algorithms

MDPI

Article

Twenty-Four-Hour Ahead Probabilistic Global Horizontal Irradiance Forecasting Using Gaussian Process Regression

Edina Chandiwana [1,2,†], Caston Sigauke [1,*,†] and Alphonce Bere [1]

[1] Department of Statistics, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa; mataree@staff.msu.ac.zw (E.C.); alphonce.bere@univen.ac.za (A.B.)
[2] Department of Applied Mathematics, Midlands State University, Private Bag 9055, Senga, Gweru 0054, Zimbabwe
* Correspondence: caston.sigauke@univen.ac.za; Tel.: +27-15-962-8135
† These authors contributed equally to this work.

Abstract: Probabilistic solar power forecasting has been critical in Southern Africa because of major shortages of power due to climatic changes and other factors over the past decade. This paper discusses Gaussian process regression (GPR) coupled with core vector regression for short-term hourly global horizontal irradiance (GHI) forecasting. GPR is a powerful Bayesian non-parametric regression method that works well for small data sets and quantifies the uncertainty in the predictions. The choice of a kernel that characterises the covariance function is a crucial issue in Gaussian process regression. In this study, we adopt the minimum enclosing ball (MEB) technique. The MEB improves the forecasting power of GPR because the smaller the ball is, the shorter the training time, hence performance is robust. Forecasting of real-time data was done on two South African radiometric stations, Stellenbosch University (SUN) in a coastal area of the Western Cape Province, and the University of Venda (UNV) station in the Limpopo Province. Variables were selected using the least absolute shrinkage and selection operator via hierarchical interactions. The Bayesian approach using informative priors was used for parameter estimation. Based on the root mean square error, mean absolute error and percentage bias the results showed that the GPR model gives the most accurate predictions compared to those from gradient boosting and support vector regression models, making this study a useful tool for decision-makers and system operators in power utility companies. The main contribution of this paper is in the use of a GPR model coupled with the core vector methodology which is used in forecasting GHI using South African data. This is the first application of GPR coupled with core vector regression in which the minimum enclosing ball is applied on GHI data, to the best of our knowledge.

Keywords: core vector regression; gaussian process; lasso; minimum enclosed ball; solar power

Citation: Chandiwana, E.; Sigauke, C.; Bere, A. Twenty-Four-Hour Ahead Probabilistic Global Horizontal Irradiance Forecasting Using Gaussian Process Regression. *Algorithms* **2021**, *14*, 177. https://doi.org/10.3390/a14060177

Academic Editor: Javier Del Ser Lorente

Received: 10 April 2021
Accepted: 12 May 2021
Published: 2 June 2021

1. Introduction

1.1. Context

Probabilistic forecasting of power grid states promotes the management of energy use and planning. The use of renewable energy sources like solar is a measure that aims at lowering effects such as that of greenhouse gas emission which harms the climate changes. It also gives a positive impact on financial costs that are associated with the use of other energy forms like thermal or hydroelectric power. Short-term forecasting has a greater impact on the safety and financial implication of the electric grid. Since solar has a stochastic and uncontrollable nature, the current study uses Gaussian processes (GP) combined with core vector regression (CVR). Application of Gaussian processes require the computation of the covariance function which is also called a kernel, Jakel et al. [1], and a key element that influences deeply the forecasting results. Hence, it is critical to focus on coming up with a more accurate kernel. We adopt the minimum enclosed ball (MEB) technique, which is a branch of core vector regression that is expected to improve

the forecasting results. The proposed Gaussian process regression (GPR) coupled with core vector methodology is expected to produce forecasts that are more accurate than the conventional benchmark models.

1.2. Literature Review

A number of studies (see, for example, Zhandire [2], Mpfumali et al. [3], Govender et al. [4] and Mutavhatsindi et al. [5]) have been carried out for forecasting global horizontal irradiance in South Africa. A unified approach to solar power forecasting remains a challenge in South Africa. Roman et al. [6], considered a modelling framework that was based on multiple quantile regression whereas Bacher et al. [7] suggested the application of quantile regression as well as statistical smoothing techniques.

Solar power forecasting depends on the use of indicators such as moistness, sun's path, temperature, solar power yard attributes that utilise energy from the sun to deliver solar power and dissipating process. Photovoltaic (PV) cells are used to come up with electric power from the sun's energy. The energy delivered relies upon radiation extracted and on the attributes of the solar panel. This study will benefit the general public by providing forecasting information that is vital for competent use, solar power trading, managing the electricity grid and, improving energy quality supplied to the grid and will assist with decreasing the costs that are identified with climate reliance.

Solar power forecasting which is based on several time horizons performs a vital role in health facilities, building conveyance frameworks, research institutions, schools, PV storage structures management and many other systems that depend on energy. It helps power grid operators in accommodating the load to be able to enhance transportation of energy, assign the necessary balance of energy from other sources during the period when solar power is inaccessible, activities to do with maintenance at the solar power plants. When considering time horizon from various minutes to a few hours, i.e., for a very short term time scale, trend models using on-site measurements are adequate. Hourly predictions containing high geographical and physical settlements that are found from ground-based images are also important.

Various authors have ventured into solar power forecasting using numerical weather prediction models, time series regression, artificial intelligence and many more. In Raza et al. [8], the authors forecasted PV yield over the range of forecasting horizon utilising diverse ordered scales. They did a study and reviewed PV forecast models where they looked at different variables used for forecasting it, that is its yield control profile and execution matrices to assess the predicted model. Their research was based on time series regression techniques and artificial intelligence. In Hong et al. [9], the authors did a study on introducing global energy forecasting in a competition that was done in 2014, a probabilistic estimation of energy using four tracks for the following variables: price, wind, load and sunlight energy-based predictions. The paper aimed at producing decade ahead probabilistic forecasts.

A day-ahead prediction of sunlight-based power yield from plants extracting solar energy in America in the Southwest was done in Larson et al. [10]. The study made use of forecasting methods for the day-ahead hourly averaged energy output from solar power plants dependent on optimisation with the use of least squares prediction on weather factors using numerical methods. Three different prediction strategies were assessed against data that were from some two tracking 1MWp plants that were in California for the years 2011–2014.

In Trapero et al. [11], the authors did an analysis applying kernel density forecasting methods, instability prediction models and a mixture of the two models with the main thrust on enhancing the predicting interim performance. An analysis was done using the two methods based on non-parametric kernel density estimation on a minute to minute solar irradiance. To incorporate volatility forecasts, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and Single Exponential Smoothing (SES) estimation were used. Ranganai and Sigauke [12] applied additive quantile regression to model

global horizontal irradiance. They used long-range dependence models on three sites based in South Africa and found that all the models were anti-persistent. Furthermore, in Govender et al. [4], the authors looked at the clustering of solar irradiance patterns related to cloud cover forecasting from climate predictions using numerical data for the forecasting of solar power or irradiance for the following day.

Amarasinghe et al. [13] used a generalised ensemble model incorporating deep learning methods to come up with solar power predictions. A comparison of the performance of this method was done against support vector regression, deep belief network and random forest regression models and their results showed that their proposed method was the best. Marizt et al. [14] applied GPR for energy predictions using the Bayesian framework. In Dahl and Bonilla [15], the authors worked on a study applying Gaussian process models to forecast solar power for 37 residential sites in a town called Adelaide in Australia. They used an integrated multi-site model without using prior data de-trending, this was achieved by capturing diurnal cycles, and discovered that the multi-site modelling was better than the single-site methods with varying weather conditions. In Hanany et al. [16], the authors did a forecasting study on GHI by applying GPR basing their study on kernel study. They applied several kernels and found that the quasi-periodic kernels outperformed most of them. Research on forecasting of solar power using grouped Gaussian processes was done by Dahl and Bonilla [17]. They applied to multitask GP models with observations being linear with several latent node functions and also based on weight functions on priors. They used grouped coupled priors to solve spatial dependence between functions. Their method improved forecasting accuracy as compared to benchmark models used.

Tipping [18] did research making use of a Bayesian framework to obtain sparse solutions to regression and classification of tasks applying linear parameter models. The research adopted a method called relevant vector machine (RVM) and a function family of the support vector machine (SVM). They found that the Bayesian method provides accurate values which make use of fewer basic functions than the SVM.

Quinonero [19] applied Gaussian processes and relevance vector machines from a Bayesian perspective not applying sampling methods of their interest in computational effection models. They worked on an improved RVM which showed that it had better predictive power. They also looked at another type of GP called reduced rank Gaussian processes (RRGPs) which are equivalent to the infinite extended linear models. Their results proved that the GPs will encounter a problem of the appropriateness of predictiveness of predictive variances, which was solved by modifying the classic RRGP. GPs and RVMs were used to derive equations for predicting uncertainty in time series that is multi-step allowing predictive uncertainties.

Martino and Read [20] studied probabilistic Bayesian analysis by applying a joint relationship between Gaussian processes and relevance vector machines. They used these to come up with a framework to view these approaches by applying kernel ridge regression and drawing connections among them by including aspects such as filtering, smoothing and interpolation. A relationship between these methods and the other methods such as linear kernel smoothers, Kalman filtering, smoothing and interpolation was studied. The results showed that the GP had a good behaviour of the predictive variance but it is restricted by the choice of a kernel function and its results of Kalman smoothing produced the same results.

