SOLAR POWER FORECASTING USING GAUSSIAN PROCESS REGRESSION

By

EDINA CHANDIWANA
18021412

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in STATISTICS

in the

Department of Mathematical and Computational Sciences
Faculty of Science, Engineering, and Agriculture.
University of Venda, Thohoyandou, Limpopo
South Africa

Supervisor: **Prof Caston Sigauke**

Co-Supervisor: **Dr Alphonce Bere**

Abstract

Solar power forecasting has become an important aspect affecting crucial day-to-day activities in people's lives. Many African countries are now facing blackouts due to a shortage of energy. This has caused the urge to encourage people to use other energy sources to rise, resulting in different energy inputs into the main electricity grid. When the number of power sources being fed into the main grid increases, so does the need for efficient methods of forecasting these inputs. Thus, there is a need to come up with efficient prediction techniques in order to facilitate proper grid management. The main goal of this thesis is to explore how Gaussian process predicting frameworks can be developed and used to predict global horizontal irradiance. Data on Global horizontal irrandiance and some weather variables collected from various meterological stations were made available through SAURAN (Southern African Universities Radiometric Network). The length of the dataset ranged from 496 to 17325 datapoints. We proposed using Gaussian process regression (GPR) to predict solar power generation. In South Africa, studies based on GPR regarding forecasting solar power are still very few, and more needs to be done in this area. At first, we explored covariance function selection, and a GPR was developed using Core vector regression (CVR). The predictions produced through this method were more accurate than the benchmark models used: Gradient Boosting Regression (GBR) and Support Vector Regression then, we explored interval estimation, Quantile regression and GPR were coupled in order to develop the modelling framework. This was also done to improve the accuracy of the GPR models. The results proved that the model performed better than the Bayesian Structural Time Series Regression. We also explored spatial dependence; spatio-temporal regression was incorporated into the modelling framework coupled with GPR. This was done to incorporate various weather stations' conditions into the modelling process. The spatial analysis results proved that GPR coupled with spatial analysis produced results that were superior to the Autoregressive Spatial analysis and benchmark model used: Linear Spatial analysis. The GPR results had accuracy measures that proved superior to the benchmark models. Various other tools were used to improve the accuracy of

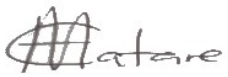
the GPR results. This includes the use of combining forecasts and standardisation of predictions. The superior results indicate a vast benefit economic-wise because it allows those who manage the power grid to do so effectively and efficiently. Effective power grid management reduces power blackouts, thus benefitting the nation economically and socially.

Keywords— Core vector regression; Gaussian process; Minimum enclosed ball; Quantile regression; Solar power; Spatio-temporal.

Declaration

I, EDINA CHANDIWANA, [18021412], hereby declare that the thesis titled: “Solar power forecasting using Gaussian process regression" for the Ph.D. degree in Statistics at the University of Venda, hereby submitted by me, has not been submitted for any degree at this or any other university, that it is my own work in design and in execution, and that all reference material contained therein has been duly acknowledged.

Signed (Student):



September 8, 2023

Dedication

To my late Dad, Samuel Charles Matare.

Acknowledgment

I thank the Almighty God, the provider of knowledge and wisdom, for enabling me to undertake my research successfully. Without His grace, I would not have made it. This work resulted from concerted efforts by many players, some of whom I might not be able to mention. However, I feel greatly indebted to the following individuals and organisations:

Firstly I would like to express my gratitude to my supervisor, Prof Caston Sigauke and my co-supervisor Dr Alphonce Bere for guiding me throughout the journey with great patience and care.

This research work would not have been possible without the help of the Mathematical Science staff for being so kind and always willing to help.

I want to thank my family and friends who supported me throughout my PhD studies.

Contents

List of Figures

List of Tables

Abbreviations

ARIMAX	Autoregressive Moving Average)
BMA	Bayesian Model Averaging
BSTS	Bayesian Structural Time Series
CRPS	Continuous Rank Probability Score
CP	Prediction Interval Coverage
DNN	Deep Neural Network
GHI	Global Horizontal Irradiance
CVR	Core Vector Regression
CSIR	CSIR Energy Centre
CUT	Central University of technology
DSS	Dawid Sebastian Score
GHI	Global Horizontal Irradiance
GBR	Gradient Boosted Regression
GBM	Gradient Boosting Method
GPR	Gaussian Process Regression
LASSO	Least Absolute Shrinkage and Selection Operator
LogS	Logarithmic Scoring
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MEB	Minimum Enclosed Ball
MLE	Maximum likelihood Estimation
MSE	Mean Square Error
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
W/m^2	Watts Per Square Meter.
QRA	Quantile Regression Averaging
QRNN	Quantile regression Neural Network

PLAQR	Partially Linear Averaging Quantile regression
Opera	Online Prediction by ExpRt Aggregation
GP Spatial	Gaussian Process Spatial Regression
GP ARSpatial	Gaussian Process Autoregressive Spatial Model
LSTR	Linear Spatial Temporal Regression
ARX	autoregressive with exogenous inputs
NN	Neural Network
RRF	random regression forest
RT	regression trees
ANN	Artificial Neural Network
SVR	Support Vector Regression
MLP	artificial neural networks
MGGP	multigene genetic programming
MDPI	Multidisciplinary Digital Publishing Institute
MIN	CRSES Mintek
NUST	Namibian University of Science and Technology
SAURAN	Southern African Universities Radiometric Network
UFH	University of Fort Hare
UNV	university of Venda
UNZ	University of Zululand
UPR	University of Pretoria
QGAM	Quantile Generalized Additive Model
QRNN	Quantile Regression Neural Networks

List of Notation and Special Symbols

$k(x, x^1)$	covariance function.
$m(x)$	mean function of X.
$k(c; r)$	minimum enclosed ball.
$log_\pi(\theta, 0\|y)$	logarithm of joint posterior
$\pi(y(s_0, t)\|y)$	Posterior predictive distribution.
$PL(q\tau, t)$	Pinball loss function.
β_0	Intercept.
β_j	Coefficients.
$\mathbf{B}$	Model matrix of basis functions.
θ	Parameter vector.
$\pi(x\|\theta)$	Likelihood function.
$d(x, x^1)$	distance between 2 datapoints
ε_{it}	Random error term.
Z_t	output vector.

Chapter 1

Introduction

1.1 Background

Electricity is one of the most important needs of people in various parts of the world. Many developing countries, like South Africa, use coal, water, and other alternative forms of energy to produce electricity. Many alternative energy sources in South Africa provide electricity to urban and rural populations. Although we have these various sources, the energy deficit is still a problem that policymakers should solve. The energy deficit has effectively stunted Africa's development, with many people who stay in sub-Saharan Africa without reliable access to power sources. According to President Cyril Ramaposa, the country has an installed capacity of 46 000MW of electricity, and the peak demand for electricity is 32,000 MW. Only 60% of the 46 000MW installed electricity is being supplied to the nation (Government, 2022). This leaves the country with a deficit of about 4 400MW, which has caused industries, schools, and health facilities to operate without electricity for some hours.

In the past, energy production mainly depended on thermal power and hydroelectricity. Using thermal power has a major problem. It produces high concentrations of greenhouse gases like carbon dioxide, carbon monoxide, and hydrocarbon sulfur dioxide. These carbon

emissions play a significant role in air pollution resulting in climatic change. The emissions trap heat in the atmosphere, which builds up, resulting in global warming effects. Air pollution also causes depletion of the ozone layer resulting in direct sun reaching the earth. Direct sun rays cause very hot temperatures, which are not good for nature. Global warming affects and changes both present and future weather patterns, thus disturbing the balance of nature. They also harm people's health. It causes diseases such as cancer. Sulfur dioxide also causes respiratory problems and contributes to acid rain. The health problems are caused by inhaling carbon emissions and high temperatures that cause heat waves.

South Africa has measures that have been put in place in order to solve the problem of carbon emissions. One major step is signing agreements that rectify practices that cause carbon emissions. The signed agreements include the Paris Agreement. This climate change agreement was signed in New York, United States, in 2015 between United Nations and South Africa. It looks at a commitment by the international community to finding significant ways of reducing the emissions of greenhouse gases. This is achieved by promoting using clean energy (Ministry of forestry and the environment, 2016). Other policies have been put in place. These include the IRP (Integrated Resource Plan), which aims to make electricity cheaper, reduce water usage, and encourage electricity generation from different sources. The Ministry of Resources and Energy formulated the plan in March 2011 (Ministry of minerals and energy, 2019). Another agreement that promotes reducing greenhouse gases was signed through the World Bank. The documents provide various ways the country will use to preserve the environment. The report is called Country Climate and development report. It is a diagnosis plan that reflects the plans to be used by the nation to prioritise actions that reduce the emission of these gases, especially promoting low carbon emission (WorldBank, 2022). Another important agreement is South Africa's Just Energy Transition Investment Plan (JETIP), which should run from 2023 to 2027. It is an agreement signed between South Africa and other European countries. The JETIP

looks at a commitment by SA to reduce gas emissions by considering the impact of the energy transition on people's lives.

Other efforts have been made in South Africa in order to promote the use of other power sources. These include the establishment of solar power plants to boost the power supply. Several solar plants have been established in South Africa. An example is the De Aar Solar power plant. Figure 1.1 shows the Northern Cape Province's De Aar Solar Power plant. It is located outside the De Aar town and has 167 580 panels (DeAar, 2022). All the solar plants in SA are linked to Eskom, the major electricity supplier, to facilitate control of the energy supply from such facilities.

South Africa can engage in projects that promote the use of solar power. It receives significant solar radiation; thus, we focused on solar energy. South Africa receives quite a substantial amount of solar radiation compared to other countries, which can be generated into electricity (Resources and Energy Department, 2016). Daily, Africa receives more hours of sunshine than other continents do yearly. This has made SA an ideal place for generating solar energy. Solar power has gained popularity in Africa since it is less costly and readily available.

Figure 1.1: De Aar Solar Power plant in the Northern Cape Province. Source `https://deaarsolar.co.za/`.

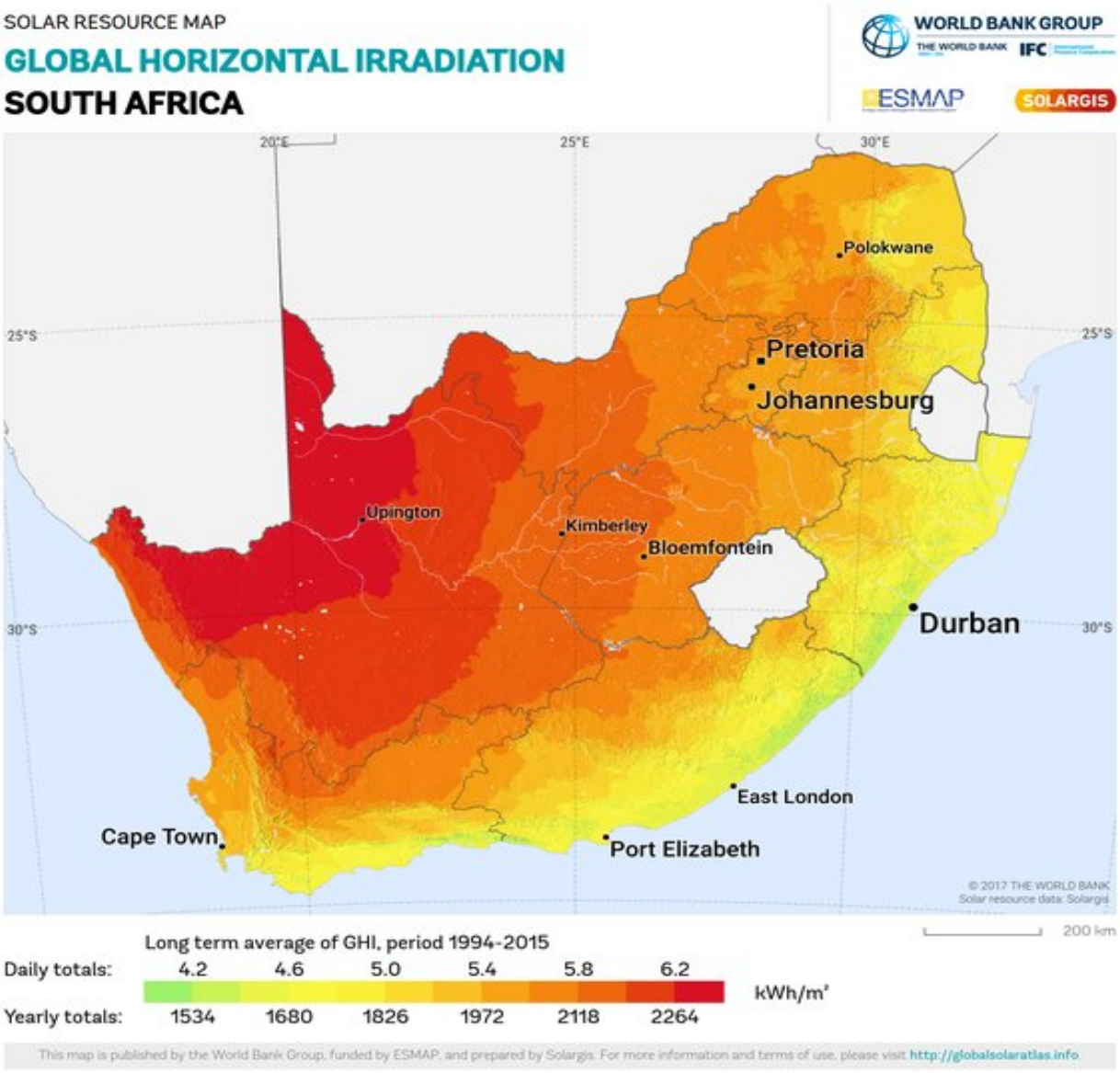


Figure 1.2: SA solar map. Source: Global Solar Atlas 2.0, Solar resource data: Solargis. `https://solargis.com/maps-and-gis-data/download/south-africa`.

Figure 1.2 shows a summary of GHI(Global Horizontal Irradiance) which is solar radiation distribution in SA. It gives average values of daily and yearly average radiation. Daily

average values range from 5.0-6.2, and yearly totals range from 1826 to 2264, with only a few coastal areas receiving less irradiance than these averages. It reflects that SA receives much solar radiation that can be used for powering the country. This makes solar energy one of the largest sources of power in South Africa.

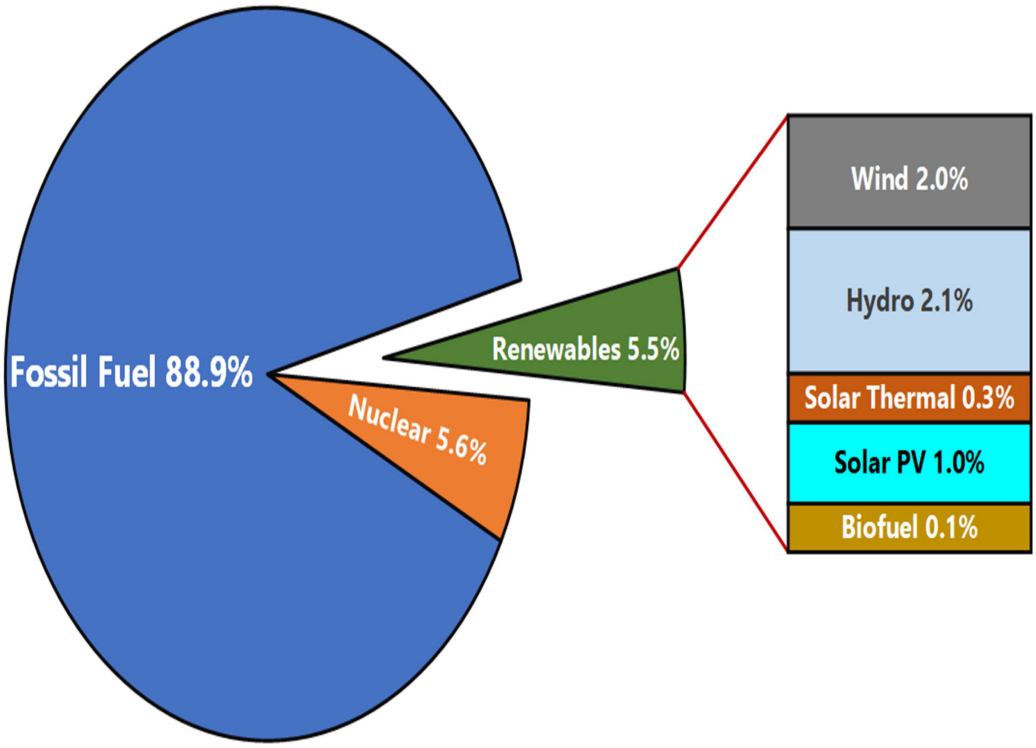


Figure 1.3: Pie Chart of energy distribution in SA. Source:IAE `//www.iea.org/data-and-statistics`.

Figure 1.3 shows that only a little of the generated solar is used. President Ramaposa pointed out in his speech Government (2022) that about 27000W of energy is being supplied. From the pie chart, only 5.5% is coming from renewable sources, and 1.1% is coming from solar power, which is about 15.18W, but 2264 can be used instead, easing the problem of electricity shortage.

Solar power is the radiation obtained through harnessing energy from the sun's rays.

It is converted into electrical energy that can be used for heating, lighting, water, and commercial, individual, or domestic use. It is a clean, reliable, and cheap source of power. The main energy producer in South Africa, Eskom, has implemented several solar energy projects. They have installed heating equipment for solar energy in private buildings. Evaluating these locations is an ongoing identifying process to quantify attainable reserve funds from solar power and the establishment of water heating tools. Solar power is produced as Photovoltaic (PV), and the measurement unit is Kilowatt peak (kWp), the voltage yield generated by solar cells to produce electricity. PV forecasting over time can help PV plants efficiently supply solar energy for power frameworks without underestimating their ability to utilise solar power boards. South African produced solar power is measured based on photovoltaic power. Photovoltaic relates to the production of electricity after exposure to sunlight.

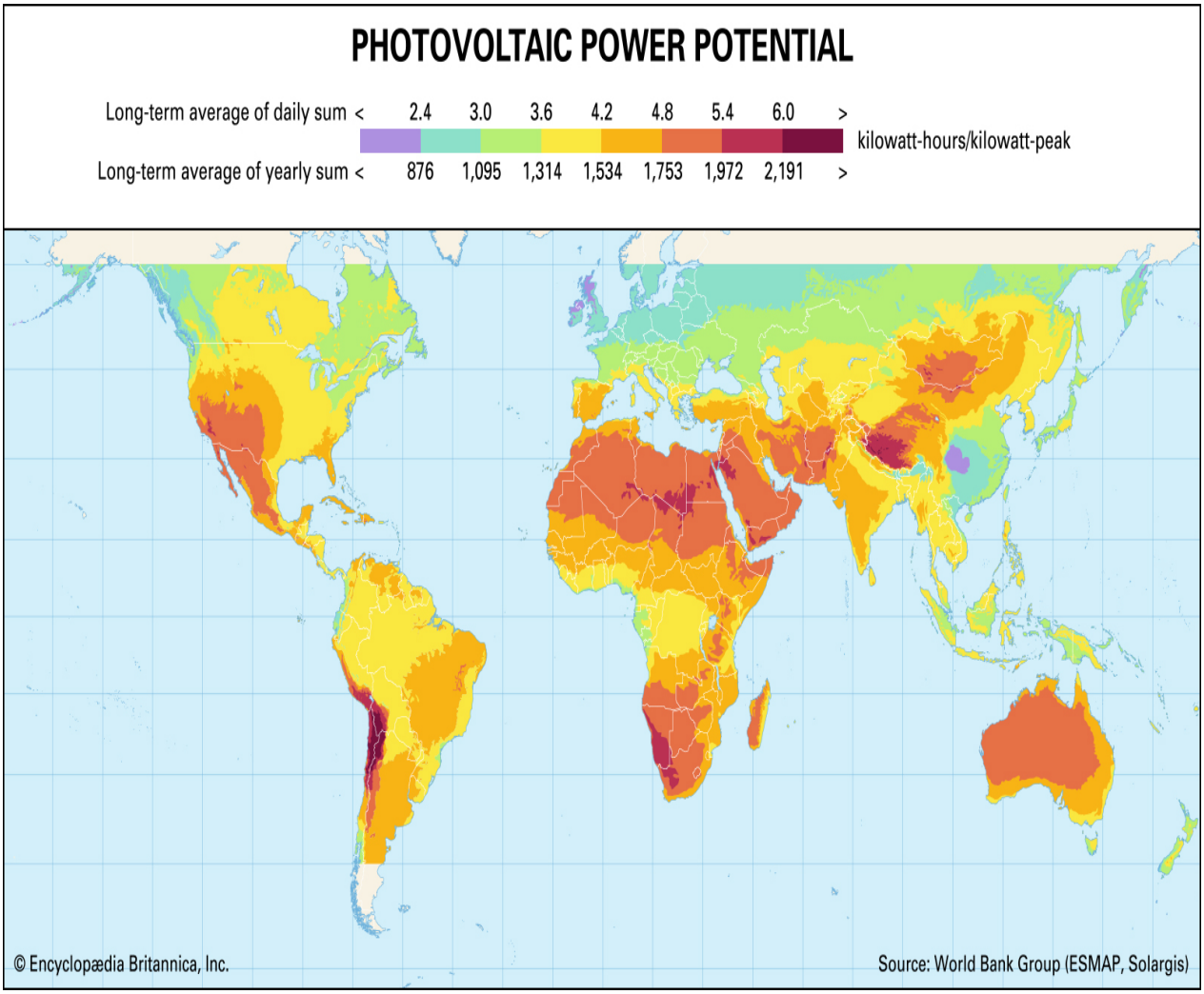


Figure 1.4: World solar power map. Source: World Bank Group (ESMAP, Solargis). `https://solargis.com/maps-and-gis-data/download/world`.

Figure 1.4 shows an aggregated and harmonised view of the world's potential photovoltaic(PV) power. It can be seen from the map that for South Africa, the amount of solar radiation received is more than most countries in the world. In order to efficiently incorporate solar power into the power grid, we need to forecast it accurately.

We used Global Horizontal Irradiance (GHI) measurements for this study, which gives

the actual yield. GHI is the solar energy measured concerning time, produced by all components of the atmosphere and reaching the Earth's surface. It combines Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance (DHI), and reflected ground radiation. It is measured in W/m^2(watts per square meter). According to NERSA (National Energy Regulator of South Africa), 1329 MW has been recorded as the installed solar power of South Africa, and the country's capacity is expected to be at 8400 MW by 2030 (NERSA, 2016). Policymakers have been trying to put measures in place to promote the use of solar power. The measures include the expense-motivating force of tax payments through the South African Revenue Service as of 1 January 2016 for establishing photovoltaic sun-powered vitality frameworks. Other than the administration motivation, certain regions have additionally structured motivators for private and business clients, for example, feed-in tariff and net metering. Solar power comes from the sun captured through heat and light, which is captured using different types of equipment such as solar architecture, photovoltaic, and solar power. It is a crucial energy source that is renewable. Technologies are active solar and passive solar, which rely upon the technique used to capture sun-oriented energy. In 2000 the United Nations, through its development program called World Assessment, revealed that the yearly solar energy that can be produced was 1575-49837 exajoules(EJ), which is higher than the consumption of world energy, which was 559.8 EJ (Smart-Energy, 2022).

1.2 Statement of the problem

The need for additional power from other sources should be addressed, and solar power has been chosen as a potential power source for this research. This was done to promote the use of solar power by indicating the amount of solar power available to boost energy in SA. This aims at addressing the shortage of electricity as well as reducing carbon emissions.

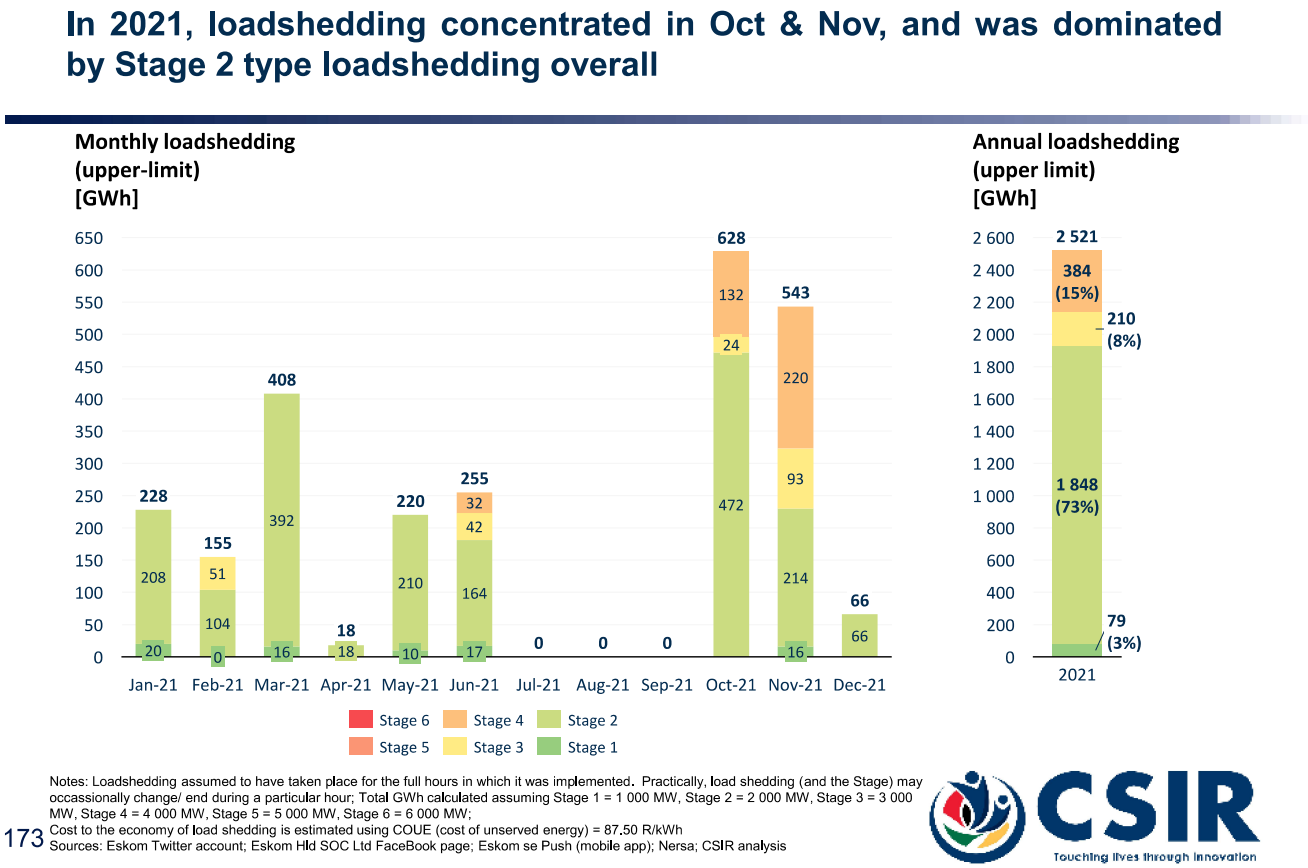


Figure 1.5: Load shedding dominating. Source:`https://www.csir.co.za`.

Figure 1.5 shows that SA has been facing electricity challenges resulting in load shedding, with some parts of the country experiencing Stage 4 load-shedding in 2021. This means you will be scheduled for load-shedding 12 times over four days for two hours or 12 times over eight days for four hours at a time. Hence the need to look at ways of solving the problem of electricity. A method rarely used for predicting solar power in South Africa was constructed and applied. Solar energy needs to be integrated into the main grid (Ismail et al., 2015). However, planning and operating power grids reliably are af-

fected by fluctuations in the penetration of solar energy (Sun et al., 2021). Solar power forecasting traditionally relies on data mining methods to progressively predict the power output over a given period as reflected in historical data, such as time series modelling and artificial neural networks. There are limitations to adopting such methods, as accuracy is restricted by data availability when data recordings are missing due to several reasons. One of the reasons may be that data collection may be interrupted by information estimation failure or incorrect recordings, and such errors may affect forecasting accuracy. Therefore, there is a need to improve PV forecasting accuracy by capturing such complexities. Solar power prediction has the challenge of being unpredictable due to changes in weather patterns and other unforeseen factors. Unlike hydroelectricity and thermal electricity, solar power cannot be planned or controlled because of inherent features such as volatility, randomness, and intermittence, which is a challenge in integrating solar power into the grid. This calls for a prediction method that captures uncertainty and complexity in the dataset used. Hence GPR was chosen for this research. Selecting Kernel functions forms a crucial part of GPR. Not much has been done to find an appropriate method of selecting kernel functions effectively. Williams and Rasmussen (2006) has explored covariance function selection in Chapter 4 of their book, they indicated that the functions are chosen based on cross-validation. On the other hand, Abdessalem et al. (2018) used Approximate Bayesian Computation(AB)algorithm for kernel selection.

Researchers have been concentrating on point estimations, paying little attention to the differences in the weather conditions of various meteorological stations. As a result, Quantile Regression was used for predictions that are interval based in order to improve prediction accuracy. Weather conditions at times vary and change with distance, so there is a need to check the relatedness of datasets from different weather stations before generalising results to the whole country. Dependences of meteorological dependencies were used to capture the variation in weather conditions through spatial analysis. Another important aspect of

forecasting is accuracy. Thus, the purpose of this study is to identify appropriate models that captures uncertainity and variability. These approaches include standardisation and combining forecasts. All the tools were applied to make the GPR model robust and accurate.

1.3 Research questions

This study will answer the following main research questions:

1) Does the Core vector regression provide a better way of selecting a kernel function for short-term Global Horizontal Irradiance forecasting?

2) Does the combination of Quantile regression and GPR improve GHI predictions?

3) Does capturing spatial dependencies improve GHI predictions?

1.4 Aim and objectives of the study

Aim

This study aims to construct and evaluate Gaussian process regression models for solar power forecasting.

Objectives

The objectives are to:

1) construct a GPR model coupled with Core vector regression for short-term forecasting of global horizontal irradiance.

2) evaluate and compare the predictive abilities of GPR models combined with Additive Quantile Regression (AQR) and Bayesian Structural Time Series (BSTS) models in the short-term forecasting of global horizontal irradiance.

3) construct spatio-temporal GPR models coupled with autoregressive models for multi-site prediction of global horizontal irradiance.

1.5 Significance of the study

This study's main contribution is the development of GPR models for forecasting solar energy production. Due to the integration of large solar farms into the power grid, solar energy forecasting becomes important to system operators responsible for making decisions concerning the operations of the power grid and the operators of the electric market. This study will help the power grid system operators to take optimal decisions.

1.6 Contributions

Solar power forecasting was done using Gaussian Process Regression using South African data. A crucial aspect of GPR modelling was addressed. In order to effectively choose an appropriate kernel function, core vector regression was coupled with GPR. This was done to improve the performance of Gaussian Process Regression; choosing the appropriate kernel yields robust results. Applying the minimum enclosed ball produces approximately optimal solutions because of core sets. Core vector regression uses the 2-norm error that makes the model robust against outliers. CRV enabled the optimisation of the kernel function parameters and removed artificial factors in the prediction process, having higher precision. Another important aspect of forecasting was addressing variable selection through various modern methods. Most research articles make use of just one particular method for variable selection. In one of the publications produced in this research, many variable selection methods were adopted, and just one, which proved to be more accurate based on the dataset used, was selected. Variables were selected using various variable selection techniques, which included shrinkage methods like lasso and elastic net, model-based methods like GBR

and MARS, and Boruta. Also, other methods like OPERA were used. The method that produced the minimum value of MAE was used for variable selection. Another important aspect applied in this research was the improvement of forecast accuracy through forecast combination and standardisation. Forecasts combination was applied in order to improve the predictions.

GPR was combined with the following models: Core vector regression, quantile regression, and Opera. Various authors like Bates and Granger (1969) and Gaillard et al. (2021) did studies on forecast combinations and proved that they improved forecasts accuracy. Standardisation of data scales forecasts to the same range of values to improve the predictions' quality. This creates consistency in the dataset by avoiding wider ranges from the data. Spatial Regression through temporal-spatial dependencies between meteorological stations in different geographic locations. This is another unique contribution. Three problem areas that have prompted this research that something beyond the usual GPR model was done is the need to improve forecasts and develop more robust forecasts through core vector regression, quantile regression, and multiple sites.

1.7 Outline of thesis

The structure of the thesis is given in eight chapters. The highlights of the chapters follow. The next chapter is Chapter 2, the Literature Review. We will review the existing literature on renewable energy forecasting, solar energy forecasting, and forecasting using Gaussian Process Regression. Chapter 3 provides a general theory of the methodologies adopted and the research design implemented in this study. Chapter 4 is on Data and study sites. Chapter 5, Chapter 6, and Chapter 7 provided how the objectives were achieved through different publications produced during this research. Firstly, in Chapter 5, Gaussian process regression coupled with Core vector regression was used to predict GHI. We adopted

a core vector regression approach called the minimum enclosing ball (MEB) technique to select the best possible kernel for the analysis. The variables were selected via hierarchical interactions using the least absolute shrinkage and selection operator. The results showed that the GPR models gave the most accurate predictions compared to those from gradient boosting and support vector regression models. Secondly, in Chapter 6, Gaussian Process Regression (GPR) combined with Quantile Regression and Bayesian Structural Time Series (BSTS) model in short- and long-term forecasting was used to forecast GHI. Four methods were adopted for variable selection, Lasso, Elastic net, Boruta, and GBR (Gradient Boosting Regression). The variables selected using GBR were used because they produced the lowest value of MAE (Minimum Absolute Errors). In order to improve the forecasting accuracy of the models, forecasts were combined with the following forecasting models: Linear QRA(Quantile Regression Averaging), QRNN(Quantile Regression Neural Network), PLAQR(Partial Linear additive Quantile Regression), and Opera(Online Prediction by ExpRt Aggregation). The GPR model was selected as the best model. Lastly, Chapter 7 explored Spatial Gaussian Process Regression (SGPR) and the Spatial GPR Autoregressive (SGPR-AR) models to predict GHI using data from seven radiometric stations from South Africa and one from Namibia. The benchmark model used was the linear Spatial Temporal Regression (LSTR) model. Five validation sets, each comprising three stations, were chosen. For each validation set, the remaining five stations were used for training. The GP model gave the most accurate forecasts across the validation sets based on root mean square error and statistical evaluations. One of the study's contributions was using standardised forecasts and including a nonlinear trend covariate, which improved the accuracy of the forecasts. In order to improve the forecasts, a monotone composite quantile regression neural network and a quantile generalised additive model were combined. The results showed that Gaussian Spatial Temporal Regression produced superior results. Chapter 8 is based on the Conclusions and recommendations from the research.

Chapter 2

Literature review

2.1 Introduction

This chapter provides reviews of the main concepts that were used as the basis of this thesis. A thorough examination of other researchers' key and recent contributions to research periodicals was done. This chapter cited other researchers' views and findings in the field of Global horizontal irradiance, renewable energy forecasting and Gaussian Process Regression. This research was based on forecasting solar power generation using Gaussian Process Regression.

2.2 A review of renewable energy forecasting on South African data

Various authors have looked at forecasting renewable energy sources using different methodologies. A review of some of them follows. Artificial neural networks (ANN) are one of the techniques widely used to forecast GHI, and some of the approaches follow. Leholo et al. (2019) did a comparative analysis based on the training functions of artificial neural networks. The training functions used are scaled conjugate Gradient, Quasi-Newton, Levenberg-Marquardt and Resilient Backpropagation. Based on the mean absolute percent-

age error and root mean square error, Leevenberg-Marquardt's function performed better. The study was based on South African data. Daniel et al. (2020) forecasted wind speed using South African data. They used artificial neural networks, Bayesian analysis, Stochastic gradient boosting and generalised Additive models to predict wind speed. Additive quantile regression averaging proved to perform better in terms of validity, reliability, quality and accuracy. Their proposed methodology was very effective, and they found that a combination of point and interval estimation improved the forecasts. Exploring other machine learning algorithms would have improved the accuracy of the predictions. The two studies by Leholo et al. (2019) and Daniel et al. (2020) were based on Neural networks (NN), which has its disadvantages. NN are black boxes, meaning the technique does not give much information about the independent variables. A GPR model's results can explain the effects of the independent variables.

Time series analysis is also one of the widely used techniques, and some of the reviews follow. Milligan et al. (2003) investigated wind power. The analysis was based on United States data. They used autoregressive moving average models (ARMA) to forecast wind speed and power output. The performance of the ARMA models was investigated over different periods. Based on periods of 10 minutes, they predicted wind power. However, time series analysis handles the dynamics of sequences to predict future values. This modelling approach has its drawbacks. Using time series techniques is not easy to capture prior knowledge. GPR can handle prior knowledge, which makes it more robust. Dealing with a multivariate dataset is also very complex. In our proposed methodology, we used BSTS (Bayesian Structural Time series), one of the methods that dealt with this problem. GPR also deals well with this kind of problem.

