Robust Modelling Framework for Short-Term Forecasting of Global Horizontal Irradiance

*Edina Chandiwana, Caston Sigauke * and Alphonce Bere*

Department of Mathematical and Computational Sciences, University of Venda, Thohoyandou, Limpopo, South Africa

Abstract: The increasing demand for electricity and the need for clean energy sources have increased solar energy use. Accurate forecasts of solar energy are required for easy management of the grid. This paper compares the accuracy of two Gaussian Process Regression (GPR) models combined with Additive Quantile Regression (AQR) and Bayesian Structural Time Series (BSTS) models in the 2-day ahead forecasting of global horizontal irradiance using data from the University of Pretoria from July 2020 to August 2021. Four methods were adopted for variable selection, Lasso, Elasticnet, Boruta, and GBR (Gradient Boosting Regression). The variables selected using GBR were used because they produced the lowest MAE (Minimum Absolute Errors) value. A comparison of seven models GPR (Gaussian Process Regression), Two-layer DGPR (Two-layer Deep Gaussian Process Regression), bstslong (Bayesian Structural Time Series long), AQRA (Additive Quantile Regression Averaging), QRNN(Quantile Regression Neural Network), PLAQR(Partial Linear additive Quantile Regression), and Opera(Online Prediction by ExpRt Aggregation) was made. The evaluation metrics used to select the best model were the MAE (Mean Absolute Error) and RMSE (Root Mean Square Error). Further evaluations were done using proper scoring rules and Murphy diagrams. The best individual model was found to be the GPR. The best forecast combination was AQRA (AQR Averaging) based on MAE. However, based on RMSE, GPNN was the best forecast combination method. Companies such as Eskom could use the methods adopted in this study to control and manage the power grid. The results will promote economic development and sustainability of energy resources.

Keywords: Additive quantile regression; Bayesian structural time series; Forecast combination; Gaussian processes; Solar irradiance.

1 Introduction

South Africa has suffered from an energy crisis due to climatic changes over the past years, which has resulted in power companies like Eskom resorting to alternative energy sources like solar power. Solar power has been preferred as an alternative energy source that is inexhaustible, sustainable, highly abundant, cheap and does not pollute the environment. This has led to solar power, also called Global Horizontal Irradiance (GHI) forecasting, to become an important aspect of the energy management system since solar power generation is directly linked to the management of the power grid Yang et al. [1]. The United Nations, Sustainable Development Goal (SDG) number 7, on affordable and clean energy, UNDP [2], encourages the use of renewable energies to mitigate climatic changes, which makes the prediction of GHI vital. The goal is to promote using renewable energy sources such as solar and wind because they are affordable, reliable and adequate.

Various methods have been proposed to predict GHI generation at different scales. The methods explored are statistical, machining learning, and numerical weather prediction models. The current study uses a 2-days-ahead probabilistic modelling framework to predict short-term solar power generation. A comparison is made between Bayesian Structural Time Series (BSTS), Deep Gaussian Regression (DGP), Gradient Boosting Regression (GBR) and the Gaussian Process Regression coupled with Quantile Regression.

Several authors have researched applying Gaussian regression, [3,4,5,6]. Billionis et al. [3] proposed a recursive Gaussian Process approach reduces the input space of satellite-based observations to perform iterated predictions. They first applied factor analysis for dimensionality reduction to develop two maps, one for reconstruction and the other for reduction. Their results proved that the proposed method performed worse than the ground-based model. Tolba et al. [4]

considered the prediction of GHI using Gaussian Process Regression(GPR). They applied GPR to forecast horizontal data, ranging from 30 minutes to 5 hours. Their main thrust was on the selection of the kernel functions. They used simple and complex kernels and compared their results to the persistent model. Their results showed that when the quasiperiodic kernel was applied, the GPR model outperformed the other GPR models and the persistent model. One of the few studies that looked at GPR was done by Wang et al. [5]. They proposed a hybrid model combining LSTM (Long Short-Term Memory) and Gaussian Process Regression. The method is compared to neural networks. The results showed that the proposed method produced superior results.

Zhang et al. [6] predicted solar power using k means time series and Gaussian Process Regression. They clustered solar power radiation using different feature input categories. Different covariance functions of GPR were applied, and a comparative analysis was done. Kernel functions were selected using deterministic and probabilistic categories. The results of the proposed GPR proved to be superior to the Artificial Neural Network benchmark model used. Yu et al. [7] developed a method of extracting trajectories called the Gaussian process factor analysis. They looked at spike trains that were first smoothed using a smoothing technique that accounted for spiking variability. Stonski [8] did a comparative analysis of feedforward layered neural networks on stochastic and Gaussian Process analyses. The results showed that the two had similar forecasting accuracy and better performance than the linear regression model. Al-Shedivat et al. [9] applied long, short-term memory recurrent networks coupled with GPs. Using expressive closed-form kernels, they used marginal likelihood estimation on a convergent semi-stochastic gradient method and exploited how kernels should be structured for forecasting. The models were applied to self-driven car applications, system identification, and power forecasting, and the results showed that the proposed method was efficient and convergent as required. Tsymbalov et al. [10] applied a dropout-based model based on Bayesian Regression and Neural network coupled with Gaussian process on a chemical and physical real-life dataset. The results showed that the methods performed well for neural architectures involving dropout and Bayesian neural networks.

Chandiwana et al. [11] did a study on predicting global horizontal irradiance (GHI) using Gaussian process regression. They predicted GHI based on GPR coupled with core vector regression. The performance of this method was compared with that of two benchmark models: support vector regression (SVR) and gradient boosting regression (GBR). The proposed methodology performed better than the benchmark models.

Over the past few years, researchers have been exploring the use of machine learning techniques which produce accurate forecasts. Most GPR methods highlighted in the literature that has looked at predicting solar power have proved to produce accurate predictions, but these models need to be improved. However, predicting solar power is difficult because of the uncertainty and variable nature of the datasets. It is, therefore, a critical aspect of forecasting to produce reliable and accurate predictions. To our knowledge, little has been done on Gaussian Process Regression on South African data. Hence there is a need to explore this area. The technique used in this research, GPR coupled with quantile regression, addresses that aspect of dealing with uncertainty, and the combination of forecasts was done to improve forecasting accuracy. GPR was combined with additive QR (AQR), quantile regression neural network (QRNN), and partial linear additive quantile regression (PLAQR) models together with the use of a convex combination method.

This research aims to develop a forecasting model that accurately predicts 2-day-ahead solar power. We compared two models, BSTS and GPR, to select the best among these two models. Finding solutions to uncertainty and variability is crucial in energy forecasting. We propose the GPR to solve this problem, combining it with quantile regression. Prediction models are being developed daily, but an inseparable part needs attention: error reduction and accurately capturing uncertainty. The bigger the error is, the poor the prediction outcomes will be. Thus it is crucial to reduce errors to improve the accuracy of results. We expect to develop an improved modelling framework for GHI by adopting the combined hybrid models, thus expecting a reduction of errors and, at the same time dealing with variability and uncertainty.

1.1 Research highlights

The first contribution of this research is the coupling of the Gaussian process with QR to develop a modelling framework for medium-term forecasting of GHI. This approach provides estimates catering to the dataset's uncertainty and scalability, hence its strong performance.

Secondly, we applied a modelling framework based on GHI data's Bayesian Structural time series. This model's performance is superior because it incorporates other variables rather than historical predictors and avoids over-fitting. BSTS avoids over-fitting and captures the correlation between many state components and multiple time series. This model is a combination of time series and Bayesian techniques.

Lastly, the combination of forecasts was done using GPNN, PLAQR, QRA and Opera. These were combined with BSTSlong and GPR. In their research, Bates and Granger [12] concluded that combining forecasts reduces errors. The rest of the paper is organized as follows: The first section describes the methods, followed by the section on empirical data analysis, a section representing the discussions and finally, the conclusion.