1.3. Contributions and Research Highlights

The main contribution of this paper is in the use of a GPR model coupled with the core vector methodology which is used in forecasting GHI using South African data. To the best of our knowledge, this is the first application of GPR coupled with core vector regression in which the minimum enclosing ball is applied on GHI data. The other contribution is the application of MEB to select an appropriate kernel to be used coupled with GPR, it finds the ball that contains a given set of points with a minimum radius. The selection of kernels using MEB gives a description of the data domain that is accurate for a given

data set. The CVM algorithm is much faster than the benchmark models, it can manage a bigger dataset than the SVM and GBM. The minimum enclosed ball is the computation of the smallest circle that accommodates the list of sets of points in the Euclidean plane. GPR takes into consideration real-time system forecasting and this improves the accuracy of the forecasts. This enables one to easily identify changes in the immediate, especially when dealing with weather conditions that are constantly evolving. This is an aspect that is ignored by most researchers when forecasting energy. The last contribution is that of combining forecasts, this was done by combining GPR and CVR. Bates and Granger [21] analyzed the effects of combining forecasts and demonstrated that a combination of two coupled models improves forecasts since it reduces errors and this makes the forecasts more accurate.

Lasso via hierarchical interactions was applied to select appropriate variables. Application of the GPR technique was used to determine the parameters of the models, which was done using the Bayesian approach to estimation. CVR is used for kernel selection by applying the concept of minimum enclosing ball to choose the appropriate kernel to use. The ratio between margin and radius of the minimum enclosed ball (MEB) is used to calculate a measure of goodness of a kernel to produce a new minimization formulation of kernel learning. After coming up with the Gaussian process regression model we then compared the results with two benchmark models, which are the gradient boosting method (GBM) and support vector modelling (SVM). The model evaluation was done using the following metrics: root mean square error (RMSE), mean absolute error (MAE), mean average percentage error (MAPE) and percentage bias (Pbias) RSME and the MAE for predictive models in each of the two different areas.

The sections that follow are arranged as follows: in Section 2, models are presented starting with GPR modelling followed by the benchmark models, which are the GBM and SVR, respectively. In Section 3, empirical results are presented. Section 4 presents the discussions while Section 5 concludes the paper.

2. Methods and Materials

The Gaussian process regression (GPR) coupled with the core vector regression are discussed in this section including the estimation methods and the forecast evaluation metrics. This method is rarely used in forecasting renewable energy such as solar power. GPR captures model uncertainty by giving the predicted value as a distribution, this uncertainty is not directly captured by other models. GPs make use of prior information to produce kernels to describe GHI data to enable direct optimisation, hence there is a better trade-off between predicting the data and smoothing it.

2.1. Gaussian Process Regression Models

The flow chart of the proposed GPR based core vector regression method is shown in Figure 1.

GPR is a modelling technique that is non-parametric named after Carl Friedrich Gauss because it is adopted from the Gaussian distribution. It uses the Bayesian approach to perform the estimation and inference of the parameters (Gershman and Blei [22]).

Let the function representing the global horizontal irradiance (GHI) be $f(X)$, where X is a finite subset $x_1, x_2, \ldots, x_n$ of the independent variables such as temperature, barometric pressure, wind speed, among others. The proposed method assumes that GHI follows a Gaussian process (GP). It is described by a mean function, $m(X)$ and covariance function, $k(X, X^1)$ and the data are stochastic over given periods. The GPR model is given in Equation (1).

$$f(x_i) \sim \mathrm{GP}(m(X), k(X, X^1)), \tag{1}$$

where $m(X)$ and $k(X, X^1)$ are the mean and covariance functions, respectively, and X is the input function of the weather variables evaluated against another input function X^1.

A GPR model is typically formulated by defining the parameters of Equation (1), which is defined by the mean and covariance function calculated as follows

$$m(X) = f(X)^T \beta, \tag{2}$$

$$k(X, X^1) = \sigma^2 r(X, X^1), \tag{3}$$

where β is a vector of coefficients, σ is the covariance and r is the basic function. An example of the basic function is the radial basis function, whose kernel is given in Equation (4)

$$k(x, x^1) = \sigma^2 \exp(-\frac{1}{2\ell^2} \mid\mid x - x^1 \mid\mid^2), \tag{4}$$

where $\ell > 0$ is a length scale parameter which determines the length of the 'wiggles' in the function. Therefore, in this case the basic r function is

$$r(x, x^1) = \exp(-\frac{1}{2\ell^2} \mid\mid x - x^1 \mid\mid^2). \tag{5}$$

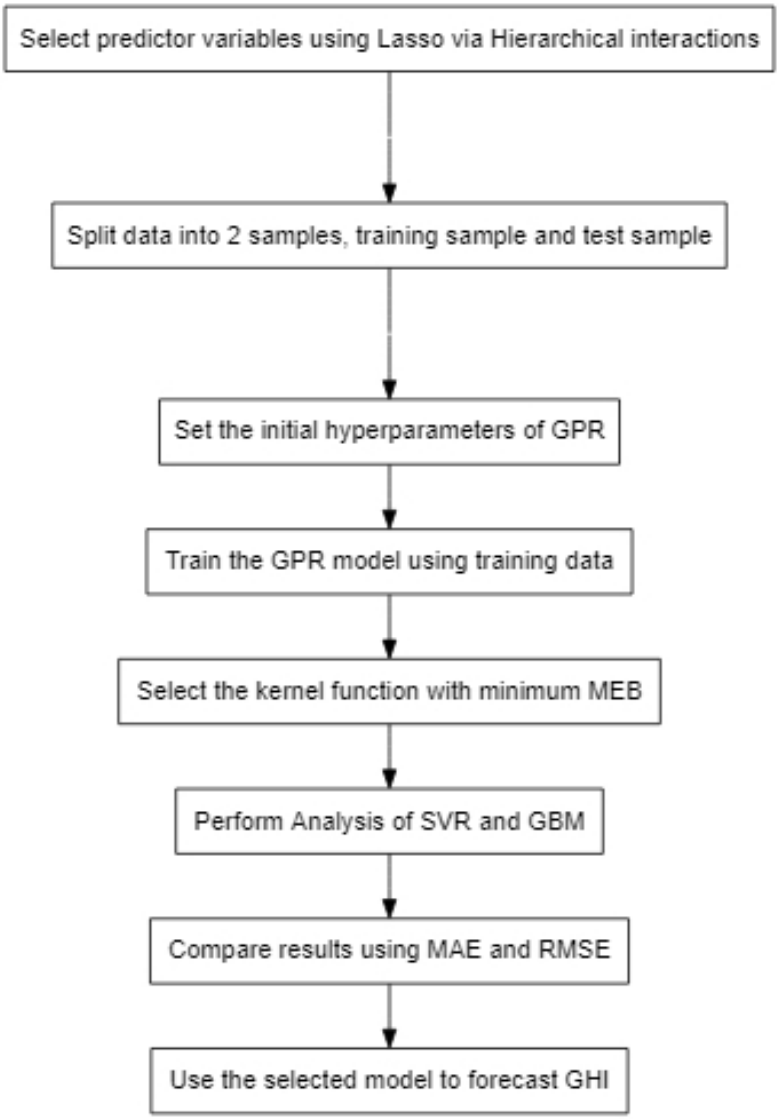

Figure 1. Flowchart of Gaussian process core vector regression model.

2.1.1. Bayesian Inference

The function $f(X)$ of the GPR is characterised by the covariance or kernel function, which is controlled by parameters Θ which is a vector of hyper-parameters (Jakel et al. [1]). Bayesian inference computes the posterior of the hyper-parameters. We compute the posterior as $p(\Theta \mid y) = \frac{p(y|\Theta)p(\Theta)}{p(y)}$, where $p(y \mid \Theta)$ is the likelihood and $p(\Theta)$ is the prior and $p(y)$ is a normalising constant, $p(y) = \int p(y \mid \Theta)p(\Theta)d\Theta$. We used the informative approach to make the inference because it uses priors from the kernel simulations that we developed.

2.1.2. Covariance Functions

Usually, to come up with the most suitable kernel to be used, we use predefined kernels to model a variety of processes. The kernel function is the form of a probability model in which any factors that are not functions of the variables in the domain are omitted. There are several common GP kernels described by Rasmussen et al. [23], some of them are called linear, squared exponential, rational quadratic and Matern. In this study, we are going to use the most widely used covariance functions, which are: Matern, rational quadratic, dot product and radial basis functions.

2.1.3. Radial Basis Kernel

The radial basis function is given in Equation (6)

$$k(x, x^1) = \sigma^2 \exp(-\frac{1}{2\ell^2} \mid\mid x - x^1 \mid\mid^2), \tag{6}$$

where $\mid\mid x - x^1 \mid\mid^2$ is the squared euclidean distance and σ^2 is the variance which determines the average squared distance of the function away from its mean.

2.1.4. Matern Kernel

It is a covariance function given in Equation (7).

$$k(x, x^1) = \sigma^2 \frac{2^{1-v}}{\Gamma(v)} \cdot (\sqrt{2v}\frac{d(x, x^1)}{\rho})^v \cdot k_v \cdot (\sqrt{2v}\frac{d(x, x^1)}{\rho}), \tag{7}$$

where $d(x, x^1)$ is the distance between points, ρ and v are parameters of covariance, k_v modified Bessel function and Γ is the gamma function.

2.1.5. Dot Product Kernel

This kernel can be derived from linear regression by standardising the priors on the coefficients $x_d(d = 1, 2, 3, \ldots, D)$ and $N(0, \sigma_0^2)$ prior on the basis. The function is given in Equation (8).

$$k(x, x^1) = \sigma_0^2 + x \cdot x^1, \tag{8}$$

where σ_0^2 is parameter of the kernel.

2.1.6. Rational Quadratic Kernel

It is a kernel function known as a scale mixture, which is given by Equation (9).

$$k(x, x^1) = (1 + \frac{d(x, x^1)^2}{2\alpha\ell^2}), \tag{9}$$

where $d(x, x^1)^2$ is the distance between points, α is the scale parameter and ℓ is the length scale parameter.