Various techniques were also used in the past to predict energy. Chen et al. (2013) forecasted wind power by combining numerical and probability techniques based on a South African dataset. Gaussian process regression and numerical methods were applied to fore-

cast one day ahead of wind power. Gaussian process regression was applied to establish the relationship between power output and wind speed. The proposed method output was compared to the benchmark model, Artificial Neural Networks (ANN) and proved to improve forecasting accuracy based on Mean absolute error. Historical data was used to build the model. Gaussian processes were used to develop a speed power model, and the variables for numerical weather prediction were chosen using Automatic Relevance Determining (ARD). ARD is a method that fits a regression equation using Bayesian Ridge regression. Two ways were used to predict wind speed: GP direct and GP based on corrected speed. The results showed that the simulation study that improved the forecasts was the one with the corrected speed. The shortcomings of applying numerical methods are that the equations used by the models to simulate the weather variables are not exact, which leads to some errors in the forecasts. Another research by Chikobvu and Sigauke (2013) analysed the influence of temperature on electricity daily demand using South African data. They used generalised extreme value distribution and a piecewise linear regression modelling framework to fit the temperature data. The results showed that temperature is an important variable explaining electricity demand. The results produced by this study are robust when dealing with extreme events though when dealing with uncertain conditions, it is complex to work with. GPR models are nonlinear, can cater for many mean functions, and are interpretable. GPR captures uncertain conditions.

In another study, Sigauke (2017) forecasted hourly electricity demand on a medium-term basis. A generalised additive lasso model was used. The analysis was further extended to GAM-te-Lasso by including the tensor product interactions. The results of this model were compared to the gradient boosting method. Based on the pinball loss function, the GAM-te-Lasso performed better. The pinball loss function is a metric used to assess a quantile forecast's accuracy. The generalised additive model has a limitation in that it tends to overfit the predictions because it is nonparametric. The fusion of GPR, a nonparametric

method, with Lasso, a regularisation technique, deals well with this limitation. Another study by Adedeji et al. (2022) looked at the prediction of electricity consumption based on South African data. Comparison analysis between the adaptive neuro-fuzzy inference system and particle swarm optimisation was performed based on multi-campus university data. Adaptive neuro-fuzzy is a method that combines the Fuzzy Logic method and Artificial Neural networks. Fuzzy inference is a method that interprets the values in the input vector and is based on some rules that assign values to the output vector that use if-then logic operators. Neural networks are a collection of simple computational units that interlink connections. Based on the evaluation metrics, RMSE(root mean square error), MAD(mean absolute deviation) and MAPE(mean absolute percentage error), the adaptive neuro-fuzzy inference system model performed better. The methodology is efficient but has a challenge of computational complexity when applied to a large dataset with many variables.

Finally, Masike and Vermeulen (2022) did a study on predicting electricity demand based on time-varying elasticity. A time-varying and parameter-based model was used based on a South African dataset. The evolution of the elasticities was estimated using the Kalman filter. The results showed that the coefficient of elasticity increased when prices increased. The results also showed that consumers might move to other energy sources if the electricity price continues. The Kalman filter has a non-realistic assumption that the system and models are linear. Linear relationships do not occur in all living conditions. GPR captures nonlinear relationships. The next section looks at GHI forecasting literature.

2.3 Review of global horizontal irradiance forecasting literature

This section looks at reviews of studies that are based on GHI forecasting. Previous researches show increased studies that try to solve energy generation-related problems. This is due to the vast increase in energy generation to support the power grid, especially wind and solar energy. The increase in energy production has created the need for reliable energy forecasting since uncertainty will rise due to variability in the production of the demand energy sources. Various methods have been applied to forecast solar power. We shall review some of them.

The following authors used Artificial Neural networks to forecast neural networks to predict solar power. Artificial Neural Network(ANN) is a popular method for forecasting solar power. We will look at a few reviews on the application of Neural Networks. Cervone et al. (2017) applied ANN and an Analog Ensemble (AnEn) to come up with 72-hour forecasts of solar energy using input from a numerical weather prediction model, and they also used astronomical variables. They applied ANN and ANEn separately and then combined the two to develop forecasts for three different stations in Italy. The proposed technique is tested using synthetic data simulated power station. The results are based on US time. The results show that combining the two methods gave the best results.

A comparative analysis was done by Gbemou et al. (2021), Gaussian Process Regression, Support Vector Regression and Artificial neural networks were used for the analysis. A benchmark model called persistence on the clear sky index was also applied. The results showed that the GPR model using a rational quadratic kernel performed better. Their analysis was based on measuring the goodness of fit of the models used. Aslam et al. (2020) did an analysis comparing various models using solar power data. The comparison on neural networks, long short-term memory, gated recurrent units, support vector regression

and feedforward neural networks. The research is based on a Korean dataset. The results which implemented gated recurrent units is the one whose performance proved to be better than the other models though the performance of the other models was also good. The Gated recurrent unit is a recurrent neural network with specialised memory elements. It contains a reset gate that is replaced with an update gate to reduce the data's redundancy and complexity. The developed models outperformed the traditional methods, thus proving their efficiency. Applications of neural networks have their challenges. They tend to be black boxes. Using the method on different datasets and complex when dealing with large datasets will be difficult.

Other statistical methods have also been used to predict solar power. Mpfumali et al. (2019) did a probabilistic forecasting analysis on solar power prediction. The study was based on analysis twenty-four hours ahead, using data from the Tellerie radiometric station in SA. They applied pinball loss function and quantile regression averaging (QRA). Prediction interval coverage probability was used to select the best forecasts. The results showed that Quantile regression averaging produced robust results. The proposed approach produced results that optimised the integration of solar power into the national grid. Quantile Regression has many merits but is limited because estimating its parameters is harder to estimate than other regression techniques. Calibration and centred intervals' ability is limited by pinballs causing poor quantiles.

The following studies are based on deep learning techniques. Amarasinghe et al. (2020) applied an ensemble method which used deep learning methods to predict solar power forecasts accurately. They applied the method to data from solar power facilities in Germany. The weather parameters used in the study were selected using the feature selection technique, and the weather classification method was used to cluster the data. Different ensemble approaches are assigned to each cluster. An ensemble approach is developed based on prediction errors. The mean square error is used to evaluate the models, and the

proposed method performed better than the deep belief network, random forest regression models and support vector regression. The ensemble approach can be improved by applying more machine learning algorithms and using short forecasting resolutions like one hour. Gensler et al. (2016) did a study on deep solar power forecasting. They applied deep learning methods and artificial Neural networks like Auto-Encoder and Deep Belief Networks. They applied a combination of the stated method to forecast solar power and compared the results. The results of the deep learning methods proved to perform better than the Artificial Neural Networks and other physical methods. Dairi et al. (2020) used a variational Auto-Encoder (VAE) Deep learning technique to forecast solar power. VAE are probabilistic generative models that make use of neural networks. It generates new content, detects anomalies and removes noise from data. A short-term approach was developed to develop an accurate and efficient way of forecasting solar power. Single and multi-step ahead predictions were produced in this research. The data was collected from one plant in the United States and another from Algeria. The results of the VAE model were compared to Bidirectional LSTM(Long short-term memory), Convolutional LSTM, long short-term memory, Logistic and Support vector regression. Mabasa et al. (2020) studied six climatic zones from South Africa. Annual regression coefficients of the Angstrom-Prescott(AP) model were applied to solar data. This method was used because the network of the radiometric stations was on a landscape with dynamic climate and weather shifts. The results showed that the observed and estimated GHI values were strongly correlated. The proposed AP model proved effective in prediction, as was proved by the valuation metrics, the Mean Absolute Error and relative Mean Bias errors. These methods have been based on deep machine-learning techniques. The advantage of using deep learning techniques is that they are not dynamic, they will not perform well under unfamiliar conditions, and that is when conditions change.

Another study by Pattanaik et al. (2020) applied Genetic Algorithm(GA) to forecast

solar photovoltaics based on two important factors: temperature and amount of solar insolation. GA is an optimisation algorithm that works with an objective function band and its boundary value to find the most suitable solution. The study was done on a dataset collected from Odisha, India. The results showed that this method is accurate and convenient based on the goodness of fit, MAE, RMSE and MAPE measures. The research was based on an analysis based on two variables only, but other weather variables can also affect GHI, compromising the accuracy of the predicted values. The other variables that can be in our research we added other variables that also affect it.

A study by Lawin et al. (2019) looked at a variability and trends analysis of Burundi's solar power and temperature. Mann-Kendall and t-tests were used to detect trends and changes in the future climate. A standardised index was used for variability analysis. Variability was analysed monthly, and the results showed that solar irradiance was in excess in the dry season, which showed that it might be a high solar power production period. Another study by Sigauke et al. (2022) was based on modelling the dependence of Global Horizontal irradiance with temperature and relative humidity based on a radiometric station in South Africa. They applied multivariate adaptive regression splines, extreme value theory and copula models. Their results showed that temperature and relative humidity had a negative extremal dependence on GHI. The Copulas used are Clayton, Frank and Gumbel. Frank Copulas produced the best model. Lawin et al. (2019) and Sigauke et al. (2022) looked at predicting solar power based on one and two variables. They would have produced better results if they had included many weather variables. In this research, we included all variables we found in the literature.

In another study, Jang et al. (2016) looked at a model-based approach on various satellite images and support vector regression based on a Korean dataset. Satellite images were used on datasets inputs and outputs for Support Vector Regression. The proposed model's performance was compared with artificial neural networks and time series modelling and

performed better than the benchmark models. Mutavhatsindi et al. (2020*a*) looked at hourly solar power forecasting using South African data from Pretoria radiometric station. A comparative analysis was done based on predictive long-short term memory, feedforward neural networks and Support Vector Regression. Principle Component analysis was used as the benchmark model. They also used forecasts combination using quantile regression averaging and convex combination. Quantile regression averaging was the best-performing model based on Diebold Mariano and Giacomini-White tests. Jang et al. (2016) and Mutavhatsindi et al. (2020*a*) applied Support Vector Regression. SVR only works well on large datasets and when the dataset has much noise. So using this method on such a dataset compromises the accuracy of the models.

Prediction of solar energy can be made using time series models. Isaksson and Karpe (2018) did a comparative study using time series analysis against some learning machine methods on Sweden's solar data. They used an artificial neural network (ANN) and GBR. Their results showed that the application of time series was complicated because the energy data was nonstationary. The two machine learning techniques proved to be easier to work with. Time series modelling is very effective in terms of simplicity and accuracy. However, the major problem with this technique is that it is difficult to generalise the results from one study to another.

Finally, Sharma et al. (2021) applied quantile regression (QR) to predict solar power on the Australian dataset. They used long short-term memory (LSTM) and a fully connected neural network. QR produced better results than the polynomials, but its results were similar to those of the fully connected neural network. Quantile regression has properties similar to GPR, but its major drawback is that estimating its parameters is hard. The section that follows looks at a review of GPR techniques.

2.4 Review of studies applying Gaussian process regression on renewable energy

This section is based on a review of literature based on GPR. Leith et al. (2004) forecasted electricity load using GPR using an Irish dataset. They used GPR models based on Basic Structural Models (BSMs) and Seasonal Auto-Regressive Integrated(SARI). Their research showed that BSM and SARI-based models improved the accuracy of the traditional GPR using MAPE as a measure of accuracy. The use of GPR coupled with structural models is very noble, although it has its limitations. Structural models simultaneously apply various relationships, leading to the methodology being computationally intensive and the computation of iterative algorithms to reach optimal solutions, resulting in parameter estimation problems Hair et al. (2021). Another study by Lourenço and Santos (2010) applied GPR models as STLF(Short Term Load Forecasting) method to distribution systems of forecasting electricity. The study considered studying input predictors from contiguous values to include the strictly necessary instances of endogenous predictors. Real data from 3 electricity substations in Portugal were used for the study. The proposed results proved accurate in forecasting electricity for different climatic areas. The limitation of this methodology is that because of the short prediction horizons taken in modelling, it is difficult to perform the predictions using a larger dataset. The predictions can only handle a dataset for a few months because of the forecasting horizons, which might compromise the accuracy of the predicted values.

Chen et al. (2013) applied Gaussian processes on the outputs of Numerical weather prediction. They forecasted wind power using Numerical weather prediction to correct censored GPs using a dataset from South Africa. The results of the proposed method were compared to Artificial Neural Networks and the persistence method. The proposed model produced an 11% improvement in forecasting accuracy compared to ANN and 5% compared

to the persistence model. The advantage of applying numerical weather forecasts is that the equations produced to simulate GHI need to be more precise.

In another study, Tolba et al. (2019) used GPR to forecast GHI over 30min to 5h based on a dataset from France. GPs with different kernels have been applied on two datasets of 45 days. The results showed that when applying the quasiperiodic kernel, the GPR outperformed the persistence and GPR models based on other kernels. The performance of the GPR would have been better if various kernels had been explored. Ma et al. (2020) used a physics-informed Gaussian process technique to forecast and estimate wind generation using a US dataset. Stochastic equations were used to compute prior statistics for power grid dynamics. The equations called swing equations were solved with Monte Carlo Simulation techniques. The results showed that the Phi GPR model accurately forecasted wind power. The Phi GPR also produced more accurate results than the Autoregressive integrated moving average. The flaw of this methodology is that the application of Monte Carlo simulation is computationally inefficient. It could perform better when there are few variables.

Lubbe et al. (2020) developed a Gaussian process regression model to predict solar radiation based on South African data. They were concerned with developing an appropriate kernel for the GP for proper interpolation and prediction of GHI data. The study was done using a forecasting horizon of 5 days. The interpolation and prediction were done using a multi-in-single-out GPR adopting the Radical Quadratic and Periodic kernels on hourly averaged data. The results showed that the combination of the periodic and rational kernels produced the best results. Their results may have been better if they had tried different kernels and chosen the best. Cao et al. (2021) applied Gaussian Process regression on electricity load forecasting using a limited dataset from Italy. They used a probabilistic approach to Anomalous events allowing extreme scenarios to be dealt with, thus causing short-term load behaviour changes. A probabilistic approach for anomalous

events is a probability method that models outliers with spatial and temporal anomalous behaviour. A comparison analysis was done using support vector regression, Double Stochastic Variational, Backpropagation neural inference networks, recurrent Neural Networks and convolution Neural Networks. Support Vector regression is a machine learning algorithm used to solve classification problems, separating two data classes. It aims to find a hyperplane that classifies data inputs. A double stochastic variational method is used to find a posterior means approximation using a Monte Carlo evaluation of the expectations. Backpropagation neural inference networks are a process of tuning neural networks to improve the accuracy of predictions. Recurrent Neural Networks is an artificial neural network that uses sequential and time series data. Convolution Neural Networks are distinguished from other neural networks by their superior performance with image, speech or audio signal inputs. The proposed method outperformed the other benchmark models.

Zhang et al. (2021) looked at the probabilistic prediction of solar power using K-means time series clustering. A comparison was done between k-means time series clustering and astronomy and K-means methods. Various Gaussian Process kernel functions were compared, and the best one was selected using deterministic and probabilistic forecasting. The research used data from Tibet province, China, to verify the validity and practicability of this research model. Two benchmark models were used, Artificial neural networks and Support Vector Regression. The results of the GPR model based on k-means time series clustering performed better than the other models. K means to have limitations; they do not perform well when clusters have different sizes.

Ferkous et al. (2021) used a Gaussian process model coupled with a wavelet model to predict solar power. An Algerian three-year dataset was used. Various wavelets and data combinations were used based on relative humidity, temperature and solar radiation. The results showed that the proposed method performed better than the ordinary GPR based on root mean square error (RMSE), relative root mean square error (rRMSE), mean

absolute error (MAE) and determination coefficient (R^2). Although Wavelet is a powerful tool, it has the disadvantage that it does not work well on a larger data set and has much redundancy.

Gbémou et al. (2022) developed a time-based GPR model and an observation-based GPR model to forecast GHI. A comparison is done between a horizon-specific and multi-horizon-specified model. A scaled persistence model is used as a benchmark model. The horizons used are 10 minutes, 1 hour, 3 hours, 5 hours and 24 hours. The best model chosen was the one with a quasiperiodic kernel and time-based. On horizon based, the best model chosen is the one that used the automatic determination rational quadratic kernel. Finally, a two-staged model was developed by Mitrentsis and Lens (2022). The Natural Gradient Boosting method was first developed to develop probabilistic predictions, and then the Shapley additive explanation was calculated. The research made use of using a dataset from Germany. A comparison analysis was done based on two Algorithms, the Gaussian Process and lower bound estimation, based on overall model performance. The forecasting accuracy was evaluated using interactive and force plots. The results showed that the proposed model had higher accuracy, and the predictions were sharper. This method has its limitations, the application of Shapley additive is computational inefficiency, and when the model is not additive, the results may be misleading. The use of Natural Gradient Boosting tends to overfit the data. This means that the use of these methods distorts the accuracy of the models.

2.4.1 Twenty-four-hour ahead probabilistic global horizontal irradiance forecasting using Gaussian process regression

Probabilistic forecasting of power grid states promotes energy use and planning management. The measure of the use of renewable energy sources like solar aims at lowering effects such as that greenhouse gas emissions. This harms climatic changes and positively

impacts financial costs associated with using other energy forms like thermal or hydroelectric power. The impact is visible, especially when predictions are based on short intervals. Short-term forecasting has a greater impact on the safety and financial implication of the electric grid. Regardless of considering short terms in predicting energy, it is necessary to use efficient statistical techniques. This study uses Gaussian processes (GP) to predict solar power because solar power has a stochastic and uncontrollable nature. Applying Gaussian processes requires the computation of the covariance functions, also called a kernel and a key element that deeply influences the forecasting results (Jäkel et al., 2007). Hence, selecting an efficient method to develop a more accurate kernel is critical. Thus, the MEB (Minimum Enclosed Ball) technique was applied to select the best kernel used in Gaussian Process Regression. Minimum Enclosed Ball Technique is a core vector regression class expected to improve the forecasting results. The proposed Gaussian Process Regression coupled with core-vector methodology is believed to produce more accurate forecasts than the conventional benchmark models.

Various authors have ventured into solar power forecasting in areas such as numerical weather models, time series regression, artificial intelligence, and many more. Authors like Raza et al. (2016) reviewed Photovoltaic forecasting methods and a review of factors affecting predicting it. The factors considered are its yield control profile and execution matrices to assess the predicted model. They discussed factors such as forecast horizon, forecast model performance, input data pre-processing, and prediction model inputs that affect PV forecasting. Various methods were reviewed. Those include Artificial neural networks (ANN), persistence models, and Radial Basis Function Neural Networks (RBFNN). Their research was mainly based on time series regression techniques and artificial intelligence. The time series methods that were reviewed are the Autoregressive(AR), Moving Average(MA), Autoregressive Moving Average(ARMA), and Autoregressive Integrated Moving Average(ARIMA). The data used for this research was collected from a university in Aus-

tralia, Queensland. The methods used are good prediction techniques, but they have flaws, some of which we have mentioned in earlier reviews. This makes them less superior because they cannot handle datasets' uncertainties. In Hong et al. (2013), the authors did a study on introducing Global Energy Forecasting in a competition in 2014, a probabilistic estimation of energy using four tracks for the following variables price, wind, load, and sunlight energy-based predictions.

The study aimed to summarise decade-ahead probabilistic forecasts of competition results. Various methods were used for wind power forecasting for the competition. The methods included: Gradient boosting regression, regularised least squares linear regression, neural networks, AR Models, Sparse Bayesian technique, and kernel-based regularised regression. The persistence model-based random walk model was used as the benchmark model. A day ahead prediction of sunlight-based power yield extracted from solar plants from Southwest America was done by Larson et al. (2016). Predictions were done for the day ahead hourly averaged energy output from solar power plants based on weather factors. The predictions made use of least squares optimisation of Numerical weather. Three different numerical variations for prediction strategies were assessed for four years against data from two tracks 1MWp plants in California. The results of the proposed method were compared to the benchmark model, the persistent model. The results showed that the model performed better than the benchmark model. Spatial averaging was done between two meteorological stations, and the results showed reduced error.

Kernel density approaches were applied by Trapero (2016) using a dataset from Spain. The application of the methods was compared to instability prediction models. Combining the two techniques was also done with the main thrust of enhancing the predicting interim performance. An analysis was done using the two methods based on nonparametric kernel density estimation on a minute-to-minute solar irradiance. GARCH (Generalized Autoregressive Conditional Heteroskedasticity) and SES(Single Exponential Smoothing)

estimations incorporated volatility forecasts. The results showed that the combined approach to forecasting produced good results based on prediction interval tests. Ranganai and Sigauke (2020) applied Additive quantile regression to model global horizontal irradiance. Predictions were done based on data from three radiometric stations from South Africa. Models based on long-range dependent models, models incorporating seasonality components, hybrid models, and those incorporating stochastic components were used.

The models were combined with quantile regression in order to improve their robustness. The results showed that the models with combined forecasts had better levels of accuracy based on the Diebold-Mariano test. Furthermore, in Govender et al. (2018), the authors looked at the clustering of solar irradiance patterns. The predictions were related to cloud cover and were done using numerical data of solar power or irradiance for the following day. K-means clustering was used to develop diurnal patterns based on weather conditions. Predictions were based on two methods. One was based on k-means clustering of daily cloud cover profiles predicted daily. The other method was based on profiles of cloud cover average being above or below 50%. The results showed that the performance of cloud cover was better on sunny and cloudy days, for the 50%

Amarasinghe et al. (2020) used a generalised ensemble model incorporating deep learning methods to develop solar power predictions. The data is clustered using the weather classification approach, and each cluster is assigned an ensemble approach. This method's performance was compared against the benchmark models, support vector regression, Deep belief Network, and random forest regression models. The results showed that their proposed method performed better, and results with clustered data gave the best results. Maritz et al. (2018) applied GPR for energy predictions using the Bayesian framework. A Gaussian Process Regression framework was constructed for measurement and verification. The periodic kernel multiplied by the squared exponential kernel was selected as the most appropriate one. The methodology adopted could have produced superior results if the

horizon of forecasts were hourly based. Also, adopting a more efficient kernel selection method and a multi-dimensional analysis would have added more value. Dahl and Bonilla (2019), the authors worked on a study applying Gaussian process models to forecast solar power for 37 residential sites in Adelaide, Australia. They used an integrated multi-site model without using prior data detrending. This was achieved by capturing diurnal cycles. They applied to multitask GP models with linear observations with several latent node functions 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 the benchmark models used. It discovered that the multi-site modelling was better than the single-site methods with varying weather conditions. In Tolba et al. (2019), the authors did a forecasting study on GHI by applying GPR based on kernel study. The dataset used in this research is from southern France. They applied several kernels and found that the quasiperiodic kernels outperformed most. The results generally showed that the GPR performed better than the persistent model. GPR structure produces better results through omnipresent periodic components. Reviewing these studies has shown that much research has been done using various time series approaches, numerical analysis, neural networks and deep machine learning techniques. For research done on Gaussian Process regression in South Africa, many things needed that we had to work on. These improvements include using multi-dimensional analysis and adopting a superior method for kernel selection, among others.

2.4.2 Probabilistic forecasting of global horizontal irradiance using ensemble GPR-AQR-BSTS

Over the past years, South Africa has suffered an energy crisis due to climatic changes. This has led to a major provider of electricity, Eskom, opting for other energy sources like solar power. Solar power is another source of energy that is clean, in-exhaustive, sustainable, and

extremely abundant, relating to different energy sources. Solar energy is highly preferred because it is cheap and does not pollute the environment. Accurate Global Horizontal Irradiance (GHI) forecasting has become an important part of the energy management system. It ensures its integration with the economy's security since solar power generation is directly connected to the control of the power grid (Yang et al., 2013). The United Nations, Sustainable Development Goal (SDG) number 7, "Affordable and clean energy" Nations. (2021), encourages the use of renewable energy as a solution to mitigating climatic changes. The goal encourages the utilisation of renewable energy sources such as solar and wind because they are affordable, reliable, and appropriate.

Thus, it is crucial to encourage the use of solar power and forecast it accurately. This calls for efficient ways of predicting the commodity. This research uses a probabilistic modelling approach to forecast short-term solar energy generation. A comparison is done between Bayesian Structural Time Series (BSTS) and the Gaussian Process Regression coupled with Quantile Regression. BSTS is a high-performing fitting algorithm that provides probabilistic estimation. The advantage of using this method is that the Bayesian technique in the multivariate analysis helps to avoid over-fitting. The results from BSTS were compared to the Gaussian Process Regression (GPR) results. BSTS is a feature selection technique and time series forecaster (Brodersen et al., 2015).

Over the past few years, researchers have been looking for machine learning techniques that perform accurately. Various methods have been proposed to predict multivariate solar power generation at different scales. Methods explored to date are Linear regression, gradient boosting regression, neural networks, time series analysis, and support vector regression.

Brodersen et al. (2015) and Peters et al. (2017) looked at a method applied to a multivariate dataset with various components. This method better handles uncertainty, a key aspect of predicting future values. It also eliminates the complexities of the Bayesian approach

because of its ability to handle unpredictable data. On the other hand, GPNN(Gaussian Process Neural Networks) was used since the parameters of this model are estimated using probabilities and predictions. GPNN enables updating uncertainty values of forecasts effectively by removing overconfidence on out-of-sample points by neural network retraining, thus improving uncertainty prediction.

Several authors have researched Gaussian regression, Gal et al. (2017), Hafner et al. (2018) and Yu et al. (2008) among others. Gal et al. (2017) combined Bayesian deep learning into the active learning framework by developing one with higher-dimensional data. They applied image data on Bayesian convolutional neural networks. Thus, there was a great improvement in the RMSE and the loglikelihood of existing learning approaches. Hafner et al. (2018) did a study whose main goal was to improve the specification of the noise contrastive prior by applying the contrastive noise priors to enable reliable predictions. They applied the method on a flight delay dataset, which produced results that improved scalability and avoided overfitting outside the training dataset. Yu et al. (2008) 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. Słoński (2011) did a comparative analysis on feedforward layered neural networks on stochastic and Gaussian Process analysis. The research showed that the two had similar forecasting accuracy and better performance than the linear regression model. Al-Shedivat et al. (2017) applied long, short-term memory recurrent networks coupled with GPs. They used marginal likelihood estimation on a convergent semi-stochastic gradient method and exploited how kernels should be structured for forecasting using expressive closed-form kernels. The models are applied to self-driven car applications, system identification, and power forecasting, and the results showed that the proposed method is efficient and convergent. Gal and Ghahramani (2016) developed a modelling framework adopting dropout neural networks using Bayesian inference. They assessed many non-linearities

in regression and classification using neural networks. Tsymbalov et al. (2019) applied a dropout-based model based on Bayesian Regression and Neural network coupled with Gaussian Process on chemical and real physical life. The results showed that the methods performed well for neural architectures involving dropout and Bayesian neural networks.

Chandiwana et al. (2021) did a study on predicting Global Horizontal Irradiance 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 2 benchmark models: Support Vector Regression and Gradient boosting regression. Their results showed that the proposed methodology was the best compared to the benchmark models.

Prediction of solar energy can be done using time series models. Tsymbalov et al. (2019) did a comparative study using time series analysis against some learning machine methods on Sweden's solar data. They used the learning machine methods of Artificial Neural Networks and Gradient Boosting Regression. Their results showed that the application of time series was complicated because the energy data was nonstationary. The two machine learning techniques proved to be easy to work with.

Sharma et al. (2021) applied Quantile Regression which they used on long-short term memory neural network and polynomial and a fully connected neural network. The results showed that Quantile Regression produced better results than the polynomials, but its results were similar to the fully connected neural network. Finally, Palm et al. (2022) applied Gaussian Process Regression modelling to optimise the design of power systems. The application was based on multi-objective optimisation. A mathematical algorithm was designed to come up with the modelling framework. The Pareto front algorithm, NSCA-11(Non-dominated Sorting Genetic Algorithm), and pure Latin Hyper Cube Sampling method were used. The proposed method proved to be better in terms of efficiency and effectiveness.

Most methods highlighted in the literature do not address the issue of uncertainty. Gaussian process regression addresses that aspect and the combination of forecasts, im-

proving forecasting accuracy. Thus, we propose this methodology since we believe it will produce superior results.

2.4.3 Spatio-temporal forecasting of global horizontal irradiance using Bayesian inference

The use of energy sources that are clean and renewable has been on the rise, hence the need to efficiently manage the power grid. This increase has led to the need to come up with predictions of the available energy sources to ease the management of the power grid. Power grid planning is a challenging process requiring efficient and accurate power prediction inputs. This research focused on probabilistic, short-term forecasting to model GHI efficiently. Probability forecasting is an emerging research area that handles weather variables' uncertainty, producing more comprehensive results Hong and Shahidehpour (2015). A probability approach, Gaussian Process regression, was combined with spatial analysis to assess the relatedness of meteorological stations. Spatial analysis is a very important technique in regression, and in this study, it was used to explore spatial dependence between various meteorological stations. Spatial analysis was used because it can solve complex location-based problems. Some authors have applied various methods incorporating spatial analysis. Reviews that follow look at a discussion of some authors who applied spatial analysis to predict solar power.

The world meteorological organisation produced a guide highlighting the major sources of uncertainty in forecasted values related to the weather as the process of producing a forecast, its interpretation, and atmospheric unpredictability (Gill et al., 2022). Most of the studies reviewed in this research on forecasting global horizontal irradiance (GHI) using spatial regression considered only the spatial feature, leaving out weather variables and the variability in GHI. Most classical approaches for predicting GHI rely mainly on a single power plant. The current study uses a multi-site approach to modelling and forecasting

GHI using Gaussian process models, including a nonlinear trend covariate.

André et al. (2016) used a spatial–temporal model for short-term solar irradiation forecasting. They used a spatio-temporal vector autoregressive model, which was designed in such a way that it could handle sparse spatiotemporal data. An iterative strategy in the model process selected related stations and eliminated insignificant predictor variables. This approach is a linear approach to forecasting. It would have been better if they had applied a nonlinear approach. It would have produced superior results. This research used a nonlinear approach to forecasting, giving it superior results.

In a study by Liu et al. (2019), an ensemble-temporal deep learning approach incorporating multi-sites in a spatial power grid was used to predict solar power. Variational modelling was used to predict uncertainty, and the proposed methods estimated uncertainty very well. Although Bayesian analysis is a powerful tool in forecasting, the approach has a disadvantage: the variational inference uses the variance of the underestimated real posterior. The approach used in this research tends to look for a posterior that optimises the results through covariance functions used in Gaussian Process regression. Yang (2018) used the ultra-fast pre-selection method to solve the lasso problem, that of having an insufficient degree of freedom and curse of dimensionality. The variables selected via ultra-fast were then used to develop a lasso-temporal model. Their results showed that their proposed algorithm did not need meteorological priors and provided the best forecasts. The main idea behind this research was to develop an effective subset of variables for forecasting, which is crucial in modelling. Our approach also tried to produce subsets of variables that efficiently predict GHI by using various variable selection methods and selecting the best. We further included other methods of improving the predictions, like forecasts combination, standardising forecasts and applying superior ways of selecting the covariance function.

Another study by Eschenbach et al. (2020) was based on forecasting solar irradiation using various machine-learning methods. The methods used were ARX (autoregressive

with exogenous inputs), NN (Neural Network), RRF (random regression forest), and RT (regression trees). Their results showed that NN produced more accurate results for a short lead time and dense-temporal input data. The approach is based on nonlinear approaches to modelling. They would have produced better results with a nonlinear approach. The NN method produced better results. The methods we used also adopted it as a benchmark model, proving to be inferior to our proposed methodology.

Kim and Suh (2020) applied a temporal method that combined public and satellite data. Based on a satellite from South Korea, SVR (Support Vector Regression), ANN (Artificial Neural Network), ARIMAX (Autoregressive Moving Average), and DNN (Deep Neural Network) were used based on short interval forecasts of GHI data. Their results showed that the models based on temporal and spatial characteristics gave better forecasts than the other models that were based on numerical data using weather variables. Time series analysis is a very simple forecasting tool, although when the data has many predictor variables and complex relationships, it tends to have no techniques to handle them. This research applied Gaussian process regression, which deals very well with complex relationships. Neural networks and SVR are not suitable for larger datasets, unlike the methods used in this thesis.

Zhang et al. (2023) applied a spatio-temporal Gaussian process state space model with the Kronecker structure on weather data from Colorado and Global Historical Climatology Network. The other objective was to develop a kernel that will be used for the Colorado and GHCN (Global Historical Climatology Network) data. To estimate the hyperparameters of the spatio-temporal Gaussian model, the Kalman Filter and smoother were used. The results showed that the forecasting performance improved from the results by Todescato et al. (2020) based on the desirable properties of the Matern kernel used for both datasets. Zhang et al. (2023) applied GPR based on Kronecker structure, but we decided to adopt a different approach coupled with core vector regression, spatial analysis and quantile regression.

Hamelinjck et al. Hamelijnck et al. (2021) applied variational spatio-temporal regression coupled with Gaussian processes. A non-conjugate GP technique produced a sparse state model using separable Markov kernels. The results of the proposed methodology proved to be more accurate and efficient due to filtering parallelisation of spatial locations, sparsity, and application of variational Gaussian process analysis. The proposed method outperformed the baseline methods.

Agoua et al. (2018) forecasted photovoltaics using a probabilistic spatio-temporal model approach using datasets from nearby plants. This technique incorporates short dry forecasting of periods of 0–6 h. Quantile regression was combined with Lasso, which was used for variable selection. The proposed methodology showed superior performance compared to KDE (Kernel Density Estimation), which was used as the benchmark model.

Banerjee et al. (2008) developed a temporal, spatial model using predictive process modelling. The method reduced computational burden by reducing the modelling space to a lower dimensional subspace. The results of the proposed methodology addressed the problem of misspecification of the model using a larger dataset, which was achieved by applying the induced specification predictive process. The main aim of their study was to reduce dimension to allow likelihood evaluation and to allow predictions based on simulation to avoid misspecifications, which produced fair results. Applying Bayesian techniques and probabilistic and simulation approaches is also crucial. The combination of these tools surely resulted in superior results.

Luttinen and Ilin (2009) did a Bayesian analysis using sea surface temperature data. They used a probabilistic factor analysis using spatiotemporal data. Gaussian process priors were used for the factors and loading matrix. According to their results, the Gaussian Process Factor Analysis outperformed the Bayesian Principal Component analysis. A drawback in their analysis will be caused by using factor analysis. Factor analysis cannot produce meaningful output if variables are not correlated and assume that the variables are

linear. This makes their methodology inferior.

Tomizawa and Yoshida (2022) applied Gaussian Process regression with various Gaussian random fields to data with problems of spatial variability. The maximum likelihood method was used to estimate the random and fluctuating fields' scale. Their results showed that the model with the Whittle Matern kernel was the best for the random component. Comber et al. (2022) used Gaussian Process splines regression considering the variational of geographical areas. The technique is a smoothing parameter used in splines regression combined with Gaussian Processes, optimising the GP splines regression. The model showed predictions that were more accurate because of accommodating heterogeneity.

Another research was done by Najibi et al. (2021*b*). They used probabilistic GPR to forecast solar power using meteorological data. Short-term forecasting was used based on k-means clustering. A Matern 5/2 covariance function was used. A 5-fold validity test set, holding out 30 random days, validated the applied method. Root mean square error was reduced due to the application of the proposed methodology. This might have been caused by the inability of the method to capture outliers. In another study, Najibi et al. (2021*a*) predicted solar power using weather variables. Eight different partitions were used to cluster the data using k-means clustering, and GPR was used based on the Matern 5/2 covariance function. The Elbow and Gap techniques were used to develop optimal clusters, and the results showed that the forecasting error was reduced. The use of k-means clustering might make the accuracy of the predicted values to be low because it does not capture outliers.

Mendonça et al. (2020) performed a comparative analysis on multilayer perceptron (MLP) artificial neural networks and multigene genetic programming (MGGP). The assessments indicated that MGGP gave more accurate and fast results in single predictions, and the ANN performed better for ensemble forecasts. Wang et al. This method has its drawbacks, the major one being that application of Neural networks is complex when dealing with larger datasets. Another drawback of the methods applied is that the multi-

genetic programming uses a linear approach to modelling, which makes it inferior to the GPR, which is a nonlinear approach. Wang et al. (2022) used a cluster-based analysis on ultra-short-term wind power using a hierarchical directed graph method and dynamic–temporal correlation. They first defined three nodes based on wind power, wind speed, and target nodes. They defined temporal-based input samples and correlation matrices to evaluate the correlation of neighbouring wind farms. They also used directed edges to connect various nodes to obtain a hierarchical-based graph form, which was later applied to train the prediction model. The proposed model outperformed the other benchmark models used. The method proved superior to the benchmark models used, but the method is very complex since graphs use many pointers.

A summary of previous studies on modelling solar radiation based on spatial analysis is given in Table 2.1.

Table 2.1: Summary of some previous studies on temporal dependence modelling of GHI.