2 Materials and methods

2.1 Schematic presentation of methodology

Fig 1 shows the schematic presentation of the adopted methodology, with all the stages laid out step by step. The data is split into training and testing sets. The training data set is used in training the models to determine the most appropriate fitting model. This is followed by using the trained models to produce two-day ahead predictions. Evaluation metrics are then used to determine the individual model with the highest predictive capability. The forecasts produced by the unique models are combined using four robust forecast combination methods. Proper scoring rules and Murphy diagrams are then used to determine the forecast combination method that produces robust medium-term forecasts of GHI.
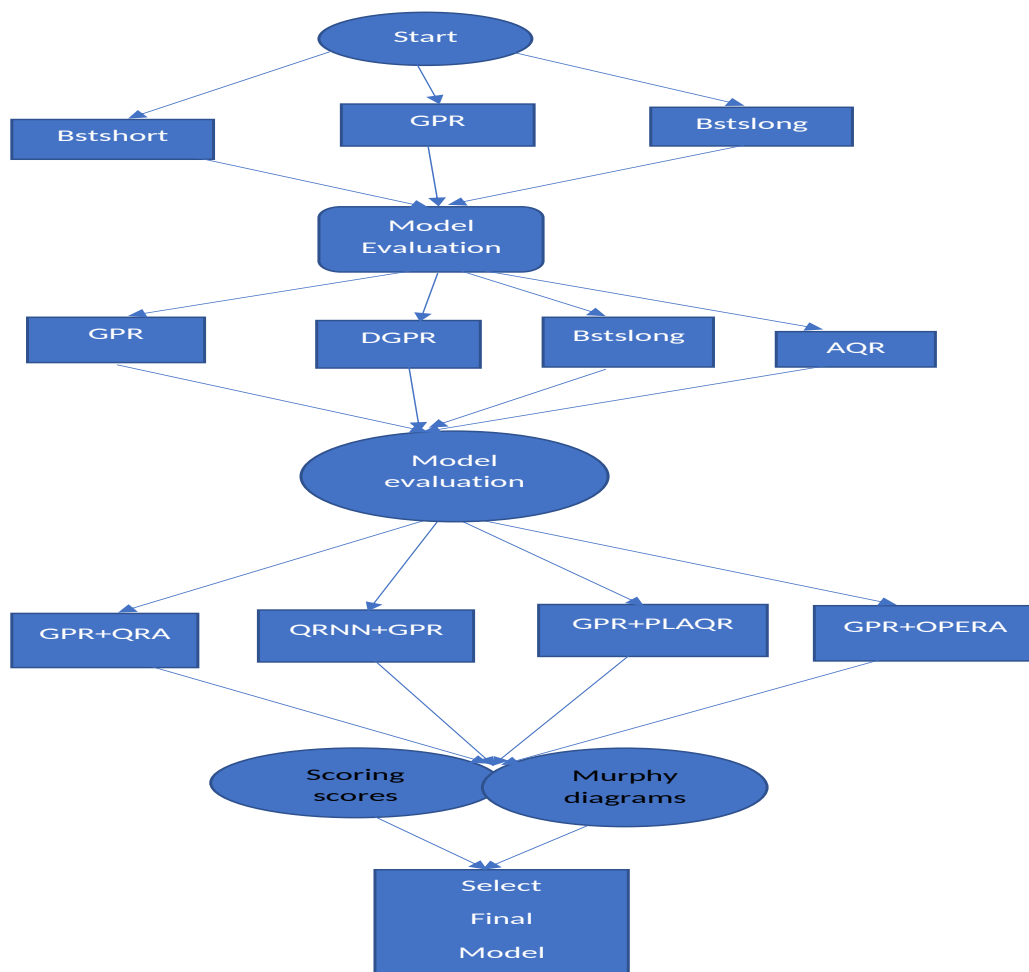

Fig. 1: Methodology flow chart.

2.2 Models

2.2.1 Gaussian process

A Gaussian process (GP) was used in this research. It is a nonparametric probabilistic stochastic machine learning technique based on multivariate normal distribution kernels. A GP is based on a function with a continuous domain and defined by its mean $m(x)$ and covariance $k(x_i, x_j)$ where x_i and x_j are two points that influence each other. If we have a

GP described by a function defined by $f(x(t)), x(t) \in \mathbb{R}^d$, where $x(t)$ are the weather variables like air temperature, wind speed and relative humidity, indexed over time t, $f(x)$ is the response of these weather variables into a d dimensional space. GP has a mean given in equation (1),

$$E(f(x)) = m(x) \tag{1}$$

and covariance function,

$$k(x_i, x_j) = E[\{f(x_i) - m(x_i)\}\{f(x_j^1) - m(x_j^1)\}]. \tag{2}$$

Therefore, the function describing the GP of GHI is given in equation (3).

$$g(x) \sim GP(m(x), k(x, x^1)) \tag{3}$$

The method is named after the famous Carl Fredrich Gauss because it is based on Gaussian distribution, commonly known as the normal distribution.

Gaussian process regression (GPR) is a powerful tool used for regression and classification, and it performs better than other forecasting methods since it predicts uncertainty directly, giving reliable estimates. The forecasts from GPR are reliable because GP uses prior information, and different specifications of the covariance function are used.

2.2.2 Covariance functions

A GPR is defined by a mean and covariance function. A covariance function is a key component of the GPR model. There are various covariance functions: Matern, Linear, periodic, Radial basis function, Polynomial, Bessel including Spline. This study uses the radial basis function (RBF). The RBF is given in equation (4)

$$RBF = \sigma^2 \exp\left(\frac{(x_i - x_j)^2}{2l^2}\right) + \sigma_{ij}\sigma_{\text{noise}}^2, \tag{4}$$

where l is the length parameter, σ^2 is the variance and σ_{noise}^2 is the variance of the noise term.

2.3 Deep Gaussian process regression

The deep Gaussian process (DGP) is defined in Saucer et al. [13] as a hierarchical layering of GPs. Each layer is assumed to follow a multivariate normal distribution (MVN).

Let X_n represent $n \times d$ inputs with $Y_n = f(X_n)$ denoting the functional evaluations. That is $X_n \to Y_n$, $f : \mathbb{R}^d \to \mathbb{R}$ and $Y_n \sim N_n(\mu, \Sigma)$. See Saucer [13] for details. For the inputs, X_n, to reach the response, Y_n has to pass through several intermediate GPs. This study will consider the two-layer model presented in Figure (2).
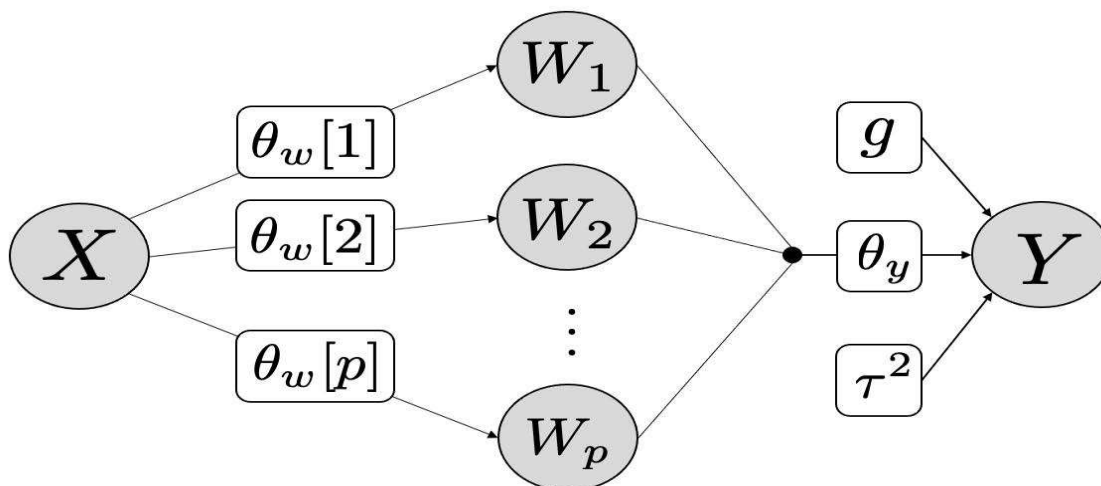

Fig. 2: Two-layer deep GP model with p latent nodes. Source: Saucer [13].

In Figure (2), W denotes the hidden GPs (latent nodes) and the structure is described hierarchically as Saucer [13]:

$$Y_n | W \sim N_n(0, \Sigma(W))$$
$$W \sim N_n(0, \Sigma(X_n))$$

The present study will consider two layers only since increasing the number of layers will not significantly improve the results. However, it will only result in high computational costs ([13, 14]). Estimating the parameters is through a hybrid Gibbs-ESS-Metropolis algorithm discussed in detail in Saucer, [13].

2.4 Bayesian structural time series

Bayesian structural time series (BSTS) was also used in this research. It is a modelling approach classified as the machine learning technique used in time series for forecasting ([15, 16]). BSTS is constructed as a state-space technique, and there are two pieces to the model: the observation equation in equation (5).

$$Y_{it} = \beta^T x_{it} + Z_t^T \alpha_t + \varepsilon_{it} \tag{5}$$

Furthermore, a transition equation defines how the latent states evolve.

$$\alpha_{a+1} = T_t \alpha_t + R_t \eta_t, \tag{6}$$

where Y_{it} is GHI, t represents time, x_{it} are the meteorological variables at time t, $\eta_t \sim N(0, \sigma_\eta^2)$, α_t, Z_t, T_t and R_t are structural parameters and $\beta^T X_{it}$ is the regression component.