2.2. Core Vector Regression Models

In this study, we proposed the adoption of a minimum enclosed ball to find the goodness of a kernel, which is a branch of core vector regression. Computing the MEB has a problem of computational geometry (Tsang et al. [24]) and this can be solved by incorporating core vector regression. Let the MEB be the minimum enclosed ball of a set X. The MEB is given by the expression (10).

$$K(c, r) \tag{10}$$

with $c, r \in \mathbb{Z}$, c is the centre and r is the radius of MEB.

To apply this algorithm, algorithms by Badoiu and Clarkson [25] and Yildirim [26] are used. They are extracted based on the technique of ϵ-core set of X implying that there exist $C \subset X$ whose MEB is given as $(1 + \epsilon)$ of X. An algorithm by [25] provides an ϵ-core set belonging to X in less than $O(1/\epsilon)$ iterations. Tsang et al. [24] introduced core vector regression which supports the reduction of the MEB problem.

2.3. Variables, Data and Software

Let y_{th}, the dependent variable denote global horizontal irradiance (GHI) measured in W/m^2, where t represents day and h the hour of the day, having $t = 1, \ldots, n$ and $h = 1, \ldots, 24$. The covariate are denoted by $\mathbf{x} = x_{th1}, x_{th2}, \ldots, x_{thp}$, where p is the number of the covariates. In this study, we considered solar hours or sunlight hours only. The data used in this study were hourly-averaged for the period 1 March 2018 to 31 March 2020, for both stations. Data for the period 1 March 2018 to 26 October 2019 were used for training the models, while the remaining data (27 October 2019 to 31 March 2020) were used for testing the models.

Figure 2 shows the map of the distribution of GHI in South Africa. The information on the map is from Solargis website [27].

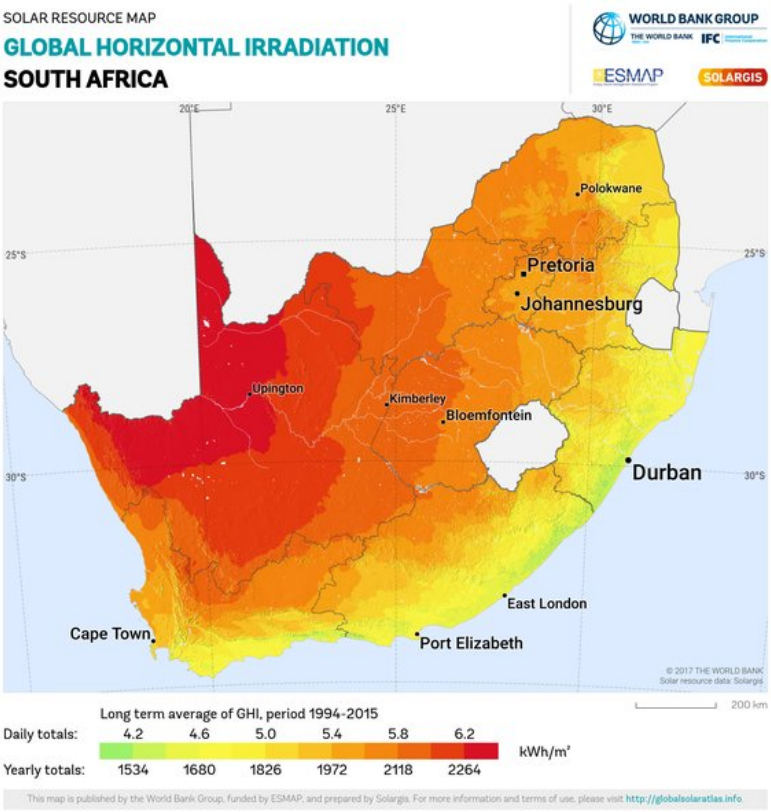

Figure 2. World bank, Global Solar Atlas 2.0, solar resource data, Solargis. Source: https://solargis.com/maps-and-gis-data/download/south-africa, accessed on 23 October 2020.

The data used in this study are from (Southern African Universities Radiometric Network (SAURAN) website https://sauran.ac.za/, accessed on 17 June 2020), one from a coastal area (Stellenbosch University) is referred to as SUN, Western Cape latitude: −33.9281° (E), longitude: 18.8654° (S) and elevation: 119 m. This study focuses on two radiometric stations due to different weather conditions at coastal and inland locations. The instruments are positioned where there is good solar exposure. The pyranometer of SUN which is shown in Figure 3 is on top of a roof of an Engineering building at Stellenbosch University.

Figure 3. Weather instruments at SUN station. Source: https://sauran.ac.za, accessed on 17 June 2020.

Another one is from an inland area USAid Venda referred to as VEN, Vuwani, latitude: −23.1310° (E), longitude: 30.4239° (S), elevation: 628 m. The pyranometer of VEN which is shown in Figure 4 is at Vuwani.

Figure 4. Weather instruments at UNV station. Source: https://sauran.ac.za, accessed on 17 June 2020.

The variable of interest is GHI and the independent variables with their respective abbreviations, as used in the dataset, are given in Table 1 for VEN data and Table 2 for SUN data:

Table 1. Independent variables: VEN station.

Name	Description	Measuring Units
Air Temperature	Temp	°C
Relative Humidity	RH	%
Wind Speed	WS	m/s
Wind Speed Maximum	WS_Max	m/s
Barometer Pressure	BP	mbar
Wind Direction	WD	°
Wind Direction Standard Deviation	WD_Stv	°
Rain Total	Rain_Tot	mm
Wind Vector	WVec	m/s

Table 2. Independent variables: SUN station.

Name	Description	Measuring Units
Temperature	AirTC_Avg	°C
Relative Humidity	RH	%
Wind Speed	WS_ms_S_WVT	m/s
Wind Direction	WindDir_D1_WVT	°
Wind direction standard deviation	WindDir_SD1_WVT	°
Barometric pressure	BP_mB_Avg	mbar

R version 4.0.4 and Python version 3.7.6 are the statistical packages that were used in this study. The package GaussianProcessRegressor from sklearn.gaussian_process was used for Gaussian process regression when Python was used. Missing data points were imputed using multiple iterative imputations.

GHI in this research is the variable of interest that is used as a measure of solar irradiance. It is defined as the aggregate sum of short wave radiation that is gotten from over a surface which is even to the ground and is made up of two forms of irradiances called the Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI).

2.4. Variable Selection and Parameter Estimation

2.4.1. Variable Selection

The variable selection has a vital role in the model building process because in some cases modelling will result in several candidate independent variables hence there is a need for variable selection. Variables are selected to avoid over-fitting, problems of multi-collinearity, and this makes it easier to interpret the model and reduction in computational time. Lasso, Ridge and elasticnet are the types of regularisation techniques that can be used for variable selection. The method adopted for the selection of variables for this research is called Lasso which was first developed by Tibshirani [28]. It performs both shrinkage and variable selection which is used to find the subset of variables that will minimise prediction error. Lasso is a powerful shrinkage method that performs feature selection and regularisation and this tackles the problem of over-fitting, which enhances its predictive accuracy. Tibshirani [28] even goes on to state that Lasso provides probabilities of conducting statistical estimation. Lasso does eliminate the coefficients thus it automatically selects the models. It implements a constraint on model parameters that affects regression coefficients for the variables so that they will shrink towards zero.

In this study, the variables will be selected using Lasso via hierarchical interactions ([29]). Considering a regression model with pairwise interactions between the predictors $x_1, \ldots, x_p$ (i.e., temperature, wind speed, barometric pressure, etc.) for a response variable Y (i.e.,

hourly GHI). The pairwise interaction model is given in Equation (11) which follows a standard GP and discussed in detail in ([29]).

$$y_{t,h} = \beta_0 + \sum_j \beta_j x_j + \frac{1}{2} \sum_{j \neq k} \Theta_{jk} x_j x_k + \varepsilon_{t,h}, \tag{11}$$

where $y_{t,h}$ is GHI on day t at hour h, x_j and x_k are predictor variables, β_0 is the intercept and $\varepsilon_{t,h}$ is the error term, β_0 is the intercept, β_j, Θ_{jk}, are the parameters, respectively. The second term after the intercept β_0 is referred to as the main effects and the third as the interaction terms.

2.4.2. Parameter Estimation

The parameters to be estimated are obtained using either the maximum likelihood estimation (MLE) approach or the Bayesian approach. We have adopted the Bayesian approach to estimation because it captures uncertainty as compared to the MLE and also the framework upon which we are forecasting is based on Bayesian assumptions. Bayesian estimation makes use of either informative or non-informative priors. In this study, we used the informative priors which are from the kernel simulations we developed.

Bayesian approach is more attractive than the frequentist MLE technique since it combines prior information to data, allows analysis with small samples and the forecasting is robust. The Bayesian analysis does not depend on asymptotic estimation and follows the likelihood principle that involves applying a subset of the selected data on the two selected probability models basing on whether they have the same likelihood outcome producing the same assumptions.

2.5. Benchmark Models

2.5.1. Support Vector Regression

Support vector regression (SVR) in general is a regression algorithm with properties of learning machines that are positioned to allow them to generalise data that is yet to be seen. Given some training data of the points $\{(x_1, y_1), \ldots, (x_i, y_i)\}, x \subset \mathbb{R}$, with x being the space inputs. The idea will be to come up with a model $f(x)$ with not more than ε deviation of actual y_i. This means we are not concerned with the errors as long as they are less than ε. Thus, the line function of y is given in Equation (12).

$$f(x) = < w, x > + b = \sum w_j x_j + b, \tag{12}$$

where $w \in X, (y, b) \in \mathbb{R}$ and $< ., . >$ is the dot product of X. SVR makes use of kernels and control of marginal space solutions. It is a regression analysis that works on a generalisation of classification problems, where the model returns a continuous-valued output as compared to an output from an exact set.

2.5.2. Stochastic Gradient Boosting Regression

Stochastic gradient boosting regression (SGBR) is a method that was developed by Friedman [30]. It is a technique that involves coming up with additive regression and fitting a parametrised function to pseudo residuals which are achieved by applying the least-squares at each iteration allowing optimisation of the functional loss.