Ref.	Data	Models	Main findings
André et al. (2016)	Solar power data	Temporal vector autoregressive model	Results show that the significant best order was equal to 1.
Yang (2018)	Solar irradiation data	Ultra fast pre-selection algorithm	The method produced results superior to those that adopted classical methods.
Liu et al. (2019)	Solar irradiation data	Variational Bayesian conditional gate recurrent unit network	Results of the study provided efficient uncertainty estimation of solar irradiation prediction.
Kim and Suh (2020)	Numerical weather data and satellite-based images	Spatial, temporal prediction using ARIMAX, SVR, ANN, and DNN	The results based on spatial data produced results that outperformed those based on numerical data.
Zhang et al. (2023)	Temperature data	Gaussian Process Analysis, Spatial-Temporal Analysis, Kalman Filter and smoother	There was a confirmation of improvement of model predictions performance through spatial analysis.
Hamelijnck et al. (2021)	Air quality data	Variational GP, Markov kernel analysis, Spatial-Temporal Analysis	The results produced showed a great improvement as a result of combining variational inference and spatial-temporal filtering.
Agoua et al. (2018)	Photovoltaic plants data	Lasso, Quantile Regression, Spatial Analysis, and Kernel Density Function	The model produced results that were superior to those of KDM.
Banerjee et al. (2008)	Point referenced biomass data	Gaussian Process Regression, Hierarchical modelling, Markov chain, Monte Carlo methods	Findings addressed the problem of fitting hierarchical spatial modelling on large datasets.
Luttinen and Ilin (2009)	Sea surface temperature data	Gaussian Process Regression, Factor Analysis, Principal Component Analysis	Findings proved that the performance of GPR improved as a result of combining it with factor analysis.
Tomizawa and Yoshida (2022)	Real ground measure dataset	Gaussian Process Regression, Gaussian random fields Analysis, Markov Analysis	The results showed that GPR with random fields using a Whittle Marten produced the best results.
Comber et al. (2022)	Social, economic data	Gaussian Process Regression, Spatial Analysis, Splines Regression	Predictions were improved by making use of GAMs model that accommodated spatial heterogeneity.
Najibi et al. (2021*a*)	Photovoltaic data	Gaussian Process Regression, feature selection, k-means clustering	The GPR model's prediction accuracy improves from the k-means clustering.
Najibi et al. (2021*b*)	Photovoltaic data	Gaussian Process Regression, k-means clustering, Elbow and Gap method	Optimal clusters used are four, and forecast error was reduced by optimising the clusters.
Wang et al. (2018)	Wind	Graph structured spatial-temporal analysis	The results indicated that the method outperformed the benchmark models.
Eschenbach et al. (2020)	Solar irradiance data	ARX, NN, RRF, and RT	Results showed that the short term is more effective on data from a dense enough network.

2.5 Conclusions from literature

This chapter provided a critical review of the literature survey. The literature explored and assessed the work of various authors on Global Horizontal Irradiance and its predictor variables. Applications of Gaussian Process Regression on solar power and renewable energy sources and the forecasting of GHI using other methods were also reviewed.

In summary, a huge variety of Global Horizontal Irradiance forecasting methods has its strengths and weaknesses. Regardless of these methods being used on other datasets comprising different characteristics, it is impossible to determine the most effective method of forecasting GHI. A comparison of various models and methods is necessary to assess the effectiveness of the predictions.

Many GPR studies have been adopting a methodology where various kernel functions are applied, and the best method will be selected based on measures of goodness of fit, Gbémou et al. (2022); Aslam et al. (2020); Leith et al. (2004). The method containing the best measures will be regarded as the one containing the best kernel function. This research has taken a different approach where the kernel function is selected first, and then that kernel function will be used for GPR forecasting. The method that was used is Minimum Enclosed Ball, which is a branch of Corevector regression. The selection of the most appropriate kernel will make the GHI forecasting effective. The effectiveness of the predictions of the model produced this way was compared to the benchmark models used.

Most of the GHI forecasting studies are based on point estimates, Leholo et al. (2019); Cervone et al. (2017); Amarasinghe et al. (2020). To improve the modelling framework of GPR, quantile regression was also used for forecasting. By doing this, it broadens the forecasting accuracy. This improves the predictions made.

Most researches on GPR forecasting is based on analysis using a single site, Mpfumali et al. (2019); Dairi et al. (2020); Pattanaik et al. (2020). We had to explore GPR further,

including multiple sites. This was done in order to capture spatial dependencies between meteorological stations.

From the reviewed literature, we can see that little has been done in South Africa regarding using the Gaussian Process in predicting energy forecasting. This research's main thrust is on effectively predicting solar power usingFrom the reviewed literature, we can see that little has been done in South Africa regarding using the Gaussian Process in predicting energy forecasting. This research's main thrust is on effectively predicting solar power using GPR in South Africa.

For various reasons that have been mentioned, it is desirable to make use of the proposed innovations in order to improve GHI prediction. The following chapter is going to discuss the methods that were used in this research.

Chapter 3

Methodology

3.1 Introduction

This chapter discusses various methods that were used to achieve the objectives that were set in this research. The variable selection methods, prediction methods and evaluation metrics used in the articles published for this study were discussed.

3.2 Gaussian process regression

Gaussian process regression is a concept named after Carl Friedrich Gauss because it is based on Gaussian distribution. It was originally developed as a probability technique by Wiener and Kolmogorov in the 1940s. The technique originated in geostatistics as kriging Krige (1976) and was later used in spatial statistics by Ver Hoef and Cressie (2020) and in some machine learning researches by Williams and Rasmussen (2006); MacKay and Neal (1997). Gaussian processes are applied for classification and non-linear regression Williams and Rasmussen (2006).

Gaussian Process Regression (GPR) is the main method explored in this research. It is a machine-learning technique used to solve regression and classification problems. GPR is a class of nonparametric Bayesian Machine learning regression methods incorporating

prior information. The prior information is further converted to a posterior distribution. Due to the uncertainty and complex nature of Global Horizontal Irradiance, we choose to forecast it using GPR. The model describes a time-based variable X with a function $f(x)$ that does not relate to any specific distribution. A GPR is built over the assumption that it comprises a collection of continuous random variables $X_1, X_2, ..., X_n$, with n observations of the joint random variables. The components of a GPR are a combination of a covariance function, conditional joint distribution and multivariate normally distributed variables. A mean and covariance function describes the model. The mean function $m(x)$ is assumed to be equal to zero. A GPR is given in Equation (6.2.7). It is described over a real domain of X over a function $f(X)$.

$$f(X) \sim GP(m(x), k(x, x^1)), \tag{3.2.1}$$

where X is a subset of independent variables $X_1, X_2, ..., X_n$, that is meteorological variables, namely relative humidity, temperature, wind speed, and air pressure. $f(X) = f(X_1), f(X_2, ..., f(X_n)$, x and x^1 are any input points(locations) of any one of the variable X. $f(X)$ is a function of the dependent variable, GHI. $m(x)$ is the mean vector of μ, that is, $E(f(x)) = m(x)$, $k(x, x^1)$ is the covariance matrix, also called the kernel function. Various kernel functions can be used when using GPR. We will consider them in the next section.

GPR was used in this research because it has several advantages. It gives probabilistic predictions, is versatile and flexible, and handles uncertainty in its predictions and complex conditions. Another important property of GPR models is that it is nonparametric. This allows the dataset to explore itself without imposing a specific distribution. As a result, the model is more accurate than parametric models and allows them to learn the training dataset freely without limitations. GPR provides uncertainty measures over forecasts Verrelst et al. (2013). This makes them flexible on many datasets. They also use prior in-

formation as kernel functions; this allows the modelling technique to handle the uncertainty component in the dataset.

Another advantage of GPR is that it is probabilistic. GPR makes use of probabilistic techniques when making predictions. One will be able to decide the appropriate region to fit the model. A GPR model is versatile Erickson et al. (2018). That is, it can develop different functions under different conditions. Various kernels are used depending on the nature of the explored variable. Custom kernels can be used, but the kernels can be explored using a different method and the appropriate one selected, Williams and Rasmussen (2006). It is flexible for modelling data with heterogeneous conditional distribution. It is more robust to outliers than other techniques of regression. The method can handle certain outlier data points effectively, Williams and Rasmussen (2006).

3.2.1 Bayesian approach to parameter estimation

The Bayesian approach for parameter estimation is used in this thesis since it gives more full-density information than the MLE, and the framework is based on Bayesian assumptions. The Bayesian approach infers the distribution of the data over all the values that fit the data.

When using the Bayesian approach to estimation, we first consider the information regarding parameters. This includes considering past related data or related studies, Marlin et al. (2003). The modelling process makes use of a prior distribution, given by $\pi(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the space of the parameters, that is $\boldsymbol{\theta} = \{\theta_1, ..., \theta_p\}$. Secondly, we collect data and use it to come up with a likelihood function, $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ where $\boldsymbol{x} = \{x_1, ..., x_n\}$. The likelihood relates to how the data $\boldsymbol{x}$ hangs on the parameter $\boldsymbol{\theta}$. The third step is to combine the likelihood function with the prior distribution to come up with the posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{x})$, which is the distribution of the parameter $\boldsymbol{\theta}$ given the data $\boldsymbol{x}$. The posterior distribution

is given in equation (3.2.2).

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \quad = \quad \frac{\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} \pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{3.2.2}$$

which can also be given as

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \quad \propto \quad \pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{3.2.3}$$

The posterior distribution is proportional to the likelihood function multiplied by the prior distribution. The normalising constant is the denominator in equation 3.2.2, and the Markov Chain Monte Carlo (MCMC) techniques are used to compute it. The statistical inferences in the research will be based on the posterior distribution because it gives all the information on parameters (Marlin et al., 2003). For example, point estimation gives the mean in equation (3.2.4),

$$E(\boldsymbol{\theta}|\boldsymbol{x}) = \int_{\boldsymbol{\theta}} \boldsymbol{\theta}\pi(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta}, \tag{3.2.4}$$

can serve as an estimate of the parameter $\boldsymbol{\theta}$. Future values of $\boldsymbol{x}$ can also be predicted by using the posterior predictive distribution $\pi(x_{n+1}|\boldsymbol{x})$ which is given in equation (3.2.5).

$$\pi(x_{n+1}|\boldsymbol{x}) = \int_{\boldsymbol{\theta}} \pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{x})d\boldsymbol{\theta}. \tag{3.2.5}$$

For solar power forecasting, the Bayesian approach gives a full range of sample inferences, including the predictive distribution of solar demand. This is not easy using classical statistics. When using large samples, the Bayesian approach based on informative priors gives results for parametric models similar to results from frequentist methods. However, the Bayesian approach is superior to other modelling methods because it combines prior information with data. The estimation based on Bayesian analysis is robust and allows predictions using small samples (Ardia and Hoogerheide, 2010). The Bayesian analysis gave an efficient modelling framework based on conditions that are uncertain (Marlin et al., 2003). Bayarri and Berger (2004) shows that the Bayesian technique is superior in many ways.

3.2.2 Gaussian process prior

The inputs are defined by a variable related to the input domain variables as described by the mean and covariance matrix. The priors are based on specifying the priors of the distribution. This is done by incorporating the specified information on the mean and selecting the appropriate covariance function. Consider any input points(locations), x and x^1, i for a covariance function $k(x, x^1)$ to be choosen. When this covariance function is chosen, we need to set prior conditions on the distribution. The Gaussian process prior is given by equation (3.2.6).

$$f(X) \mid Y \sim GP(m(X)) \equiv 0, \tag{3.2.6}$$

where X are the predictor variables, Y is GHI, and $m(X)$ is the mean function. GP priors are evaluated based on mean and covariance functions on X and are applied to a Gaussian distribution. For most regression computations, the mean function is usually set to zero and uses the central tendency of the function.

For Bayesian estimation, we can either use the informative or noninformative approach. The informative approach was used because it uses priors from the kernel simulations.

Informative priors

An informative prior is a prior that expresses specific, definite information about a variable. When a prior distribution dominates the likelihood, then it is an informative prior. The kernel function depends on different hyperparameters whose prior distribution should be specified. These sorts of distributions must be indicated with consideration in the actual application. Then again, the best possible utilisation of prior distributions illustrates the intensity of the Bayesian technique: historical data, previous experience, or expert opinion can be joined with present data ordinarily.

3.2.3 Gaussian process posterior

The posterior is based on Bayes and is drawn from prior data. A GP posterior is given in equation (3.2.7).

$$\text{Posterior} = \frac{\text{likelihood} * \text{prior}}{\text{marginal likelihood}} \tag{3.2.7}$$

$$= \frac{P(y \mid X, x) \times p(x)}{P(y \mid x)},$$

where y- output values, $\sum_p$ is the covariance matrix.

3.3 Gaussian process kernel regression

As alluded to in the previous sections, the GP is parametrised by the mean function given by μ_x and a covariance function given by $k(x, x^1)$. The mean function is assumed to be approximately equal to zero. The covariance function is the kernel function, and it should be computed. This has parameters that are interpretable and can be learned from the data. The kernel function describes how a feature of the dependent variable of a function change depending on how all other features of the independent variable change. The kernel function should be positive definite, and specified. It is said to be positive semi-definite if the function in (3.3.1) is satisfied.

$$\int k(x, x^1) f(x) f(x^1) d\mu(x) d\mu(x^1) \geq 0, \tag{3.3.1}$$

where $k(x, x^1)$ is the covariance function and (x, x^1) are the sample pairs.

3.3.1 Kernel regression

Kernel methods for regression are the deterministic regression model and the probability regression model. The deterministic regression model is given by

$$y = f(x), x \in \mathbb{R}^N, y \in \mathbb{R}, \tag{3.3.2}$$

and the probabilistic regression model is given by

$$y = f(x + \varepsilon), \tag{3.3.3}$$

with ε an estimate of the noise model $\varepsilon \sim N(0, \sigma^2)$ and f is the support vector.
The kernel functions depend on different hyperparameters whose prior distribution should be specified. It is the form of the probability function in which any factors that are not functions of the variables in the domain are omitted. There are many common functions. These are: Constant, Gaussian Noise, Squared exponential, Ornstein-Unlenbeck, Matern, Periodic, and Rational quadratic

3.3.2 Common kernel functions

A kernel is used to develop a covariance kernel of a Gaussian process of a stochastic random process. In order to come up with the most suitable kernel to be used, we use predefined kernels to model various processes. The kernel function is 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 Williams and Rasmussen (2006). Some are called Linear, Squared Exponential, Rational Quadratic, and Matern. This study used several popular kernel functions: Matern, Rational quadratic, Dot product, and Radial Basis Function.

Radial basis function kernel

The radial basis function is given in equation (3.3.4)

$$k(x, x^1) = exp(\frac{|| \ x - x^1 \ ||^2}{2\sigma^2}), \tag{3.3.4}$$

where σ^2 is the variance, l is the length scale, α is the scale mixture, and $|| \ x - x^1 \ ||^2$ is the Euclidean distance, also known as an infinite sum.

Matern kernel

It is a covariance function given in equation 3.3.5

$$k(x, x^1) = \sigma^2 \frac{2^{1-v}}{\Gamma(v)} \cdot (\sqrt{2}v\frac{d(x, x^1)}{\rho})^v \cdot k_v \cdot (\sqrt{2}v\frac{d(x, x^1)}{\rho}), \tag{3.3.5}$$

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

Dot product kernel

It is a non-stationary kernel attained by putting standard normal priors on linear regression coefficients. The priors of the coefficients are such that $X \sim N(0, \sigma_0^2)$, parameterised by σ_0^2, where σ controls the kernels homogeneity. The kernel function is given in equation (3.3.6)

$$k(x, x^1) = \sigma_0^2 + x \cdot x^1, \tag{3.3.6}$$

where σ_0^2 is a parameter to the kernel.

Rational quadratic kernel It is a kernel function known as a scale mixture. It is given in equation (3.3.7)

$$k(x, x^1) = (1 + \frac{d(x, x^1)^2}{2\alpha l^2}), \tag{3.3.7}$$

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

Gaussian kernel

It is given by

$$k(x, x^1) = \exp(\frac{-\mid\mid x - y \mid\mid^2}{2\sigma^2}), \tag{3.3.8}$$

where σ is the adjustable parameter that plays an important role in the performance of a kernel.

x and x^1 are the input points of the observations of the independent variables: temperature, humidity, dew point, wind speed, wind direction, and rainfall.

3.4 Core vector regression modelling

In this research, we used the minimum enclosed ball (MEB) to find the best kernel function for GPR use. MEB was coupled with GPR. Minimum Enclosed Ball (MEB) is a branch of core vector regression. The MEB algorithm is a method used to find a centre c such that the minimum distance to input points realisations. This is done to fully contain all relevant data points that accurately describe the dataset. The Minimum enclosed algorithm was applied to several kernel functions. The kernel function that gave the minimum radius was the one selected for Gaussian process modelling. The ball that gave the minimum radius is the one that had the minimum enclosed ball.

The MEB method computes the ball with a minimum radius that captures a set of points as a ball. The method was first developed by Badoiu and Clarkson (2003), who discovered that the approximation of the ball was $(1+\epsilon)$ of the MEB obtained through core sets. The MEB has a computational geometry problem, which can be solved by incorporating core vector regression (Tsang et al., 2005), which is done by observing the approximation of the optimal solution by an iterative strategy. Let the MEB be the minimum enclosed ball of a set X. It is given in equation (5.2.2)

$$K(c;,r) \tag{3.4.1}$$

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 (2003) and Yildirim (2008) are used. They are extracted based on the technique of ϵ-core set of X, implying that there exists $C \subset X$ whose MEB is given as $(1+\epsilon)$ of X. An Algorithm by Badoiu and Clarkson (2003) provides a ϵ-core set belonging to X in less than $O(1/\epsilon)$ iterations. Tsang et al. (2005) introduced core vector regression which supports the reduction of the MEB problem.

There are many advantages of using the core vector regression modelling technique. Core vector regression reduces time and space complexities. The method uses an asymp-

totically efficient algorithm in handling space and time. Another advantage of this method is that it uses various numerical routes with an iterative strategy that estimates the optimal solution. The method has a ratio that computes how bad the approximate solution is; it is then compared to the optimal solution. This makes the method efficient. An efficient method, MEB, calculates the ball with the minimum ball that scales well on dimensional points.

3.5 Gaussian spatio-temporal predictive modelling

Analysis was also done to model spatial dependences between stations. Spatial analysis is the process of performing an analysis that involves modelling spatial features of data like the location and their relatedness to geographical data. In this study, we extended the research in Gaussian Processes to spatial GP analysis. The advantage of using spatial analysis is that since spatial data is selected from multiple sources and consolidated, it improves the accuracy of the models since it captures variations of data from different sources. The prediction of GHI depends on how far apart are the meteorological stations. Therefore, these predictions need to incorporate the spatial dependence of stations. This was achieved by considering the spatial correlation of the stations. Two spatio-temporal analysis methods were used: Gaussian Spatial analysis and autoregressive spatial analysis. The results of the two were compared to Linear Spatial analysis,

3.5.1 Spatio-temporal Gaussian process regression model

Gaussian processes were developed by Williams and Rasmussen (Williams and Rasmussen, 2006). Gaussian processes can be presumed as elements of spatial-temporal modelling because these stochastic processes are defined, in this case, over a given region where the stations have different locations. The spatial analysis aims to develop the best model to

produce a set of outputs using inputs from various locations at different time frames. A -temporal Gaussian process (GP) model is given in equation (7.4.1).

$$Y(s_i,t) = \boldsymbol{x}^T(s_i,t)\beta + w(s_i,t) + \varepsilon(s_i,t), \tag{3.5.1}$$

where $w(s_i,t)$ is the spatially-dependent error term (spatio-temporal process), $Y(s_i,t)$ represents the response variable (GHI) at station s_i at time t, for $i = 1,2,...,n$ and $t = 1,2,...,T$; with $\boldsymbol{x}(s_i,t)$ representing the covariates, β is a vector of constants and $\varepsilon(s_i,t)$ is the pure random error term. Gaussian processes are characterised by the covariance (kernel) functions. Various kernel functions can be used, but the Matern and squared exponential are commonly used. In this study, the Matern covariance kernel was used because it better balances the roughness and smoothness of the developed model Stein (1999). The family Matérn kernels were first explored by Michael Stein, who took the name based on initial work by statistician Bertil Matérn, Matérn (2013). The Martern kernel was used for spatial analysis because it was discovered to suit spatial data Matérn (2013); Stein (1999). The temporal process $w(s_i,t)$ is assumed to be a GP with the Matern covariance function given in equation (7.4.2), (Sahu, 2022).

$$C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{\rho}\right)^\nu K_\nu \left(\sqrt{2\nu}\frac{d}{\rho}\right), \tag{3.5.2}$$

where d represents the distance between two points, K_ν is the modified Bessel function of the second kind and ρ and ν are parameters. That is (7.4.3)

$$\text{Cov}\left(w(s_i,t), w(s_j,t)\right) = \sigma_w^2 \rho\left(d_{ij}|\boldsymbol{\nu}\right). \tag{3.5.3}$$

The GP model can be represented hierarchically as shown in equation (7.4.4).

$$\begin{cases} Y_{it} = \boldsymbol{O}_{it} + \varepsilon_{it}, \\ \boldsymbol{O}_{it} = X_{it}\beta + w_{it}, \end{cases} \tag{3.5.4}$$

where $i = 1, ..., n;\ t = 1, ..., T$ assuming $\varepsilon_{it} \sim N(O, \sigma_{\varepsilon}^2 I)$ and $w_{it} \sim N(O, \sigma_w^2 S_w)$ with S_w representing the sample covariance matrix.

The prior distribution assumes a normal distribution as shown in equation (7.4.5).

$$W_t \sim N(0, \sigma_w^2 S_w). \tag{3.5.5}$$

The posterior distribution is given in equation (7.4.6).

$$\begin{cases} Y(S_0, T+1) = O(s_0, T+1) + \epsilon(s_0, T+1), \\ O(s_0, T+1) = x^1(s_0, T+1)\beta + w(s_0, T+1), \end{cases} \tag{3.5.6}$$

where $Y(s_0, T+1)$ is the 1-step ahead forecast and s_0 is an unobserved location at time $T+1$. The hyperparameters θ are solved by maximising the marginal likelihood function in equation (7.4.7).

$$L(\theta, \sigma_{noise}^2) \quad = \quad -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log K_\theta + \sigma_{noise}^2 I - \frac{1}{2}Y^T(K_\theta + \sigma_{noise}^2 I)^{-1}, \tag{3.5.7}$$

where σ_{noise}^2 is the noise variance, θ represents the hyperparameters and K_θ is the covariance function. The posterior predictive distribution of $Y(s_0, T+1)$ given y is given in equation (7.4.8).

$$\begin{aligned} \phi(y(s_0, T+1)|Y) = \int \pi(y(s_0, T+1)|\theta, \mathbf{O}, O(s_0, T+1), y) \times \\ \pi(0(s-0, T+1)|\theta, y)\pi(\theta, O|y)dO((s_0, T+1)d\theta 0 d\theta), \end{aligned} \tag{3.5.8}$$

where $\phi(\theta, 0|y)$ is the joint posterior distribution of $\mathbf{O}$ and θ. This study uses the GP-temporal approach to forecasting solar irradiance.

The predictive modelling process of the Gaussian spatial model is flexible. It reduces the dimensions of the stations' subspace, reducing the computational burden. Incorporating spatial dependence enables the model to use datasets from different stations with different conditions, making the model very effective in modelling GHI for the whole country.

3.5.2 Gaussian process regression based autoregressive model

Another approach to forecasting solar irradiance used in this research is the autoregressive (AR) Spatial-temporal model. The Gaussian process autoregressive model is given hierarchically as shown in equation (7.4.9) (Sahu et al., 2020).

$$\begin{cases} Y_{it} = \boldsymbol{O}_{it} + \varepsilon_{it} \\ \boldsymbol{O}_{it} = \rho \boldsymbol{O}_{it-1} + X_{it}\beta + w_{it}, \end{cases} \tag{3.5.9}$$

where $i = 1, ..., n;\ t = 1, ..., T$ assuming $\varepsilon_{it} \sim N(O, \sigma_{\varepsilon}^2 I)$ and $w_{it} \sim N(O, \sigma_w^2 S_w)$ with S_w representing the sample covariance matrix.

A direct autoregression on the response variable Y_{it}, which is $Y_{it} = \rho Y_{it-1} + \varepsilon_{it}$ gives challenges when there are missing values in Y_{it}. Parameters of the spatio-temporal AR model are computed based on prior distributions. Let $\boldsymbol{\theta} = (\beta, \rho, \sigma_{\varepsilon}^2, \sigma_w^2, v, \mu_0, \sigma_0^2)$ denote all the parameters with $\pi(\theta)$ as the joint prior distribution. Equation (7.4.10) presents the logarithm of the joint posterior distribution of the parameter.

$$\begin{aligned} \log\pi(\boldsymbol{\theta}, \boldsymbol{o}|\boldsymbol{y}) \propto -\frac{nT}{n}\log(\sigma_{\varepsilon}^2) - \frac{1}{2\sigma_{\varepsilon}^2}\sigma_{t=1}^{T}(y_t - \theta_t)^T(y_t - \theta_t) \\ -\frac{T}{2}\log|\sigma_w^2 S_w| - \frac{1}{2\sigma_w^2}\sigma(o_t - \rho o_{t-1} - \mu_t)^T S_w^{-1}(o_t - \rho o_{t-1} - \mu_t) \\ -\frac{1}{2}\log|\sigma_0^2 S_0| - \frac{1}{2\sigma_0^2}\left(\boldsymbol{O}_0 - \boldsymbol{\mu}_0\right)_T S_0^{-1}\left(\boldsymbol{O}_0 - \boldsymbol{\mu}_0\right) + \log\pi(\boldsymbol{\theta}) \end{aligned} \tag{3.5.10}$$

The posterior predictive distribution of $Y(\boldsymbol{s}_0, t)$ is given in equation (7.4.11).

$$\begin{aligned} \pi(y(\boldsymbol{s}_0, t)|\boldsymbol{y}) = \int \pi(y(\boldsymbol{s}_0, t)|o(\boldsymbol{s}_0, [t]), \sigma_{\varepsilon}^2)\pi(o(\boldsymbol{s}_0, [t])|o(\boldsymbol{s}_0, 0), \boldsymbol{\theta}, \boldsymbol{o}) \\ \pi(o(\boldsymbol{s}_0, O)|\boldsymbol{\theta}, \boldsymbol{O}_0)\pi(\boldsymbol{\theta}, \boldsymbol{o}|\boldsymbol{y})do(\boldsymbol{s}_0, [t])dO(\boldsymbol{s}_0, o)d\boldsymbol{\theta}d\boldsymbol{o} \end{aligned} \tag{3.5.11}$$

One of the advantages of Spatial Autoregressive predictions is that it accommodates time series dependence, that is, being able to handle cyclical variations in GHI. This enables

it to handle various time series trends. Another advantage is that the spatial error accommodates spatial dependence. The model is flexible and accommodates multiple variables with lagged values of the dependent variable as its predictor variables.

3.6 Bayesian structural time series

Bayesian structural time series (BSTS) is one of the methods used in this research to predict global horizontal irradiance (GHI). The approach has been applied by Brodersen et al. (2015) and Scott and Varian (2015), and they discovered that it had several advantages. It is a statistical method for feature selection and modelling time series data. A BSTS model is given in 6.2.8.

$$y_t = Z_t^T \alpha_t + \epsilon_t, \tag{3.6.1}$$

where y_t are the observed values of GHI, Z_t is the output vector, ϵ_t is the error term and α is the state vector.

The transition model 6.2.9 gives the latent state that changes over time.

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \tag{3.6.2}$$

where η_t is the q by q dimension error term which is assumed that $\eta_t \sim N(0, \sigma_\eta^2)$, T_t and R_t are the structural parameters, α_t is the latent variable and η_t is the noise error term. BSTS incorporates both long and short-term forecasting. Both these two methods were explored in this research. The model was fitted using an R package called *bsts*, a tool for fitting Bayesian Structural series functions. BSTS comprises Kalman filter, spike and slab, and Bayesian model averaging (BMA).

Kalman filter

It is an efficient and recursive method to select variables inferring parameters from uncertain observations. It produces good results because of its optimality nature. It is named after

Kalman (1960). Its advantage is that it is computationally inexpensive and has well-designed recursive properties. BSTS assumes that the state at time k is evolved from $k-1$, and the prediction at time k is given in equation (3.6.3).

$$x_k = F_k x_{k-1} + B_k U_k + W_k, \tag{3.6.3}$$

where F_k is the state transition model applied to previous state vector x_{k-1}, B_k control input model applied to control vector U_k, W_k is the white noise process, x_k is the current state.

At time k an observation z_k of the true state x_k is made such that

$$z_k = H_k x_k + v_k, \tag{3.6.4}$$

where H_k is the noiseless connection model and v_k measurement errors, v_k is the measurement errors.

Spike and slab

It is a method to select a subset of predictors retained in the model. This method's advantage is that it uses prior information when selecting variables. The Spike and Slab model was considered by George and McCulloch (1997) and is given in expression (3.6.5).

$$\begin{aligned} \beta_\ell|\gamma_\ell &\sim (1-\gamma_\ell)\mathrm{N}(0,\tau_\ell^2) + \gamma_\ell \mathrm{N}(0, c_\ell^2\tau_\ell^2) \\ \gamma_\ell|\pi_\ell &\sim \text{Bernoulli}(\pi_\ell), \end{aligned} \tag{3.6.5}$$

where β_ℓ is the vector of coefficients, γ_ℓ is the vector of latent variables, τ_ℓ^2 is a scale parameter, is chosen small, and c_ℓ^2 is a scale parameter, is chosen to be large.

Bayesian model averaging

Bayesian Model Averaging (MBA) produces forecasts accounting for uncertainty. Authors like Raftery et al. (1997), Lee and Carlin (2010), and (Fernandez et al., 2001), discovered

that the BMA has superior predictive ability. It is found by considering equation 3.6.6.

$$P(\Delta|y = \sum_{j=1}^{J} P_{\Delta|y,M_j}(M_i|y))), \tag{3.6.6}$$

where Δ is the quantity of interest, y is the observed quantity, M_j The advantage of using BSTS is that it uses priors by BSTS to enable the model to handle uncertainty. It uses spike and slab priors that reduce correlated variables, creating a parsimonious model. This is because slab and spike priors comprise two parts, spike prior and slab prior. Spike prior gives the probability of each variable selected, and slab shrinks the coefficients. It uses parameters obtained through Bayesian techniques. Thus, it is flexible. Also, BSTS can incorporate variation from multiple sources.

3.7 Variable selection

The variables used for analysis in this research were chosen using the methods discussed in the following sections. Variable selection has a vital role in the model-building process because, in some cases, modelling will result in quite a several candidate independent variables; hence there is a need for variable selection. Variables are selected to deal with the following problems encountered in modelling. One of the problems it handles is that it avoids overfitting by avoiding including too many variables. The process selects a subset of significant variables to develop a better model removing insignificant and irrelevant variables, thus improving model accuracy. Leaving insignificant variables in a model will result in less precise estimates of coefficients. This results in producing more accurate models. There will be a great improvement in predicted values and reduced errors. It deals with problems of multicollinearity. Multicollinearity is when predictor variables are highly correlated. This creates overfitting problems by producing a more reliable model by eliminating multicollinearity. Removing irrelevant variables eliminates unnecessary computations, thus reducing computation time.

Various methods were used for model selection. This includes shrinkage methods and feature selection methods. The following sections are based on the methods applied.

3.7.1 Shrinkage methods

Shrinkage methods were used in this research to select variables for modelling. Shrinkage refers to reducing variation in sampling when data points are shrunk towards a central point, enabling the minimisation of overfitting and underfitting of a dataset through introducing the penalty term, thus reducing the MSE error.

3.7.2 Least absolute shrinkage and selection operator

The least absolute shrinkage and selection operator (Lasso) was one of the shrinkage methods used for variable selection. Lasso was first developed by Tibshirani (1996), its main objective is to improve the predictive accuracy of regression models, and it comes up with a selected reduced set of covariates used for modelling. Lasso is a penalised regression technique to develop a subset of independent variables that gives us a minimum error. It performs L_1 regularisation, which results in sparse models with few coefficients. Those that become zeros are eliminated from the model. It performs shrinkage and variable selection to find the subset of variables that minimises prediction error. Lasso is a powerful shrinkage method that performs feature selection and regularisation, and this tackles the problem of overfitting, which enhances its predictive accuracy. Tibshirani (1996), even states that Lasso provides probabilities for 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 toward zero. The method minimises Equation (6.2.4).

$$\sum (y_i - \sum x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{3.7.1}$$

subjected to constraint

$$\sum_{j=1}^{p} \mid \beta_j \mid \leq t, \tag{3.7.2}$$

where λ is the turning parameter, y_i are the observed values of the dependent variable of GHI, t is a parameter that determines the quantity of regularisation, x_{ij} is the covariance matrix of the independent weather variables, and β_j is the column vector of regression coefficients.

3.7.3 Lasso via hierarchical interactions

Another shrinkage method used for variable selection in this research is Lasso via hierarchical interactions. Lasso via hierarchical interactions has been actively researched in many areas of science, Bien et al. (2013); Lim and Hastie (2015) and proved very effective. It is a technique that uses pairwise interaction regression incorporating a strong hierarchy. Sparse models are created by adding a set of convex constraints that keep hierarchy restrictions. For this research, we used pairwise interactions that if k=2, The model we used to select the variables is given in equation (5.2.1).

$$Y = \beta_0 + \sum_{j} \beta_j X_j + \frac{1}{2} \sum \Theta_{jk} X_j X_k + \varepsilon_j, \tag{3.7.3}$$

where Y is the independent variable GHI, β_0 is the slope, β_j are the regression coefficients, $X_j X_k$ are the predictors with pairwise interaction and $\Theta \in \mathbb{R}^{p \times p}$.

3.7.4 ElasticNet

In order to select significant variables for the analysis, ElasticNet, a shrinkage regularisation technique, was used. ElasticNet is a technique that produces a hybrid penalty by

combining penalties from ridge regression(L_2) and Lasso regression(L_1) to reduce the number of variables in a regression model. This makes it superior. It combines both ridge and Lasso methods to overcome their limitations. Elastic net minimises the function in equation (7.3.1).

$$min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+\lambda\left(\frac{1-\alpha}{2}||\beta||_2^2+\alpha||\beta||_1\right)\right), \tag{3.7.4}$$

where N is the number of observations, y_i is the dependent variable, GHI. β_0 is the intercept, β is the vector of coefficients, and x is the vector of the independent variables, weather variables. Alpha is the mixing parameter between Lasso $\alpha = 1$ and ridge when $\alpha = 0$ and λ is the regularisation penalty parameter.

3.7.5 Feature selection methods

Feature selection techniques isolate the most non-redundant, relevant, and consistent features, also known as variables, to use in modelling. The objective of feature selection is to improve the accuracy of predictive models by selecting the best set of variables to use. Feature selection methods in this research are given in the sections that follow.

3.7.6 Multivariate adaptive regression spline

Multivariate adaptive regression spline (MARS) is a nonparametric regression method developed by (Friedman, 1991). MARS uses piecewise linear or cubic splines to select the significant variables. MARS uses piecewise linear basic functions given in equation (6.2.1).

$$(x-t)_+ = \begin{cases} x-t & , \text{ if } \ x>t \\ 0 & , \ otherwise \end{cases} \quad \text{and} \quad (t-x)_- = \begin{cases} t-x & , \text{ if } \ x<t \\ 0 & , \ otherwise \end{cases} \tag{3.7.5}$$

where $+$ is positive part, x are the weather predictor variable, y is the response variable GHI, t is the break point or knot point.

Mars uses reflected pairs for predictor variables with knots at each observed value, which are given as a set S given in Equation (3.7.6).

$$\mathcal{S} = \{(x_j - t)_+, (t - x_j)_+ | \, t \in \{x_{1j}, x_{2j}, \ldots, x_{nj}\}, \, j \in \{1, \ldots, p\}\}. \tag{3.7.6}$$

The MARS model incorporating knots is given in equation (3.7.7). Equation (3.7.7) is the one that was used for variable selection.

$$f(\mathbf{x}) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(\mathbf{x}), \tag{3.7.7}$$

where $B_m(\mathbf{x}$ is the piecewise linear function, M is the number of piecewise linear function, β_0 is the intercept and β_m is the coefficient of the regression equation.

3.7.7 Boruta

Boruta is one of the feature selection methods used to develop a set of variables for analysis. Feature selection is a method that is used to reduce the number of features which is done by choosing features that have a great influence on the variables under study. It is a wrapper feature selection method that uses Random Forecast. The algorithm makes use of the following steps

- makes copies of all variables, which are called shadow feature
- removes any correlations by shuffling the added attributes by training a random forecast classifier.
- a random classifier is run, and a z score is computed and checked if the feature has higher importance.
- the algorithm stops when all chosen features have higher importance.