BSTS is a technique that introduces a regression aspect developed from the structural model's independent variables. It deals with many independent variables by coming up with latent factors. It induces sparsity on the coefficients, putting a spike and slab prior to distribution. The model uses three components, spike and slab, Kalman filter and Bayesian model averaging.

2.4.1 Spike and slab

Spike and slab is a variable selection technique that uses the Bayesian technique. The idea was first developed by Mitchel [17]. It was further developed by Madigan [18]. This Bayesian variable selection approach is made by mixing priors using a spike and slab for the effects subject to selection. Variable selection is a major step in regression analysis. The model parameters of the regression model $p(\mu, \sigma^2, \alpha)$ are specified with priors having a structure in equation (7).

$$p(\mu, \sigma^2, \alpha) = p(\mu, \sigma^2) . p(\alpha \mid \sigma^2, \mu) \tag{7}$$

with the probability of mean and error variance being $p(\mu, \sigma^2) = \frac{1}{\sigma^2}$. The spike and slab priors are given in equation (8).

$$p(\alpha \mid \delta) = p_{slab}(\alpha_j) . \prod_{j : \delta_j = 0} p_{spike} \alpha_j, \tag{8}$$

where δ_j represents an indicator variable and α_j is the regression effect.

2.4.2 Kalman filter

The time series feature is solved using a Kalman filter ([19, 20]). A Kalman filter can be defined as an equation that gives a recursive solution to the least-squares method. It is an optimal iterative estimation algorithm whose objective is to minimize error. A Kalman filter is given in equation (9)

$$K = P_j C^T (C P_j C^T + R_t)^{-1} \tag{9}$$

Equation (10) is called the Kalman gain.

$$P_{j+1} = P_j - K C P_j, \tag{10}$$

where P is the state error, C is the observational model and R is the noise covariance matrix.

2.4.3 Bayesian model averaging

The Bayesian model averaging (BMA), developed by [21], is a parameter estimation method obtained from averages of the predictions. BMA is a method that analysts prefer because it reduces underestimation of uncertainty, produces optimal forecasts under several loss functions, and is robust to model misspecification. The average posterior distribution of each model is given in equation (11).

$$p(GHI \mid D) = \sum_{j=1}^{J} p(GHI \mid A_J, D) \times P(A_j \mid D), \tag{11}$$

where $A_1, ..., A_J$ are the models considered. The posterior probability of the model A_J is given in equation (12)

$$p(A_J \mid D) = \frac{p(D \mid A_l) \times p(A_J)}{\sum_{l=1}^{J} p(D \mid A_J) \times P(A_l)}, \tag{12}$$

where the likelihood of A_j is given in equation (13)

$$p(D \mid A_J) = \int p(D \mid \theta_j, M_j) \times p(\theta_j \mid M_j) d\theta_j, \tag{13}$$

where θ_j is a vector parameter of model A_J.

2.4.4 Gradient boosting regression

Gradient Boosting regression is one of the Benchmark models used in this research. It is a machine learning method used in regression analysis and classification tasks. It is used to build forecasting models by using decision trees to generate negative gradients using samples and introducing a weak learner to relate to the available weak learners. The function $F(x)$, which approximates the best values of the predicted output variables, is given by

$$\hat{F} = arg_F \ minE_{x,y} [L(y, F(x))] \tag{14}$$

where $L(y, F(x))$ is the loss function. GBR aims at minimizing the loss function in (14).

$$L(y, F(x)) = (Y, F(x))^2 \tag{15}$$

we minimize the function in equation (14) by adjusting $F(x_1), F(x_2), F(x_3), ..., F(x_n)$.

Table 1 summarises some of the advantages and disadvantages of the proposed models: GPR, two-layer DGPR and BSTS.

Table 1: Model comparisons.

Models	Strengths	Weaknesses
GPR (M1)	1. It is robust and versatile. 2. It accurately captures model uncertainty. 3. It offers a probabilistic approach convenient for solving stochastic problems. 4. Allows addition of prior knowledge and specification. 5. Allows scalability of large datasets. 6. Tune hyper-parameters by maximizing the marginal likelihood.	1. It is computationally expensive. 2. They are not sparse.
BSTS (M2)	1. Handles uncertainty well. 2. It can handle a large number of variables. 3. Can predict stochastic patterns of the time series accurately. 4. The parameters of the model change over time. 5. Uses prior information on the variables.	1. Computational analysis is extremely difficult for the posterior distribution.
Two-layer DGP (M3)	1. Nonstationary flexibility. 2. Ability to cope with abrupt regime changes in training data. 3. Nonlinear predictive capability.	1. Computational cost is high. 2. MLE point estimtes lead to overfitting, treating noise as a signal.
GBR(M4)	1. It is flexible and sensitive to outliers 2. No data pre-processing required 3. Conversion of weak learners into strong learners	1. The model may cause overfitting

2.5 Forecasts combination

2.5.1 Additive quantile regression model

The study used the quantile generalized additive model (quantGAM) based on the work of [22] and extended by [23], defined as:

$$y_{t,\tau} = \sum_{j=1}^{p} s_{j,\tau}(x_{tj}) + \varepsilon_{t,\tau}; \;\; \tau \in (0,1). \tag{16}$$

The smoothing function, s, is written as:

$$s_{j\tau}(x) = \sum_{k=1}^{q} \beta_{kj} b_{kj}(x_{tj}), \tag{17}$$

where β_{kj} denotes the jth parameter and $b_{kj}(x)$ represents the jth basis function with the dimension of the basis being denoted by q. The parameter estimates of equation (16) are obtained by minimizing the function given in equation (18) as:

$$q_{Y|X}(\tau) = \sum_{t=1}^{n} \rho_{\tau} \left(y_{t,\tau} - \sum_{j=1}^{p} s_{j,\tau}(x_{tj}) \right), \tag{18}$$

where $q_{Y|X}(\tau)$ is the conditional quantile function of $\tau \in (0,1)$ and $\rho_{\tau}(u) = u[\tau - \mathbf{I}(u < 0)]$ is the pinball function. In this study, we are interested in estimating extreme conditional quantiles, i.e., in combining forecasts, we shall call it quantile regression averaging (QRA).

2.5.2 Quantile regression neural network

A Quantile regression neural network(QRNN) is a model composed of quantile regression coupled with neural networks. A QRNN is given in equation (19)

$$
\begin{aligned}
\min_{v,w} \sum \theta |y_t - f(x_t - v - w) \\
+ \sum (1-\theta)|y_t - f(x_t - v - w) \\
+ \lambda_1 \sum w_{ji}^2 + \lambda^2 \sum_i v_i^2,
\end{aligned}
\tag{19}
$$

where λ_1 and λ_2 are regularisation parameters and

$$
f(x_t - v - w) = g_2 \left(\sum_{j=0}^{m} v_j g_1 \left(\sum_{i=0}^{n} W_{ji} x_{it} \right) \right).
\tag{20}
$$

Equation (20) is the neural network model where $g_1(.)$ and $g_2(.)$ are activation functions, which are sigmoid and linear, respectively. The weight parameters are v and w, m is the number of hidden layers, and n is the number of inputs. Theta (θ) is the θth Quantile, and x_{it} are independent variables which are the weather variables, and y_t is the GHI.

2.5.3 Partially linear additive quantile regression

The two techniques, the Generalized Additive Model (GAM) and QR(Quantile Regression) models are combined to come up with the partially linear additive quantile regression (PLAQR). A GAM is a generalized linear model where the dependent variable GHI depends on a smooth function of the independent variable. GAMS was developed to come up with a combination of the properties of the generalized linear models with additive models. The GAM model is given in equation (21)

$$
g(E(Y)) = \beta_0 + s_1(x_1) + s_2(x_2) + ... + s_m(x_m),
\tag{21}
$$

where Y represents GHI which is the response variable and, x_i denotes the covariates, s_j are functions of the independent variable. These functions might differ; for example, $s_1(x_1)$ might be a factor model, $s_2(x_2)$ a weighted mean, etc.