The additive model can take the form given in Equation (13) ([31]).

$$f(x) = \sum_{m=1}^{M} \beta_m b(x; \gamma_m), \tag{13}$$

where $b(x; \gamma_m) \in \mathbb{R}$ are functions of x which are characterised by the expansion parameters γ_m, β_m. The parameters β_m and γ_m are fitted in a stage-wise way, a process which slows down over-fitting ([31]). Stochastic gradient boosting (SGB) is a modification of GB in which a random sample of the training dataset is taken without replacement.

2.6. Evaluation Metrics

The mean absolute error (MAE), root mean square error (RMSE) and percentage bias (Pbias) will be used for model prediction evaluation. The error measures are given by

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2} \tag{14}$$

Equation (14) is the RMSE which is a measure of goodness of fit which measures the variation, $y_i - \hat{y}_i$. The measure in Equation (15), MAE measures the errors between paired observations which is the predicted against actual values of GHI.

$$\text{MAE} = \frac{1}{n}\sum \mid y_i - \hat{y}_i \mid. \tag{15}$$

Pbias measures the average tendency of the predicted values to be bigger or smaller than the observed values. It is given in Equation (16).

$$\text{PBIAS} = \frac{\sum_{i=1}^{n} y_i - \hat{y}_i}{\sum y_i} \times 100. \tag{16}$$

3. Empirical Results

Results of forecasting solar power hourly data were done on two different datasets from two radiometric stations. A detailed discussion of the radiometric stations was done in Section 2.3.

3.1. Exploratory Data Analysis

A summary of the datasets of the two radiometric stations was done based on their main characteristics through observing descriptive statistics. The summary statistics in Table 3 show that the maximum GHI for the VEN data is 1179.16 W/m^2. The coefficient of skewness is 1.31 indicating the distribution is right-skewed. The coefficient of kurtosis is greater than zero meaning that the distribution has light tails. As for the SUN data, the maximum GHI is 1106.53 W/m^2 and the mean is 220.58. The coefficient of skewness is 1.29, the value is positive, indicating that the data are skewed to the right. The coefficient of kurtosis is 0.398 which means the distribution appears moderately distributed.

Table 3. Descriptive statistics for VEN and SUN data.

	Min	Q1	Median	Mean	Q3	Max	Skew	Kurt
GHI(VEN)	0.0005	238.99	368.39	368.39	378.31	1179.16	1.31	0.84
GHI(SUN)	0.0002	15.73	28.19	220.58	387.88	1106.53	1.29	0.398

Table 4 shows the correlations between GHI of each of the radiometric stations with the important weather variables which are used as covariates in this study. There is a fairly strong positive correlation between GHI and temperature. Similarly, there is strong negative correlation between GHI and relative humidity, while the relationship between GHI and barometric pressure suggests a weak negative correlation.

Table 4. Correlations between GHI and weather variables.

	Temp	BP	RH	WD	WD_SD	WS
GHI(VEN)	0.589	−0.100	−0.528	−0.221	0.492	0.168
GHI(SUN)	0.626	−0.104	−0.640	0.251	0.189	0.242

3.1.1. VEN—Real-Time Analysis

Figure 5 shows a periodogram, which identifies dominant cycles in the time series, for $t = 60$, that is a measure every hour the period is equal to 15, which means that the frequency is 1/15, implying that it takes 15 min periods to come up with a cycle.

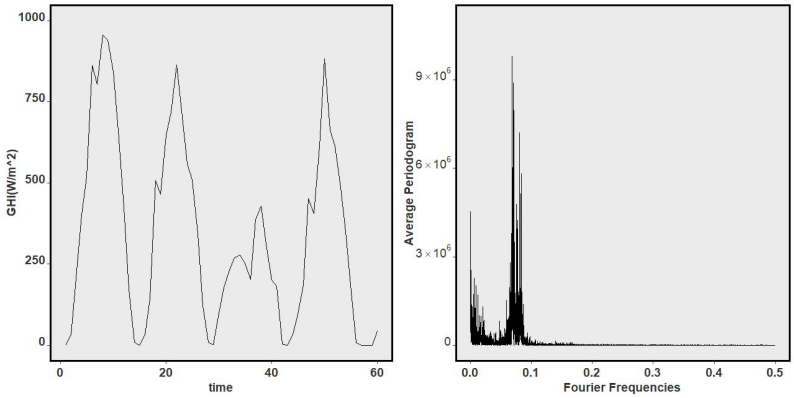

Figure 5. Tsplot real-time data-VEN. **Left panel**: Time series plot. **Right panel**: Periodogram plot.

Figures 6 and 7, respectively show the multiple histograms and scatter diagrams for independent variables. Figure 8 shows a matrix of scatter diagrams showing the relationship between GHI and each independent variable, all the variables except for AirTC _Avg appear to be not linearly related to GHI.

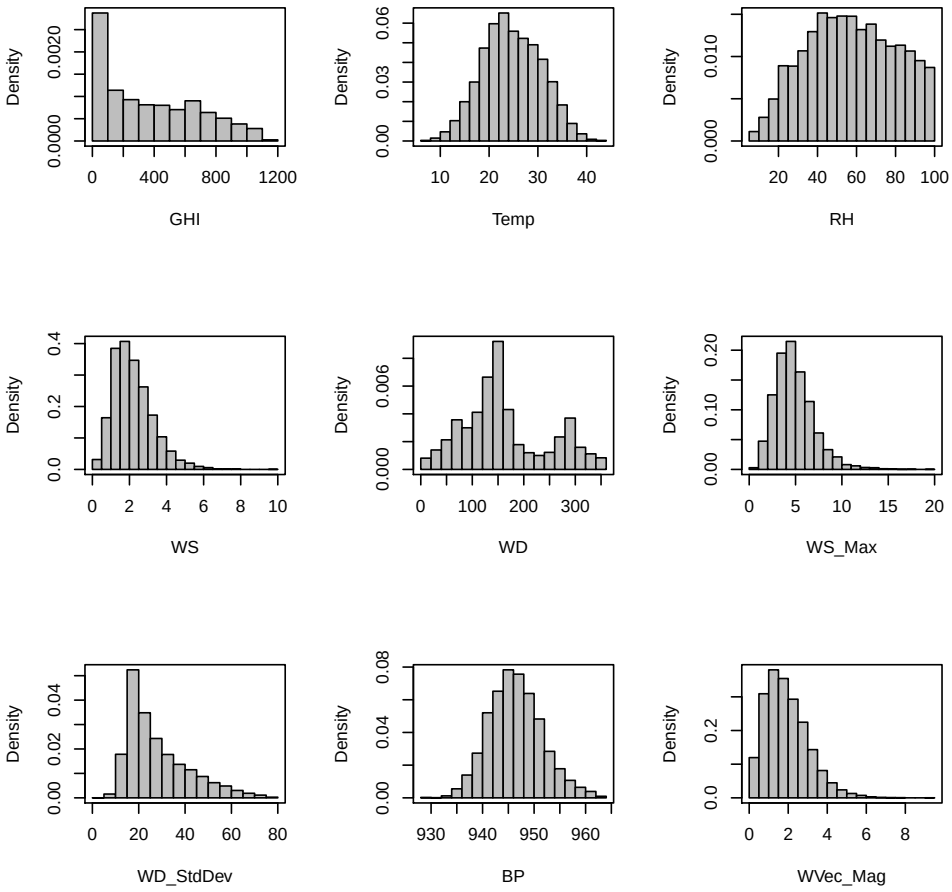

Figure 6. Multiple histograms for VEN data.

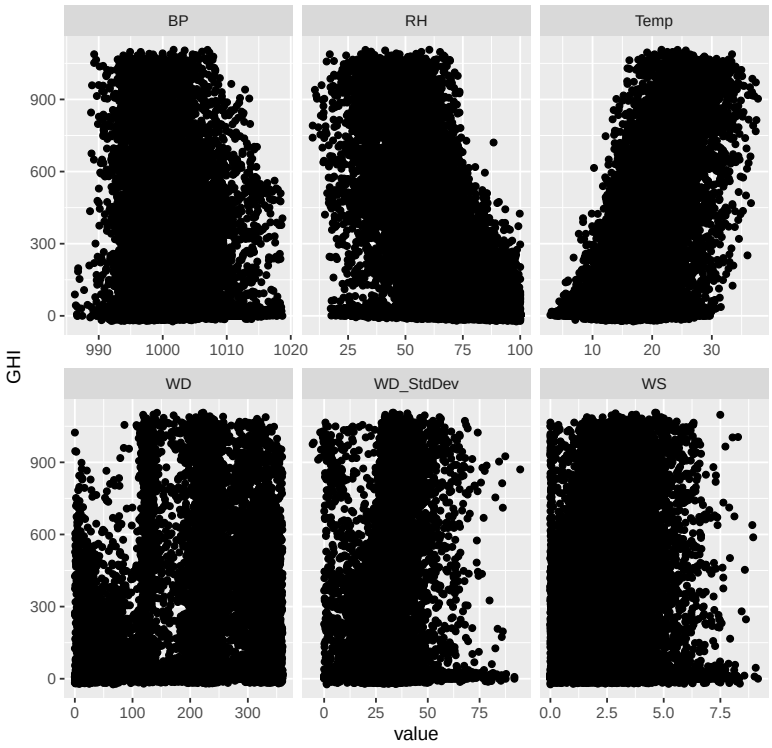

Figure 7. Multiple scatter diagrams for VEN data.

3.1.2. SUN—Real-Time Analysis

A real-time plot was done to show the trend of GHI over a period of 1 h.

