation. In *33rd IEEE Photovoltaic Specialists Conference*, 2008.
- A. Ranganathan, M. Yang, and J. Ho. Online sparse gaussian process regression and its applications. *IEEE Transactions on Image Processing*, 20(2):391–404, 2011.
- C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- G. Reikard. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3):342–349, 2009.
- C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105:569–582, 2017.