2.5.4 Online prediction by expert aggregation

Lastly, Online Prediction by ExpRt Aggregation (OPERA) was one of the methods used for forecasts combination and was developed by [22]. Given a set of observed values for a particular variable Y, GHI, with a sequence of values $y_1, y_2, ..., y_n$, are the predicted values. For a given time step $t = 1, 2, ..., n$, there are predictions from independent variables, which are the weather variables given by $x_{k,t}$, where k is a set of finite methods $k = 1, ..., K$ combines a multiple of algorithms of online learning, and they predict forecasts. The OPERA forecasts are given by equation (22).

$$
\hat{y}_t = s \sum_{k=1}^{K} P_{k,t} x_{k,t}
\tag{22}
$$

Table 2 summarises some of the advantages and disadvantages of this paper's four forecast combination models.

Table 2: Model comparisons of forecast combination models.

Models	Strengths	Weaknesses
AQR	1. Estimates are more robust to non-normal errors. 2. Estimates are robust against outliers. 3. Characterization of the data is richer. 4. It is invariant to monotonic transformations.	1. Parameters are harder to estimate than other regression techniques.
PLAQR	1. It is flexible to model data that is complex. 2. It is more parsimonious than other regression models. 3. Very stable and flexible for large-scale interpolation.	1. Challenge of the existence of non-smoothness of the loss function.
QRNN	1. It is flexible to model data on modelling dependencies. 2. Scalability is relatively good with large datasets.	1. Not very well developed for tasks that are not classified. 2. Sometimes gives overconfidence in predictions when applying training samples.
OPERA	1. Combines multiple individual techniques to build a more powerful technique. 2. Better prediction. 3. Improved results as compared to individual models.	1. Finding feature importance is a challenge.

2.6 Data and variables

This research made use of hourly data from SAURAN (Southern African Universities Radiometric Network) website https://sauran.ac.za/). The data are from a radiometric station based at the University of Pretoria, South Africa, with Latitude:-25.75308 ○ (E), Longitude: 28.22859 *circ* (S) and Elevation: 1410m and for the period July 2020 to August 2021, giving us a total of 8522 observations. The data was split into two, the training set and the test set. The ratio used for the training was 80:20.

The response variable is GHI(Global Horizontal Irradiance) and the explanatory variables with their specific abbreviations and measuring units are given in Table 3.

Table 3: Independent variables: UPR Station.

Name	Description	Measuring units
Air Temperature	Temp	^oC
Relative Humidity	RH	$\%$
Wind Speed	WS	m/s
Barometer Pressure	BP	mbar
Wind Direction	WD	^{o}
Wind Direction Standard Deviation	WD_Stv	^{o}
Rain Total	Rain_Tot	mm
Maximum wind speed	WS_Max	m/s

The SAURAN UPR station is at the University of Pretoria, South Africa. The pyranometer is on top of a science building, giving good solar exposure. Figure 3 shows the position of weather instruments at the top of a building.

Fig. 3: UPR Weather instruments. Source https://sauran.ac.za/.

2.7 Evaluation metrics

The models' performances were assessed using RMSE (Root Mean Square Error) and MAE (Mean Absolute Error). Further evaluations used probability evaluation metrics, proper scoring rules, and Murphy diagrams.

2.7.1 Mean absolute error

The mean absolute error (MAE) measures the average of all magnitude of errors. MAE is a measure of the accuracy of a given model. Absolute error is a measure of the amount of error in a model. It is defined in equation (23).

$$\text{MAE} = \frac{1}{m} \sum_{j=1}^{m} |y_i - \hat{y}_j| \tag{23}$$

2.7.2 Root mean square error

Root Mean Square Error(RMSE) measures a quadratic scoring rule for the average magnitude of errors. It is defined in equation (24)

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum (y_i - \hat{y}_j)^2}, \tag{24}$$

where y_i are the observed values, $\hat{y}_j$ are the predicted values and m denotes the number of observations in the testing set.

2.8 Proper scoring rules

The proper scoring rules that were used in this research are logarithmic score (LogS), continuous ranked probability score (CRPS), and Dawid-Sebastian score (DSS). 'ScoringRules' R package developed by Jordan, [?] is used to compare

and evaluate the methods used. A scoring rule assigns a penalty score, given by a function $S(F,y)$, used to evaluate a distribution F over outcomes $y \in \mathbb{R}$ observed values. The scoring rules were used to assess the models. The lower the score, the better the forecasting accuracy.

2.8.1 Continuous ranked probability Score

The continuous ranked probability score (CRPS) evaluates probabilistic systems. It generalizes the MAE to a probabilistic scenario. The score is calculated using equation (25).

$$\text{CPRS}(F,x) = \int_{-\infty}^{\infty} (F(y) - \mathbf{1}(y-x))^2 \, dy, \tag{25}$$

where $\mathbf{1}(.)$ is the indicator function, it is one of the functions that is positive or zero and 0; otherwise, $F(y)$ is the cumulative distribution function of X, which are weather variables, y is the response variable and x are the independent variables.

2.8.2 Logarithmic score

It is a measurement that assesses how well the given variables perform on a certain data set. The score is given in equation (26).

$$\text{LogS}(F,y) = -\log f_y, \tag{26}$$

where f_y is the probability function and y are the observed values.

2.8.3 Dawid-Sebastian score

It is a measure for evaluating the accuracy of multivariate forecasts. The score is given in equation (27)

$$\text{DSS}(F,y) = \left(\frac{y - \mu_F}{\sigma_F}\right)^2 + 2\log\sigma_F, \tag{27}$$

where $y - \mu_F$ is the square error score, μ_F and σ_F are the mean and variance, respectively.

2.8.4 Pinball Loss

The pinball loss function measures the accuracy of a quantile forecast. It is given by equation (28).

$$\text{PL}(q\tau,t) = \begin{cases} 2(1-\tau)|y_t - q_{\tau,t}| & if \quad y_t < q_{\tau,t} \\ 2\tau|y_t - q_{\tau,t} & if \quad y_t \geq q_{\tau,t} \end{cases}, \tag{28}$$

where y is the observed values, q-quantile forecasts at the τth-is quantile.

3 Results

3.1 Exploratory data analysis

The top panel of Figure 4 shows the hourly solar irradiance data distribution. It shows that the sunlight hours are from 07:00 to 20:00. The largest amount of GHI is harvested between 11:00 and 16:00. The bottom panel, Figure 4, shows a plot of the joint distribution of GHI and temperature. As temperature increases, GHI also increases and seems to reach a peak around $29^o C$ then drops.
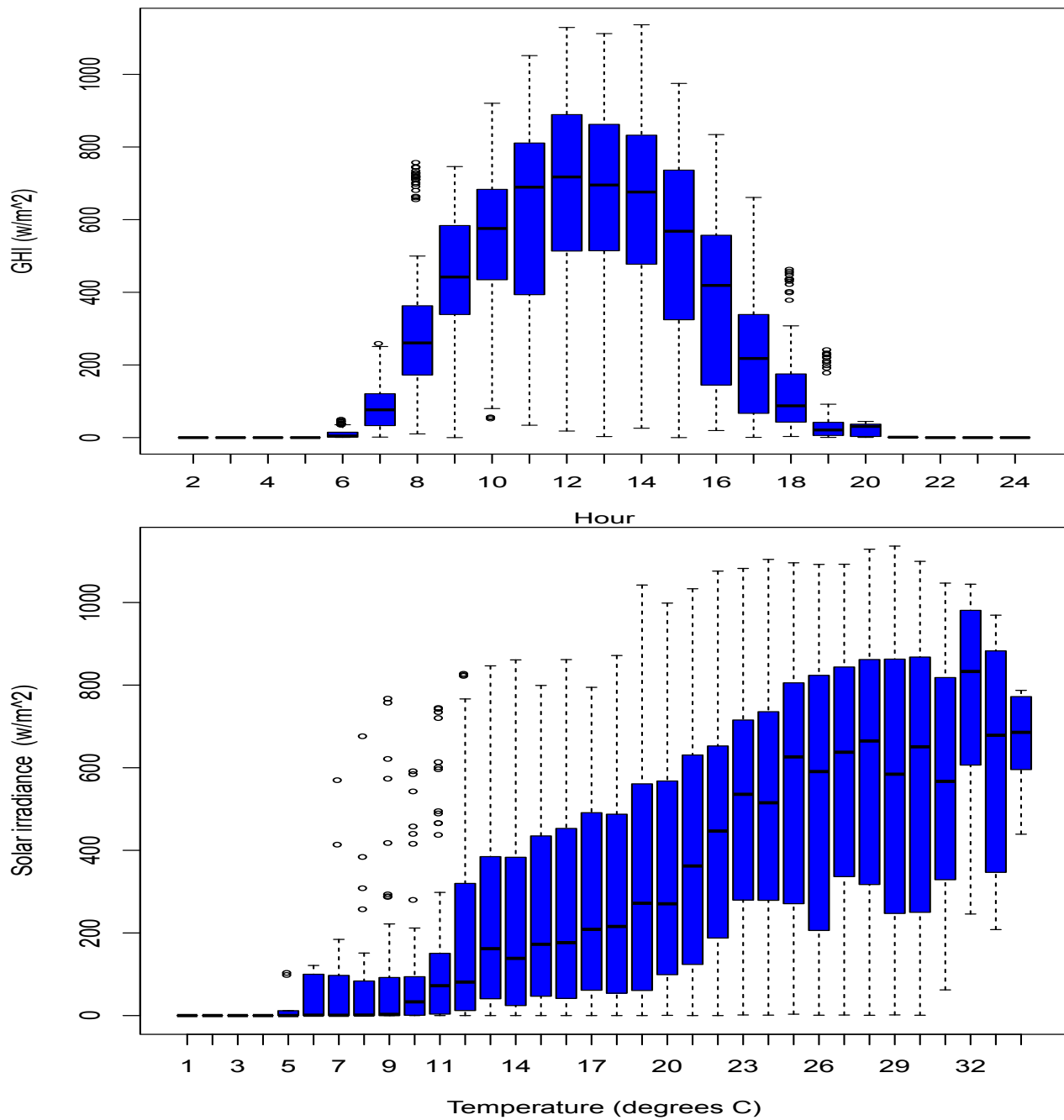
Fig. 4: Top panel: Distribution of hourly GHI. **bottom panel:** Joint distribution of GHI and temperature.