Figure 8 shows a periodogram for SUN data, which identifies dominant cycles in the time series, for $t = 60$ the period is equal to 15 which means that the frequency is $1/15$, implying that it takes 15 min periods to come up with a cycle. Figures 9 and 10, respectively show the multiple histograms and scatter diagrams for variables for SUN data.

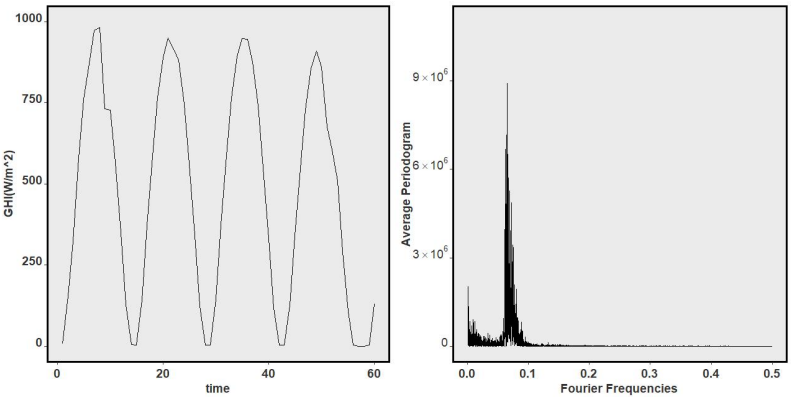

Figure 8. Tsplot real-time data. **Left panel**: Time series plot. **Right panel**: Periodogram plot.

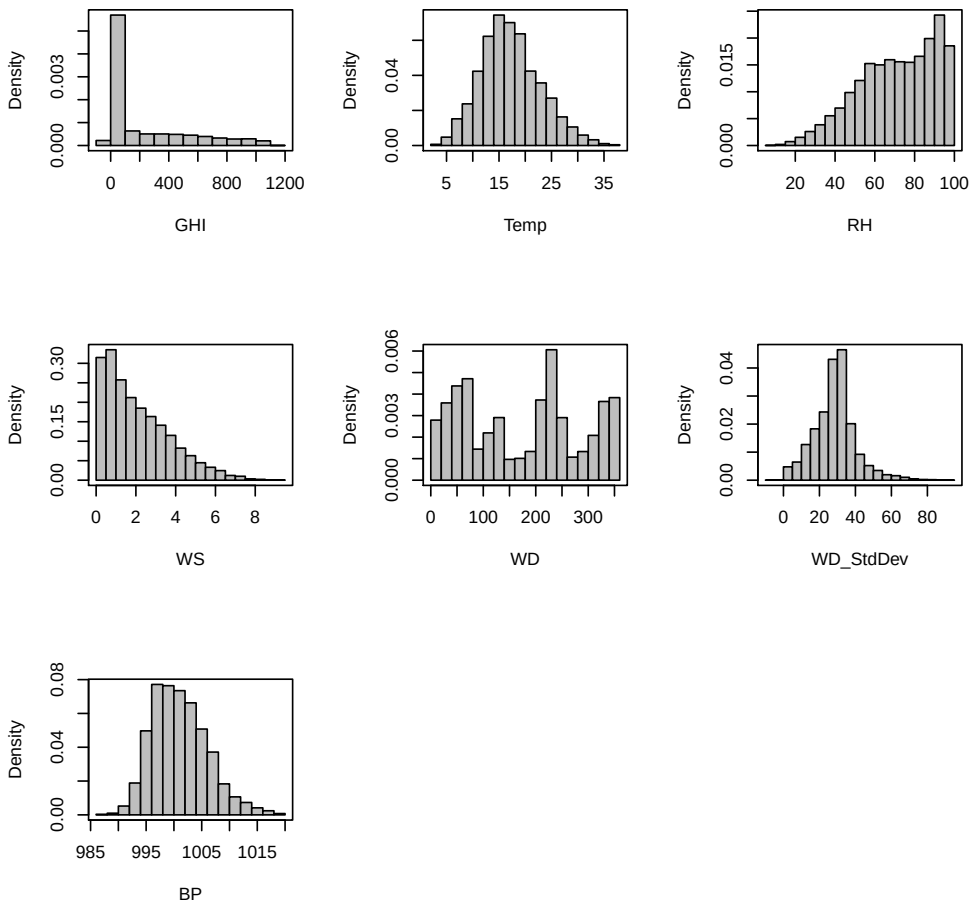

Figure 9. Multiple histograms for SUN data.

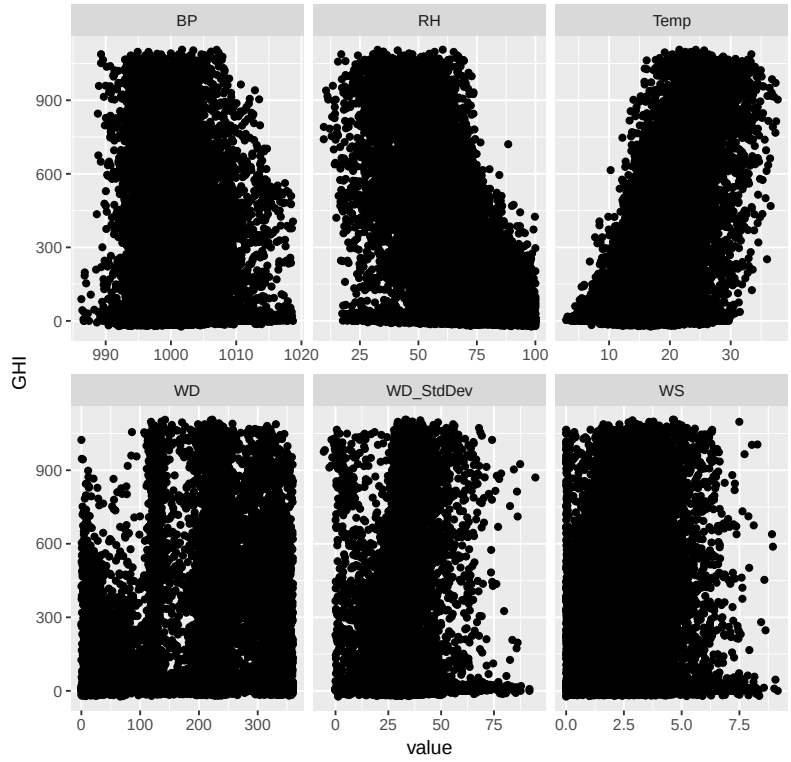

Figure 10. Multiple scatter diagrams for SUN data.

3.2. Forecasting Results

3.2.1. Variable Selection Using Lasso via Hierarchical Interactions

VEN Data

Selection of variables was done for VEN data using Lasso via hierarchical interactions and a summary of the results is presented in Figure 11. Twelve variables were selected and used in all the models in this research.

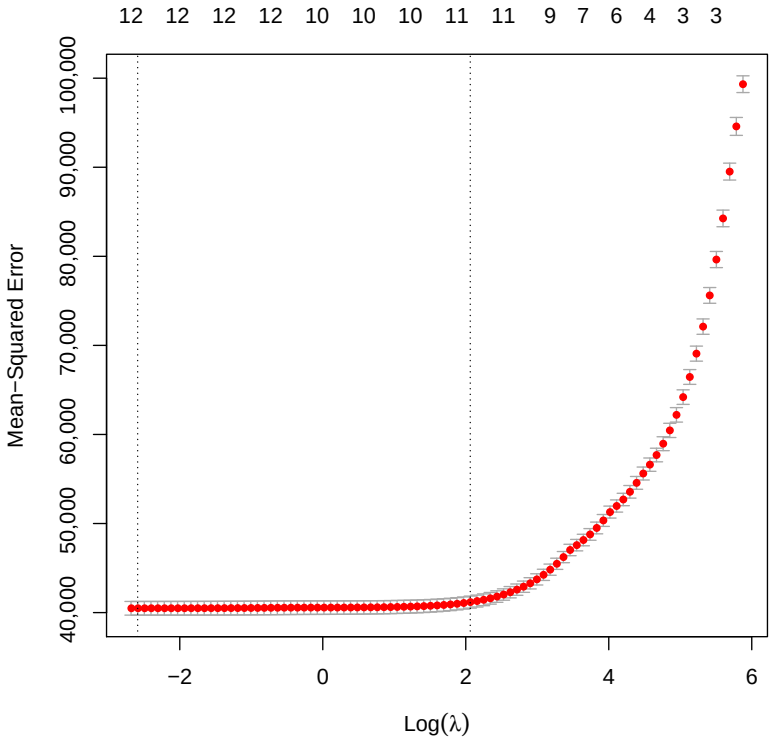

Figure 11. Lasso regression plot for VEN data.

Figure 11 shows the Lasso results for VEN data, with the graph showing the 12 selected variables. The plot helps us visualise MSE which is on the y-axis with log(λ) which is given on the x-axis, it is the cross-validation curve (with red dots) and the variables which are included at these points are on the top. The lower and upper curve limits are the 2 dotted lines which are the λs. The error is very high when we have a few variables but as the number of variables increases the error approaches zero.

SUN Data

Variable selection was also done using SUN data. Figure 12 shows the Lasso results, the graph helps us visualise MSE on the vertical axis with log(λ) on the horizontal axis and we can see that nine variables were selected. In the beginning, the error is very high whereas the coefficients are very small and then at some point, it levels off, this implies the inclusion of good variables.

3.2.2. Gaussian Process Regression Results

The analysis that follows is based on GPR results which were done in two parts, one without interactions and the other with interactions. GPR analysis was done based on the four kernels: Matern, rational quadratic, dot product and radial basis function.

VEN Data without Interactions

Figure 13 shows a static plot of the predicted solar irradiance values against the observed values of VEN data without interaction. The predicted values are in blue which are following closely the observed values in orange. The appropriate kernel was chosen

using the MEB technique and the radial quadratic kernel had the minimum enclosed ball, hence it was used in the forecasting process.