The advantages of using the Boruta package are that it does both classification and regression analysis and deals with the interaction between variables.

3.8 Forecasts combination

In order to get the superior performance of the model, a forecast combination was performed. It is achieved by combining two sets of forecasts based on two hybrid models making the forecasts robust. The idea of forecast combination was first used by Granger (1989) and Crane and Crotty (1967). Various authors like Clemen (1989), Timmermann (2006) and Nowotarski et al. (2014), have used a combination of forecasts and the predicted values showed that it improved forecasting accuracy. The advantage of combining forecasts is that it reduces prediction errors and variability of various combination accuracy.

3.8.1 Partially linear additive quantile regression

Partially linear additive quantile regression (PLAQR) is a hybrid model produced from the combination of Generalised Additive and Quantile Regression models. The proposed methodologies in this research were combined with PLAQR to produce a hybrid model that would improve the predictions. The PLAQR is given by Equation (6.3.1).

$$y_{th} = \beta_{oh,\tau} + \sum_{j=1}^{p1} s_{jh,\tau}(x_{thj}) + \sum_{j=1}^{p2} \beta_{hj,\tau} Z_{thj} + \varepsilon_{th,\tau}\tau \in (0,1), \tag{3.8.1}$$

where y_{th} is the GHI on day t and hour h, β_{oh} is the intercept, $\beta_{hj,\tau}$ are the model parameters h-hours, $s_{jh,\tau}$ smooth function is the independent variables and z_{thj} is the linear variables. $\varepsilon_{th,\tau}\tau$ is the quantile error. The advantages of applying this method are that it is robust to outliers, the sparsity level can be changed with quantiles, and conditional distribution can be analysed thoroughly, Mpfumali et al. (2019).

3.8.2 Quantile regression neural network

A quantile regression neural network (QRNN) is a hybrid model combining Quantile regression and Neural Networks. It has the advantage of capturing non-linear patterns in

datasets and overdispersion and underdispersion in the dataset. The QRNN model was improved by Zhang et al. (2018) and is given in Equation 6.3.2.

$$f(x_t, v, w) = g_2\left(\sum_{j=0}^{m} v_j g_1\left(\sum_{i=0}^{n} w_{ji} x_{it}\right)\right), \tag{3.8.2}$$

where n-number of inputs, m-units of the hidden layer, x_{it} are the predictor weather variables, $g_1(.)$ and $g_2(.)$ are activation functions, w_{ij} and v_j are variable weights of parameters to be estimated.

3.8.3 Linear quantile regression averaging

Linear quantile regression averaging (LQRA) was used for forecasting combinations in this research. It is a linear modelling approach that makes use of prediction intervals. Quantile Regression is a method that uses least squares regression to create a conditional mean of the dependent variable GHI over different values. It was first introduced by Koenker and Bassett (1978) and is widely used for energy forecasting, Liu et al. (2015); Elamin (2018); Uniejewski and Weron (2021). The Linear QRA model is given in Equation (6.3.3)

$$Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + ... + \beta_p(\tau)x_{ip}, \tag{3.8.3}$$

where p is the number of independent variables, which are the weather variables, τ is the τ^{th} quantile, y_i is GHI, and x_{ij} are the weather variables.

The best Linear QRA model is given by (6.3.4)

$$\min(\rho_\tau(Y_i - (\beta_0(\tau) + \beta_1(\tau)x_{i1} + ... + \beta_p(\tau)x_{ip})), \tag{3.8.4}$$

where $\rho_\tau(z)$ is a loss function. QRA is when forecast combinations are done on prediction intervals given on point predictions done on individual models. The combined forecasts produced by LQRA are given in Equation (6.3.5).

$$\hat{y}_{iq} = \beta_0 + \sum_{k=1}^{k} \beta_k \hat{y}_{ij} + \epsilon_{iq}, \tag{3.8.5}$$

where β_0 is the intercept, β_k are the coefficients of the independent variables, $\hat{y}_{iq}$ is the combined forecasts, $\hat{y}_{ij}$ is the forecast for the $j_t h$ technique and ϵ_{iq} are the residuals. It is used to explore the conditional distribution of the variable of interest GHI incorporating explanatory variables, which are the weather variables.

3.8.4 Online prediction by ExpRt aggregation

Lastly, online prediction by ExpRt aggregation (OPERA) is one of the packages used for combining forecasts Gaillard et al. (2021). It is a method that is based on regression-oriented time series forecasts. Given a set of observed values for a particular variable Y, GHI, with a sequence of values $y_1, y_2, ..., y_n$, 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 (6.3.6).

$$\hat{y}_t = s \sum_{k=1}^{K} P_{k,t} x_{k,t} \tag{3.8.6}$$

The method uses various online learning literature algorithms that combine experts' forecasts based on historical performance. It gives three important features, mixture, prediction and oracle. The mixture is used to build the algorithm's object, and prediction is used to make predictions using the algorithm. Oracle is for performance evaluation of the experts and comparing the algorithm.

3.9 Parameter estimation

This section looks at the parameter estimation of the Gaussian processes. The parameters to be estimated are β, the regression coefficients, σ^2, the scaling parameters, and δ, the correlation length. We can either start with the maximum likelihood of these estimates by marginalising β and σ^2 out or using a Bayesian approach to estimate parameters. We

then focus on correlation lengths to investigate the effect of several design points of the estimation. This is considered a function of the inputs dimensionally and the correlation length of the data.

3.9.1 Bayesian approach to parameter estimation

The Bayesian approach for parameter estimation will be used. The Bayesian approach is more attractive than the frequentist maximum likelihood estimation (MLE) approach as it combines prior information with data, enables robust analysis with small samples, and makes the estimation robust. The Bayesian approach does not rely on asymptotic approximation. It obeys the likelihood principle, which states that any two probability models with the same likelihood function yield the same inference for a given sample of data.

3.10 Forecast evaluation metrics

Predicting future values is very useful, especially for decision-makers. Energy forecasting helps power grid managers to make informed decisions. When dealing with prediction models, the forecasted values should be evaluated. Prediction accuracy is very important in the model-building process to come up with accurate predictions. Forecast evaluation metrics are measurements that compute the reliability of the forecasts.

3.10.1 Deterministic forecast evaluation metrics

Many metrics can be used to evaluate forecasts. In this research, we used metrics such as the mean absolute error (MAE), root mean square error (RMSE), mean absolute error (MAE), mean average percentage error (MAPE) and percentage bias (PBIAS) will be used for model evaluation to come up with models for each of the two different areas, Diebold and Lopez (1996).

Mean absolute error

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 (6.3.7).

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_i - \hat{y}_j| \tag{3.10.1}$$

It is a measure of the mean forecast errors expressed in percentages. This shows the level of inaccuracy we should get on the prediction mean. When the value is equal to zero, the predicted values are correct. The lower the value is, the more accurate the model is.

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 (6.3.8).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_j)^2}, \tag{3.10.2}$$

where y_i-are the observed values, $\hat{y}_j$- are the actual values and n represents the number of observations. The lower the value of RMSE, the better the predictions.

Mean absolute percentage error

The mean absolute percentage error (MAPE) is a measure that expresses accuracy as a percentage of error. It is given in Equation (5.2.8).

$$\text{MAPE} = \frac{100}{m}\sum_{i=1}^{m}\mid \frac{\varepsilon_i}{\hat{y}_i} \mid. \tag{3.10.3}$$

The low the value of MAPE, the higher the model performance.

Percentage bias

Percentage bias (PBIAS) measures the average tendency of the predicted values to be bigger or smaller than the observed values. It is given in Equation 3.10.4

$$\text{PBIAS} = \frac{\sum_{i=1}^{n} Y_i - \hat{Y}_i \times 100}{\sum Y_i}, \tag{3.10.4}$$

3.10.2 Probabilistic evaluation metrics

We need to choose among different data models by selecting a model among several models that can best capture the underlying regularities. So, the question will be, how do we identify such a model? So, we want to select a model that solves problems with complex conditions and complexity. Complexity is a model's inherent flexibility that allows it to fit a wide range of data patterns. We need to compare several models using various selection methods. Making predictions from probability distributions for future events is an interesting research area. Various authors have looked at probabilistic forecasting, Clements (2004), Gneiting and Raftery (2007), among others. Many authors are showing interest in probabilistic forecasting; hence there is a need to develop methods for evaluating the appropriateness of forecasting techniques. Various tools were also used to evaluate forecasts further. The tools used in this research are: proper scoring rules, pinball losses, and Murphy diagrams were used to evaluate the appropriateness of modelling techniques.

Pinball Loss

The pinball loss (PL) is a measure that assesses the accuracy of a quantile forecast. It is given by (6.3.9).

$$PL(q\tau, t) = \begin{cases} 2(1-\tau)|y_t - q_{\tau,t}| & if \quad y_t < q_{\tau,t} \\ 2\tau|y_t - q_{\tau,t} & if \quad y_t \geq q_{\tau,t} \end{cases}, \tag{3.10.5}$$

where y is the observed values, q-quantile forecasts and τ-is the selected quantile and t is time. More accurate predictions produce a lower pinball loss.

Proper scoring rules

Proper scoring rules were used in this research to evaluate the effectiveness of probability models. Scoring rules are functions used to evaluate the accuracy of the probability function. Scoring rules were introduced by Jordan et al. (2017) as functions that give functions that are used to compute scoring rules for a variety of distribution functions. These were produced in order to measure how good a probabilistic forecast is. It gives a numerical score that is based on a predictive distribution. The numerical score is the expected score. It is given in equation (6.3.10).

$$S(P, Q) = E_{Y \sim P}(S(F, Q)), \tag{3.10.6}$$

where F is the probabilistic forecast, X is a random variable, Q is the probabilistic distribution. The forecast is minimised in the expectation by the true probability for a proper score. Many scoring rules can be used: continuous ranked probability score, logarithmic, spherical, pseudospherical, Dawid Sebastian score, and quadratic scores. We have used the Logarithmic score(LogS), Dawid Sebastian Score(DSS), and Continuous ranked Probability score(CRPS).

The uses of scoring rules are to

- evaluate the forecasting performance of a model,
- eliciting probabilities,
- estimating parameters.

The scoring rules in this study were used to evaluate the forecasting performance of a model. The R package *scoringRules* was used to develop the scoring scores (Gneiting and Raftery,

2007).

Logarithmic score Logarithmic scoring (LogS) is a scoring rule used to measure a probability function's effectiveness in prediction models. A logarithmic score was used in this research to evaluate probabilistic models. Let Y be a random variable over a sample of size n assuming values $x_1, x_2, ..., x_n$. Suppose we are given the probability values as $p_1, p_2, ..., p_n$ belonging to a forecasting distribution F with $P_i \in [0, 1]$ and $\sum P_i = 1$. The logarithmic scoring is given in equation (6.3.11)

$$S(F, y_{obs}) = logP(y_{obs}), \tag{3.10.7}$$

where y_{obs} are the observed values.

Continuous rank probability score Continuous rank probability score(CRPS) is a scoring rule used mainly in meteorological studies. The score is given in equation (6.3.12).

$$S(F, y_{obs}) = \int_{-\infty}^{\infty} (F(y) - 1\{y \geq y_{obs}\})^2 dy, \tag{3.10.8}$$

where $F(y)$ cumulative function of the predictions.

Dawid-Sebastian score Dawid-Sebastian score (DSS) is a measure for evaluating the accuracy of multivariate forecasts. The score is given in (6.3.13).

$$DSS(F, y) = \left(\frac{y - \mu_F}{\sigma_F}\right)^2 + 2log\sigma_F, \tag{3.10.9}$$

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

3.10.3 Murphy diagrams

Murphy diagrams were used for comparing predictions made by various models. The technique was proposed by Ehm et al. (2016). In order to evaluate probability forecasts, we

plot a function of θ given in equation (6.3.14).

$$S(\theta) = \frac{1}{n}\sum_{i=1}^{n} S_\theta(y_i, y_i^*), \tag{3.10.10}$$

where y_i is the point estimate and y_i^* is the actual GHI observation and $S_\theta(y_i, y_i^*)$ is a function given in equation (6.3.15).

$$S_\theta(y_i, y_i^*) = \phi(y_i^*) - \phi(y_i) - \phi^1(y_i)(y_i^* - y_i), \tag{3.10.11}$$

where ϕ is convex, and ϕ^1 is the gradient.

A plot of the function $S(\theta)$ is compared against the predictions to check if it fits perfectly.

3.10.4 Averaging forecasts

It is well established in the literature that combining forecasts improves forecast accuracy, Armstrong (2001); Granger (1989); Batchelor and Dua (1995). This study combines forecasts with two models, the monotone composite quantile regression neural network (MCQRNN) and the QGAM.

Given the forecasts $\hat{y}_k$, from models $k = 1, \ldots, K$ and the combined forecasts $\hat{y}^{\text{comb}}$, MCQRNN seeks to minimise the loss function given in equation (3.10.12).

$$L(\boldsymbol{\theta}) = \frac{1}{k}\sum_{k=1}^{K} L(\tau_k) = \frac{1}{NK}\sum_{k=1}^{K}\sum_{i=1}^{N} \rho_{\tau_k}\left(\hat{y}_i^{\text{comb}} - \hat{y}_{ik}\right), \tag{3.10.12}$$

where

$$\rho_{\tau_k}(u) = \tau_k u I(u \geq 0) + (\tau_k - 1) u I(u < 0)$$

is the conditional quantile of y at τ_k and $\boldsymbol{\theta}$ is a vector of unknown parameters. For a detailed discussion of MCQRNN, see Zou and Hastie (2005).

Table 3.1: Model comparisons of the proposed techniques.

Models	**Strengths**	**Weaknesses**
Spatial Linear	1. It is easy to implement, train and interpret. 2. The algorithm is less complex then other models are used. 3. Uncertainty is captured directly.	1. It is sensitive to outliers
Spatial GP	1. It uses kernel functions, making the model computationally cheaper, giving minimum loss. 2. Can handle complex datasets. 3. The kernel function uses an inner product to combine features from different locations.	1. Spatial GP is computationally expensive.
Spatial AR	1. They allow the investigations of the associativeness of an individual. independent variables, adjusting the non-independent variables. 2. The model can determine if randomness is lacking in the spatial function. 3. It can predict patterns that are related to recurring datasets.	1. Robustness is limited against non-stationarity and additive noise.

QGAM is a hybrid model combining a generalised additive model (GAM) and a quantile regression (QR) model. QGAMs were developed by Fasiolo et al. (2020). The QGAM model

is given in equation (3.10.13).

$$q_{Y|X}(\tau) = \sum_{i=1}^{N} \rho_\tau \left(\hat{y}_i^{\text{comb}} - \sum_{k=1}^{K} s_k \hat{y}_{ik} \right), \tag{3.10.13}$$

where ρ_τ is the pinball loss function.

3.11 Descriptive analysis

Various descriptive statistics measures are applied in this research. Both graphical and numerical measures were used. To describe GHI, the numerical measures used were median, mean, quartiles, minimum, maximum values, kurtosis, and skewness. Correlations between variables were calculated in order to establish the kind of relationship that existed between variables. Real-time analysis was also done to see if there were any cycles in the time series plots. A Matrix of scatter diagram and histogram was also produced to visualise the distribution of the datasets of the variables used.

3.12 Forecasting

Forecasting is the estimation of the value of a variable (or set of variables) at some future point in time that is done to aid planning and decision-making. In this research, we are estimating Global horizontal irradiance using meteorological irradiance. Various types of forecasting methods are classified according to the interval it is estimated. The forecasting horizon is classified as very short forecasting, which looks at seconds up to 1-hour; short-term forecasting ranges from 1hr to 168hrs(one week), while medium-term forecasting ranges from 1 week to 6 months. Last is long-term forecasting, which is done between six months and several years.

3.13 Conclusion

This chapter highlighted several stages in data analysis and models used. It covered descriptive statistics and data visualisation, which intends to generate understandable results. It also reviewed data training and how data is split into train and test sets, variable selection and forecast combination methods. Lastly, various evaluation methods were also discussed in this study, including probabilistic evaluations to check for any irregularities. The following chapter will discuss the data and study sites.

Chapter 4

Data and study sites

4.1 Introduction

Secondary data was used in this study, though first-hand data is intrinsically good, as it most likely provides trustworthiness and transparency about the phenomena researchers focus on, is expensive to obtain or gather, and, overall, is limited. Thus, in this study, a secondary dataset was used (Martins et al., 2018). This chapter discusses the data for the study, outlining how the data was used and manipulated to obtain the best possible results that answered the research questions.

4.2 Overview of data

The data used in this study is South African hourly data from SAURAN (Southern African universities radiometric network). A Namibian dataset was also used in this research as part of the dataset that was used for spatial analysis. SAURAN provides solar radiometric data, accessible at `http://sauran.ac.za`. SAURAN currently has 19 South Africa, Botswana, and Namibia radiometric stations. The radiometric stations use Kipp and Zonen CMP11 pyranometers for global irradiance measurements. The data that was obtained from SAURAN had measurements of various predictor variables. Only weather variables were

used for this research as the predictor variables. The predictor variables are Air temperature, total rainfall, humidity, wind direction, wind speed, and barometric pressure. Other variables were added to the weather variables to improve the predictions' forecasting accuracy. These included lagged dependent variables and adaptive trend variables. GHI(Global Horizontal Irradiance) was predicted based on weather variables, and it was the dependent variable. The following section will discuss various variables used in this research.

4.2.1 Global horizontal irradiance

The variable of interest is the Global Horizontal Irradiance (GHI), the dependent variable. It is the variable of interest that is used to measure solar power. GHI is the total solar radiation, the short wave received above the earth. It is the radiation of the sun that reaches the Earth. When the sun's rays hit the solar cells inside the solar panels, these cells transform into electricity. GHI value includes direct normal irradiance (DNI) and Diffuse Horizontal Irradiance (DHI). It is recorded as the amount of solar energy per unit area per unit time in W/m^2. The next section is going to discuss the predictor variables.

4.2.2 Meteorological variables

The energy that comes from solar power depends on several meteorological variables. The performance of a solar cell is believed to be affected by them. We shall explore the variables: relative humidity, temperature, wind speed, air pressure, wind speed, rainfall, Clearness index, nonlinear trend, and lagged variables.

Air temperature

Air temperature affects the transference of GHI. Air temperature is a physical quantity that expresses how cold or hot the air is. It is measured using a Fahrenheit scale (denoted as °F). Temperature contributes to the production of voltage. Too much of it destabilises it.

Solar power generation works well at low temperatures. When the temperatures are high, there is a shift in semiconductor properties which results in a slight increase in current, but it decreases in voltage. A temperature rise will cause damage to the cells leading to a reduction in efficiency and the cells' lifetime. The temperature dropping increases the solar panels' efficiency, and the voltage between cells drops when there is a decrease in temperature.

Barometric pressure

Barometric pressure is one of the predictor variables used in this study. It is also called atmospheric pressure. The pressure at any location exerted on the Earth is proportional to the gravitational force. It is measured using an instrument called Barometer. The units of measurement are given in millibars(mb). Low-pressure results in cloudiness and precipitation, which harm the solar panels.

Relative humidity

Humidity is the quantity representing the amount of water vapour in the air. It is measured as a percentage. An increase in relative humidity reduces solar cell output. Humidity might cause fog or cloud formation, blocking irradiance and reducing voltage if the air outside becomes warm due to increased humidity. Relative humidity brings down the utilisation of solar power panels. A temperature that is constantly hot and humid can degrade solar panels. Humidity can result in water droplets collecting on solar panels, and these reflect sunlight away from solar cells.

Rainfall

Another important variable affecting solar irradiance production is Rainfall, recorded as the total rain. Rainfall is the rain falling in a given time and area, usually expressed in mm. Rainfall itself does not have an impact on the panels but the rain clouds. The proportion of

solar irradiance is reduced when it reaches the Earth due to rainfall and overcast conditions caused by it. Cloud cover due to rainfall might be too thick to allow the panels to produce electricity. Cloud-like fluffy cumulus that brings rain that passes in front of sunlight, the wispy edges will act as a magnifying glass that will cause a strong beam of sunlight to reach the solar panels to rainfall.

Wind direction

Wind direction affects the amount of solar power received on the earth's surface. The wind is the air moving relative to the Earth's surface, which controls the weather. The effect of wind boosts the efficiency of solar power. When the panel is hot, efficiency is reduced; when its cooler efficiency is improved, the wind cools it. When the panel is cooled, more energy gets through it.

Wind speed

Wind Speed also affects the solar radiation received on the earth's surface. As wind speed increases, the amount of solar radiation is reduced.

Clearness index

The clearness index is a measure of the clearness of the atmosphere. It is a measure between 0 and 1, a fraction of the solar radiation transported through the atmosphere to reach the Earth. When the atmosphere has clear conditions, the index is high and has a low value when it is cloudy. It is recorded on an hourly and monthly basis. The data for the clearness index was collected from the Homer Pro website. The equation that defines the average clearness index is given by equation (4.2.1)

$$K_t = \frac{H_{ave}}{H_{0,ave}}, \tag{4.2.1}$$

where H_ave is the monthly average radiation received on earth. $H_{0,ave}$ is the extraterrestrial radiation received on earth given in $[kWh/m^2]$ per day.

Lagged variables

Lagged variables were used in this research. This was done in order to incorporate feedback over time. A lag variable is based on the past values of the time series. Lagged variables were created in this research as explanatory variables and added to regression models. Lagging of explanatory variables is necessary because it helps predict future values t using knowledge of what happened before $t-1$. Lags at 1,12,24,48 were used, 1 for one hour, 12 for 12 hours, 24 for a day, and 48 for two days.

Lagged variables are used when hybrid models are constructed to show important interactions between periods. Lags for distinct periods in the data were used. This was done to check whether there was a causal effect between the lag and the dependent variable. Lagged explanatory variables are used when a researcher suspects the existence of a statistical relationship between the lags in time, like hourly or daily effects. Variables are lagged to remove autocorrelation in variables. After the lagged variables have been added, whether the model has benefitted from them is checked. Adding lagged variables produces a more accurate parameter estimate (Wilkins, 2018). A dependent variable Y_t that is lagged will be a lagged variable with a lag of one period Y_{t-1}. In this research, Y_{t-1} is the variable that is 1 hour. This is done to check whether 1 hour will affect the forecasting results of the model being affected by the next period.

Nonlinear trend

A nonlinear trend variable was used as a predictor variable in this study. Time series nonlinear trend covariate is fitted in this research to capture non-normality in the dependent variable. The simplest way of fitting the nonlinear trend is to use the highest order trends

or quadratic forms by considering $X_{1.t} = t$ and $X_{2,t} = t^2$. The nonlinear trend variable is added to transform a forecast variable before estimating the regression model. In this research, we used regression splines to fit a nonlinear trend. It mainly transforms a nonlinear dataset before predictions to reduce residual confounding.

Regression splines

Regression Splines Regression is a technique that is used to model nonlinear relationships. It performs interpolation smoothly between points. They are a series of polynomial segments that are combined. The regression spline model is given in Equation 4.2.2,

$$Y_i = \begin{cases} \beta_0 + \beta_{11}x_i + \beta_{21}x_i^2 + \beta_{31}x_i^3 + \epsilon_i & x_i < 0, \\ \beta_0 + \beta_{12}x_i + \beta_{22}x_i^2 + \beta_{32}x_i^3 + \epsilon_i & x_i \geq 0. \end{cases} \tag{4.2.2}$$

The advantage of splines is that they are more stable because they are piecewise, making them flexible.

4.3 Data training

The data used in this study was first trained before analysing it. Training data is the initial data set used to train machine learning models that act as a base for forecasting. They are fed to machine learning algorithms to teach them how to make predictions for particular tasks. Training data is either supervised or unsupervised. Supervised training depends on the person performing the analysis. They either tag or annotate the data to their criteria. Unsupervised learning involves the models finding patterns in the data and then making inferences to reach conclusions.

This study splits the data into a training data set and a test set used for model validation. A model built using a training dataset is usually evaluated using a test dataset. There are various methods of splitting the dataset. Some use the Pareto ratio of 80:20, and

others use 70:30. In this study, we used the Pareto ratio. The advantages of data training are that the model is more robust and over-fitting is avoided.

4.4 Single site modelling

Single sites analyses were done on the first two research articles. Paper 1 used two stations; each analysis was independent of the other. The stations used were from Stellenborcsh University (SUN) and the University of Venda(UNV). The two were selected because one was from a coastal area and the other was from an inland.

SUN station is located in a coastal area in Stellenbosch, South Africa and has a latitude of -33.9281, longitude of 18.8654 and Elevation is 119m, and the solar equipment is located on top of the roof of an Engineering building. Stellenbosch University is well known for its dedication to solar-related research, producing highly efficient equipment for recording information related to solar measurements. They have different projects in place. This includes CPS(Concentrating Solar Power) project and SUNSPOT(Stellenbosch Solar Power Technical project). The university also has a group of researchers that belong to a group called STERG(Solar Thermal Research Group), which is a firm that concentrates on solar and thermal energy research. The university is the first to build solar research on the roof of a building. Figure 4.1 shows the solar research equipment.

Figure 4.1: Source STERG website `https://sterg.sun.ac.za/about/`.

VEN is an inland station located in Vuwani, South Africa, with a latitude of -23.1310, a longitude of 30.4239 and an Elevation of 628m. The solar equipment is located at the Vuwani Science Research Centre, as shown in Figure 4.2.

Figure 4.2: Source: SAURAN website `https://sauran.ac.za/`.

The second paper used a dataset from the University of Pretoria(UPR) radiometric station. The SAURAN UPR station is at the University of Pretoria, South Africa. The pyranometer is on top of a science building, giving them good solar exposure. Figure 4.3 shows the position of weather instruments at the top of a building, giving them good weather exposure.

Figure 4.3: UPR Weather instruments.

4.5 Multi-site modelling

This research relates to more than one site performing research based on data from different sites. Paper 3 made use of multiple site analysis. Eight radiometric stations were selected for the study. These were the only radiometric stations from SAURAN that had clean data for the sampling period. However, future research will include the use of satellite data and more stations, including the use of gridded data covering the whole country.

The stations are CSIR - CSIR Energy Centre (1), CUT-Central University of Technology (2), UFH-University of Fort Hare (3), UNV-University of Venda (4), UNZ-University of Zululand (5), UPR-University of Pretoria (6), MIN-CRSES Mintek (7) and NUST-Namibian University of Science and Technology (8).

4.6 Descriptive Statistics: All Meteorological Stations

Table 4.1 shows the descriptive statistics for the meteorological stations used in this research. The dataset used to achieve objective 1 was obtained from UNV(University of Venda) and SUN(Stellenbosch University) stations, and it was an hourly solar dataset. The second dataset was used to achieve objective number 2, and it was obtained from UPR(University of Pretoria); it was an hourly dataset. The remaining dataset was used to achieve objective number 3, and it was obtained from 8 meteorological stations, CSIR - CSIR Energy Centre, CUT-Central University of Technology, UFH-University of Fort Hare, UNV-University of Venda, UNZ-University of Zululand, UPR-University of Pretoria, MIN-CRSES Mintek and NUST-Namibian University of Science and Technology. The dataset used was a monthly daily dataset.

Table 4.1: Descriptive statistics for all stations.

	Datapoints	**Min**	**Q1**	**Median**	**Mean**	**Q3**	**Max**	**Skew**	**Kurt**
VEN	9858	0.0005	238.99	368.39	368.39	378.31	1179.16	1.31	0.84
SUN	17325	0.0002	15.73	28.19	220.58	387.88	1106.53	1.29	0.39
UPR	8791	0.0001	13.46	97.50	245.59	443.19	1142.51	1.11	-0.03
CSIR	62	37.53	187.16	267.44	250.35	331.92	386.41	-0.53	-0.68
CUT	62	43.45	257.17	318.27	301.15	365.21	504.55	-0.70	0.43
FRH	62	50.86	204.34	272.86	260.89	339.02	389.10	-0.54	-0.58
UNV	62	49.89	178.34	245.59	233.37	302.89	356.90	-0.53	-0.84
UNZ	62	62.69	167.36	269.01	245.12	321.84	350.45	-0.48	-1.13
UPR	62	31.16	191.20	254.57	246.36	332.57	379.90	-0.58	-0.58
MIN	62	60.22	208.79	289.37	259.85	336.06	377.33	-0.69	-0.52
NUST	62	148.60	264.90	323.80	317.60	381.10	397.60	-0.69	0.52

4.7 Conclusion

The chapter highlighted the data and study sites. It covered the dataset description, the study's variables, and the weather stations used. The following chapter is based on the first publication. It looks at a method that couples GPR and Core vector regression.

Chapter 5

Twenty-four-hour ahead probabilistic global horizontal irradiance forecasting using Gaussian process regression

5.1 Introduction

In this study, we ventured into various methods adopted to forecast solar power by other researchers. Solar power forecasting can be used as a venturing area to eliminate certain difficulties, for example, reduction in power cuts and carbon emissions-related challenges. Solar power forecasting depends on indicators such as moistness, the sun's path, temperature, and solar power yard attributes that utilise energy from the sun to deliver solar power and dissipating process. The forecasted values will benefit policymakers by providing forecasting information that is vital for competent use. The other important benefits are that it will promote solar power trading, manage the electricity grid, and improve the energy quality supplied to the grid. This will help decrease the costs that are caused by climate reliance.

Solar power forecasting based on several time horizons is vital in health facilities, pow-

ering homes, research activities, educational purposes, PV storage structures management, and many other activities that depend on energy. It helps power grid operators accommodate the load to enhance energy transportation, assign the necessary balance energy from other sources during the period when solar power is inaccessible, and perform activities to do with maintenance at the power manufacturing locations.

The main contribution is using a GPR model coupled with the core-vector methodology in forecasting GHI using South African data. According to our knowledge, this is the first application of GPR coupled with core-vector regression in which the minimum enclosing ball is applied to GHI data.

This chapter discusses Gaussian process regression (GPR) coupled with core vector regression for short-term hourly global horizontal irradiance (GHI) forecasting. The major challenge this chapter addresses is predicting solar power within short periods. GPR is a powerful Bayesian nonparametric regression method that works well for small data sets and quantifies the uncertainty in the predictions. Choosing the appropriate kernel characterising the covariance function is crucial in Gaussian process regression. This chapter adopts the minimum enclosing ball (MEB) technique to select the best kernel function used in Gaussian Process Regression. The MEB improves the forecasting power of GPR because the smaller the ball is, the shorter the training time and captures uncertainty. Hence the 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 (VEN) station in the Limpopo Province. Variables were selected via hierarchical interactions using the least absolute shrinkage and selection operator. The Bayesian approach using informative priors was used for parameter estimation. The GPR models were compared with its benchmark models based on the following evaluation metrics: root mean square error, mean absolute error, and percentage bias. This research's main contribution is using a GPR model coupled with the core vector methodology for 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 to GHI data, to the best of our knowledge.

5.1.1 Research highlights and contributions

This research used a modelling framework of the Gaussian process Regression coupled with core vector regression (CVR). 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 using the informative approach. CVR is a method that was used for kernel selection by applying the concept of the minimum enclosing ball to choose the appropriate kernel to use. The ratio between the margin and radius of the minimum enclosed ball (MEB) is used to measure a kernel's goodness to produce a new minimisation formulation of kernel learning. After developing the Gaussian process regression model, we compared the results with two benchmark models: the gradient boosting method (GBM)and support vector modelling (SVM). The model evaluation was done using the following metrics: root means 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.

This research's first contribution is applying GPR to forecast solar power energy. This method is rarely used to forecast energy. It captures model uncertainty by giving the predicted value as a distribution. Other models do not directly capture this uncertainty. GPs use prior information to produce kernels to describe GHI data, enabling direct optimisation. Hence there is a better trade-off between predicting the data and smoothing it.

The second contribution is the application of MEB to select an appropriate kernel 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 us an accurate description of the data

domain for a given data set. The CVM algorithm is much faster than the benchmark models. It can manage a bigger data set than the SVM and GBM. The minimum enclosed ball is the smallest circle's computation that accommodates the set of points in the Euclidean plane.

The last contribution is that of combining forecasts. This was done by combining GPR and CVR. Bates and Granger (1969) analysed the effects of combining forecasts. Combining the two coupled models improves forecasts since it reduces errors, making them more accurate.

This chapter discusses the data analysis and findings of the research study based on Gaussian process regression coupled with core vector regression. The first segment was to develop exploratory data analysis of the GHI dataset, subdivided into two meteorological stations. Gradient boosting regression and Support vector regression was the benchmark models used to compare with the proposed method. A comparison was done based on RMSE(Root Mean Square Error), MAE(Mean Absolute Error), and Pbias(Probability errors) metrics.

5.2 Methodology

5.2.1 Lasso via hierarchical interactions

Lasso via hierarchical interactions was used for variable selection. It is a technique that uses pairwise interaction regression for this research, and we used pairwise interactions that if k=2, The model we used to select the variables is given in equation (5.2.1).

$$Y = \beta_0 + \sum_j \beta_j X_j + \frac{1}{2} \sum \Theta_{jk} X_j X_k + \varepsilon_j, \tag{5.2.1}$$

where Y is the independent variable GHI, β_0 is the slope, β_j are the regression coefficients, $X_j X_k$ are the predictors with pairwise interaction and $\Theta \in \mathbb{R}^{p \times p}$.

5.2.2 Minimum Enclosed Ball

In this research, we used the minimum enclosed ball (MEB) to find the best kernel function for GPR use. The kernel functions explored are Matern, Radial Basis Function, Rational Quadratic and Dot Product. MEB was coupled with GPR. Minimum Enclosed Ball (MEB) is a branch of Core vector regression(CVR). The MEB algorithm is a method used to find a centre c such that the minimum distance to input points realisations. This is done to fully contain all relevant data points that accurately describe the dataset. The Minimum enclosed algorithm was applied to several kernel functions. The kernel function that gave the minimum radius was the one selected for Gaussian process modelling. The ball that gave the minimum radius is the one that had the minimum enclosed ball.

Let the MEB be the minimum enclosed ball of a set X. It is given in equation (5.2.2)

$$K(c;,r) \tag{5.2.2}$$

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 (2003) and Yildirim (2008) are used.

5.2.3 Gaussian process regression

After selecting the appropriate kernel function, GPR was used to forecast GHI using the selected kernel. The model describes a time-based variable X with a function $f(x)$ that does not relate to any specific distribution. A GPR is built over the assumption that it comprises a collection of continuous random variables $X_1, X_2, ..., X_n$, with n observations of the joint random variables. The components of a GPR are a combination of a covariance function, conditional joint distribution and multivariate normally distributed variables. A mean and covariance function describes the model. The mean function $m(x)$ is assumed to be equal to zero. A GPR is given in Equation (6.2.7). It is described over a real domain of

X over a function $f(X)$.

$$f(X) \sim GP(m(x), k(x, x^1)), \tag{5.2.3}$$

where X is a subset of independent variables $X_1, X_2, ..., X_n$, that is meteorological variables, namely relative humidity, temperature, wind speed, and air pressure. $f(X) = f(X_1), f(X_2, ..., f(X_n)$, x and x^1 are any input points(locations) of any one of the variable X. $f(X)$ is a function of the dependent variable, GHI. $m(x)$ is the mean vector of μ, that is, $E(f(x)) = m(x)$, $k(x, x^1)$ is the covariance matrix, also called the kernel function.

5.2.4 Benchmark models

Various Benchmark models were used in this study. Benchmarking is a technique used to make comparisons between models. The benchmark models used in this research are given in the following sections.

Support vector regression

Support Vector Regression (SVR) is one of the benchmark models used. It is a supervised machine learning technique involving statistical learning theory and structural risk minimisation principle. SVR was introduced by Drucker et al. (1997) and extended from the SVM model. Support Vector Machine (SVM) is one of the most used supervised learning methods for classification problems and regression. Its main purpose is to come up with the best line that can be used to come up with predictions. SVM is a robust forecasting method based on statistical learning. An SVM comes up with a hyperplane in a high or infinite-dimensional space that can be used for regression classification or detecting outliers.

Stochastic gradient boosting regression

Stochastic Gradient Boosting Regression (SGBR) was one of the regression techniques used as a benchmark model in this research. The model is solving and minimising the loss

function $L(x, f(x))$ using Equation (6.2.2)

$$\hat{f(x)} = \arg min \mathbb{E}_{x,y} \left[L(y, f(x)) \right], \tag{5.2.4}$$

where y are the observed values, and $f(x)$ in Equation (6.2.3)

$$f(x) = \sum_{m=1}^{M} \beta_m b(x; \gamma_m), \tag{5.2.5}$$

where the base function $b(x; \gamma_m) \in \mathbb{R}$ are functions of x which are characterised by the expansion parameters γ, β_m. The parameters β_m and γ_m are fitted stage-wise, a process which slows down over-fitting.

5.2.5 Forecast evaluation metrics

The metrics used for forecast evaluation are the mean absolute error (MAE), root mean square error (RMSE), mean average percentage error (MAPE) and percentage bias (PBIAS).