Scatter plots of the covariates against the response variable GHI are given in Figure 5. The plots show the GHI relationship as each covariate increases in size.
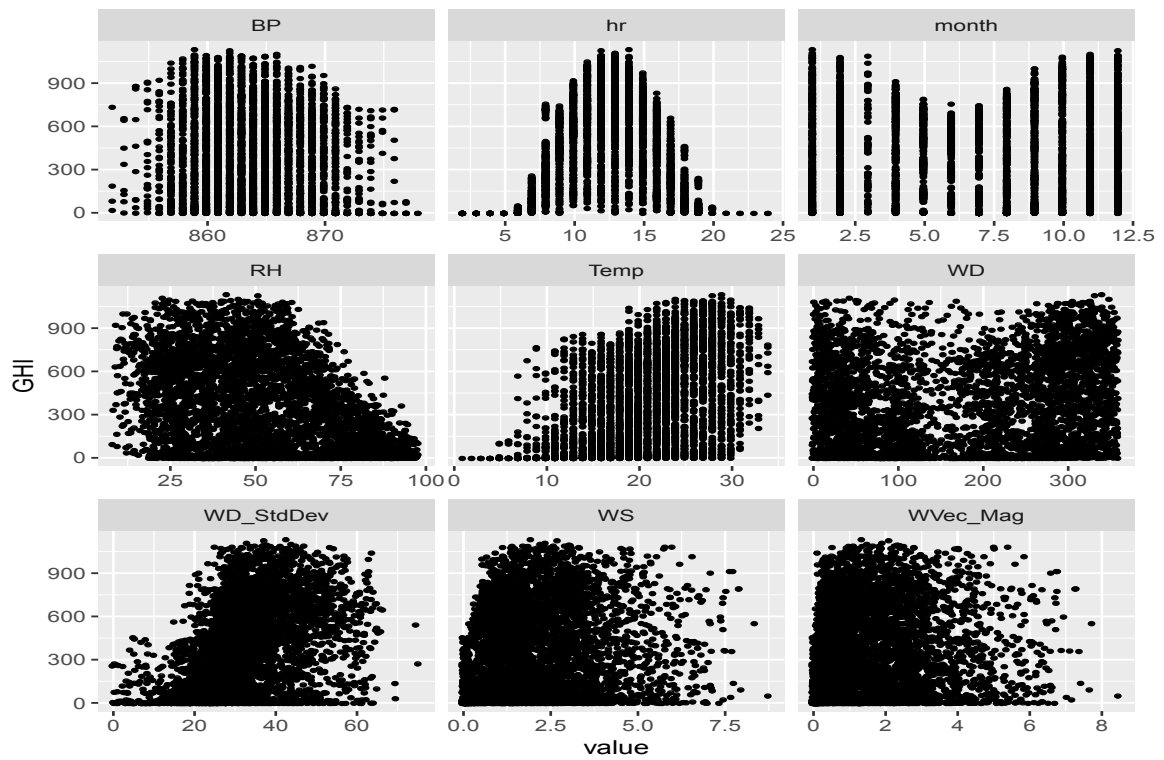

Fig. 5: Scatter plots of covariates against GHI.

3.2 Variable Selection

Variable selection is one of the most important challenges in statistical analysis. We considered five modern variable selection techniques: Lasso, Elasticnet, Boruta, GBM (Gradient Boosting Method) and MARS (multivariate adaptive regression splines). GBM (Gradient Boosting Method) was selected for variable selection because it had the lowest MAE compared to the other methods.

A training dataset was used for variable selection, and then Linear Quantile Regression (LQR) was applied to the different methods. The evaluation results from the LQR models concerning the variable selection method are shown in Table 4.

Table 4: Variable Selection.

Method	MAE
Lasso	100.48
Elasticnet	105.43
MARS	111.54
BORUTA	100.55
GBR	100.36

The models were then used for two-day ahead forecasting of GHI. The two-day ahead forecasts superimposed with observed GHI are given in Figure 6. The plots show a smooth fit of the predicted in red and GHI in black.
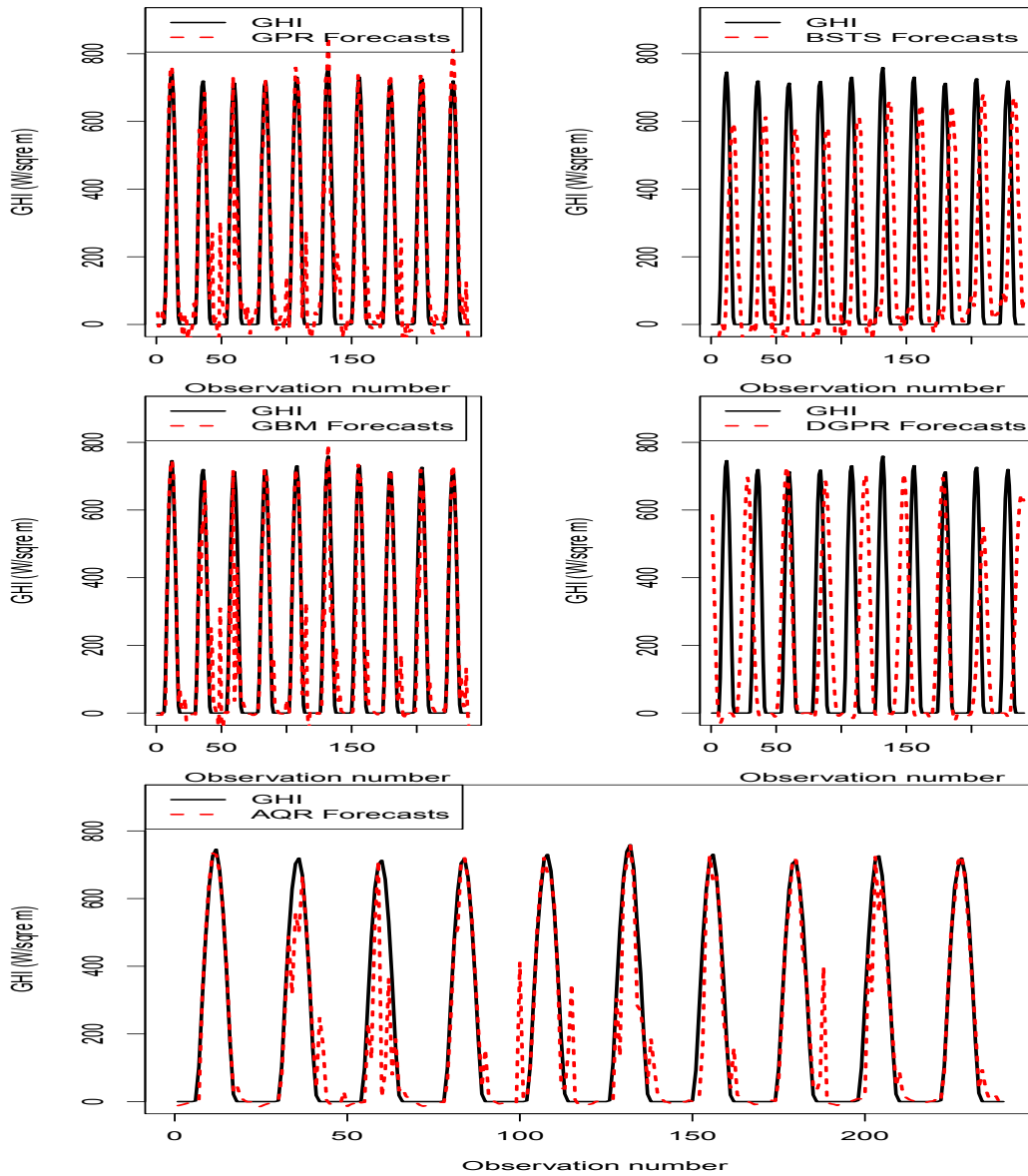
Fig. 6: GHI vs predicted using the developed models.