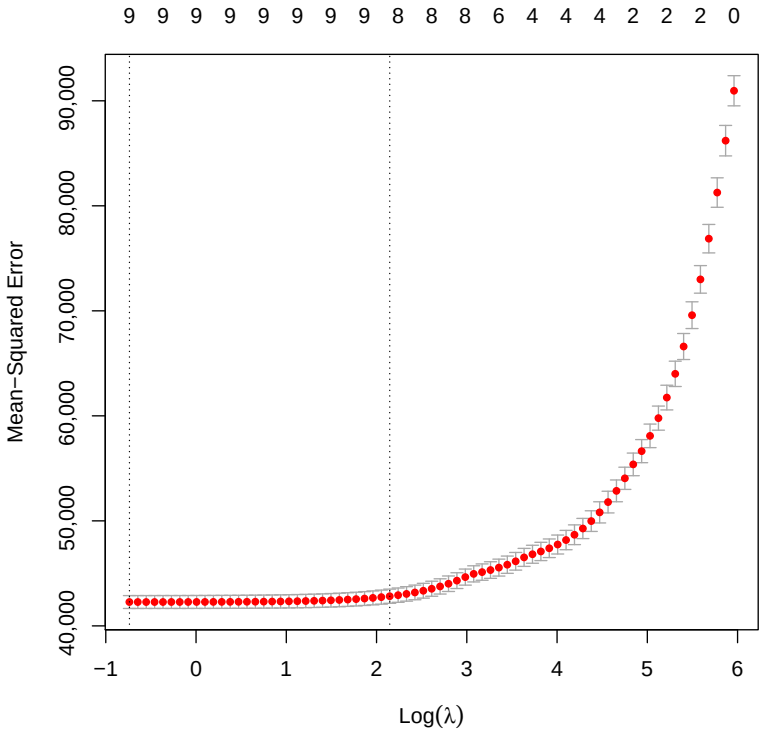

Figure 12. Lasso regression plot for SUN data.

Figure 14 shows a real-time plot for VEN data, the observed data are in red and predicted in blue which were observed over 10 h. On real-time analysis, forecasting was done for 10 h using 150 observations because the period was found to be equal to 15 from the periodogram for 1 h.

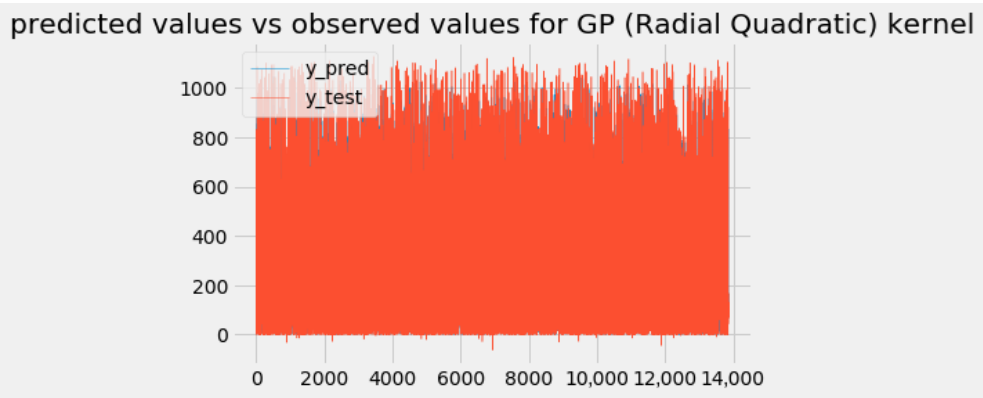

Figure 13. GPR with interactions.

VEN Data with Interactions

Forecasting was also done applying interactions, again the radial quadratic kernel produced the minimum radius, hence it was used to come up with the forecasts. Figure 15 shows a static plot for VEN data with interactions, the observed data are in red and predicted in blue. On real-time analysis, forecasting was done for 10 h using 150 observations and is shown in Figure 16.

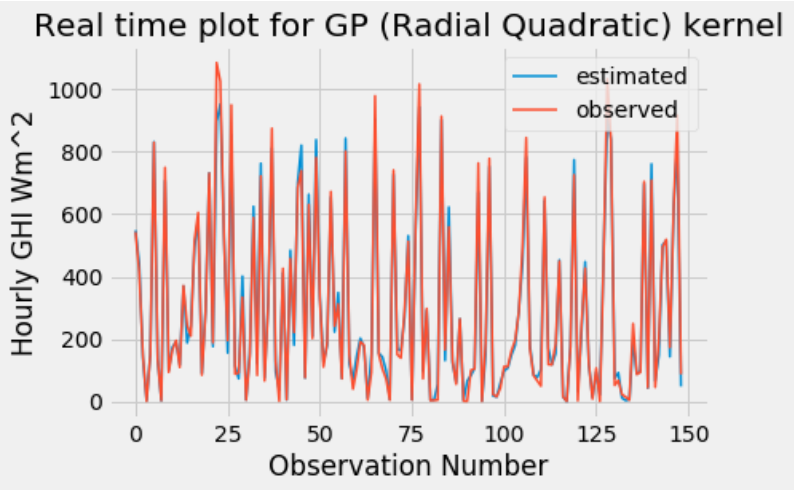

Figure 14. GPR no interaction relative influence plot.

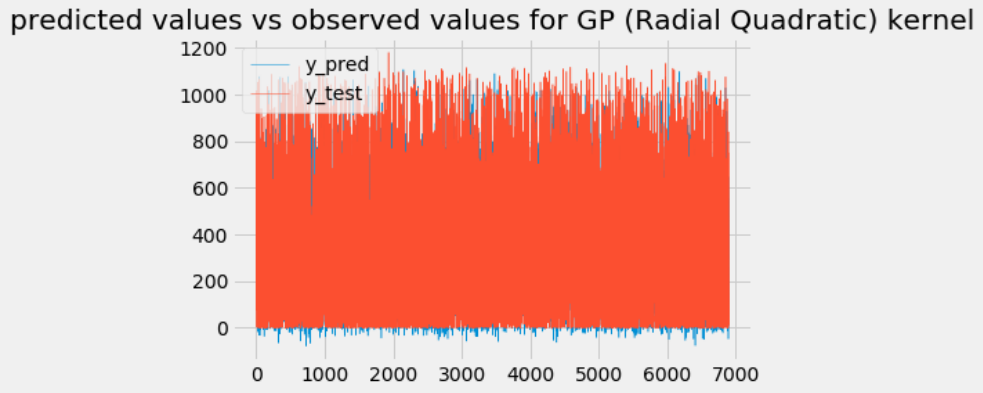

Figure 15. GPR VEN with interactions.

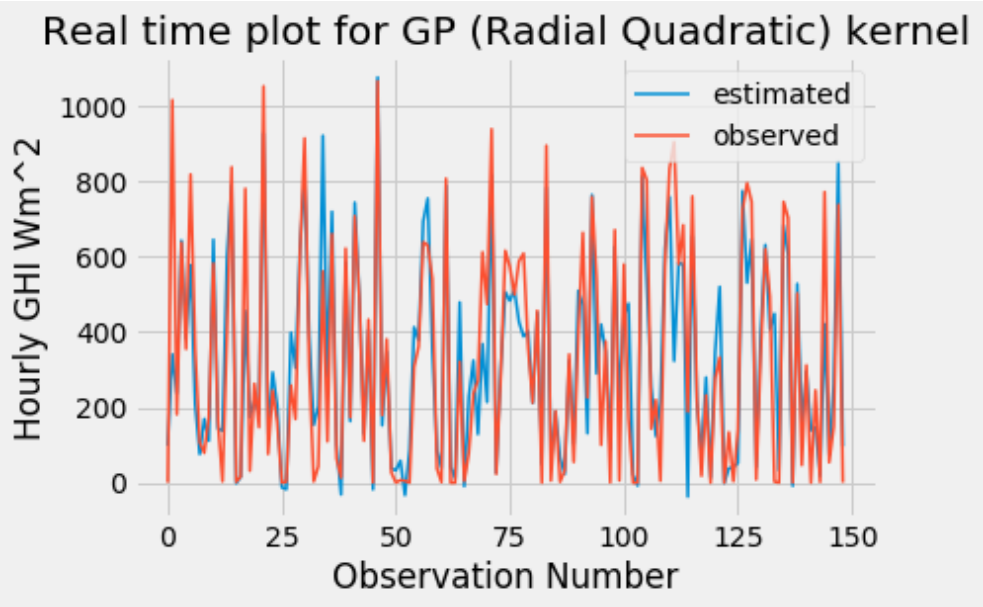

Figure 16. GPR VEN with interaction relative influence plot.

SUN Data with No Interactions

Forecasting was also done for SUN data with no interactions included. The appropriate kernel was chosen using the MEB technique and the radial quadratic kernel had

the minimum enclosed ball, hence it was used for forecasting. Results of the forecasts are shown in Figure 17, which is a static plot of SUN data, the predicted solar irradiance in blue follows closely the observed values in orange.