Mean absolute error

Mean absolute error (MAE) measures the average of all magnitude of errors, and it is defined in Equation (5.2.6).

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^{n} |y_i - \hat{y}_j|, \tag{5.2.6}$$

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 (5.2.7).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_j)^2}, \tag{5.2.7}$$

where y_i-are the observed values, $\hat{y}_j$- are the actual values and n represents the number of observations.

Mean absolute percentage error

The mean absolute percentage error (MAPE) is a measure that expresses accuracy as a percentage of error. It is given in Equation (5.2.8).

$$\text{MAPE} = \frac{100}{m} \sum_{i=1}^{m} \mid \frac{\varepsilon_i}{\hat{y}_i} \mid. \tag{5.2.8}$$

Percentage bias

Percentage bias (PBIAS) measures the average tendency of the predicted values to be bigger or smaller than the observed values. It is given in Equation 5.2.9

$$\text{PBIAS} = \frac{\sum_{i=1}^{n} Y_i - \hat{Y}_i \times 100}{\sum Y_i}, \tag{5.2.9}$$

5.3 Results

5.3.1 Data and data pre-processing

Section 5.3.1 represents the results of the analysis. Results of forecasting solar power hourly data were done on two different datasets from radiometric stations, VEN (University of Venda) from Limpopo Province, which is an inland and SUN(Stellenbosch University) from the Western Cape, which is a coastal area, both from South Africa. The predictor variables used are Barometer Pressure, Relative Humidity, Temperature, Wind Direction, Wind Direction Standard deviation, Wind Speed, Wind Velocity Magnitude, hour, month, day and clearness Index. Lag variables were created, lag 1, lag 2, lag 12, lag 24, lag 48, and a covariate temperature×GHI. Data that was incorrect, corrupt, incorrectly formatted, and duplicate data within the dataset were removed. Data that was missing was imputed using a package called MICE in R. The data used for the analysis was split into two, training sets consisting of 80% of the dataset and the remaining 20%

was the test set. The statistical software R(4.2.1) programming performed the exploratory analysis, Lasso, and benchmark models. The packages that were used for the research are: *glmnet*-Lasso,*gbm*-gradient boosting, *forecast*-forecasts evaluation, *e1071*-SVM, *hydroGof*-Goodness of fit functions. The software Python 3.8.8 was used for GPR and MEB. The following packages were used *numphy*-multidimension array, *pandas*-datasets, *matplotlib*-visualisation,*sklearn.gaussian_process.kernels*-GPR analysis, *sklearn.gaussian_process.kernels*-incorporates kernels and *miniball*-Minimum Enclosed Ball.

5.3.2 Exploratory data analysis

Exploratory data analysis was done on the dependent variable GHI. Exploratory analysis is a method of analysing datasets to summarise their main characteristics and understand the data. A summary of the datasets of the two radiometric stations was produced, VEN(University of Venda) and SUN(University of Stellenborsh), based on their main characteristics through observing descriptive statistics through visualising graphs and descriptive numerical measures. We started with analysing GHI using the VEN dataset. Descriptive numerical values were computed and are shown in Table 5.1. The maximum value of the GHI for the VEN data is 1179.16 W/m^2. The coefficient of skewness is a measure of the asymmetry of the dataset. The dataset is said to be skewed to the right or to have a positive skew if it has a longer tail to the right.

On the other hand, it is said to be skewed to the left or negatively skewed if it has a longer tail to the left. The coefficient of skewness is 1.31, indicating the distribution is right-skewed, and the skewness can also be seen in Figure 5.1, the tail is longer on the right side. The coefficient of skewness is 1.29, and the value is positive, indicating that the data are skewed to the right.

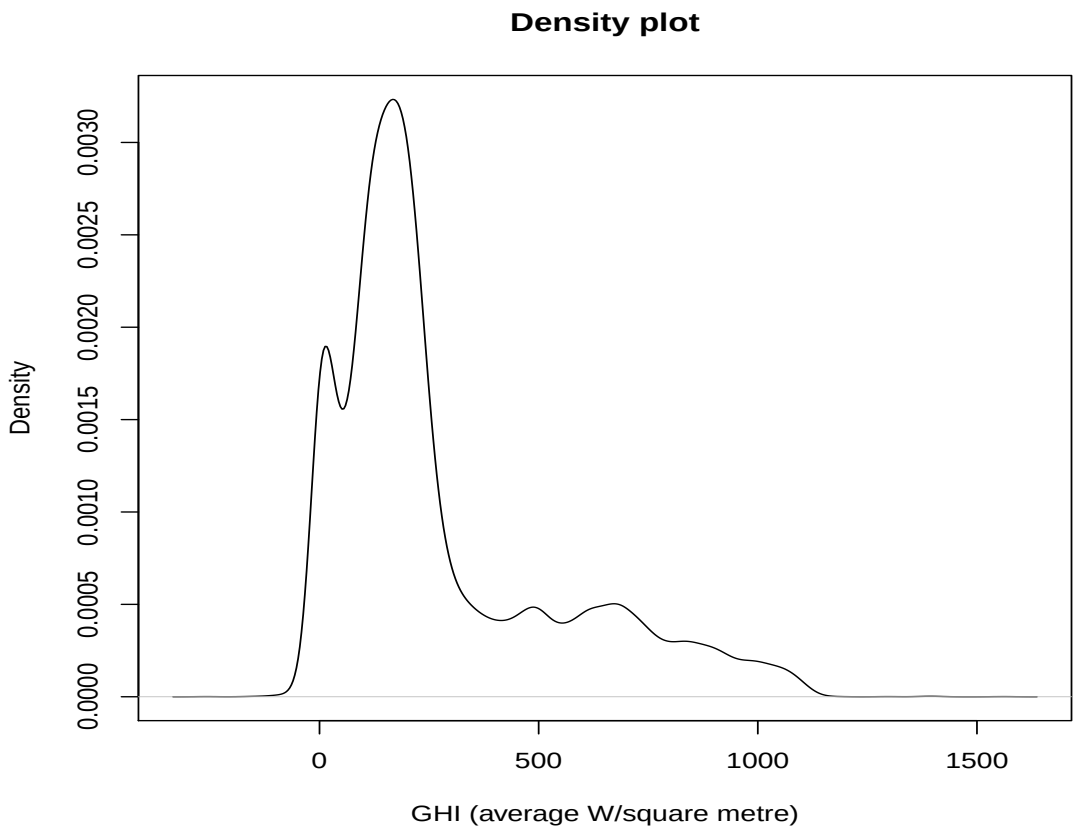


Figure 5.1: Empirical probability density function showing the distribution of GHI values (W/m^2) at the University of Venda site.

The coefficient of kurtosis is a measure of the peakedness or flatness of distribution, and it is based on the moments of the distribution, and it is based on moments of the distribution. It is used to describe the shape of the distribution. When the kurtosis is greater than 0, the distribution is peaked, and when the coefficient is less than 0 indicates a flat distribution. The coefficient of kurtosis is greater than zero, meaning that the distribution is peaked. The coefficient of kurtosis is 0.398, which means the distribution appears moderately distributed.

Looking at SUN data, the maximum GHI is 1106.53 W/m^2, and the mean is 220.58.

Table 5.1: 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

The Density plot in Figure 5.2 also show how the data is distributed. It is skewed to the right and peaked.

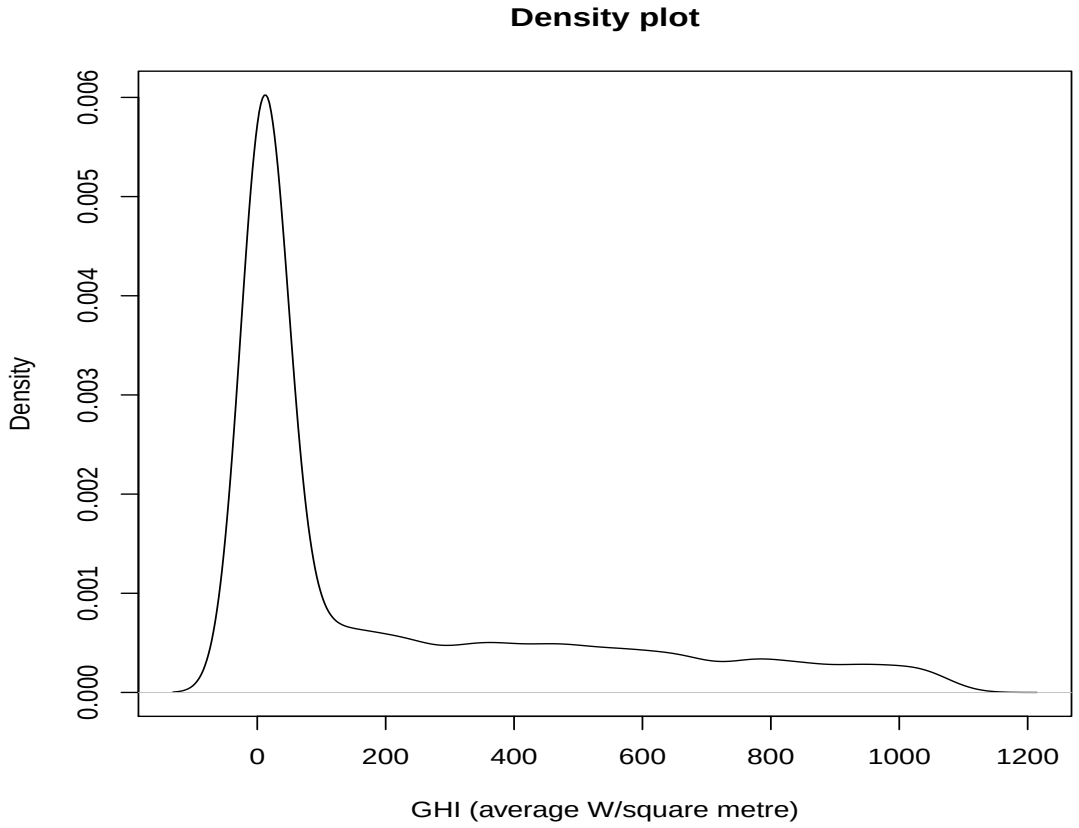


Figure 5.2: Empirical probability density function showing the distribution of GHI values (W/m^2) at the Stellenbosch University site.

Table 5.2 shows the correlations between the GHI of each radiometric station and the important weather variables used as covariates in this study. There is a fairly strong positive correlation between GHI and temperature. Similarly, there is a strong negative correlation between GHI and relative humidity, while the relationship between GHI and barometric pressure suggests a weak negative correlation.

Table 5.2: 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

VEN data

An exploratory analysis was also done for VEN data; the results are shown in Figure 5.3.

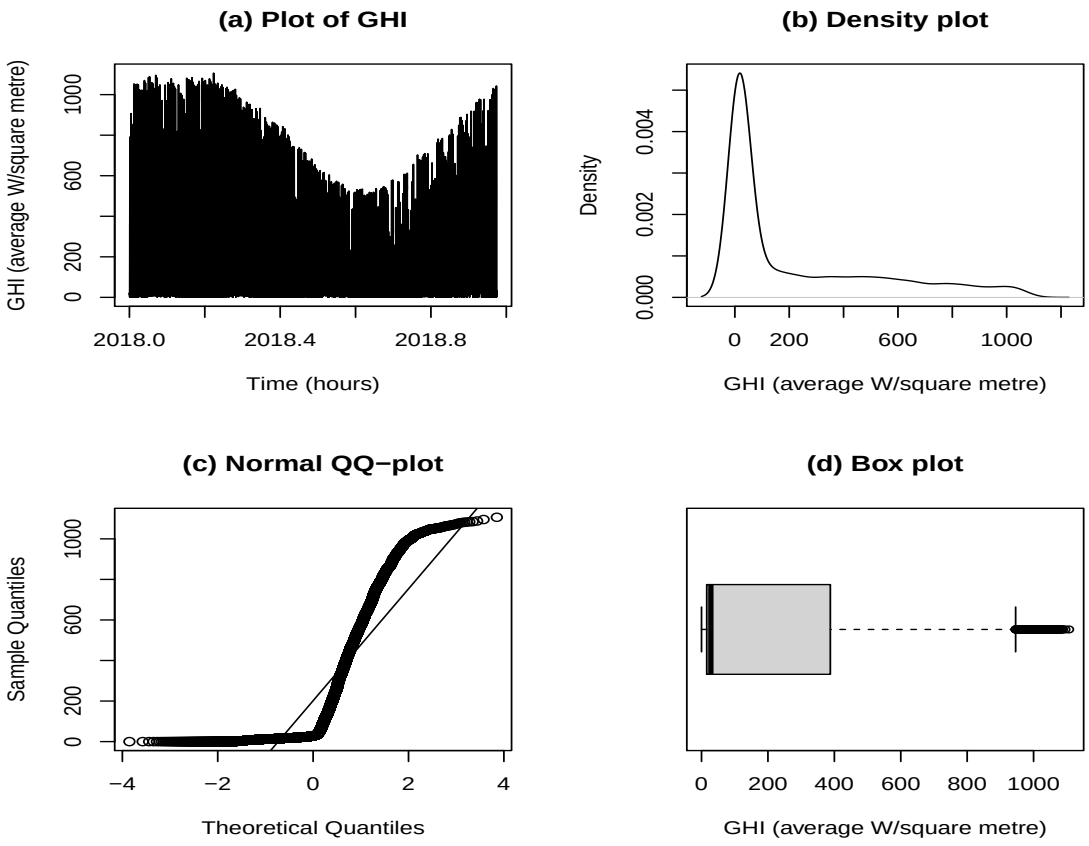


Figure 5.3: Descriptive analysis plots of Global horizontal irradiance for VEN dataset.

Plot(a) shows a time series plot of GHI averaged hourly from the beginning of 2018 to the beginning of 2019. Plot(c), the Q-Q plot, shows that the GHI data is not normally distributed, and this can also be seen from plots (b), the density plot, and (d), the box plot. Table 5.2 shows the summary statistics for the VEN data.

Plots of data over time(VEN)

A periodogram is used to identify the dominant periods (or frequencies) of a time series. This can be a helpful tool for identifying the dominant cyclical behaviour in a series, partic-

ularly when the cycles are not related to the commonly encountered monthly or quarterly seasonality. Figure 5.4 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. Thus, it takes 15 time periods to come up with a cycle.

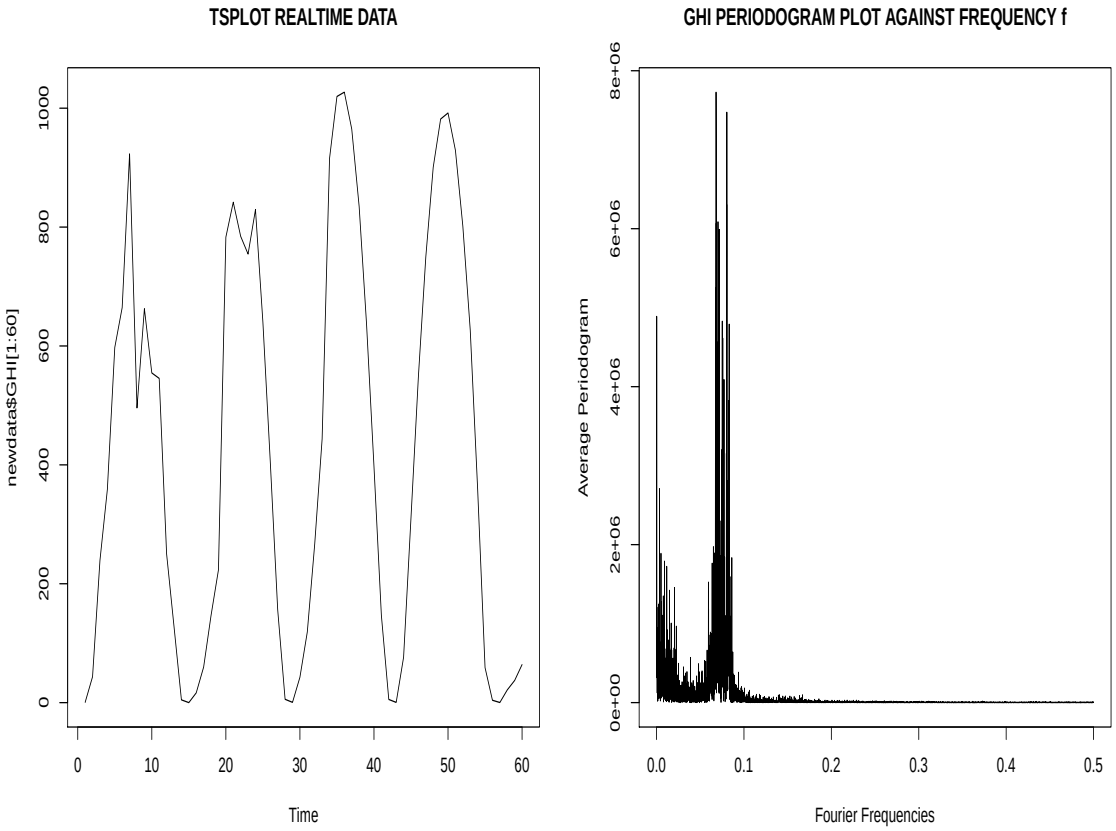


Figure 5.4: **Left panel:** Times series plot. **Right panel:** Periodogram plot.

Figures 5.5 and 5.6 show the multiple histograms and scatter diagrams for independent variables. Figure 5.6 shows a matrix of scatter diagrams showing the relationship between GHI and each independent variable. All the variables except Temperature, Relative Hu-

midity and Wind Direction appear not linearly related to GHI.

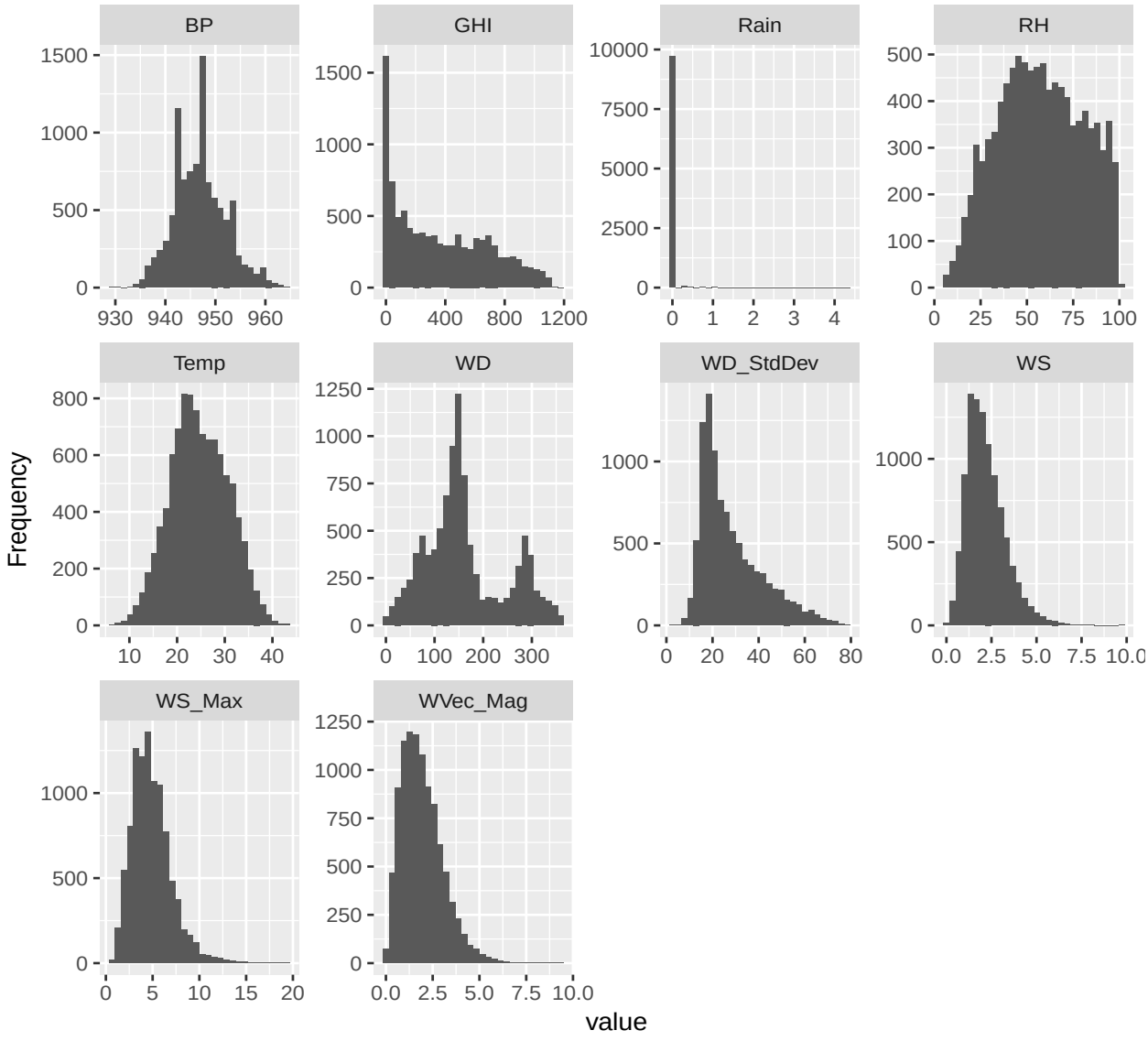


Figure 5.5: Multiple histograms for VEN station.

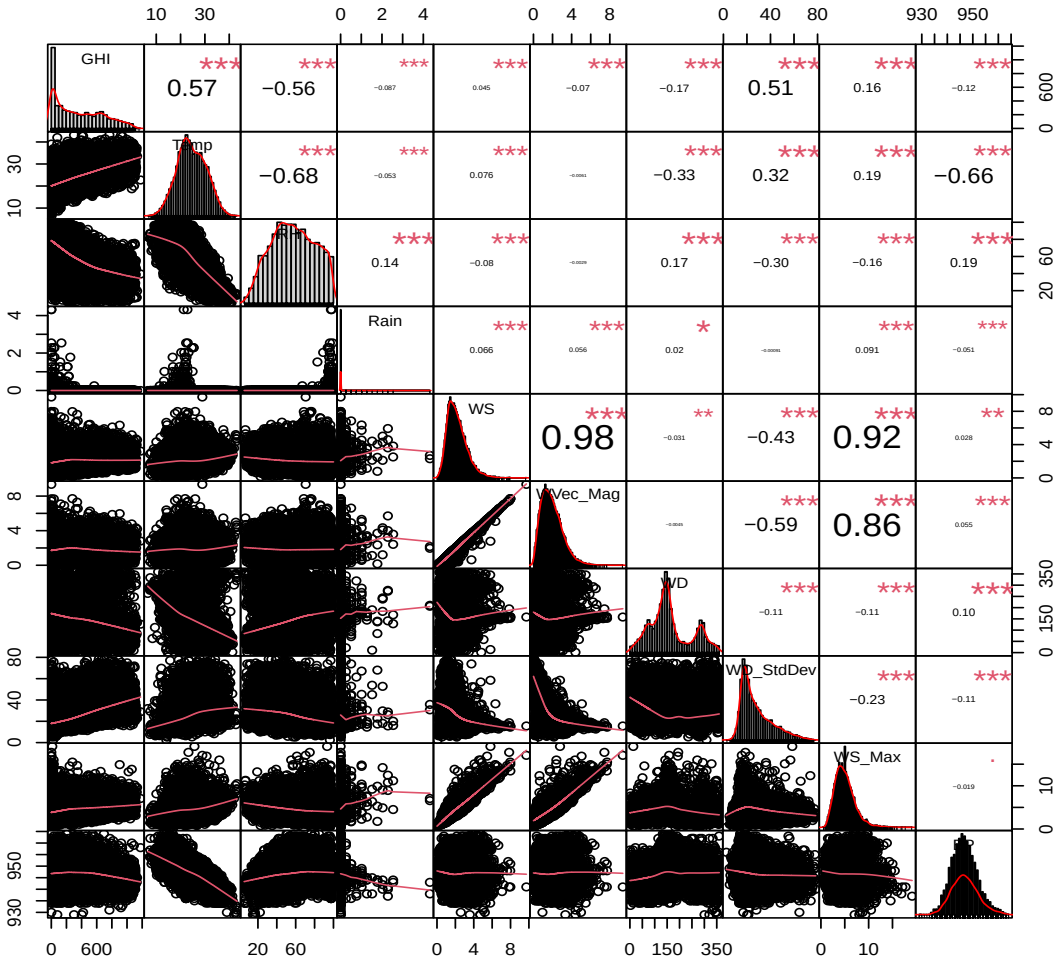


Figure 5.6: Multiple scatter diagrams for VEN data.

SUN Data

An exploratory analysis was also done for SUN data, and the results are shown in Figure 5.7.

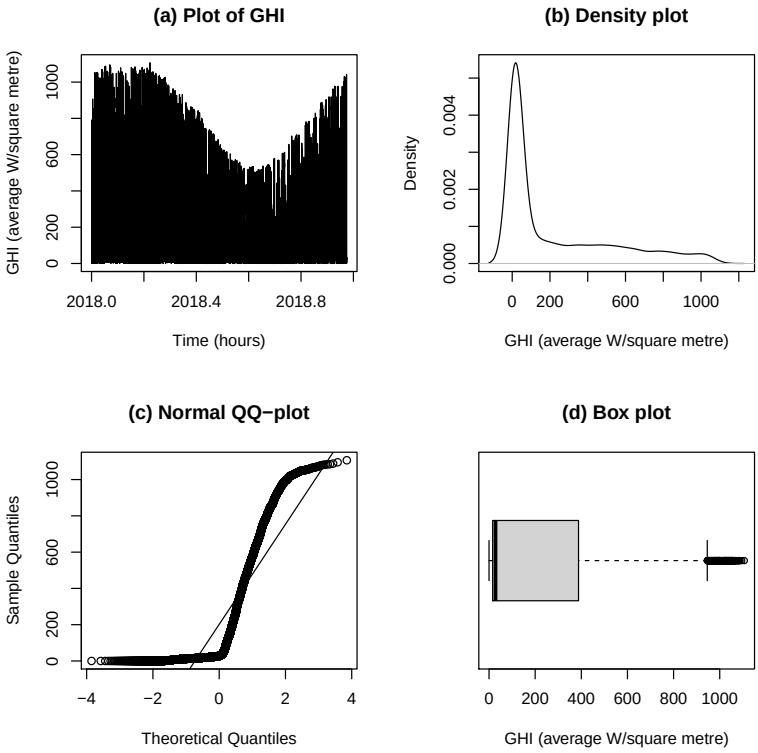


Figure 5.7: Plots of descriptive analysis of Global horizontal irradiance for SUN station.

Plot(a) shows a time series plot of GHI averaged hourly from the beginning of 2018 to the beginning of 2019. Plot(c), the Q-Q plot, shows that the GHI data is not normally distributed, and this can also be seen from plots (b), the density plot and (d), the box plot.

Plots of data over time (SUN)

Plots of data over time were produced to visualise what was happening to the data hourly. A plot was done to show the trend of GHI for one h.

Figure 5.8 shows a periodogram for SUN data, identifying dominant time series cycles. 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 develop a cycle. Figures 5.9 and 5.10, respectively, show the multiple histograms, scatter diagrams, and correlation matrix that show the relationships between variables for SUN data.

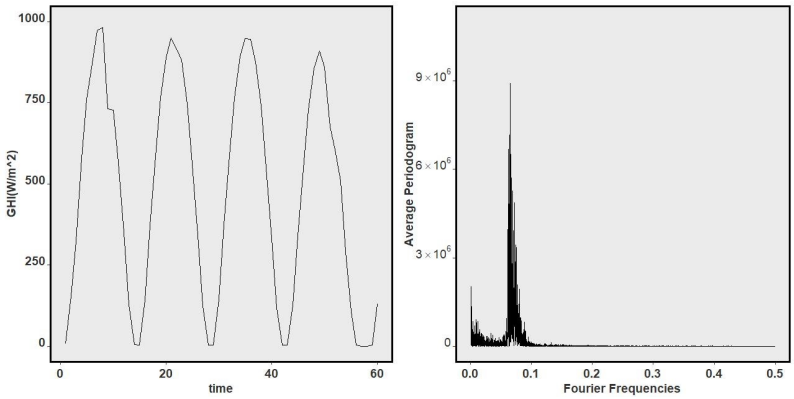


Figure 5.8: Tsplot real-time data. **Left panel**: Time series plot. **Right panel**: Periodogram plot.

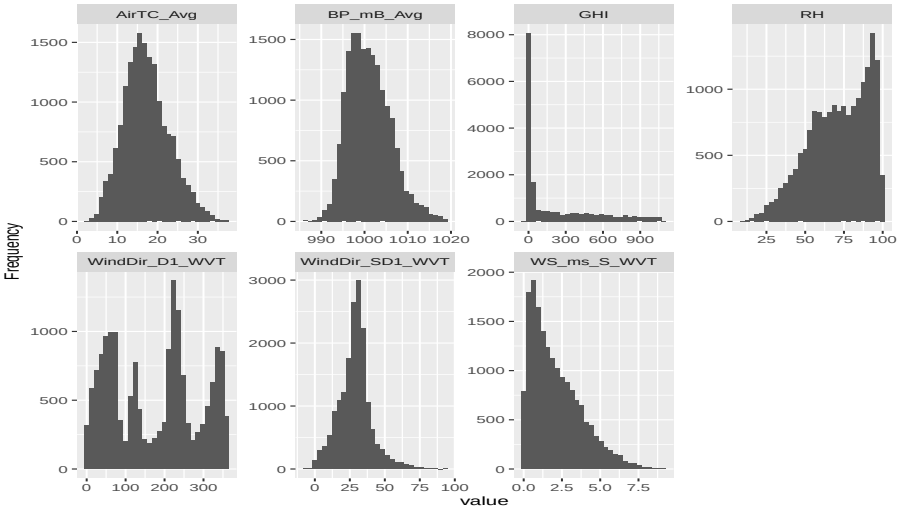


Figure 5.9: Multiple histograms for SUN data.

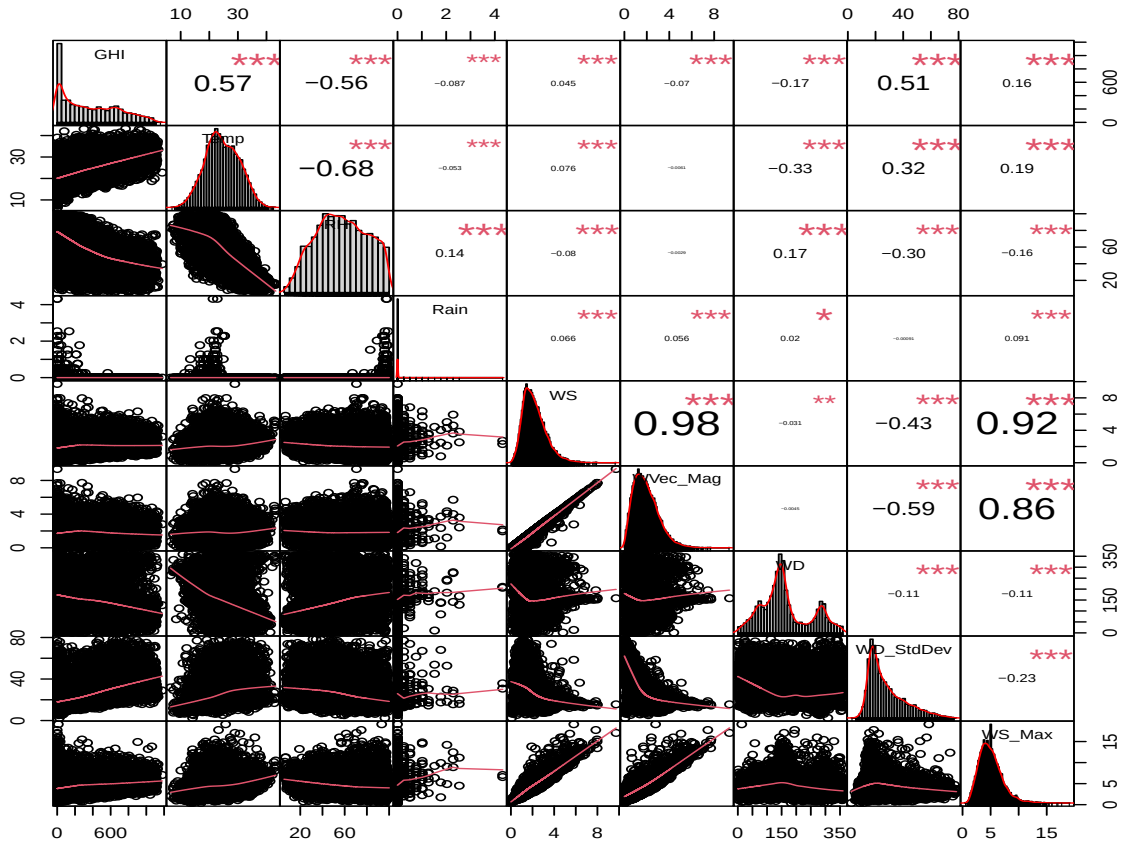


Figure 5.10: A plot showing multiple scatter diagrams and correlation matrix for SUN data.

5.3.3 Forecasting results

Variable selection using Lasso via hierarchical interactions

We used Lasso via hierarchical interaction to select the variables to be used to model the data

VEN data The selection of variables was done for VEN data using Lasso via hierarchical interactions, and a summary of the results is presented in Figure 5.11. Twelve variables

were selected and used in all the models in this research.

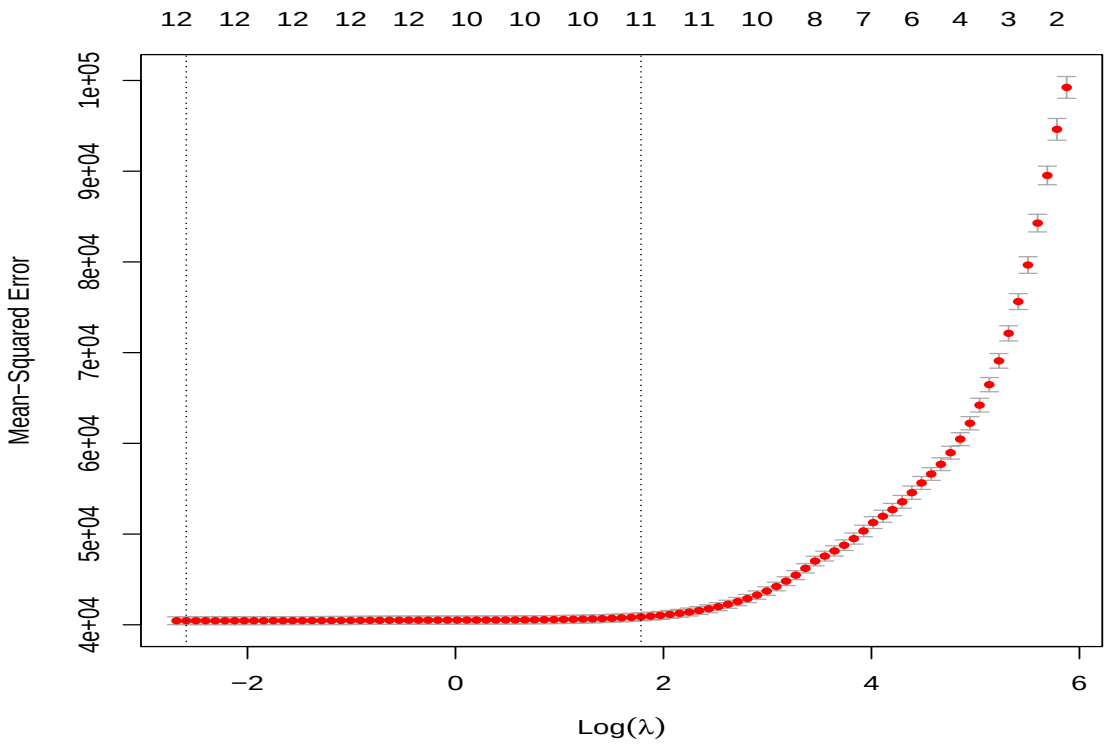


Figure 5.11: A plot showing the results for Lasso regression for VEN data.

Figure 5.11 shows the Lasso results for VEN data, with the graph showing the 12 selected variables. The variables selected are Temperature, Relative humidity, Rain, Wind Speed, Wind Velocity Magnitude, Wind Direction, Wind Direction Standard Deviation, Wind Speed Max, Barometer, noltrend, lag1 and lag2. The plot helps us visualise MSE, which is on the y-axis with $\log(\lambda)$ and given on the x-axis, the cross-validation curve (with red dots). The variables which are included at these points are on the top. The two dotted lines are lower and upper curve limits: the λs. The error is very high when we have a few variables, but the error approaches zero as the number of variables increases.

SUN data Variable selection was also done using SUN data. Figure 5.12 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. The variables selected are relative humidity, air temperature, wind direction standard deviation, wind speed, wind direction, barometer pressure, noltrend, lag 1 and lag 2. 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.