A comparative analysis of the GPR model was done with that of BSTS short and long prediction, GBM, AQR and Two-layer DGPR models and based on MAE and RSME as the evaluation metrics. The comparison results are shown in Table 5. The GPR model proved to be more efficient than the other models since it had low values of RMSE and MAE. We decided that the BSTS_short and Two-layer DGPR were inappropriate models since they produced very high MAE and RMSE. The next section is on a combination of forecasts, where we will combine the GPR with other quantile regression models and the OPERA model.

Table 5: Comparison of the models.

Model	RMSE	MAE
GPR	75.92	45.32
BSTS_short	3010.72	2533.07
BSTS_long	257.482	196.14
GBM	78.37	36.26
AQR	92.83	37.49
Two-layer DGPR	360.74	256.55

3.2.1 Combining forecasts

As a way of improving the accuracy of forecasts, forecasts were combined. Bates and Granger [12] found that combining forecasts enhances the accuracy of forecasts. We used forecasts from the following models: quantile regression averaging (QRA), QRNN, PLAQR and OPERA, a convex combination method. The first method, linear Quantile Regression, was combined with bstslong and GPR. We are going to use fQRA to represent the produced forecasts. The second method used is Quantile regression neural network combined with bstslong and GPR. The combined forecasts will be referred to as the QRNN model. The third method is Partially linear Additive Quantile Regression combined with bstslong and GPR. The combined forecasts will be referred to as PLAQR. The fourth method uses the convex combination using the OPERA R package. The combined forecasts will be referred to as OPERA. Table 6 shows the results for forecasting the accuracy of the four methods based on two evaluation metrics.

Table 6: Combined forecasts.

Model	RSME	MAE
QRA	71.92	26.59
QRNN	68.33	29.88
PLAQR	72.77	29.60
OPERA	78.53	35.51

Table 6 shows that the QRNN model has the most accurate forecasts based on evaluation metrics RMSE and MAE. GHI superimposed with combined predictions using QRNN for the testing set, 1759 observations which are approximately seventy-three days of forecasts, are shown in Figure 7.
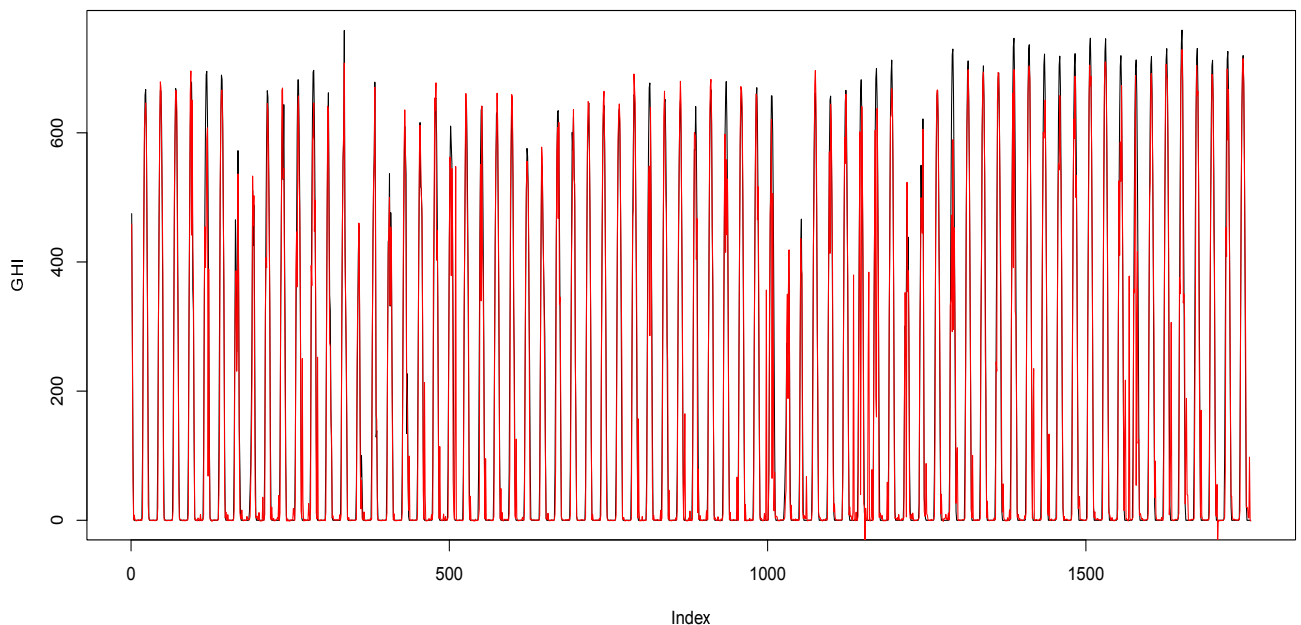

Fig. 7: GHI vs combined forecasts(using QRNN).

3.3 Murphy diagrams

The forecasting models were also evaluated using proper scoring rules, and Murphy diagrams were used to compare forecasts. Figures 8 − 11 show plots giving empirical scores and Differences in scores, and the GPR model proves superior to other models. The Murphy diagrams show that the Opera model is very good though the GPR model is a better forecaster than Opera.
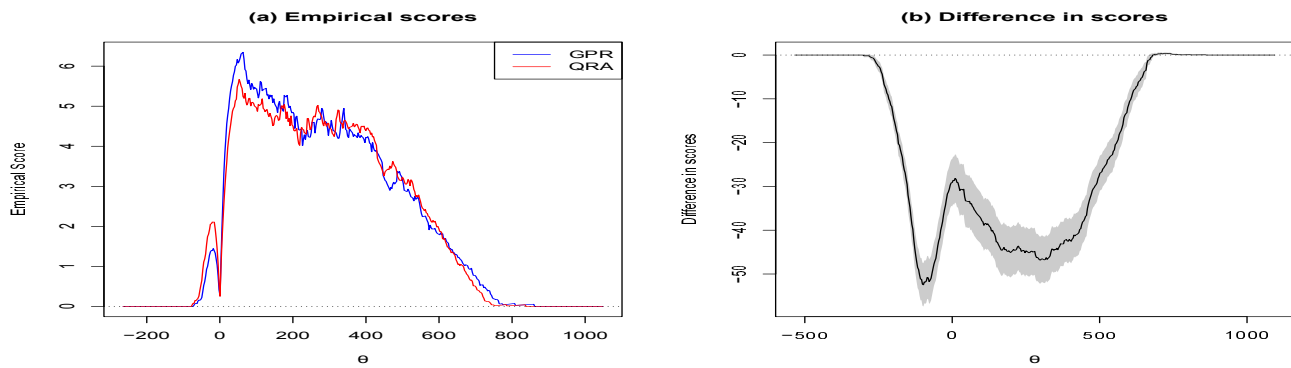
Fig. 8: Murphy diagram for QRA vs GPR. **Left panel:** (a) Empirical scores. **Right panel:** (b) Difference in scores.
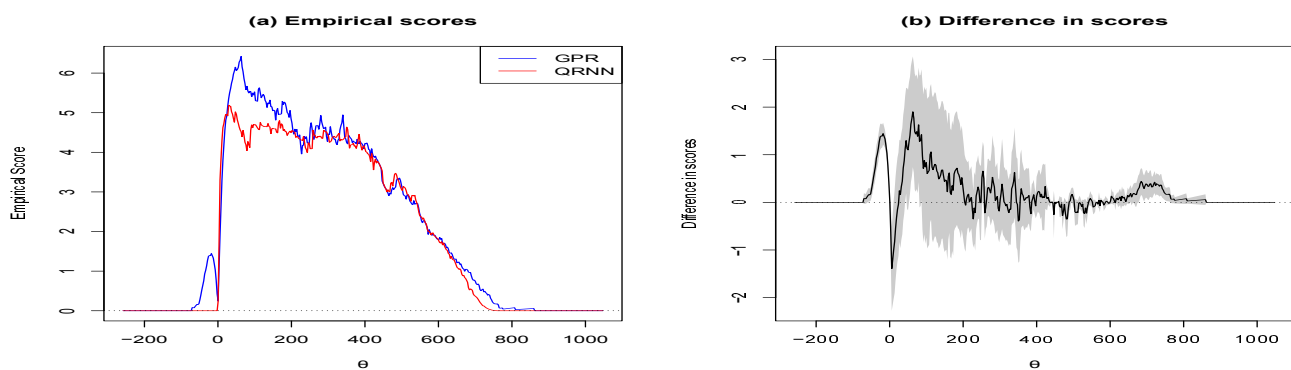