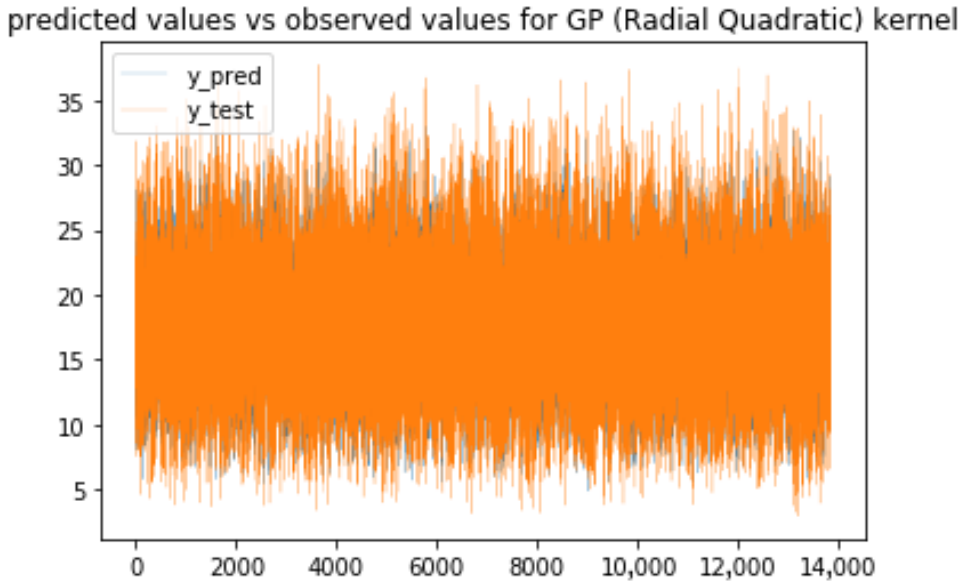

Figure 17. GPR with no interactions.

The plot on Figure 18 shows a real-time plot for SUN data, the observed data are in red and predicted in blue which were observed over 10 h and they follow each other very closely.

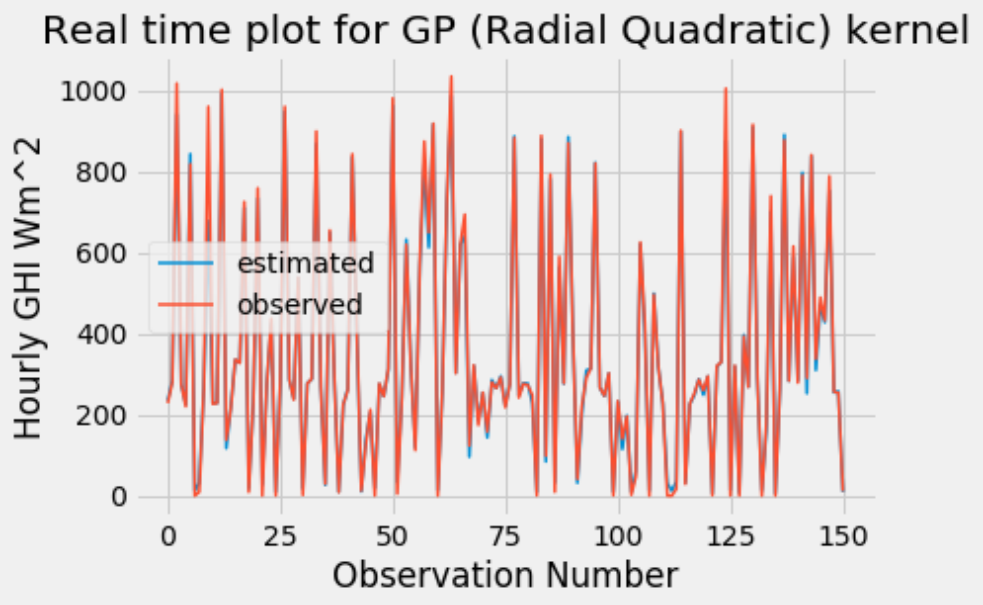

Figure 18. GPR with no interaction real-time plot.

SUN Data with Interactions

GPR analysis was also done on SUN data with interactions included, the results are shown in Figure 19, a static plot, the predicted solar irradiance in blue follows closely the observed values in orange. The appropriate kernel was chosen using the MEB technique and the radial quadratic kernel had the minimum enclosed ball, hence it was used for forecasting. The plot in Figure 20 shows a real-time plot for SUN data, the observed data

are in red and predicted in blue which were observed over 10 h and they follow each other very closely.

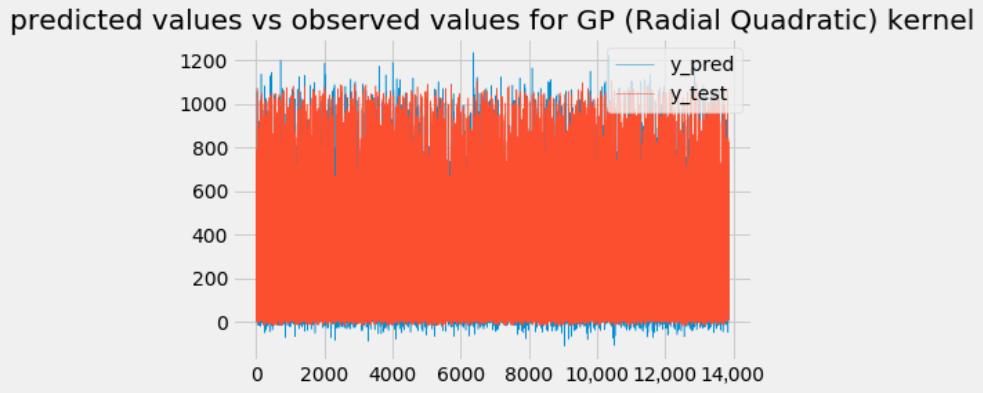

Figure 19. GPR SUN with interactions.

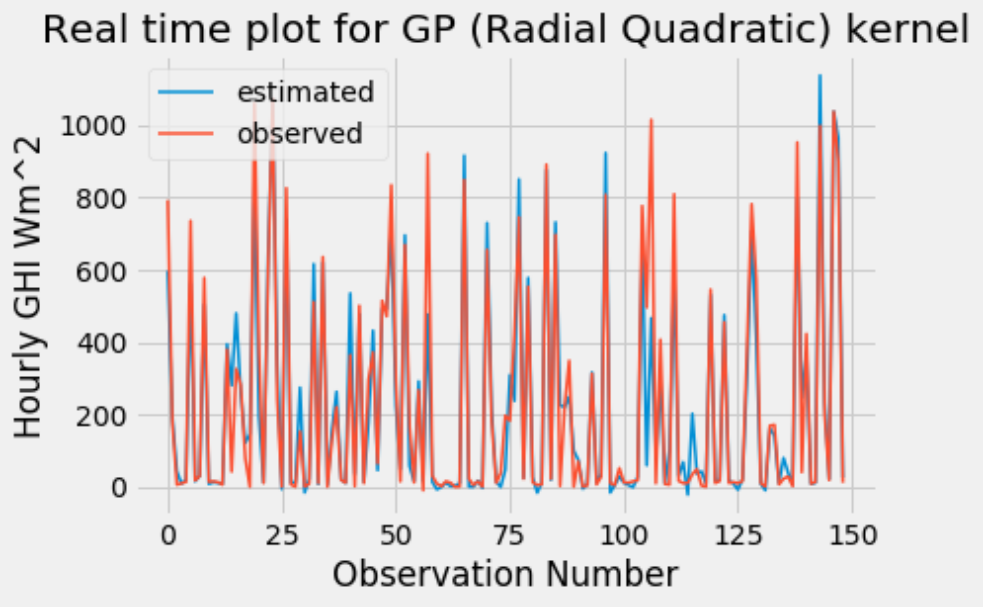

Figure 20. GPR SUN with interaction real-time plot.

3.2.3. Benchmark Models and Evaluation of Prediction Techniques

Two benchmark models, GBM and SVM, were used as a basis of comparison to the GPR model. To reduce the amount of diagnostic plots output produced, we only show the evaluation metrics of the models. All the analysis and diagnostic plots for the benchmark models are provided as Supplementary Materials.

To choose the best model, we used the MAE, RMSE and Pbias to measure the accuracy of models to make comparisons between the benchmark models and the GPR model. The idea is to come up with a model that minimizes these three measures. A summary of the results is shown in Table 5. For both stations, GPR produced the lowest values of MAE and RMSE.

The models considered are:

Main Models

M1—gpr-VEN no interaction

M2—gpr-VEN with interaction

M3—gpr-SUN no interaction

M4—gpr-SUN with interaction

Benchmark Models

M5—gbm-VEN no interaction
M6—gbm-VEN with interaction
M7—gbm-SUN no interaction
M8—gbm-SUN with interaction
M9—svr-VEN no interaction
M10—svr-VEN with interaction
M11—svr-SUN no interaction
M12—svr-SUN with interaction
where
gpr—Gaussian process regression
gbm—gradient boosting method
svr—support vector regression

Table 5. Evaluation metrics.

Model	RMSE	MAE	Pbias
M1	34.1	20.9	0.2
M2	148.4	102.8	1.3
M3	2.7	2.1	0.2
M4	122.5	64.9	43.4
M5	168.5	126.6	7.1
M6	142.7	103.8	0.6
M7	173.9	116.9	11.9
M8	164.7	110.5	10.6
M9	173.1	126.6	9.5
M10	130.5	91.1	1.2
M11	199.2	112.9	21.6
M12	185.1	104.0	18.4

4. Discussion

The present study was motivated by previous research by other authors such as Zhandire [2], Mpfumali et al. [3], Govender et al. [4] and Mutavhatsindi et al. [5]) Marizt [14], Bonilla [16], Dahl and Bonilla [17], among others, and the proposed method was developed. A new approach to solar power forecasting was done and the Gaussian process regression approach was used based on core vector regression. It produced better forecasts as a result of the best kernel which was selected.

The results produced from this study were from the application of GPR coupled with core vector regression to produce real-time results. Lasso was performed on the data incorporating interactions to select variables for both UNV and SUN data. The application of the GPR involved the computation of a kernel, which was done by choosing an appropriate kernel from the four which were used. To improve forecasting results the MEB was adopted to come up with an appropriate kernel that would yield the best results. To achieve this, different kernels were used, that is: radial basis function, Matern, rational quadratic and dot product. Empirical results showed the radial basis function to be the best kernel since it had the smallest radius in all cases considered.