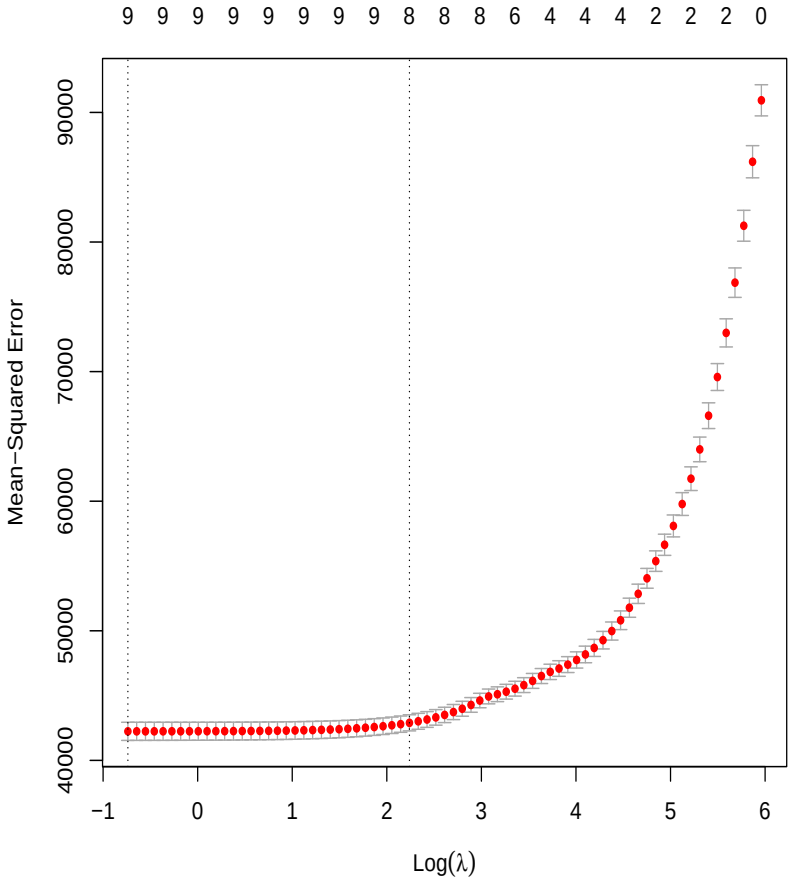


Figure 5.12: A plot showing the results for Lasso for the SUN data.

Gaussian process regression results

The following results are the results of the proposed methodology, GPR coupled with core vector regression. The analysis was done in two parts, one without interactions and the other with interactions. This was done in order to see whether interactions had effects or not on the results produced. GPR analysis was done based on the four kernels: Matern, rational quadratic, dot product, and radial basis function because these are the popularly used kernels. The results that follow were obtained using the statistical package called Python 3.8.8.

VEN data without interactions Figure 5.13 shows a static plot of the predicted solar irradiance values against the observed values of VEN data without interaction. The output shows the best-chosen results, giving the minimum enclosed ball. The predicted values are in blue, closely following 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.

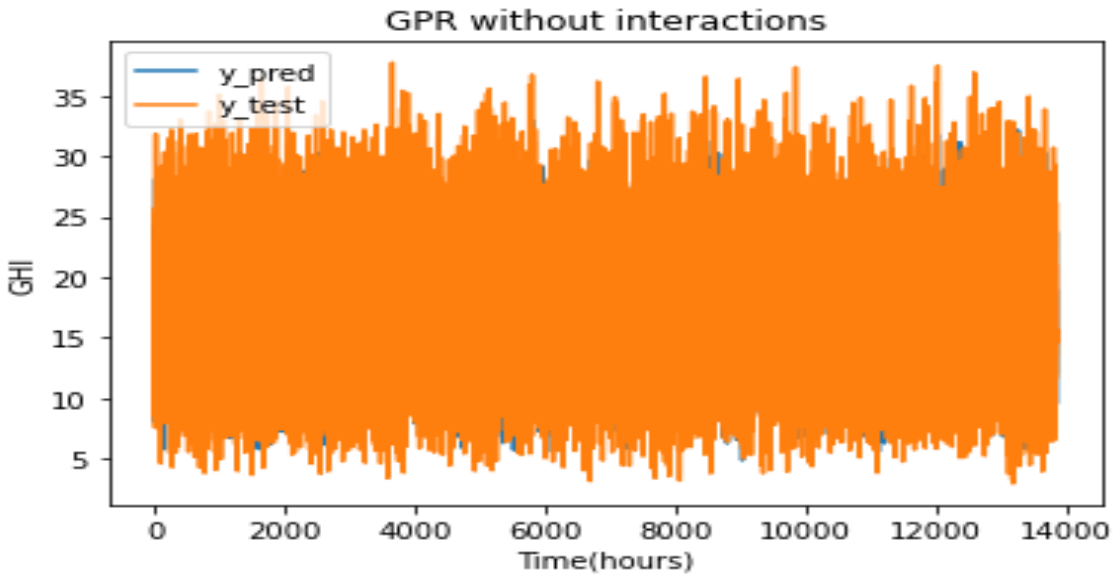


Figure 5.13: Predicted values vs observed values for GPR using a Radial Quadratic kernel.

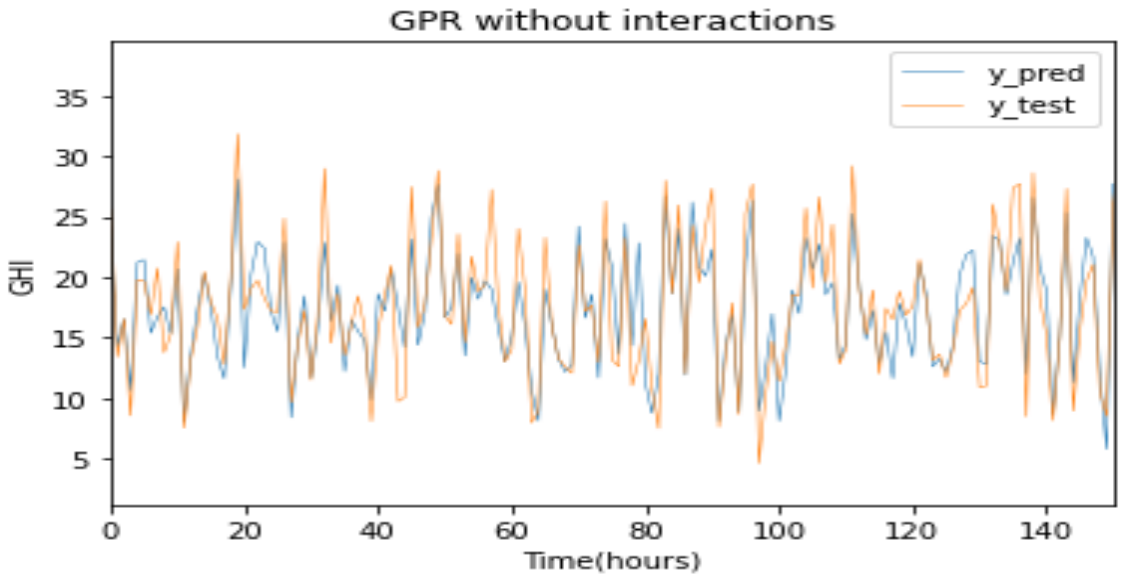


Figure 5.14: Observed values vs GPR without interactions using a radial quadratic kernel over a shorter period.

Figure 5.14 shows the observed GHI versus the predicted values using GPR with interactions using a radial quadratic kernel for a selected dataset with a shorter time frame for VEN data. The observed data are in red and predicted in blue, which was observed over ten h.

VEN data with interactions Forecasting was also done by applying interactions. Again the radial quadratic kernel produced the minimum radius. Hence it was used to come up with forecasts. Figure 5.15 shows a static plot for VEN data with interactions. The observed data are in red and predicted in blue. The observed data is also plotted against actual data using a smaller dataset, and forecasting was done for ten h using 150 observations, shown in Figure 5.16.

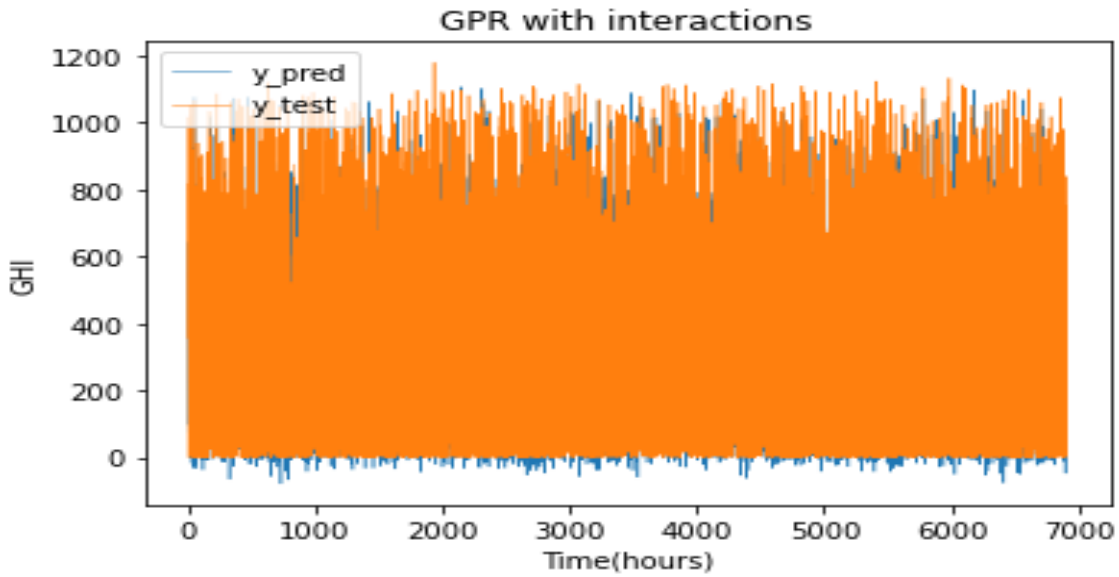


Figure 5.15: A plot of predicted values via GPR using the radial quadratic kernel using VEN data with interactions.

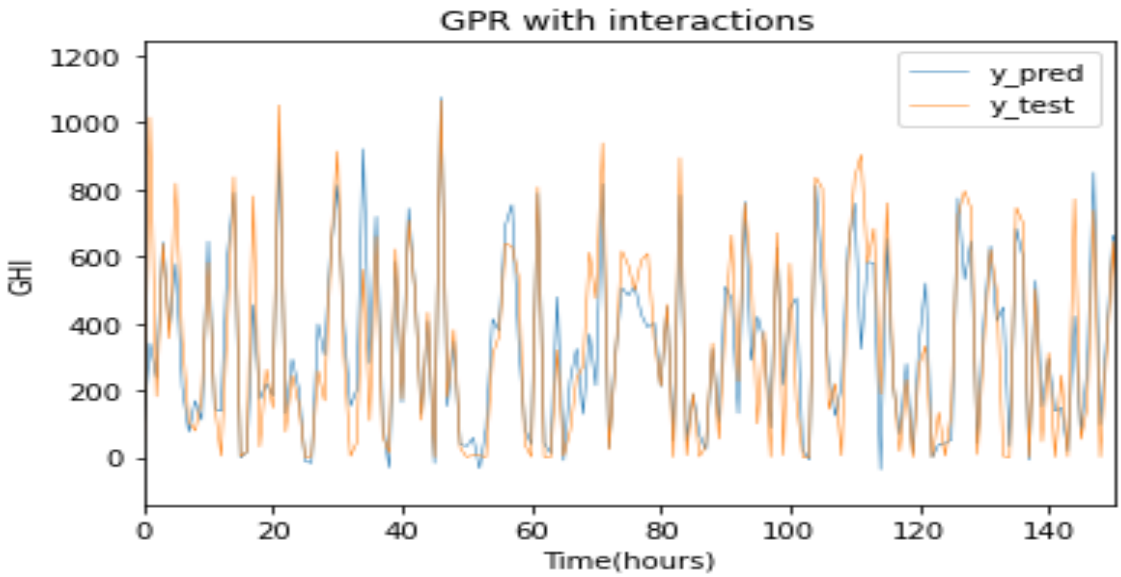


Figure 5.16: Plot of predicted values via GPR using the radial quadratic kernel using VEN data with interactions using a shorter dataset.

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 5.17, a static plot of SUN data. The predicted solar irradiance in blue closely follows the observed values in orange.

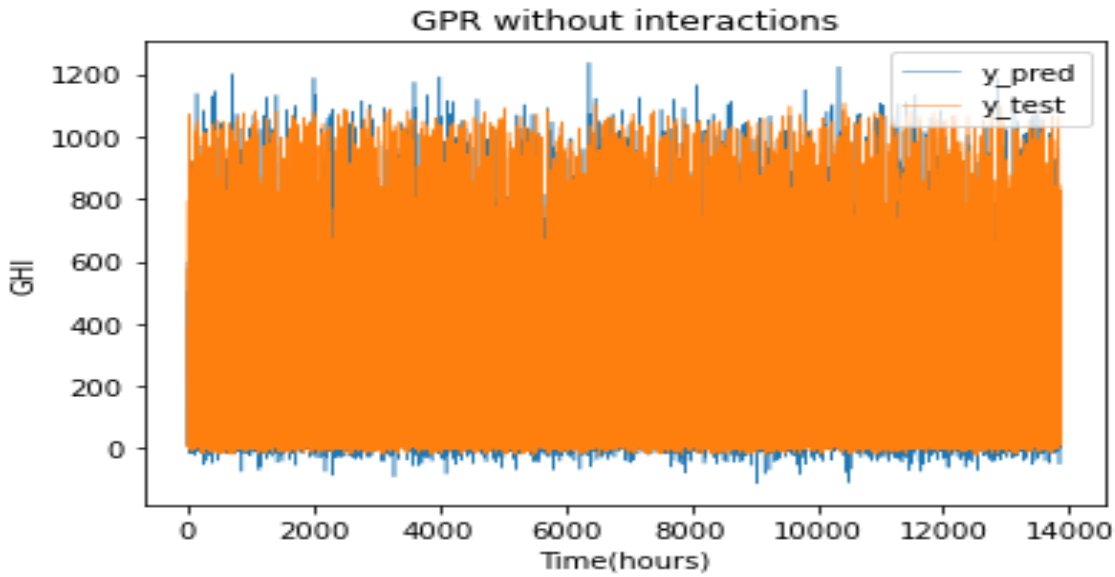


Figure 5.17: Predicted values vs observed values for GPR with no interactions using a radial quadratic kernel.

Figure 5.18 shows a plot for SUN data, only showing observations over a shorter period of 150 hours. The observed data are in red and predicted in blue, observed over ten h, and follow each other very closely.

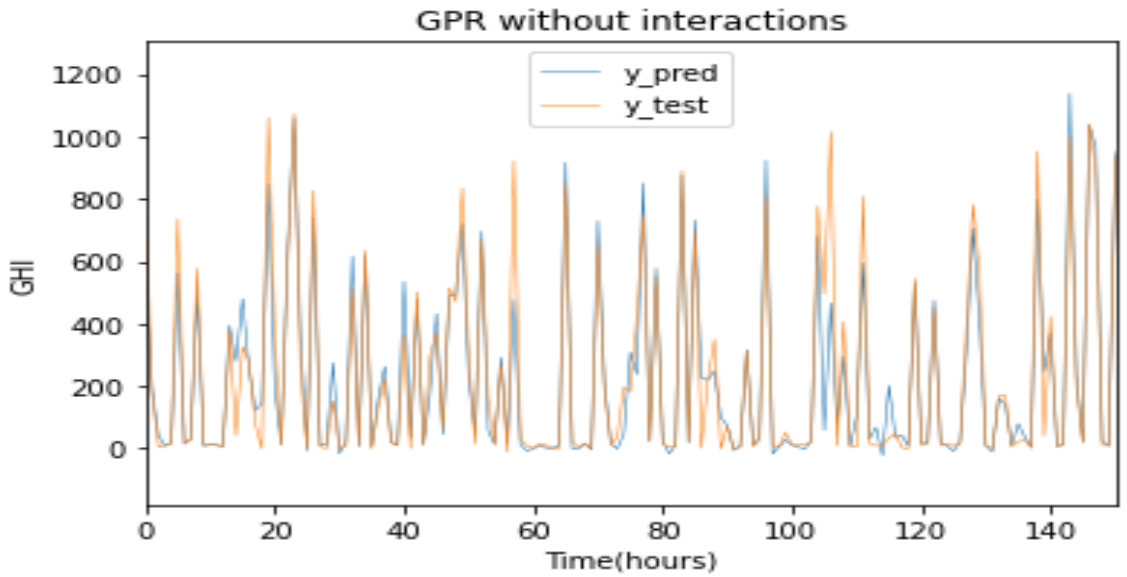


Figure 5.18: Predicted values using GPR vs actual values over a shorter period with no interaction.

SUN data with interactions GPR analysis was also done on SUN data with interactions included. The results are shown in Figure 5.19, a static plot. The predicted solar irradiance in blue closely follows 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. Figure 5.20 shows a real-time plot for SUN data. The observed data are in red and predicted in blue, observed over ten h, and follow each other very closely.

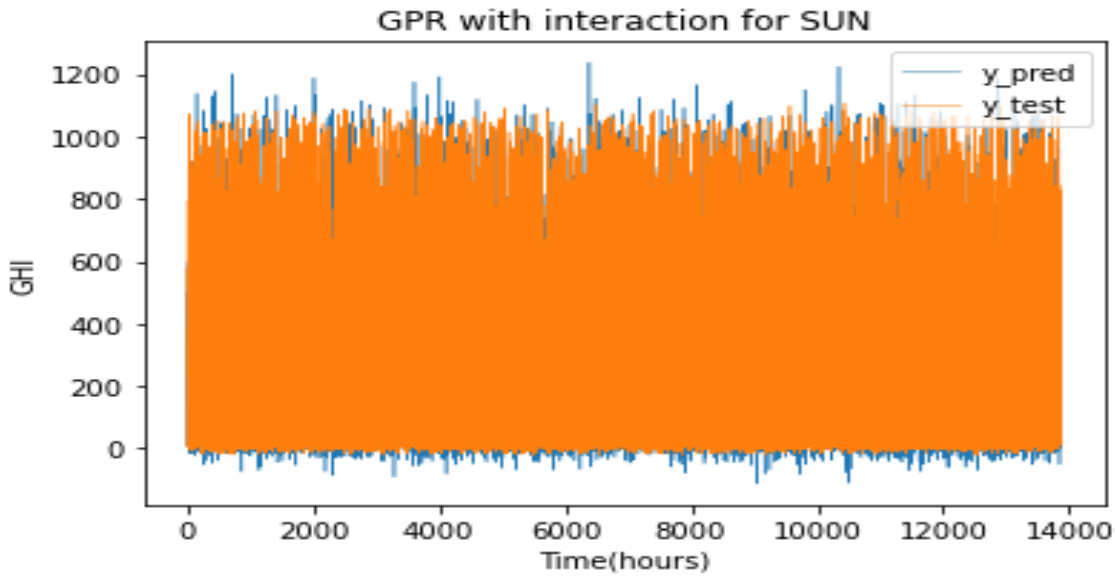


Figure 5.19: A plot of the predicted values using GPR vs actual values for the SUN dataset with interactions.

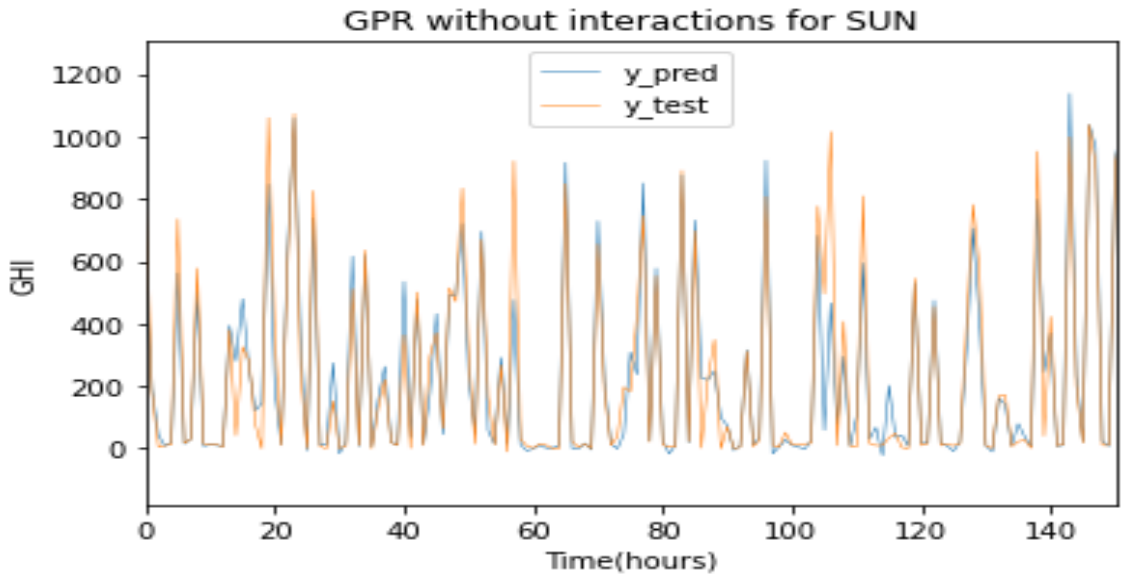


Figure 5.20: A plot of predicted vs actual observations for SUN using a shorter dataset with interactions.

Benchmark models and evaluation of prediction techniques

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

To choose the best model, we used the MAE, RMSE, and Pbias to measure the accuracy of models to compare the benchmark models, Gradient Boosting Regression and Support Vector Regression, and the GPR model. The idea was to devise a model that minimises these three measures. The results are summarised in Table 5.3 and 5.4. For both stations, GPR produced the lowest values of MAE and RMSE.

The models considered are given below

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 5.3: Evaluation metrics for VEN station.

Model	RMSE	MAE	Pbias
M1(GPR with nointeraction)	34.1	20.9	0.2
M10(SVR interaction)	130.5	91.1	1.2
M6(GBM interaction)	142.7	103.8	0.6
M2(GPR with interactions)	148.4	102.8	1.3
M5((GBM nointeraction))	168.5	126.6	7.1
M9(SVR nointeraction)	173.1	126.6	9.5

Table 5.4: Evaluation metrics for SUN station.

Model	RMSE	MAE	Pbias
M3(GPR nointeraction)	2.7	2.1	0.2
M4(GPR interaction)	122.5	64.9	43.4
M8(GBM interaction)	164.7	110.5	10.6
M7(GBM nointeraction)	173.9	116.9	11.9
M12(SVR interaction)	185.1	104.0	18.4
M11(SVR nointeraction)	199.2	112.9	21.6

The code that was used to produce the GPR models can be found on GitHub at https://github.com/edina440/Corevector-GPR/blob/main/README.md. The code for the GBM and SVR, which were done via Python, is found at https://github.com/edina440/Corevector-GPR/blob/main/SVR

5.4 Discussion

The present study was motivated by previous research by other authors such as Zhandire (2017), Mpfumali et al. (2019), Govender et al. (2018) and Mutavhatsindi et al. (2020*b*) Maritz et al. (2018), Dahl and Bonilla (2019), 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.

Lasso was performed on the data incorporating interactions effects to select variables for 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 used. To improve forecasting results, the MEB was adapted to develop an appropriate kernel that would yield the best results. Different kernels were used to achieve this: radial basis function, Matern, rational quadratic, and dot product. Empirical results showed the radial basis function as the best kernel since it had the smallest radius in all cases.

The various models' results are shown in Table 5.3 and 5.4. Table 5.3 shows the GPR, GBM and SVR results with or without interactions for the VEN station. The results show that the model with the least error measurements is GPR without interactions. This means that the predicted values produced by GPR without interactions performed better than the other models. Table 5.4 shows the results of the GPR model against the benchmark models for the SUN stations. The results show that the GPR model without interactions performed better. Based on the evaluation metrics RMSE, MAE and Pbias, the model that produced the least values is the GPR model. Interaction effects did not change the GPR model but did for the other benchmark models.

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 the radial basis kernel for both stations outperformed the other kernels producing the minimum enclosed ball. The decision was arrived at based on the radius of each kernel.

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 into the grid in balancing the 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 and electricity demand, among others. System operators can use the developed models 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. In this case, the mean function is computed using Bayesian regression but is not easily interpreted for Gaussian process models.

5.5 Conclusion

It is a crucial and hard task to forecast short-term solar power, which greatly impacts the control and management of the electric grid. This study presented an application of Gaussian process regression coupled with core vector regression on two radiometric stations from South Africa. The 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 in Gaussian

process regression produced better results than the benchmark models. Most 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 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, it will be more interesting to include non-weather variables like calendar effects and denoising the data using either wavelets or empirical mode decomposition before applying the forecasting models.

This chapter discussed the application of Gaussian process regression coupled with core vector analysis to develop solar power predictions. The predictions done in this study were point estimates. The chapter that follows is going to make use of quantile regression, which is an interval-based approach. This was expected to improve forecasting accuracy.

Chapter 6

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

6.1 Introduction

This research aims to develop a forecasting model to predict solar power more accurately. Forecasting is improved in this chapter by moving from point estimation to a probabilistic way of forecasting using interval estimation. Probabilistic modelling is done through quantile regression, which is combined with GPR. We compared two hybrid models, BSTS and GPR, to select the best among these two hybrid models. Predicting models are being developed daily, but an inseparable part needs attention: error reduction. The bigger the error is, the less accurate the predictions will be. Thus, it is crucial to reduce errors to improve the accuracy of results. Forecasts were combined using QRNN, PLAQR, QRA and OPERA to improve the forecasting models' prediction power. We expect to develop an improved modelling framework for Global Horizontal Irradiance by adopting these hybrid models. Adding quantile regression to the Gaussian Process Regression model gives a methodology that gives more efficient estimators Koenker and Bassett (1978); Koenker

(2005). Quantile regression is a way of estimating the conditional distribution resulting in efficiency being improved. This research combines information from the GPR with the Quantile regression modelling to gain efficiency. Adding quantile regression is more robust to outliers and more comprehensive than other point estimation methods Koenker (2005).

6.1.1 Contributions of the research

The first contribution of this research is the coupling of the Gaussian Process and Quantile Regression to develop a modelling framework to forecast GHI. The approach that was used in Chapter 5 is the one that was further combined with quantile regression in the chapter. This approach provides estimates catering to datasets' uncertainty and scalability, hence its strong performance. Also, a combination of two models gives us a hybrid model. Thus, we expect to produce improved results.

Secondly, a modelling framework based on the Bayesian Structural time series on GHI data. This model's performance is superior because it avoids over-fitting. It makes use of a combination of Bayesian Structural and time series analysis. 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. Thus, we expect better results since this is also a hybrid model.

Thirdly, the efficiency of the developed models, the GPR and BSTS, will be compared against each other to develop a model that best describes solar power generation, thus selecting the most excellent model. Probabilistic tools like proper scoring rules and Murphy diagrams were used to evaluate the forecasts.

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 (1969) concluded that combining forecasts reduces errors. The other sections are arranged as follows: Section 6.4 empirical data analysis, Section 6.5 presents the discussions, and

finally, Section 6.6 is the conclusion.

6.2 Methodology

Figure 6.1 shows the schematic presentation of the adopted methodology, with all the stages laid out step by step.

6.2.1 Schematic presentation of methodology

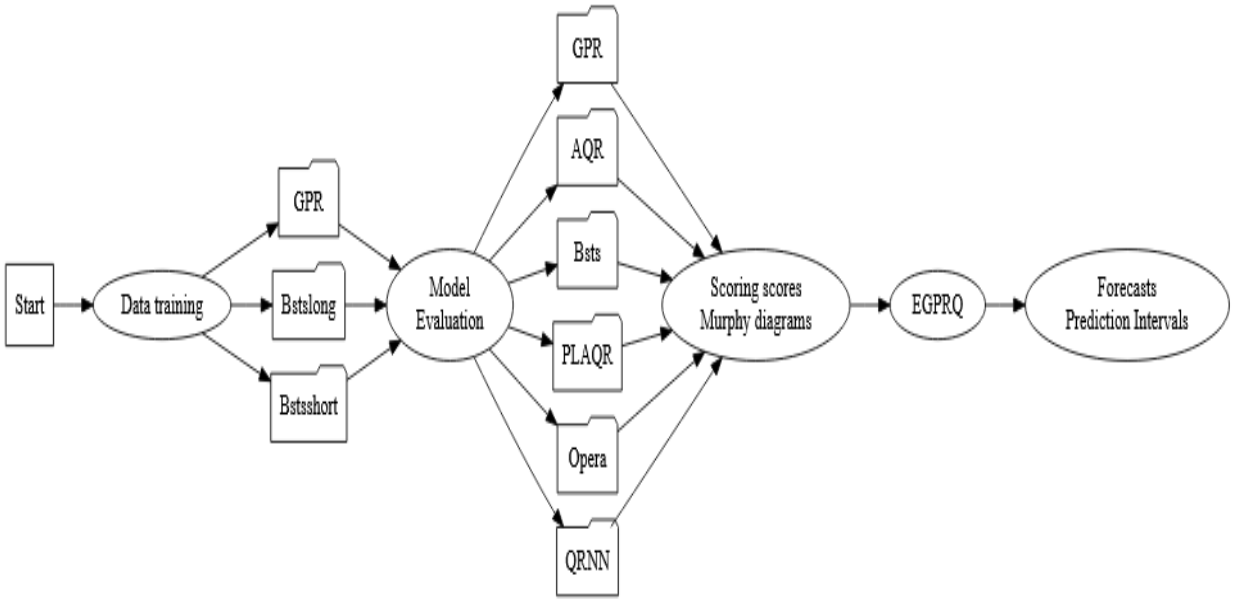


Figure 6.1: Methodology flowchart.

Various methods were adopted in this research and have been discussed in Chapter 3. The methods are briefly explained in the subsections that follow.

6.2.2 Variable selection methods

The first thing that was done before modelling was variable selection. Traditionally, for variable selection, most authors select one particular variable selection method Chen et al. (2013); Amarasinghe et al. (2020). In this research, various methods were first explored for variable selection, and one method was selected. The methods used include feature selection techniques and shrinkage methods.

6.2.3 Feature selection methods

The objective of feature selection is to improve the accuracy of predictive models by selecting the best set of variables to use. The feature selection method is explained in the next subsections.

Multivariate adaptive regression spline

Multivariate adaptive regression spline (MARS) was one of the methods explored in this research. It is a nonparametric regression method developed by (Friedman, 1991) and given in Equation (6.2.1).

$$(x-t)_+ = \begin{cases} x-t & , \text{ if } \ x > t \\ 0 & , \ otherwise \end{cases} \text{ and } \ (t-x)_- = \begin{cases} t-x & , \text{ if } \ x < t \\ 0 & , \ otherwise \end{cases} \tag{6.2.1}$$

where $+$ is positive part, x are the weather predictor variable, y is the response variable GHI, t is the break point or knot point.

Boruta

Boruta is one of the feature selection methods used to develop a set of variables for the analysis. It is a wrapper feature selection method that uses Random Forecast.

Gradient boosting regression

Gradient Boosting Regression (GBR) was one of the regression techniques used for variable selection in this research. It is a feature selection algorithm which is a flexible and straightforward method. GBR regression is solved by selecting a subset of the training set at each iteration to accommodate the base learner, a basic parameterised model. Gradient boosting (GB) model is solving and minimising the loss function $L(x, f(x))$ using Equation (6.2.2)

$$\hat{f(x)} = \arg min \mathbb{E}_{x,y} \left[L(y, f(x)) \right], \tag{6.2.2}$$

where y are the observed values, and $f(x)$ in Equation (6.2.3)

$$f(x) = \sum_{m=1}^{M} \beta_m b(x; \gamma_m), \tag{6.2.3}$$

where the base function $b(x; \gamma_m) \in \mathbb{R}$ are functions of x which are characterised by the expansion parameters γ, β_m. The parameters β_m and γ_m are fitted stage-wise, a process which slows down over-fitting. The next subsection is based on the shrinkage method.

6.2.4 Shrinkage methods

Shrinkage methods were used in this research to explore variable selection. Shrinkage refers to reducing variation in sampling when data points are shrunk towards a central point. This enables the minimisation of overfitting and underfitting of a dataset by introducing the penalty term, thus reducing errors.

Least absolute shrinkage and selection operator(LASSO)

The least absolute shrinkage and selection operator (Lasso) was one of the shrinkage methods explored for variable selection. Lasso was first developed by Tibshirani (1996). It performs L_1 regularisation, which results in sparse models with few coefficients. The method

minimises Equation (6.2.4).

$$\sum(y_i - \sum x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{6.2.4}$$

subjected to constraint

$$\sum_{j=1}^{p} \mid \beta_j \mid \leq t, \tag{6.2.5}$$

where λ is the turning parameter, y_i are the observed values of the dependent variable of GHI, t is a parameter that determines the quantity of regularisation, x_{ij} is the covariance matrix of the independent weather variables, and β_j is the column vector of regression coefficients.

ElasticNet

ElasticNet, a shrinkage regularisation technique, was also explored for variable selection. Elastic net is a technique that produces a hybrid penalty by combining penalties from ridge regression(L_2) and Lasso regression(L_1) to reduce the number of variables in a regression model. ElasticNet minimises the function in Equation (7.3.1). .

$$min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}(y_i - \beta_0 - x_i^T\beta)^2 + \lambda\left(\frac{1-\alpha}{2}||\beta||_2^2 + \alpha||\beta||_1\right)\right), \tag{6.2.6}$$

where N is the number of observations, y_i is the dependent variable, GHI. β_0 is the intercept, β is the vector of coefficients, and x is the vector of the independent variables, weather variables. Alpha is the mixing parameter between lasso $\alpha = 1$ and ridge when $\alpha = 0$ and λ is the regularisation penalty parameter.

The five methods were explored for variable selection, and the best method was found based on the accuracy measure MAE.

After variable selection, two models were explored, GPR and Bayesian structural time series (BSTS).

6.2.5 Gaussian Process Regression(GPR)

Gaussian Process regression was used in this study. The model describes a time-based variable X with a function $f(x)$ that does not relate to any specific distribution. A GPR is built over the assumption that it comprises a collection of continuous random variables $X_1, X_2, ..., X_n$, with n observations of the joint random variables. The components of a GPR are a combination of a covariance function, conditional joint distribution and multivariate normally distributed variables. A mean and covariance function describes the model. The mean function $m(x)$ is assumed to be equal to zero. A GPR is given in Equation (6.2.7). It is described over a real domain of X over a function $f(X)$.

$$f(X) \sim GP(m(x), k(x, x^1)), \tag{6.2.7}$$

where X is a subset of independent variables $X_1, X_2, ..., X_n$, that is meteorological variables, namely relative humidity, temperature, wind speed, and air pressure. $f(X) = f(X_1), f(X_2, ..., f(X_n)$, x and x^1 are any input points(locations) of any one of the variable X. $f(X)$ is a function of the dependent variable, GHI. $m(x)$ is the mean vector of μ, that is, $E(f(x)) = m(x)$, $k(x, x^1)$ is the covariance matrix, also called the kernel function.

Parameter estimation

The Bayesian approach for parameter estimation was used. The Bayesian approach is more attractive than the frequentist maximum likelihood estimation (MLE) approach as it combines prior information with data, enables robust analysis with small samples, and makes the estimation robust.

6.2.6 Bayesian structural time series

Bayesian structural time series (BSTS) is one of the methods used in this research to predict global horizontal irradiance (GHI). A BSTS model is given in Equation 6.2.8.

$$y_t = Z_t^T \alpha_t + \epsilon_t, \tag{6.2.8}$$

where y_t are the observed values of GHI, Z_t is the output vector, ϵ_t is the error term and α is the state vector. The transition model in Equation 6.2.9 gives the latent state that changes over time.

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \tag{6.2.9}$$

where η_t is the q by q dimension error term which is assumed that $\eta_t \sim N(0, \sigma_\eta^2)$, T_t and R_t are the structural parameters, α_t is the latent variable and η_t is the noise error term. BSTS incorporates both long and short-term forecasting. Both these two methods were explored in this research. The model was fitted using an R package called *bsts*, a tool for fitting Bayesian Structural series functions. BSTS_short was based on short-term forecasting, which ranges from five minutes to one week ahead. In this case one hour horizons were used. BSTS_long stands for BSTS long-term forecasting, which ranges from 6 months to many years ahead. In this case, it was 24 months ahead.

The methods were compared based on the following metrics to choose the best ones: RMSE and MAE. GPR was found to be the best model. In order to improve the prediction accuracy, the forecast combination was done using quantile regression models. The models combined with GPR are given in the following section.