Fig. 9: Murphy diagram for QRNN vs GPR. **Left panel:** (a) Empirical scores. **Right panel:** (b) Difference in scores.
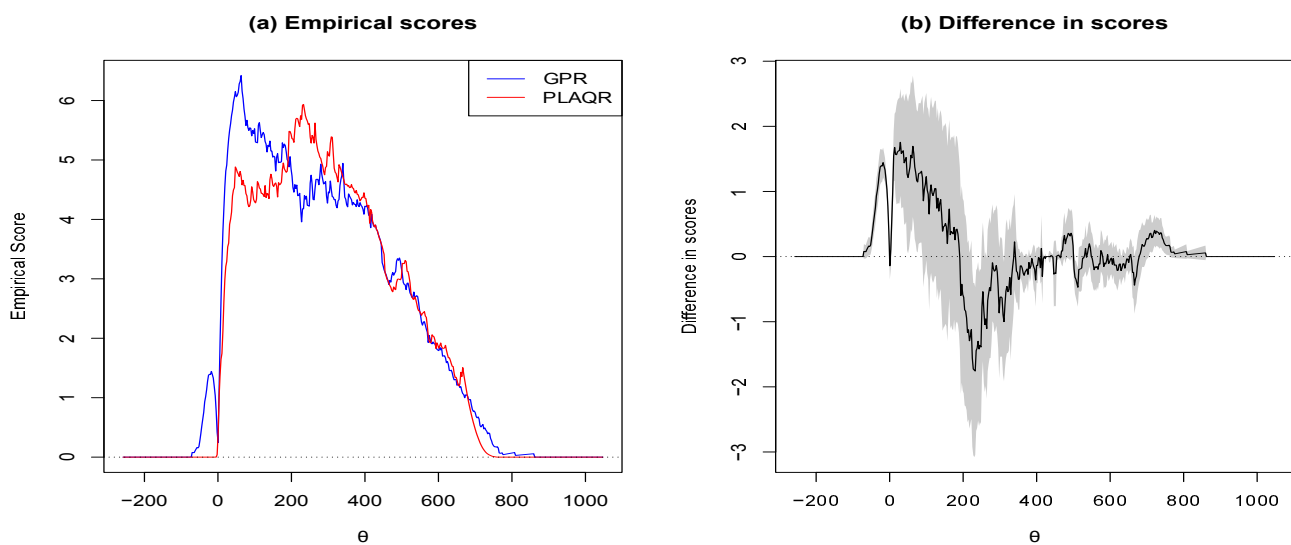

Fig. 10: Murphy diagram for PLAQR vs GPR. **Left panel:** (a) Empirical scores. **Right panel:** (b) Difference in scores.
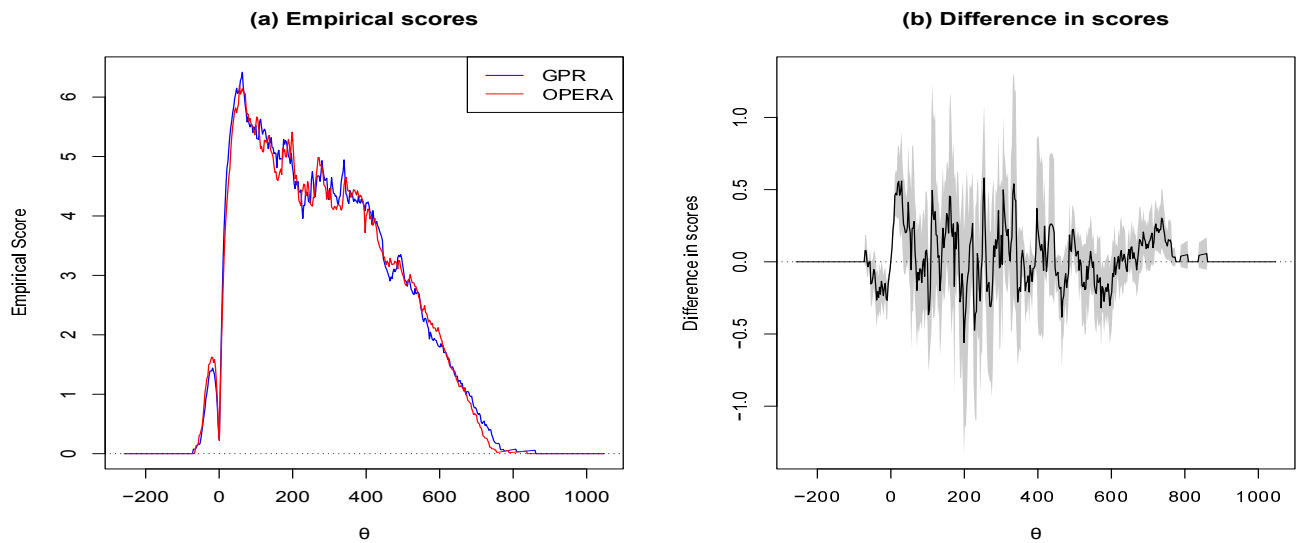
Fig. 11: Murphy diagram for OPERA vs GPR. **Left panel:** (a) Empirical scores. **Right panel:** (b) Difference in scores.
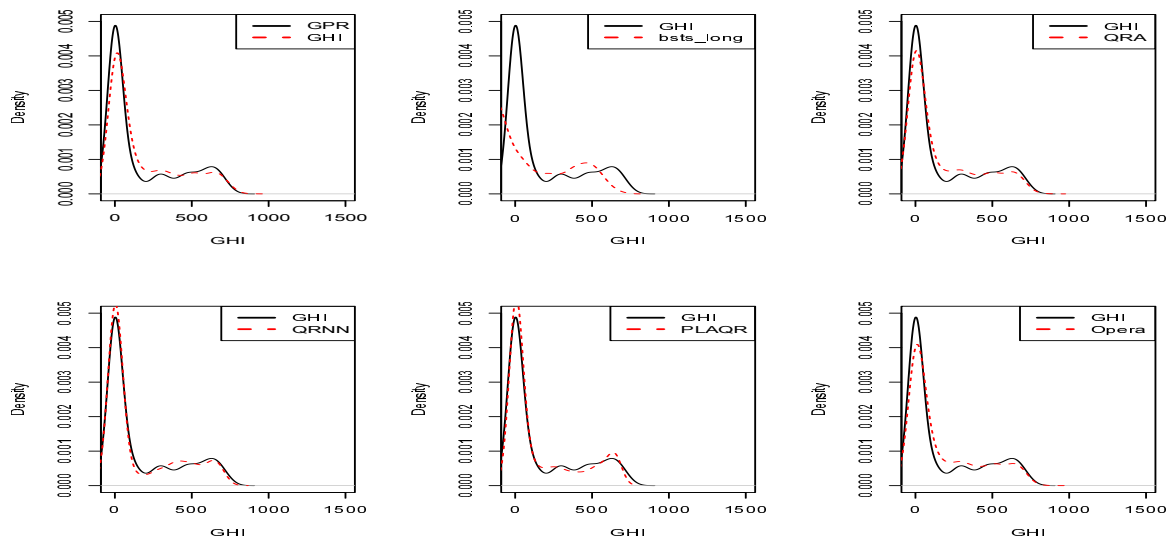

Fig. 12: Density plots.

Plots in Figure 12 show a comparison between the densities of the models(in red) and GHI(in black). The plots show that models GPR, PLAQR, and QRNN are better forecasters. The densities of the forecasts are very close to the actual values of GHI.

3.4 Evaluating models using scoring rules

A further comparison of the models was made based on proper scoring rules. The proper scoring rules used are the logarithmic score (LogS), the continuous ranked probability score (CPRS), Dawid–Sebastiani Score (DSS), and Pinball Losses (PL). The forecasts which gave the lowest values of the scoring rules were the ones that gave a better forecasting performance.