GPR models were developed using the radial basis kernel. A comparison of GHI forecasts results of benchmark models SVM and GBM with the GPR was done and the GPR had the lowest mean absolute error and root mean square error for both radiometric stations. The mean absolute error for GPR on UNV data with interactions recorded the minimum values. The results showed that for both stations the radial basis kernel outperformed the other kernels producing the minimum enclosed ball. The decision was arrived at based on the radius of each kernel.

Real-time analysis was done and dominant cycles were identified and the results show that for every hour the period was 15, this means that it took 15 min to come up with a cycle. The incorporation of real-time analysis enabled the analysis to reflect the cycle of

15 over 1 h, so the analysis was done over 10 h showing a cycle of 150. This improved forecasting accuracy in the sense that analysis was done over a small interval of 10 h. It was observed when data were analysed that including interaction effects did not improve the results thus we discarded the inclusion of interaction effects.

The modelling framework proposed in this study is important to system operators and decision-makers in power utility companies who require accurate forecasts of intermittent solar power which has to be integrated on the grid in balancing demand and supply of electricity in an effective way which is environmentally secure and also favours future economic prosperity of a country. The GPR models developed in this study are robust and as such can easily be adapted to forecasting different datasets such as wind speed, electricity demand, among others. The developed models can be used by system operators in short-term solar power forecasting.

A limitation of the Gaussian process method is that despite its ability to compute the covariance functions, there is a lack of information on the uncertainty of the forecasts. The mean function is computed in this case using Bayesian regression but is not easily interpreted for Gaussian process models.

5. Conclusions

It is a crucial and hard task to forecast short-term solar power which has a great impact on the control and management of the electric grid. This paper presented an application of Gaussian process regression coupled with core vector regression on two radiometric stations from South Africa. Computation of the covariance function is a key element in Gaussian process forecasting and a more accurate kernel was adopted by considering one with a minimum enclosed ball. The results showed that this improvement on Gaussian process regression produced results that were better than the benchmark models. The bulk of the discussions highlighted that the Gaussian process regression model coupled with core vector regression provides accurate and robust solar power forecasts. This study could be useful to system operators of power utility companies who have to integrate intermittent renewable energies such as solar power with electricity generated by other energy sources on the national grid.

Future research can extend the results of the present study by using additive quantile regression models for probabilistic forecasting including other probabilistic forecasting techniques such as Bayesian structural time series models. Furthermore, we think it will also be more interesting to include non-weather variables like calendar effects and denoising the data first using either wavelets or empirical mode decomposition before applying the forecasting models.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/a14060177/s1, analytic data used in the study and a detailed analysis of the benchmark models.

Author Contributions: Conceptualization, E.C. and C.S.; methodology, E.C. and C.S.; software, E.C.; validation, E.C., C.S. and A.B.; formal analysis, E.C.; investigation, E.C. and C.S.; data curation, E.C. and C.S.; writing—original draft preparation, E.C.; writing—review and editing, E.C., C.S. and A.B.; visualization, E.C. and C.S.; supervision, C.S. and A.B.; project administration, C.S. and A.B. All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are from Southern African Universities Radiometric Network (SAURAN), website (https://sauran.ac.za, accessed on 17 June 2020).

Acknowledgments: The authors acknowledge Southern African Universities Radiometric Network (SAURAN) for providing the data.

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

Abbreviations

The following abbreviations are used in this manuscript:

CVR	Core Vector Regression
GBM	Gradient Boosting Method
GHI	Global Horizontal Irradiance
GPR	Gaussian Process Regression
MAE	Mean Absolute Error
MEB	Minimum Enclosed Ball
MLE	Maximum likelihood Estimation
RBF	Radial Basis Function
RMSE	Root Mean Square Error
SAURAN	Southern African Universities Radiometric Network
SGBR	Stochastic Gradient Boosting Regression
SVM	Support Vector Machine
SVR	Support Vector Regression

References

1. Jakel, F.; Scholkopf, B.; Wichmann, F.A. A Tutorial on Kernel Methods for Categorization. *J. Math. Psychol.* **2007**, *51*, 343–358. [CrossRef]
2. Zhandire, E. Predicting clear-sky global horizontal irradiance at eight locations in South Africa using four models. *J. Energy S. Afr. Energy Res. Cent.* **2017**, *28*, 77–86. [CrossRef]
3. Mpfumali, P.; Sigauke, C.; Bere, A.; Mulaudzi, S. Day ahead hourly global horizontal irradiance forecasting: An application to South African data. *Energies* **2019**, *12*, 3569. [CrossRef]
4. Govender, P.; Brooks, M.J.; Matthews, A.P. Cluster analysis for classification and forecasting of solar irradiance in Durban, South Africa. *J. Energy S. Afr.* **2018**, *29*, 63–76. [CrossRef]
5. Mutavhatsindi, T.; Sigauke, C.; Mbuvha, R. Forecasting Hourly Global Horizontal Solar Irradiance in South Africa Using Machine Learning Models. *IEEE Access* **2020**, *8*, 198872–198885. [CrossRef]
6. Juban, R.; Ohlsson, H.; Maasoumy, M.; Poirier, L.; Kolter, J.Z. A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014. *Int. J. Forecast.* **2016**, *32*, 1094–1102. [CrossRef]
7. Bacher, P.; Madsen, H.; Nielsen, H.A. Online short-term solar power forecasting. *Sol. Energy* **2009**, *83*, 1772–1783. [CrossRef]
8. Raza, M.Q.; Nadarajah, M.; Ekanayake, C. On recent advances in PV output power forecast. *Sol. Energy* **2009**, *136*, 125–144. [CrossRef]
9. Hong, T.; Wilson, J.; Xie, J. Long term probabilistic load forecasting and normalization with hourly information. *IEEE Trans. Smart Grid* **2013**, *5*, 456–462. [CrossRef]
10. Larson, D.P.; Nonnenmacher, L.; Coimbra, C.F. Day-ahead forecasting of solar power output from photovoltaic plants in the American South west. *Renew. Energy* **2013**, *91*, 11–20. [CrossRef]
11. Trapero, J.R. Calculation of solar irradiation prediction intervals combining volatility and kernel density estimates. *J. Energy* **2016**, *114*, 266–274. [CrossRef]
12. Ranganai, E.; Sigauke, C. Capturing long-range dependence and harmonic phenomena in 24-h olar irradiance forecasting: A quantile regression robustification via forecasts combination approach. *IEEE Access* **2020**, *8*, 172204–172218. [CrossRef]
13. Amarasinghe, P.A.G.M.; Abeygunawardana, N.S.; Jayasekara, T.N.; Edirisinghe, E.A.J.P.; Abeygunawardane, S.K. Ensemble models for solar power forecasting: A weather classification approach. *AIMS Energy* **2020** *8*, 252–271. [CrossRef]
14. Maritz, J.; Lubbe, F.; Lagrange, L. A Practical Guide to Gaussian Process Regression for Energy Measurement and Verification within the Bayesian Framework. *Energies* **2018**, *11*, 935. [CrossRef]
15. Dahl, A.; Bonilla, E. Scalable Gaussian Process Models for Solar Power Forecasting. In *Data Analytics for Renewable Energy Integration: Informing the Generation and Distribution of Renewable Energy*; Woon, W., Aung, Z., Kramer, O., Madnick, S., Eds.; Springer: Cham, Switzerland, 2017. [CrossRef]
16. Tolba, H.; Dkhili, N.; Nou, J.; Eynard, J.; Thil, S.; Grieu, S. GHI forecasting using Gaussian Process regression: Kernel study. *IFAC Paper Online* **2019**, *52*, 455–460. [CrossRef]
17. Dahl, A.; Bonilla, E. Grouped Gaussian Processes for solar power prediction. *Mach. Learn.* **2019**, *108*, 1287–1306. [CrossRef]
18. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **2001**, *1*, 211–244.
19. Quinonero-Candela, J. Learning with Uncertainty: Gaussian Processes and Relevance Vector Machines. Ph.D. Thesis, Technical University of Denmark, Lyngby, Denmark, 2004.
20. Martino, L.; Read, J. A joint introduction to Gaussian Processes and Relevance Vector Machines with connections to Kalman filtering and other kernel smoothers. *Inf. Fusion* **2021**, *74*, 17–38. [CrossRef]
21. Bates, J.M.; Granger, C.W. The combination of forecasts. *J. Oper. Res. Soc.* **1969**, *20*, 451–468. [CrossRef]

22. Gershman, S.J.; Blei, D.M. A tutorial on Bayesian nonparametric models. *J. Math. Psychol.* **2012**, *56*, 1–12. [CrossRef]
23. Rasmussen, C.E.; Williams, C.K. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006; Volume 38, pp. 715–719.
24. Tsang, I.; Kwok, J.; Cheung, P.M. Core vector machines: Fast SVM training on very large data sets. *J. Mach. Learn. Res.* **2005**, *6*, 363–392.
25. Badoiu, M.; Clarkson, K. Smaller core-sets for balls. *Comput. Geom.* **2008**, *40*, 14–22. [CrossRef]
26. Yildirim, E.A. Two algorithms for the minimum enclosing ball problem. *J. Optim.* **2008**, *19*, 1368–1391. [CrossRef]
27. World Bank. Global Solar Atlas 2.0, Solar Resource Data, Solargis. Available online: https://solargis.com/maps-and-gis-data/download/south-africa (accessed on 22 January 2021).
28. Robert, T. Regression Shrinkage and Selection via Lasso. *J. R. Stat. Soc. B* **1996**, *58*, 265–288.
29. Bien, J.; Taylor, J.; Tibshirani, R. A lasso for hierarchical interactions. *Ann. Stat.* **2013**, *41*, 1111–1141. [CrossRef] [PubMed]
30. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [CrossRef]
31. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.