6.3 Forecasts combination

In order to get the superior performance of the model, a forecast combination was performed. It is achieved by combining two sets of forecasts based on two hybrid models

making the forecasts robust. The idea of forecast combination was first used by Granger (1989) and Crane and Crotty (1967). Various authors like Clemen (1989), Timmermann (2006) and Nowotarski et al. (2014), have used a combination of forecasts and the predicted values showed that it improved forecasting accuracy. The advantage of combining forecasts is that it reduces prediction errors and variability of various combination accuracy.

6.3.1 Partially linear additive quantile regression

Partially linear additive quantile regression (PLAQR) is a hybrid model produced from the combination of Generalised Additive and Quantile Regression models. The proposed methodologies in this research were combined with PLAQR to produce a hybrid model that would improve the predictions. The PLAQR is given by equation (6.3.1).

$$y_{th} = \beta_{oh,\tau} + \sum_{j=1}^{p1} s_{jh,\tau}(x_{thj}) + \sum_{j=1}^{p2} \beta_{hj,\tau} Z_{thj} + \varepsilon_{th,\tau}\tau \in (0,1), \tag{6.3.1}$$

where y_{th} is the GHI on day t and hour h, β_{oh} is the intercept, $\beta_{hj,\tau}$ are the model parameters h-hours, $s_{jh,\tau}$ smooth function is the independent variables and z_{thj} is the linear variables. $\varepsilon_{th,\tau}\tau$ is the quantile error. The advantages of applying this method are that it is robust to outliers, the sparsity level can be changed with quantiles, and conditional distribution can be analysed thoroughly, Mpfumali et al. (2019).

6.3.2 Quantile regression neural network

A quantile regression neural network (QRNN) is a hybrid model combining Quantile regression and Neural Networks. It has the advantage of capturing nonlinear patterns in datasets and overdispersion and underdispersion in the dataset. The QRNN model was improved by Zhang et al. (2018) and is given in equation 6.3.2.

$$f(x_t, v, w) = g_2\left(\sum_{j=0}^{m} v_j g_1\left(\sum_{i=0}^{n} w_{ji} x_{it}\right)\right), \tag{6.3.2}$$

where n-number of inputs, m-units of the hidden layer, x_{it} are the predictor weather variables, $g_1(.)$ and $g_2(.)$ are activation functions, w_{ij} and v_j are variable weights of parameters to be estimated.

6.3.3 Linear quantile regression averaging

Linear quantile regression averaging (LQRA) was used for forecasting combinations in this research. It is a linear modelling approach that makes use of prediction intervals. Quantile Regression is a method that uses least squares regression to create a conditional mean of the dependent variable GHI over different values. It was first introduced by Koenker and Bassett (1978) and is widely used for energy forecasting, Liu et al. (2015); Elamin (2018); Uniejewski and Weron (2021). The Linear QRA model is given in Equation (6.3.3)

$$Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + ... + \beta_p(\tau)x_{ip}, \tag{6.3.3}$$

where p is the number of independent variables, which are the weather variables, τ is the τ^{th} quantile, y_i is GHI, and x_{ij} are the weather variables.

The best Linear QRA model is given by Equation (6.3.4)

$$\min(\rho_\tau(Y_i - (\beta_0(\tau) + \beta_1(\tau)x_{i1} + ... + \beta_p(\tau)x_{ip})), \tag{6.3.4}$$

where $\rho_\tau(z)$ is a loss function. QRA is when forecast combinations are done on prediction intervals given on point predictions done on individual models. The combined forecasts produced by LQRA are given in Equation (6.3.5).

$$\hat{y}_{iq} = \beta_0 + \sum_{k=1}^{k} \beta_k \hat{y}_{ij} + \epsilon_{iq}, \tag{6.3.5}$$

where β_0 is the intercept, β_k are the coefficients of the independent variables, $\hat{y}_{iq}$ is the combined forecasts, $\hat{y}_{ij}$ is the forecast for the j_th technique and ϵ_{iq} are the residuals. It is used to explore the conditional distribution of the variable of interest GHI incorporating explanatory variables, which are the weather variables.

6.3.4 Online prediction by ExpRt aggregation

Lastly, online prediction by ExpRt aggregation (OPERA) is one of the packages used for combining forecasts Gaillard et al. (2021). It is a method that is based on regression-oriented time series forecasts. Given a set of observed values for a particular variable Y, GHI, with a sequence of values $y_1, y_2, ..., y_n$, 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 (6.3.6).

$$\hat{y}_t = s \sum_{k=1}^{K} P_{k,t} x_{k,t} \tag{6.3.6}$$

The method uses various online learning literature algorithms that combine experts' forecasts based on historical performance. It gives three important features, mixture, prediction and oracle. The mixture is used to build the algorithm's object, predict is used to make predictions using the algorithm, and Oracle is for performance evaluation of the experts and comparing the algorithm.

6.3.5 Forecast evaluation metrics

The evaluation was first done for the methods explored for variable selection. Mean absolute error(MAE), a deterministic approach to forecast evaluations, was used for the evaluation. The methods evaluated are Lasso, Elasticnet, MARS, Boruta and GBR. Secondly, three methods were compared: GPR, BSTS_short and BSTS_long, the deterministic methods used for evaluation are MAE, and Root mean square error (RMSE).

Deterministic forecast evaluation metrics

Generally speaking, deterministic metrics measure differences, the forecast accuracy of deterministic estimates is evaluated by measuring the discrepancy between the forecast and

actual values through several criteria. The following deterministic metrics were used:

Mean absolute error

Mean absolute error (MAE) measures the average of all magnitude of errors. It is defined in Equation (6.3.7).

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_i - \hat{y}_j| \tag{6.3.7}$$

The lower the value is, the more accurate the model is.

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 (6.3.8).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_j)^2}, \tag{6.3.8}$$

where y_i-are the observed values, $\hat{y}_j$- are the actual values and n represents the number of observations. The lower the value of RMSE, the better the predictions. The performance of the combined models was measured using the following probability evaluation metrics.

6.3.6 Probabilistic evaluation metrics

Probabilistic evaluation tools were also used in this research to evaluate probabilistic forecasts. Probabilistic metrics are used to assess the reliability and sharpness of predictive distributions. The tools used in this research are: proper scoring rules, pinball losses, and Murphy diagrams were used to evaluate the appropriateness of modelling techniques. The Quantile regression models that were combined with GPR were the ones that were evaluated using the probability evaluation metrics.

Pinball Loss

The pinball loss (PL) is a measure that assesses the accuracy of a quantile forecast. It is given by Equation (6.3.9).

$$PL(q\tau, t) = \begin{cases} 2(1-\tau)|y_t - q_{\tau,t}| & if \quad y_t < q_{\tau,t} \\ 2\tau|y_t - q_{\tau,t} & if \quad y_t \geq q_{\tau,t} \end{cases}, \tag{6.3.9}$$

where y is the observed values, q-quantile forecasts and τ-is quantile are the mean and variance respectively. More accurate predictions produce a lower pinball loss.

Proper scoring rules

Proper scoring rules were used in this research to evaluate the effectiveness of probability models, and it is given in Equation (6.3.10).

$$S(P, Q) = E_{Y \sim P}(S(F, Q)), \tag{6.3.10}$$

where F is the probabilistic forecast, X is a random variable, Q is the probabilistic distribution. The forecast is minimised in the expectation by the true probability for a proper score. Logarithmic score(LogS), Dawid Sebastian Score(DSS), and Continuous ranked Probability score(CRPS) are the scoring rules used in this research. The scoring rules in this study were used to evaluate the forecasting performance of a model. The R package *scoringRules* was used to develop the scoring scores (Gneiting and Raftery, 2007). A summary of the models used is given in the sections that follow.

Logarithmic score Logarithmic scoring (LogS) is a scoring rule used to measure a probability function's effectiveness in prediction models. Let Y be a random variable over a sample of size n assuming values $x_1, x_2, ..., x_n$. Suppose we are given the probability values as $p_1, p_2, ..., p_n$ belonging to a forecasting distribution F with $P_i \in [0, 1]$ and $\sum P_i = 1$. The logarithmic scoring is given in Equation (6.3.11)

$$S(F, y_{obs}) = logP(y_{obs}), \tag{6.3.11}$$

where y_{obs} are the observed values.

Continuous rank probability score Continuous rank probability score(CRPS) is given in Equation (6.3.12).

$$S(F, y_{obs}) = \int_{-\infty}^{\infty} (F(y) - 1\{y \geq y_{obs}\})^2 dy, \tag{6.3.12}$$

where $F(y)$ cumulative function of the predictions.

Dawid-Sebastian score Dawid-Sebastian score(DSS) is a measure for evaluating the accuracy of multivariate forecasts. The score is given in Equation (6.3.13).

$$DSS(F, y) = \left(\frac{y - \mu_F}{\sigma_F}\right)^2 + 2log\sigma_F, \tag{6.3.13}$$

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

Murphy diagrams

Murphy diagrams were used to compare probabilistic predictions. The technique was proposed by Ehm et al. (2016). In order to evaluate probability forecasts, we plot a function of θ given in Equation (6.3.14).

$$S(\theta) = \frac{1}{n} \sum_{i=1}^{n} S_\theta(y_i, y_i^*), \tag{6.3.14}$$

where y_i is the point estimate and y_i^* is the actual GHI observation and $S_\theta(y_i, y_i^*)$ is a function given in Equation (6.3.15).

$$S_\theta(y_i, y_i^*) = \phi(y_i^*) - \phi(y_i) - \phi^1(y_i)(y_i^* - y_i), \tag{6.3.15}$$

where ϕ is convex, ϕ^1 is the gradient and θ is the cost loss ratio. A plot of the function $S(\theta)$ is compared against the predictions to check if it fits perfectly. Murphy diagrams are graphical checks that check whether one forecast method dominates another under a

scoring function. They check for significant deviations of one model from another method's performance. The Murphy diagrams show the Empirical scores and different scores. For empirical scores, the model mentioned first dominates the other; for the difference scores, negative scores indicate that the first named model is preferable. Difference Scores show the 95% confidence intervals for the difference between the function. The empirical score is a plot that shows the empirical score and θ the cost-loss ratio.

6.4 Results

6.4.1 Data and data preprocessing

This research made use of hourly data from SAURAN (Southern African Universities Radiometric Network) website `https://sauran.ac.za/`). The data is from a radiometric station from the University of Pretoria, South Africa, with a Latitude:-25.75308° (E), a Longitude: of 28.22859 ° (S), and an Elevation: 1410m. The dataset used is an hourly dataset with 8791 observations.

The independent variable is Global horizontal irradiance (GHI), and the explanatory variables with their specific abbreviations and measuring units are given in Table 6.1.

Table 6.1: 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

Missing data imputation was done using the MICE (Multiple Imputation by Chained Equations) packages. The imputation method used is predictive mean matching. Various statistical R packages were used in this research. The packages include *quantreg* for quantile regression, *qrnn* for quantile regression neural networks. The other packages used are *devtools* and *gefcom2017* for pinball losses. For murphy diagrams *murphydiagram* was used, *devtools* for proper scoring rules and *kernlab* were also used for GPR modelling. The package *bsts* was used for BSTS modelling, *gbm* for Gradient boosting regression, *Boruta* for Boruta, *eart* for MARS and *gmlnet* for lasso and elasticnet.

6.4.2 Exploratory data analysis

Figure 6.2 is a time series plot of the Global horizontal irradiance. The observed values are black, while the results of fitting a penalised cubic regression spline to the data are red.

Figure 6.3 shows density plots of all the variables used in this research.

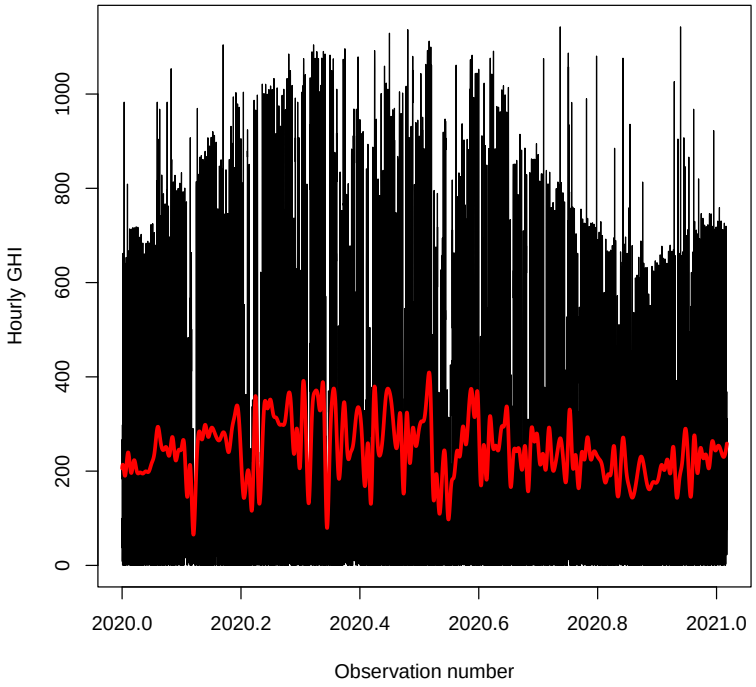


Figure 6.2: GHI Smoothing spline.

Figure 6.4 shows the descriptive analysis of GHI from the UPR dataset. Plot (a) is the time series plot of the data, which shows the distribution of the data. Plots (b) and (c) are the density and histogram, respectively, indicating that the data is skewed to the right, with the coefficient of skewness being -4.842 and Kurtosis being 23.544. Plot (d) is a normal Q-Q plot that shows that most of the quantile points line on a straight line. Plot (e) is a boxplot.

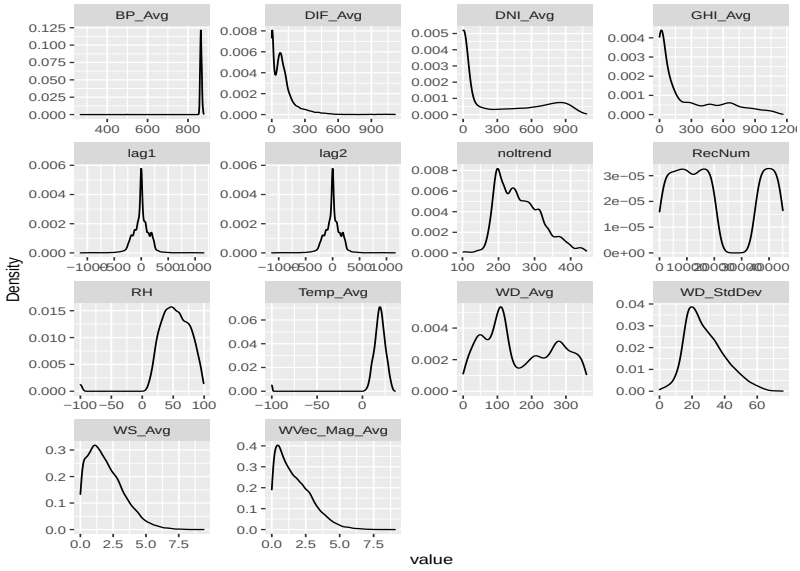


Figure 6.3: Density plots.

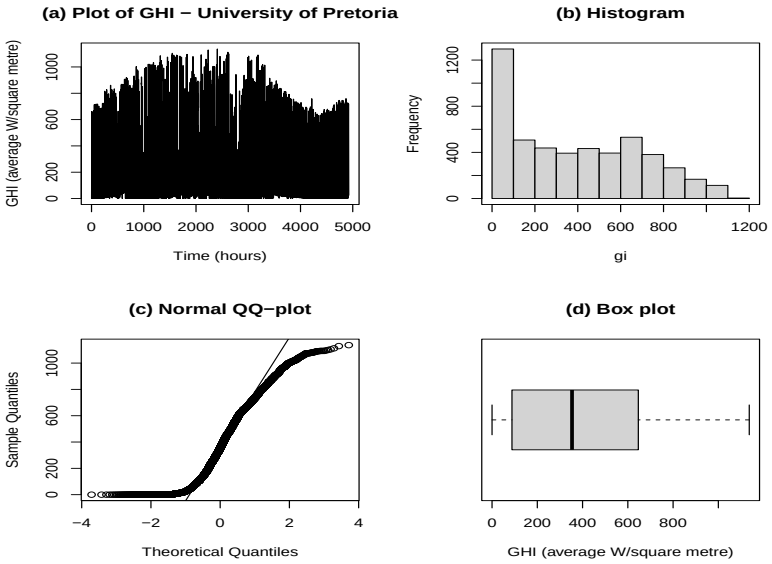


Figure 6.4: Descriptive analysis for UPR.

6.4.3 Variable selection

Variable selection is one of the most important challenges in statistical analysis that needs to be addressed. We considered five modern data 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 a lower MAE than other methods.

A Training dataset was used for variable selection. Then Linear Quantile Regression was applied to the different methods. The evaluation results from the linear quantile models concerning the variable selection method are shown in Table 6.2.

Table 6.2: Variable Selection.

Method	MAE
Lasso	100.48
Elastic net	105.43
MARS	111.54
BORUTA	100.55
GBR	100.36

The method that gave us the most accurate results was GBM (Gradient Boosting Method), which selected 15 variables in the analysis. Figure 6.5 gives the observed GHI (black) values versus predicted values of the GPR/GBM model (red). The GPR model appears to be a good fit. The testing dataset was used to check for the method's accuracy for variable selection.

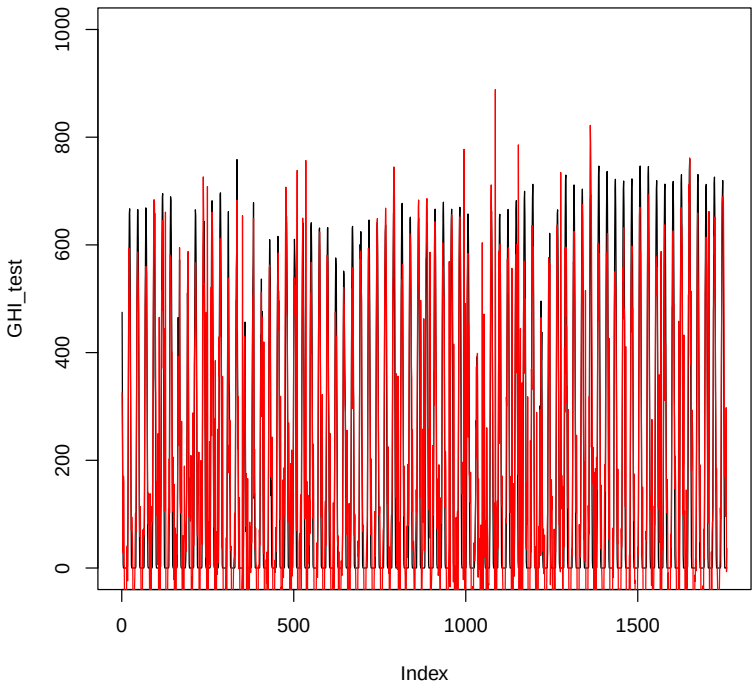


Figure 6.5: Linear Quantile predicted results based on GBM.

6.4.4 Forecasting results

A Gaussian process regression analysis was done using the variables produced from GBM.

A comparative analysis of the GPR model was done with that of BSTS short and long prediction models, and the comparison was based on MAE and RSME as the evaluation metrics. The results of the comparison are shown in Table 6.3. The GPR model proved more efficient than the BSTS models since it had low values of RMSE and MAE. The next section is on a combination of forecasts, but we decided to drop the BSTS_short model because, from Table 6.3, the model had error values higher than the GPR.

Table 6.3: Comparison of three models.

Model	RMSE	MAE
GPR	73.10	43.19
BSTS_short	3010.72	2533.07
BSTS_long	260.02	199.10

Combining forecasts

As a way of improving the accuracy of forecasts, forecasts were combined. Bates and Granger (1969) found that combining forecasts enhances the accuracy of forecasts. We used forecasts from the following models: linear Quantile Regression, Quantile Regression neural network combined, Partial Linear Additive Quantile Regression, and OPERA. 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 6.4 shows the results for forecasting the accuracy of the four methods based on two evaluation metrics.

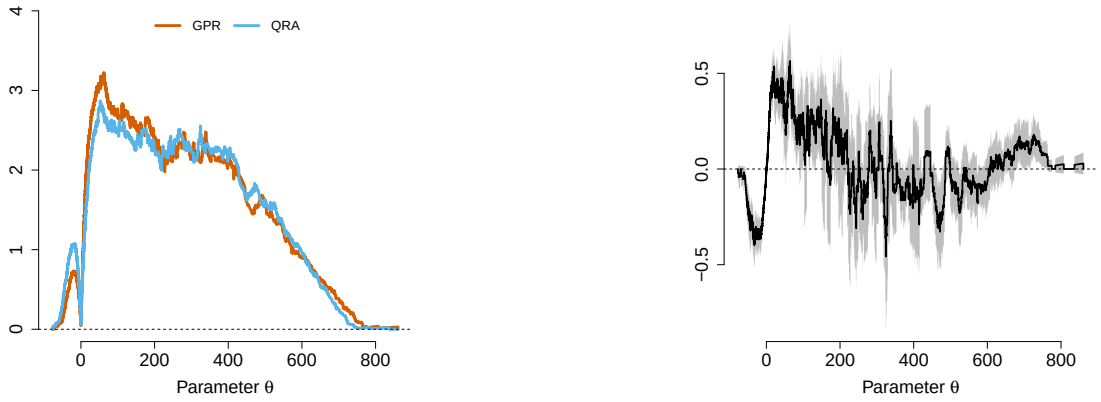
Table 6.4: Combined forecasts

Model	RMSE	MAE
QRA(GPR+bstslong)	73.18	42.13
QRNN(GPR+bstslong)	70.53	30.53
PLAQR(GPR+bstslong)	72.04	34.08
OPERA(GPR+bstslong)	73.34	42.89

Table 6.4 shows that the QRNN model has the most accurate forecasts based on evaluation metrics RMSE and MAE.

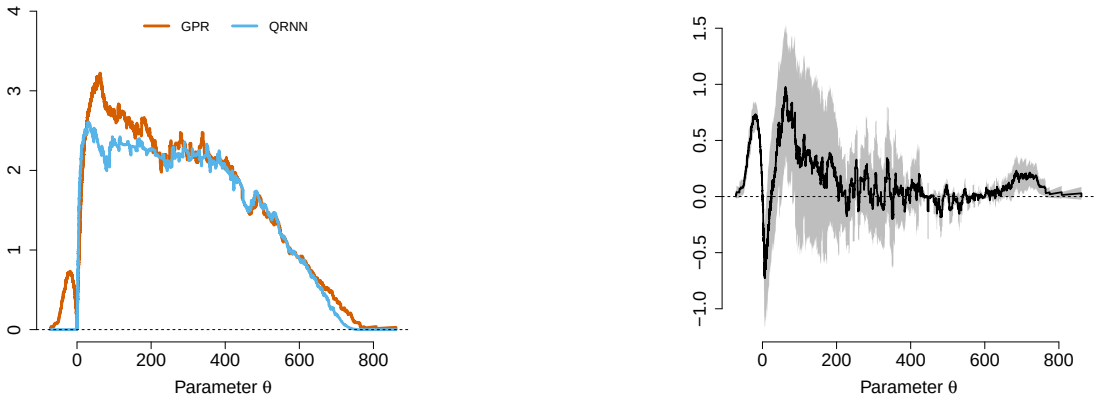
6.4.5 Murphy diagrams

The forecasting models were also evaluated using proper scoring rules, and Murphy diagrams were used to compare forecasts. Figures 6.6, 6.7, 6.8 and 6.9 show plots showing empirical scores and differences in scores, and the GPR model proves to be superior to other models. There is not much comparison that we can do using these Murphy diagrams because all these forecasts reflect that there is not much difference with the compared GPR model. The Opera model is very good though the GPR model is a better forecaster than Opera.



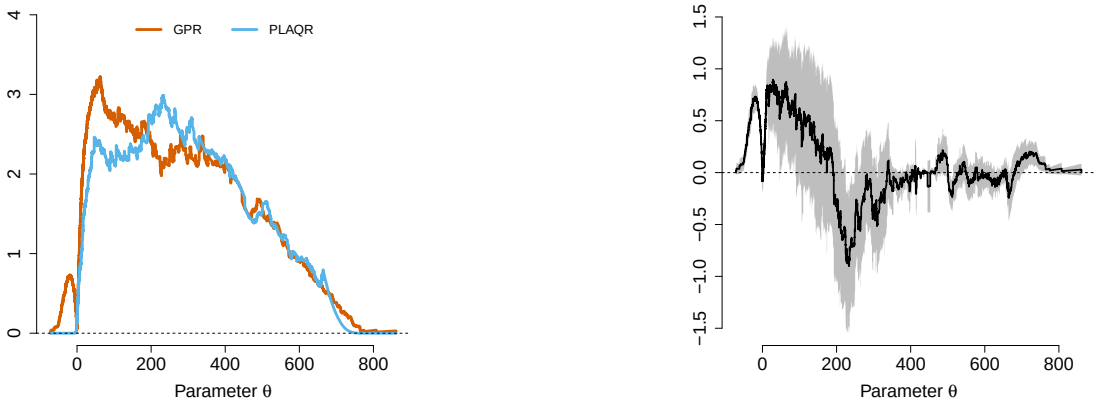
(a) Empirical scores. (b) Difference in scores.

Figure 6.6: Murphy diagram for QRA vs GPR.



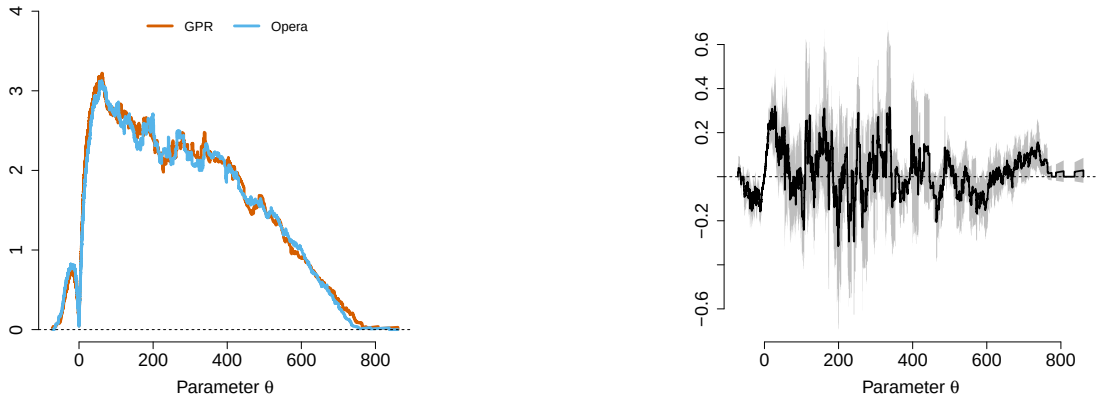
(a) Empirical Scores (b) Difference in scores

Figure 6.7: Murphy diagram for QRNN vs GPR.



(a) Empirical Scores (b) Difference in scores

Figure 6.8: Murphy diagram for PLAQR vs GPR.



(a) Empirical Scores

(b) Difference in scores

Figure 6.9: Murphy diagram for OPERA vs GPR.

Plots in Figure 6.10 show a comparison between models(in red) and GHI(in black). The plots show that models GPR, PLAQR, and QRNN are better forecasters.

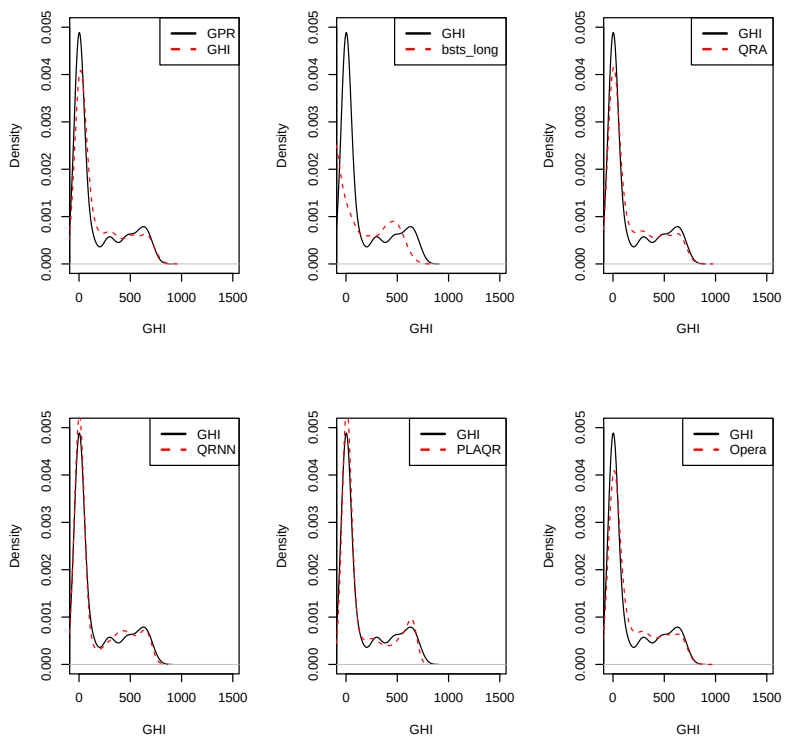


Figure 6.10: Density plots.

6.4.6 Evaluating models using scoring rules

A further comparison of the model 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 6.5 for the models GPR, bstslong,

QRA, QRNN, PLAQR, and Opera using probabilistic evaluations. The GPR model was the best-fitting model based on the probabilistic evaluation metrics CRPS, LogS, DSS, and PL.

Table 6.5: Probabilistic model evaluations.

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

6.5 Discussion

This study was based on predicting Global horizontal irradiance at UPR radiometric station in South Africa using hourly intervals data. The dataset used was from July 2020 to August 2021. GPR was the best model based on RMSE and MAE, compared to bstslong. Further probabilistic evaluations used proper scoring rules, pinball losses, and Murphy diagrams. To improve the forecasting accuracy, the models' forecasts were combined. Although the PLAQR and QRNN forecasts appeared superior to the other combined forecasts, the GPR individual forecasts outperformed them. The results of the GPR model's accuracy proved consistent based on RMSE, MAE, and probability measures of accuracy. The values of RMSE and MAE and the probabilistic evaluation metrics CRPS, LogS, DSS, and PL were lower for GPR than other models, as seen in Table 6.5.

6.6 Conclusion

The QRA, PLAQR, Opera, BSTS, and GPR models discussed in this study were based on predicting GHI. Variables were selected using the GBR method, and a comparative analysis

was done based on QRA, PLAQR, Opera, BSTS, and GPR models. Quantile regression was then used to combine the forecasts of the GHI.

Based on the weather-independent variables, we recommend using the GPR to predict GHI in Southern Africa because it proved to be a robust model that performed better than the combined models. The methods used in this study could be used by companies such as Eskom to control and manage the power grid, and the results will promote economic development and sustainability of energy resources. The first two studies were based on single-site analysis. We move on to a multiple-analysis approach. The next chapter will develop the Gaussian Process regression method by combining it with spatial analysis. This was done to incorporate spatial dependencies between radiometric stations.

Chapter 7

Spatio-temporal forecasting of global horizontal irradiance using Bayesian inference

7.1 Introduction

The current study presents an in-depth analysis and spatio-temporal predictive modelling of monthly GHI at radiometric stations in sparse geographical regions. To the best of our knowledge, this is the first study to be carried out using South African data. The study also proposes standardising predictions by normalising them, which improves forecast accuracy. Another highlight is in the selection of the validation sets. Eight radiometric stations were used in the study, with three locations in each validation set. This resulted in fifty-six possible validation sets. An in-depth analysis was conducted from the randomly selected five sets to investigate how the proposed models would perform, especially those stations far apart in Euclidean space.

The flow chart of the structure of the study outlining the proposed models and evaluation metrics is shown in Figure 7.1. Two proposed models, the spatio-temporal Gaussian (GP) regression model and GP coupled with autoregressive (GP-AR) error terms, are com-

pared with a benchmark model, the spatio-temporal linear regression model. The performance of the models is evaluated using the root mean square error (RMSE), mean absolute error (MAE), and skill score (SS).

Figure 7.1: Flow chart of the structure of the study.

The highlights and contributions made in this study are summarised as follows. The research focused on moving from a single site to a multi-site forecasting approach, that is,

using geographically sparse data. Multi-site research is expected to give a diverse or large sample that is good enough to come up with significant associations between meteorological stations; hence, it enhances the statistical power of the model. Previously, solar forecasting has been based on single-site analysis. Exploring the spatio-temporal forecasting approach is expected to improve modelling accuracy.

This study uses Bayesian inference to predict GHI using spatial regression coupled with Gaussian process modelling. Though computationally expensive, GP regression has proved to be a powerful tool for modelling the variability and uncertainty of solar irradiation (Chandiwana et al., 2021). The combination of these two methods thus produces a hybrid model that produces favourable results.

A comparative analysis was done between the proposed models, GP Spatial and GP-AR Spatial, and the linear spatial model was used as a benchmark model. A Linear spatial regression model is a spatial regression model with a linear specification. The evaluation metrics used to assess the accuracy of the models are MAE (Mean Absolute Error), RMSE (Root Mean Square Error), CP (Coverage Probability) and CRPS (Continuous Rank Probability Score).

Standardisation of the forecasts made from the selected model was also done in this research. The standardised forecasts were then added to the original forecasts, improving the forecast accuracy. We used ElasticNet, one of the shrinkage methods used in variable selection. ElasticNet is a supervised algorithm that identifies the variables strongly associated with the response variable. The nonlinear rend variable is another important variable that resulted in a significant improvement in the forecast accuracy. Finally, we combined forecasts using QGAM (Quantile Generalized Additive Model) and MCQRNN (Monotone Composite Quantile Regression Neural Networks). Combining forecasts is a very effective tool for reducing errors.

7.2 Methodology

7.3 Variable selection

The variables used for analysis in this research were chosen using the shrinkage variable methods discussed in the following sections.

7.3.1 ElasticNet

In order to select significant variables for the analysis, Elastic net, a shrinkage regularisation technique, was used. Elastic net minimises the function in equation (7.3.1).

$$min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+\lambda\left(\frac{1-\alpha}{2}||\beta||_2^2+\alpha||\beta||_1\right)\right), \tag{7.3.1}$$

where N is the number of observations, y_i is the dependent variable, GHI. β_0 is the intercept, β is the vector of coefficients, and x is the vector of the independent variables, weather variables. Alpha is the mixing parameter between Lasso $\alpha = 1$ and ridge when $\alpha = 0$ and λ is the regularisation penalty parameter. The R package called glmnet was used for performing ElasticNet. Glmnet is a package that fits generalised linear and similar models via penalised maximum likelihood. The regularisation path is computed for the Lasso or elastic net penalty at a grid of values for the regularisation parameter lambda. ElasticNet is a widely used regularisation method and is a logical pairing with GLMs. It removes unimportant and highly correlated features, hurting accuracy and inference. The main purpose is to remove the influence of predictors that are not useful. Firstly, unimportant coefficients are removed by shrinking its coefficient to zero. This is called LASSO (or $L1$) regularisation. Secondly, the coefficient size of predictors correlated with other predictors is reduced, called Ridge regression ($L2$ regularisation). The realisation of the two is called Elastic Net Regression.

7.4 Gaussian spatio-temporal predictive modelling

Analysis was also done to model spatial dependences between stations. Spatial analysis is the process of performing an analysis that involves modelling spatial features of data like the location and their relatedness to geographical data. In this study, we extended the research in Gaussian Processes to spatial GP analysis. The advantage of using spatial analysis is that since spatial data is selected from multiple sources and consolidated, it improves the accuracy of the models since it captures variations of data from different sources. Two spatio-temporal analysis methods were used: Gaussian Spatial analysis and autoregressive spatial analysis. The results of the two were compared to Linear Spatial analysis.

7.4.1 Spatio-temporal Gaussian process regression model

This study uses the GP-temporal approach to forecasting solar irradiance. Gaussian processes can be presumed as elements of spatial-temporal modelling because these stochastic processes are defined, in this case, over a given region where stations have different locations. A -temporal Gaussian process (GP) model is given in equation (7.4.1).

$$Y(s_i,t) = \boldsymbol{x}^T(s_i,t)\beta + w(s_i,t) + \varepsilon(s_i,t), \tag{7.4.1}$$

where $w(s_i,t)$ is the spatially-dependent error term (spatio-temporal process), $Y(s_i,t)$ represents the response variable (GHI) at station s_i at time t, for $i = 1,2,...,n$ and $t = 1,2,...,T$; with $\boldsymbol{x}(s_i,t)$ representing the covariates, β is a vector of constants and $\varepsilon(s_i,t)$ is the pure random error term. Gaussian processes are characterised by the covariance (kernel) functions. Various kernel functions can be used, but the Matern and squared exponential are commonly used. In this study, the Matern covariance kernel was used because it better balances the roughness and smoothness of the developed model Stein (1999). The family Matérn kernels were first explored by Michael Stein, who took the name based on initial work by statistician Bertil Matérn, Matérn (2013). The Martern kernel was used for spatial

analysis because it was discovered to suit spatial data Matérn (2013); Stein (1999). The temporal process $w(s_i, t)$ is assumed to be a GP with the Matern covariance function given in equation (7.4.2), (Sahu, 2022).

$$C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{\rho}\right)^\nu K_\nu \left(\sqrt{2\nu}\frac{d}{\rho}\right), \tag{7.4.2}$$

where d represents the distance between two points, K_ν is the modified Bessel function of the second kind and ρ and ν are parameters. That is (7.4.3)

$$\text{Cov}\left(w(s_i, t), w(s_j, t)\right) = \sigma_w^2 \rho\left(d_{ij}|\boldsymbol{\nu}\right). \tag{7.4.3}$$

The GP model can be represented hierarchically as shown in equation (7.4.4).

$$\begin{cases} Y_{it} = \boldsymbol{O}_{it} + \varepsilon_{it}, \\ \boldsymbol{O}_{it} = X_{it}\beta + w_{it}, \end{cases} \tag{7.4.4}$$

where $i = 1, ..., n;\ t = 1, ..., T$ assuming $\varepsilon_{it} \sim N(O, \sigma_\varepsilon^2 I)$ and $w_{it} \sim N(O, \sigma_w^2 S_w)$ with S_w representing the sample covariance matrix.