The scoring rules were computed: we fit a parametric distribution to the forecasts. In this case, the Gamma distribution was the one that fitted the data and then evaluated the parameters. A comparison of the models is given in Table 7 for the models GPR, bstslong, QRA, QRNN, PLAQR, and Opera using probabilistic evaluations. The GPR model best fits based on the probabilistic evaluation metrics CRPS, LogS, DSS, and PL.

Table 7: Combined forecasts.

Model	CRPS	LogS	DSS	PL
GPR	131.76	6.87	11.89	34.28
bstslong	205.93	Inf	12.85	294.27
QRA	143.79	Inf	12.50	43.45
QRNN	150.14	Inf	12.53	38.23
PLAQR	146.80	Inf	12.53	43.15
OPERA	142.15	Inf	12.48	37.40

4 Discussion

Decision-makers need robust models with high predictive capabilities in the electricity sector in medium-term forecasting solar power, which is highly intermittent and must be optimally integrated onto the grid. The GPR modelling framework, flexible in modelling nonstationary time series data and its ability to cope with abrupt regime changes, was found among the individual models to have the highest predictive ability in producing the most accurate and robust forecasts. The forecasts from the individual models were then combined using four combination methods, QRA, QRNN, PLAQR and OPERA. This increased the forecast accuracy, with the QRA being the best based on MAE and QRNN being the best based on the RMSE. Based on the evaluation measures for probabilistic forecasting, GPR has the lowest CRPS, DSS and PL scores proving to be the most superior individual model for predicting GHI. Using the Murphy diagrams, the GPR and QRNN models seem to have the same predictive abilities.

Compared to previous findings, the present study reveals an upgrade to the work of Chandiwana [11], Mpfumali [25], among others from literature who did not include an ensemble of robust methods in medium-term forecasting of GHI. The present study produces robust medium-term forecasts of up to ten days ahead using flexible and powerful techniques such as DGPR, GPR, AQR and BSTS. A limitation of this study is using a two-layer DGPR model. Future research will explore the predictive abilities of DGPR models with more layers in the DGPR model.

Conclusion

This study explored the predictive abilities of robust models in the medium-term forecasting of GHI. A key finding of this result is that the GPR modelling framework provides robust predictions of medium-term solar power. This modelling framework provides nonlinear predictive capability with flexibility with nonstationary time series data. It can also cope with abrupt regime changes in the data and allows the inclusion of prior knowledge in the model specification. These results could be useful for system operators and decision-makers in power utility companies to integrate the intermittent renewable energy source into the grid.

Acknowledgments

References

[1] D. Yang, C. Gu, Z. Dong, P. Jirutitijaroen, N. Chen and W.M. Walsh, Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging, *Renewable Energy*, **60**, 235-245 (2013). https://doi.org/10.1016/j.renene.2013.05.030

[2] UNDP, Sustainable Development Goal 7: Affordable and clean energy, [Online] Available https://www.undp.org/sustainable-development-goals, Accessed on 4 April 2022.

[3] I. Bilionis, M.C. Emil and A. Mihai, Data-driven model for solar irradiation based on satellite observations, *Solar energy*, **110**, 22-38, (2014). https://doi.org/10.1016/j.solener.2014.09.009

[4] H. Tolba, N. Dkhili, J. Nou, J. Eynard, S. Thil and S. Grieu, GHI forecasting using Gaussian process regression: Kernel study, *IFAC-PapersOnLine*, **2(4)**, 455-460, (2019). https://doi.org/10.1016/j.ifacol.2019.08.252

[5] Y. Wang, F. Bo, H. Qing-Song and S. Li, Short-term solar power forecasting: A combined long short-term memory and Gaussian process regression method, *Sustainability*, **13(7)**, 3665, (2021). https://doi.org/10.3390/su13073665

[6] Z. Zhang, W. Chao, P. Xiaosheng, Q. Hui, L. Hao, F. Jialong and W. Hongyu, Solar Radiation Intensity Probabilistic Forecasting Based on K-Means Time Series Clustering and Gaussian Process Regression, *IEEE Access*, **9**, 89079-89092, (2021). https://doi.org/10.1109/ACCESS.2021.3077475

[7] B.M. Yu, J.P. Cunningham, G. Santhanam, S.I. Ryu, K.V. Shenoy and M. Sahani, Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity, *Journal of neurophysiology*, **102(1)**, 614-635, (2009). https://doi.org/10.1152/jn.90941.2008

[8] M. Stonski, Bayesian neural networks and Gaussian processes in identification of concrete properties, *Computer-Assisted Methods in Engineering and Science*, **18(4)**, 291-302, (2017). https://cames.ippt.pan.pl/index.php/cames/article/view/108

[9] M. Al-Shedivat, A.G. Wilson, Y. Saatchi, Z. Hu and E.P. Xing, Learning scalable deep kernels with recurrent structure, *The Journal of Machine Learning Research*, **18(1)**, 2850-2886, (2017). https://dl.acm.org/doi/10.5555/3122009.3176826

[10] E. Tsymbalov, S. Makarychev, A. Shapeev and M. Panov, *Deeper connections between neural networks and Gaussian processes speed-up active learning*, Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 3599-3605, (2019). https://doi.org/10.24963/ijcai.2019/499

[11] E. Chandiwana, C. Sigauke and A. Bere, Twenty-Four-Hour-Ahead Probabilistic Global Horizontal Irradiance Forecasting Using Gaussian Process Regression, *Algorithms*, **14(6)**, 177, (2021). https://doi.org/10.3390/a14060177

[12] J.M. Bates and C.W. Granger, The combination of forecasts, *Journal of the Operational Research Society*, **20(4)**, 451-468, (1969). https://doi.org/10.1057/jors.1969.103

[13] A. Saucer, R.B. Gramacy and D. Hogdon, *Active Learning for Deep Gaussian Process Surrogates*, Available online: https://arxiv.org/abs/2012.08015v2 (accessed on 25 November 2021).

[14] M.I. Radaideh, and T. Kozlowski, Surrogate modeling of advanced computer simulations using deep Gaussian processes, *Reliability Engineering & System Safety*, **195**, 106731, (2020). https://doi.org/10.1016/j.ress.2019.106731

[15] K.H. Brodersen, F. Gallusser, J. Koehlar, N. Remy and S.L. Scott, Inferring causal impact using Bayesian structural time-series models, *The Annals of Applied Statistics*, **9(1)**, 247-274, (2015). https://doi.org/10.1214/14-AOAS788

[16] S.L. Scott and H.R. Varian, Predicting the present with Bayesian structural time series, *Int. J. Math. Model. Numer. Optim*, **5(1-2)**, 4-23, (2014). http://dx.doi.org/10.1504/IJMMNO.2014.059942.

[17] T.J. Mitchell and J.J. Beauchamp, Bayesian variable selection in linear regression, *Journal of the American Statistical Association*, **83(404)**, 1023-1032, (1988). https://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478694

[18] D. Madigan and A.E. Raftery, Model selection and accounting for model uncertainty in graphical models using Occam's window, *Journal of the American Statistical Association*, **89(428)**, 1535-1546, (1994). https://doi.org/10.1080/01621459.1994.10476894

[19] A.C. Harvey, *Forecasting, structural time series models and Kalman filter*, Cambridge university press, (1990).

[20] J. Durbin and S.J. Koopman, A simple and efficient simulation smoother for state space time series analysis, *Biometrika*, **89(3)** 603-616, (2002). https://www.jstor.org/stable/4140605

[21] J.A. Hoeting, D. Madigan, A.E Raftery and C.T. Volinsky, Bayesian model averaging: a tutorial, *Statistical Science*, **14(4)**, 382-401, (1999). https://www.jstor.org/stable/pdf/2676803.pdf

[22] P. Gaillard, Y. Goude and R. Nedellec, Additive models and robust aggregation for GEFcom2014 probabilistic electric load and electricity price forecasting, *Int. J. Forecast.* **32(3)**, 1038-1050, (2016). https://doi.org/10.1016/j.ijforecast.2015.12.001

[23] M. Fasiolo, Y. Goude, R. Nedellec and S.N Wood, Fast Calibrated Additive Quantile Regression, *Journal of the American Statistical Association*, **116(535)**, 1-12, (2020). https://doi.org/10.1080/01621459.2020.1725521

[24] A. Jordan, F. Krüger and S. Lerch, Evaluating Probabilistic Forecasts with scoringRules, *Journal of Statistical Software*, **90(12)**, 1-37, (2019). https://doi.org/10.18637/jss.v090.i12

[25] P. Mpfumali, C. Sigauke, A. Bere and S. Mulaudzi, Day ahead hourly global horizontal irradiance forecasting: An application to South African data, *Energies*, **12(18)**, 1-28, (2019). https://doi.org/10.3390/en12183569