The prior distribution assumes a normal distribution as shown in equation (7.4.5).

$$W_t \sim N(0, \sigma_w^2 S_w). \tag{7.4.5}$$

The posterior distribution is given in equation (7.4.6).

$$\begin{cases} Y(S_0, T+1) = O(s_0, T+1) + \epsilon(s_0, T+1), \\ O(s_0, T+1) = x^1(s_0, T+1)\beta + w(s_0, T+1), \end{cases} \tag{7.4.6}$$

where $Y(s_0, T+1)$ is the 1-step ahead forecast and s_0 is an unobserved location at time $T+1$. The hyperparameters θ are solved by maximising the marginal likelihood function in equation (7.4.7).

$$L(\theta, \sigma_{noise}^2) \quad = \quad -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log K_\theta + \sigma_{noise}^2 I - \frac{1}{2}Y^T(K_\theta + \sigma_{noise}^2 I)^{-1}, \tag{7.4.7}$$

where σ^2_{noise} is the noise variance, θ represents the hyperparameters and K_θ is the covariance function. The posterior predictive distribution of $Y(s_0, T+1)$ given y is given in equation (7.4.8).

$$\phi(y(s_0,T+1)|Y) = \int \pi(y(s_0,T+1)|\theta,\mathbf{O},O(s_0,T+1),y) \times \pi(0(s-0,T+1)|\theta,y)\pi(\theta,O|y)dO((s_0,T+1)d\theta 0 d\theta), \tag{7.4.8}$$

where $\phi(\theta,0|y)$ is the joint posterior distribution of $\mathbf{O}$ and θ.

7.4.2 Gaussian process regression based autoregressive model

Another approach to forecasting solar irradiance used in this research is the autoregressive (AR) Spatial-temporal model. The Gaussian process autoregressive model is given hierarchically as shown in equation (7.4.9) (Sahu et al., 2020).

$$\begin{cases} Y_{it} = \boldsymbol{O}_{it} + \varepsilon_{it} \\ \boldsymbol{O}_{it} = \rho \boldsymbol{O}_{it-1} + X_{it}\beta + w_{it}, \end{cases} \tag{7.4.9}$$

where $i = 1, ..., n;\ t = 1, ..., T$ assuming $\varepsilon_{it} \sim N(O, \sigma^2_\varepsilon I)$ and $w_{it} \sim N(O, \sigma^2_w S_w)$ with S_w representing the sample covariance matrix. A direct autoregression on the response variable Y_{it}, which is $Y_{it} = \rho Y_{it-1} + \varepsilon_{it}$ gives challenges when there are missing values in Y_{it}. Parameters of the spatio-temporal AR model are computed based on prior distributions. Let $\boldsymbol{\theta} = (\beta, \rho, \sigma^2_\varepsilon, \sigma^2_w, v, \mu_0, \sigma^2_0)$ denote all the parameters with $\pi(\theta)$ as the joint prior distribution. Equation (7.4.10) presents the logarithm of the joint posterior distribution of the parameter.

$$\begin{aligned} \log\pi(\boldsymbol{\theta},\boldsymbol{o}|\boldsymbol{y}) \propto -\frac{nT}{n}\log(\sigma^2_\varepsilon) - \frac{1}{2\sigma^2_\varepsilon}\sigma^T_{t=1}(y_t-\theta_t)^T(y_t-\theta_t) \\ -\frac{T}{2}\log|\sigma^2_w S_w| - \frac{1}{2\sigma^2_w}\sigma(o_t-\rho o_{t-1}-\mu_t)^T S_w^{-1}(o_t-\rho o_{t-1}-\mu_t) \\ -\frac{1}{2}\log|\sigma^2_0 S_0| - \frac{1}{2\sigma^2_0}\left(\boldsymbol{O}_0-\boldsymbol{\mu}_0\right)_T S_0^{-1}\left(\boldsymbol{O}_0-\boldsymbol{\mu}_0\right) + \log\pi(\boldsymbol{\theta}) \end{aligned} \tag{7.4.10}$$

The posterior predictive distribution of $Y(\boldsymbol{s}_0, t)$ is given in equation (7.4.11).

$$\pi(y(\boldsymbol{s}_0,t)|\boldsymbol{y}) = \int \pi(y(\boldsymbol{s}_0,t)|o(\boldsymbol{s}_0,[t]),\sigma_\varepsilon^2)\pi(o(\boldsymbol{s}_0,[t])|o(\boldsymbol{s}_0,0),\boldsymbol{\theta},\boldsymbol{o}) \\ \pi(o(\boldsymbol{s}_0,O)|\boldsymbol{\theta},\boldsymbol{O}_0)\pi(\boldsymbol{\theta},\boldsymbol{o}|\boldsymbol{y})do(\boldsymbol{s}_0,[t])dO(\boldsymbol{s}_0,o)d\boldsymbol{\theta}d\boldsymbol{o} \tag{7.4.11}$$

7.4.3 Benchmark model

The linear spatial model

The Spatial Linear model was used as a benchmark model. The linear spatial model is given in Equation (7.4.12).

$$Y(s_i,t) = \beta_1 x_1(s_i,t) + \cdots + \beta_p x_p(s_i,t) + \varepsilon(s_i,t),\ i = 1,\ldots,n, t = 1,\ldots,T, \tag{7.4.12}$$

where $\varepsilon(s_i,t)$ is an error term with a zero-mean spatio-temporal Gaussian process with a covariance structure given in Equation (7.4.13).

$$\text{Cov}\left\{\varepsilon(s_i,t_k),\varepsilon(s_j,t_l)\right\} = \rho_s\left(|s_i - s_j|;v_s\right)\rho_t\left(|t_k - t_l|;v_t\right), \tag{7.4.13}$$

where $\rho_s\left(|s_i - s_j|;v_s\right)$ represents an isotropic correlation function Sahu et al. (2020).

7.4.4 Combining Forecasts

This research combines forecasts with two models, the monotone composite quantile regression neural network (MCQRNN) and the QGAM.

Quantile Regression Averaging

It is a forecast combination approach to the computation of prediction intervals. It involves applying quantile regression to the point forecasts of individual forecasting models. The Quantile regression problem is given by

$$Q_y(q|X_t) = X_t\beta_q \tag{7.4.14}$$

where $Q_y(q|.)$ is the $q-th$ quantile of the dependent variable y_t and $X_t = [1, \hat{y}_{1,t}, ..., \hat{y}_{m,t}]$ is a vector of point forecast of individual models and β_q is a vector of p. The two quantile regression models used are the QGAM and MOQRNN. Let $y_{t,\tau}$ be the global GHI and K be the methods used in the next m observations where m is the total number of forecasts. The combined forecasts are given in Equation 7.4.15 (Gaillard et al., 2021; Sigauke, 2017)

$$z_{t,\tau} = \sum_{k=1}^{K} w_{t,k,\tau} \hat{y}_{(t,k)} + \epsilon_{t,\tau} \tag{7.4.15}$$

$\tau \in (0'1)$, $t = 1, ..., m$, $\hat{y}_{t,k}$ represents forecasts from method K For QGAM, the combined forecasts are given in Equation 7.4.16

$$\hat{y}_{t,\tau}^{QGAM} = \beta_{0,\tau} + \sum_{k=1}^{K} \beta_{k,\tau} \hat{y}_{t,k} + \epsilon_{t,\tau} \tag{7.4.16}$$

$\hat{y}_{t,\tau}^{QGAM}$ represents forecasts from method k and $\epsilon_{t,\tau}$ is the error term. We seek to minimise Equation

$$\arg\max_{\beta} \sum_{t=1}^{m} \rho_\tau (\hat{y}_t^{QGAM} - \beta_0 - \sum_{k=1}^{k} \beta_k \hat{y}_{t,k}) \tag{7.4.17}$$

For MCQRNN, the combined forecasts are given in Equation 7.4.18

$$\hat{y}_{t,\tau}^{MCQRNN} = \beta_{0,\tau} + \sum_{k=1}^{K} \beta_{k,\tau} \hat{y}_{t,k} + \epsilon_{t,\tau} \tag{7.4.18}$$

$\hat{y}_{t,\tau}^{MCQRNN}$ represents forecasts from method k and $\epsilon_{t,\tau}$ is the error term. We seek to minimise Equation

$$\arg\max_{\beta} \sum_{t=1}^{m} \rho_\tau (\hat{y}_t^{MCQRNN} - \beta_0 - \sum_{k=1}^{k} \beta_k \hat{y}_{t,k}) \tag{7.4.19}$$

7.5 Empirical results

7.5.1 Data used in the study

The dataset used in this study was collected from the SAURAN (Southern African Universities Radiometric Network) website `https://sauran.ac.za/`, accessed on 13 August

2022. Eight radiometric stations were selected for the study. These were the only radiometric stations from SAURAN with clean data for the sampling period 1 December 2021-31 January 2022. The covariates used in the study are temperature, relative humidity, wind speed, wind direction, wind direction standard deviation, barometric pressure, month, day, noltrend (nonlinear trend variable), elevation, and the dependent variable GHI. R statistical software was used for all the analyses in this study. The packages used in this study are *MICE*-data imputation, *bmstdr*-spatial analysis, *SpTimer*-Spatial analysis, *glmnet*-variable selection, *forecast*-forecasts evaluation, *qgam*-generalised additive regression and *qrnn*-mcquantile regression neural network.

The modelling and analysis were done using a Dell laptop with the following specifications: 8GB RAM, processor Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz 2.11 GHz, 64-bit operating system, x64-based processor.

7.5.2 Data preprocessing and exploratory data analysis

The locations of the eight radiometric stations used in this study are given in Figure 7.2.

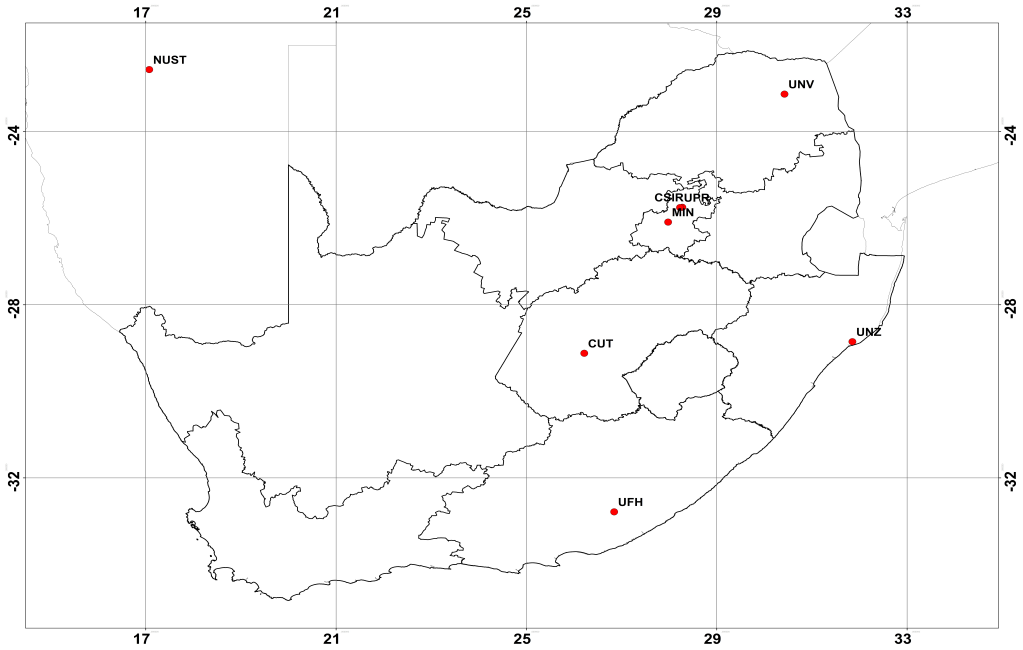


Figure 7.2: Map Showing radiometric stations. Source: Author's creation.

The eight radiometric stations' longitudes, latitudes and elevations are shown in Figure 7.2 and are presented in Table 7.1a. Table 7.1b summarises the distances between radiometric stations. The bottom left of the leading diagonal shows the distance between stations, while the top right shows the pairwise Kendall's rank correlation coefficients of GHI.

Table 7.1: (**a**) Longitude, latitude and elevation. (**b**) Distance (km) matrix for the eight radiometric stations.

(a)

Station	Longitude	Latitude	Elevation (m)
CSIR	28.279	−25.747	1400
CUT	26.216	−29.121	1397
UNZ	31.852	−28.853	90
UNV	30.424	−23.131	628
UFH	26.845	−32.785	540
UPR	28.229	−25.753	1410
MIN	27.978	−26.089	1521
NUST	17.075	−22.565	1683

(b)

	CSIR	CUT	UNZ	UNV	UFH	UPR	MIN	NUST
CSIR	0	0.42	0.34	0.11	0.08	0.9	0.48	0.1
CUT	456	0	0.14	0.04	0.01	0.41	0.3	0.13
UNZ	664	779	0	0.01	−0.03	0.30	0.24	−0.02
UNV	441	898	907	0	−0.08	0.1	0.14	0.01
UFH	986	527	878	1422	0	0.11	0.00	−0.23
UPR	6.9	460	662	440	990	0	0.51	0.06
MIN	568	410	641	493	940	54.7	0	0.06
NUST	1651	1520	2296	2087	1796	1649	1611	0

The left panel in Figure 7.3 shows the variogram cloud for GHI based on eight stations. The right panel shows the empirical variogram, the plot of averaged points of the clouds' X and Y coordinates.

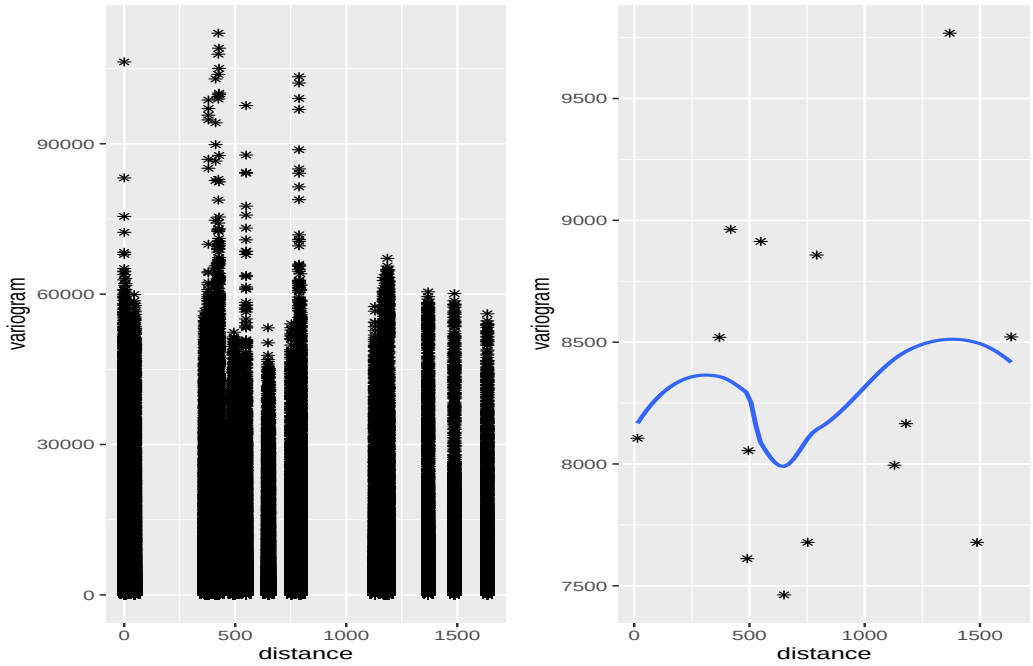


Figure 7.3: Variogram cloud (**left** panel) and an empirical variogram (**right** panel) for the GHI at the eight radiometric stations. The distance is measured in kilometres.

Daily GHI temporal variations of all eight radiometric stations over the sampling period, December 2021 to January 2022, are given in Figure 7.4. GHI dropped slightly in January compared to December for all the radiometric stations. This is possibly due to the onset of the rainy season.

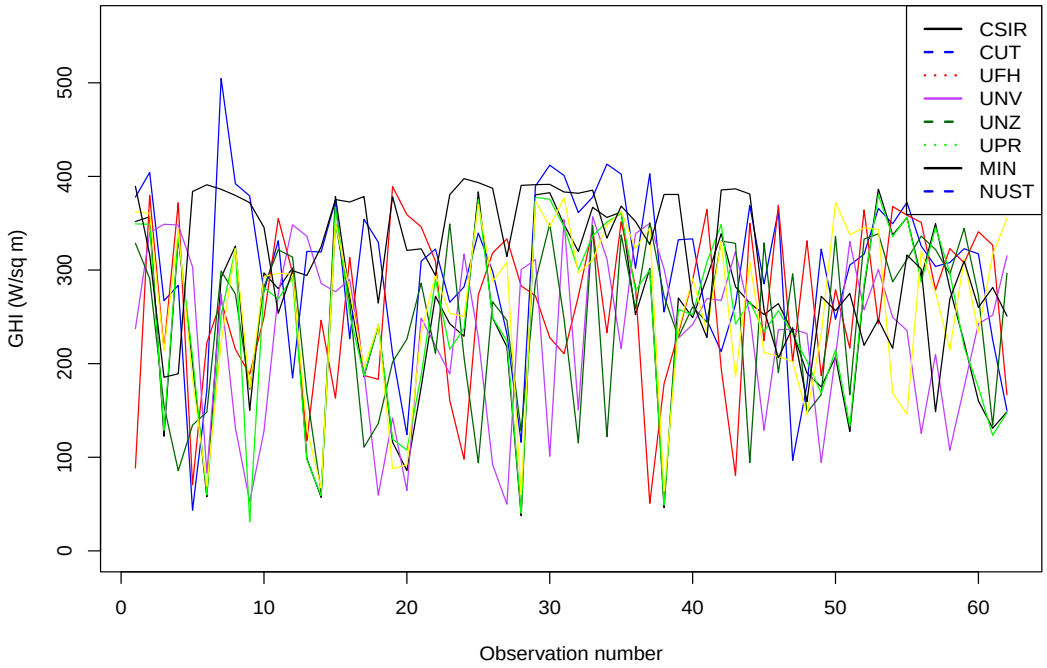


Figure 7.4: Time series plot of the GHI data at the eight radiometric stations.

In order to get some insight into the distribution of the GHI data, the box plots of daily GHI data based on eight radiometric stations are presented in Figure 7.5. Most stations have GHI values ranging between 20 and 400 W/m^2. Only CUT and NUST have measures ranging from 100 to 400 and 500 W/m^2. The spread of the distributions is almost similar except for CUT and NUST, whose spread is slightly different.

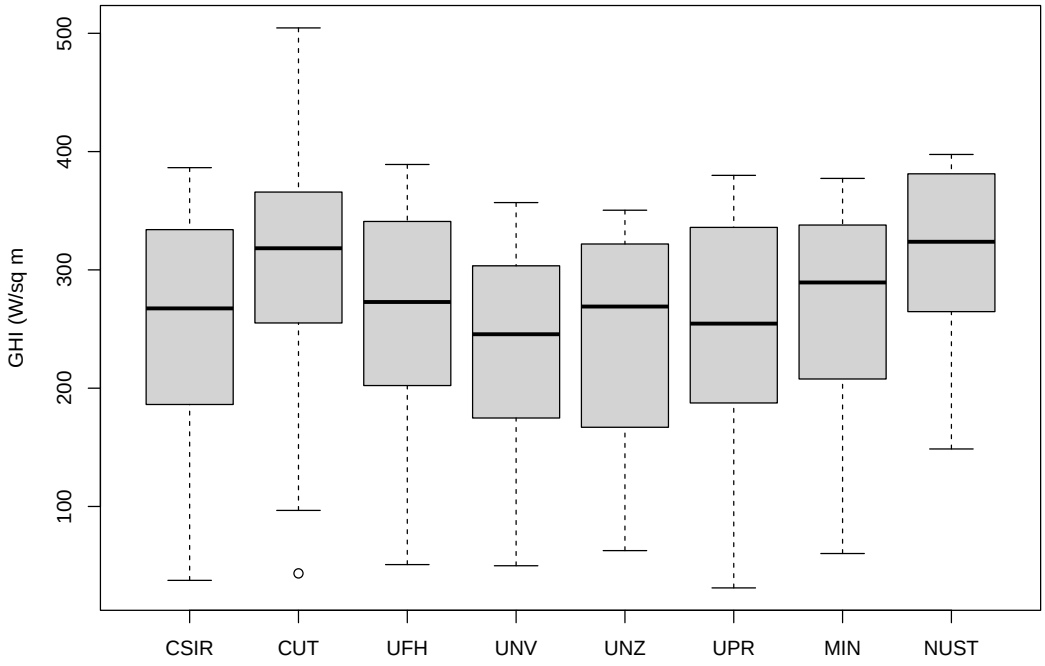


Figure 7.5: Box plots of GHI at the eight radiometric stations.

Summary statistics of the GHI data are shown in Table 7.2. The measures of skewness are negative, indicating the data are skewed to the left, and kurtosis is less than three for all the stations except CUT, whose kurtosis is greater than three, indicating that its distribution is heavy-tailed.

Table 7.2: Summary statistics of GHI for the eight stations.

Station	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Skewness	Kurtosis
CSIR	37.53	187.16	267.44	250.35	331.92	386.41	−0.5382635	2.395768
CUT	43.45	257.17	318.27	301.15	365.21	504.55	−0.719209	3.5455
UFH	50.86	204.34	272.86	260.89	339.02	389.10	−0.551794	2.503086
UNV	49.89	178.34	245.59	233.37	302.89	356.90	−0.5387249	2.228135
UNZ	62.69	167.36	269.01	245.12	321.84	350.45	−0.4954172	1.928348
UPR	31.16	191.20	254.57	246.36	332.57	379.90	−0.5952533	2.499935
MIN	60.22	208.79	289.37	259.85	336.06	377.33	−0.7116837	2.564314
NUST	148.6	264.9	323.8	317.6	381.1	397.6	−0.6146715	2.306814

Figure 7.6 displays the Kendall tau's coefficients, pairwise scatter plots, and histograms to show the association between the meteorological stations. A measurement of Kendall's τ was calculated using equation (7.5.1).

$$\tau = 1 - \frac{2 \times I}{0.5 \times n \times (n-1)} \tag{7.5.1}$$

Measure of $\tau < 0.07$ indicates weak association, $0.07 < \tau < 0.21$-medium association and $0.21 < \tau < 0.35$ fairly strong association and $\tau > 0.35$ strong association. Most of the associations between stations reflected that most of these stations have strong associations. NUST has a fairly strong negative association with UFH, while CUT and NUST have a positive medium association.

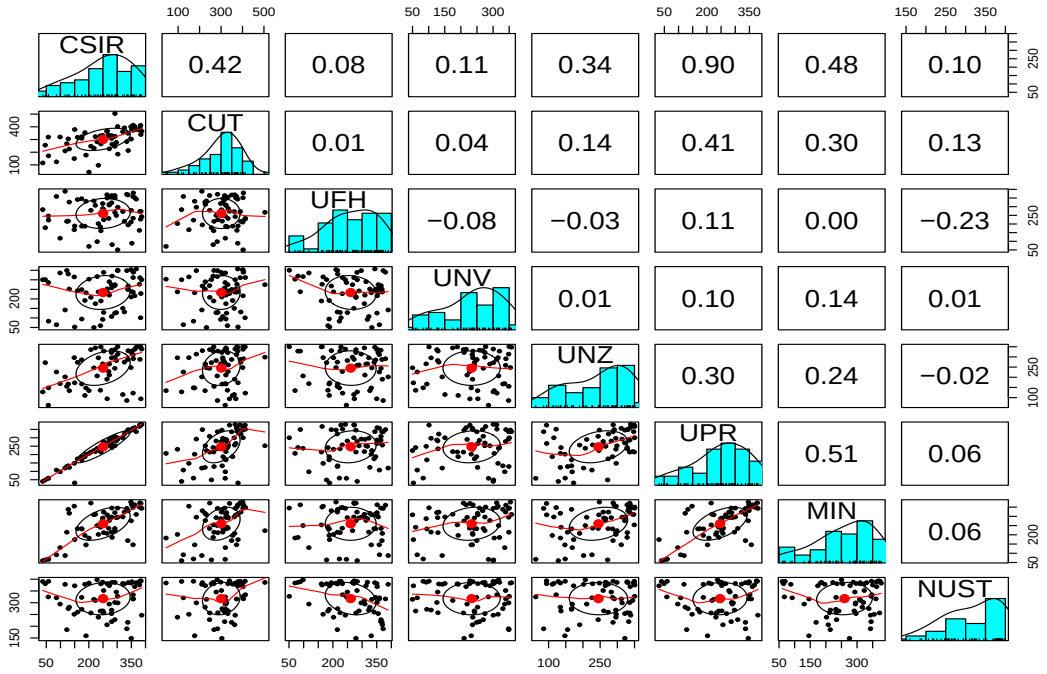


Figure 7.6: Histograms (diagonal), pairwise scatter plots (**bottom** left), and pairwise Kendall's rank correlation coefficients (**top** right) of GHI for all the eight radiometric stations.

7.5.3 Variable selection

The study used ElasticNet, one of the shrinkage methods used in variable selection. ElasticNet is a supervised learning algorithm that identifies the variables strongly associated with the dependent variable. The nonlinear rend variable is another important variable that resulted in a significant improvement in the forecast accuracy.

Figure 7.7 shows the results for variable selection using the R package 'glmnet' Hastie et al. (2021). The significance of the variables was checked using the shrinkage method, elastic net. As shown in Figure 7.7, all twelve variables were retained in the model. Figure 7.8 shows the relative importance of the independent selected variables. Relative humidity and temperature are the most significant, while the day is the least important covariate.

7.5.4 Forecasting results

The models used in this study are the linear spatial model (M1), which is the benchmark model, GP spatial regression (M2) and GP spatial autoregressive (M3). Eight radiometric stations were selected, and two months of data were used for each station. Table 7.3 shows the model evaluations of the forecasting accuracy based on RMSE, MAE, CRPS, and CP. Five stations were used as training sets, and three were used as test sets. With eight stations, the possible number of validation sets is given as $\binom{n}{k} = \frac{n!}{(n-k)!k!}$, where n is the total number of locations and k is the number of locations in a validation set. In this study, five validation sets were choosen at random, each having three locations, were used. They were chosen at random.

The running times for the three models were as follows: M1 (linear model) took 3.39 s, M2 (GP model) took 2.27 s, and M3 (GP-AR model) took 2.89 s. The results show that, overall, the spatio-temporal Gaussian regression model (M2) produces the least measures of error based on the MAE and RMSE values, proving to be the best model. Based on the CRPS, the linear spatio-temporal model (M1) is the best-fitting model. However, regarding coverage probability, M1 has the least values on all five validation sets. Overall, M2 is the best-fitting model based on all four evaluation metrics.

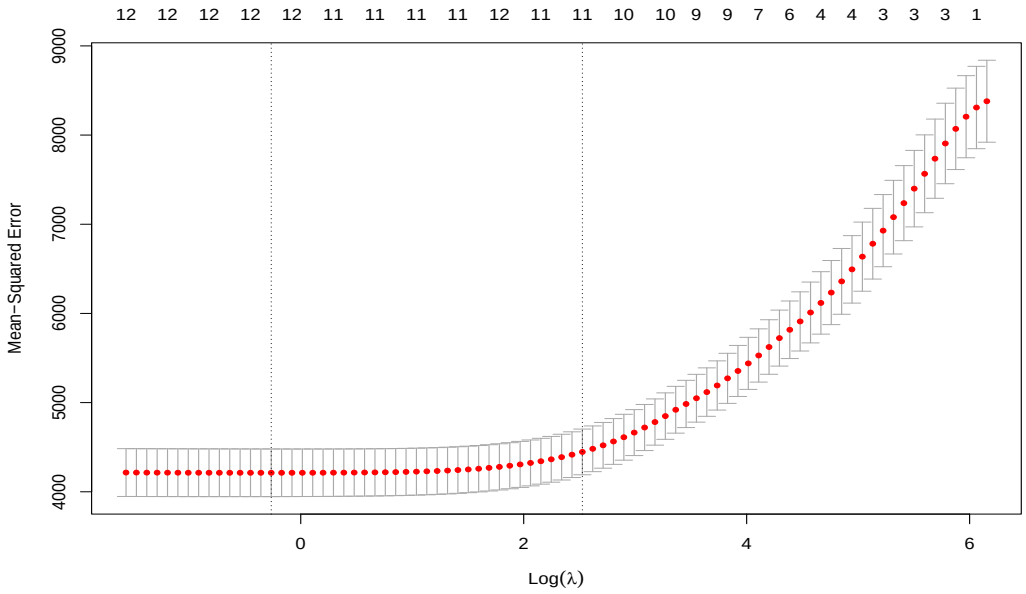


Figure 7.7: Variables selected using the elastic net.

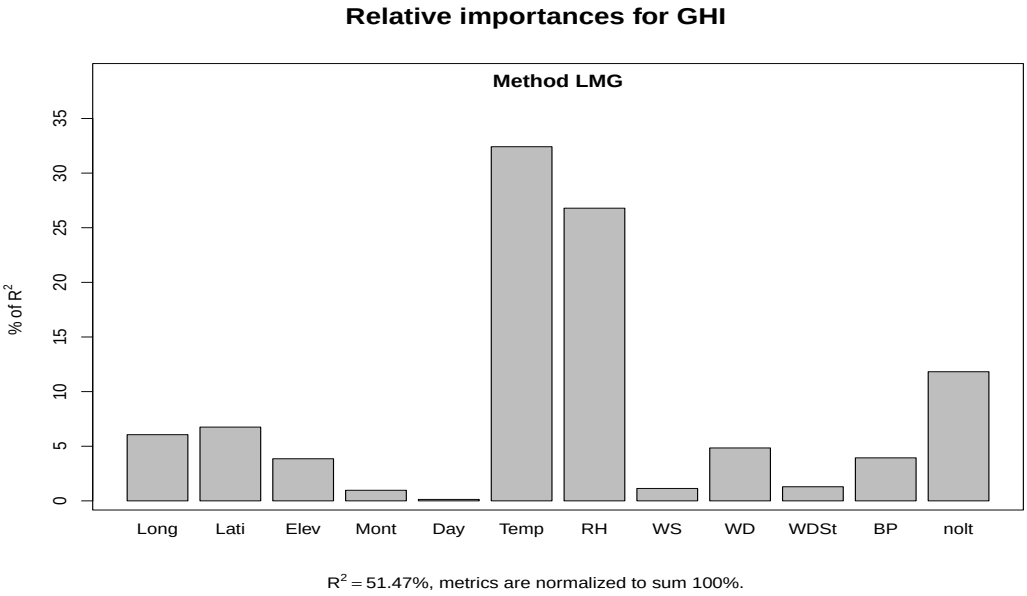


Figure 7.8: Relative importance for GHI.

Table 7.3: Overall model evaluation for the three models.

Station	RMSE	MAE	CRPS	CP
Validation set 1 (CSIR, UNZ, NUST)				
M1	162.497	113.547	**46.211**	74.731
M2	**61.893**	**49.113**	68.769	**100**
M3	64.105	50.308	65.013	**100**
Validation set 2 (CUT, UNV, UPR)				
M1	74.585	60.518	**41.410**	91.935
M2	**67.675**	**53.149**	59.323	**99.462**
M3	68.156	53.302	58.813	**99.462**
Validation set 3 (UFH, UNV, MIN)				
M1	82.741	68.085	**37.566**	90.323
M2	**81.297**	63.439	54.311	96.774
M3	83.740	**63.437**	63.496	**98.925**
Validation set 4 (UPR, MIN, NUST)				
M1	152.961	107.354	**45.989**	76.344
M2	**63.905**	53.028	67.072	**100**
M3	64.748	**53.007**	62.307	**100**
Validation set 5 (UNZ, UPR, MIN)				
M1	74.785	62.499	**40.558**	94.624
M2	**70.185**	**55.993**	69.137	**100**
M3	73.341	57.245	59.016	98.924

We also tested for the statistical significance of the models using the Diebold–Mariano and Giacomini–White tests to check the predictive accuracies of the developed models. The results are given in Table 7.4. The results show that Model M2 (Gaussian Process Spatial) performed better than all the other models.

Table 7.4: Diebold-Mariano Test

H_0	**Test Statistic**	***p*-Value**	**Result**
M2 = M3	−6.841	1.116×10^{-010}	Not equally accurate
M2 = M1	6.936	6.546×10^{-011}	Not equally accurate
M3 = M1	1.913	0.0574	Not equally accurate
Giacommini-White Test			
H_0	**Test Statistic**	***p*-Value**	**Result**
M2 = M3	3.606	0.058	Sign of mean loss is (-). M2 dominates M3
M2 = M1	38.387	5.802×10^{-010}	Sign of mean loss is (-). M2 dominates M1
M3 = M1	37.550	$8.909 \times 10^{-0.010}$	Sign of mean loss is (-). M3 dominates M1

Figure 7.9 displays a plot of the forecasts in red and the actual GHI in black. The plots are quite close to each other. We used validation set 2 (CUT, UNV, UPR).

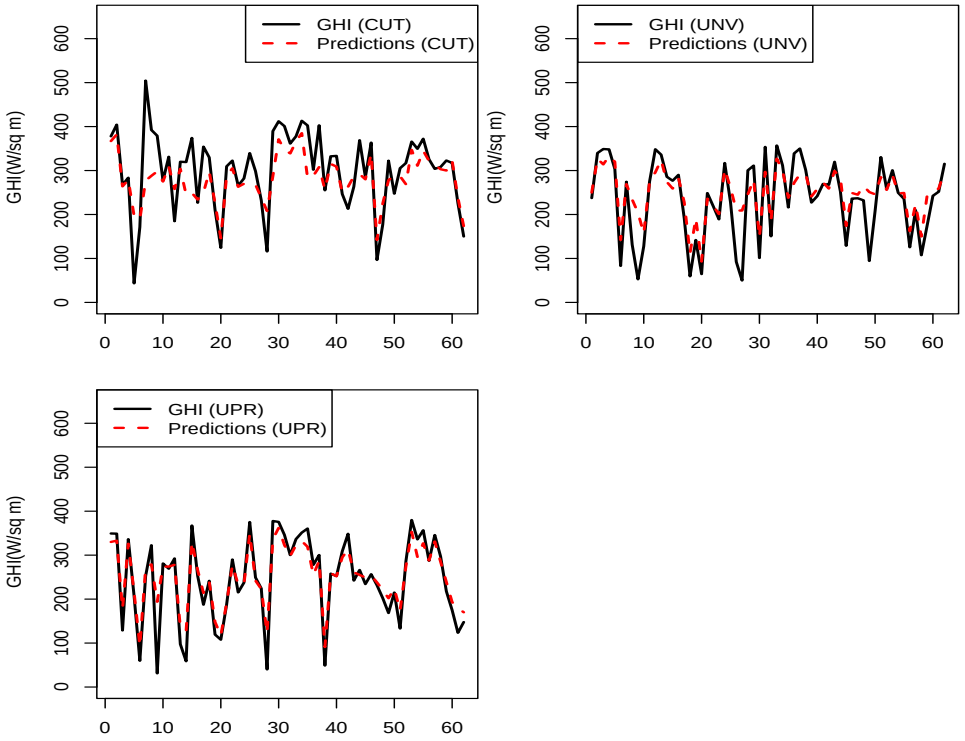


Figure 7.9: Predictions at CUT, UNV, and UPR radiometric stations.

Kendall's rank correlation coefficients were used to measure the association between the prediction errors shown in Figure 7.10. It shows pairwise scatter plots, the Kendall tau coefficients and the histograms. The computations of the Kendall tau coefficients show that errors for stations CUT, UNV and UPR are close to zero except between CUT and UPR, which is 0.17.

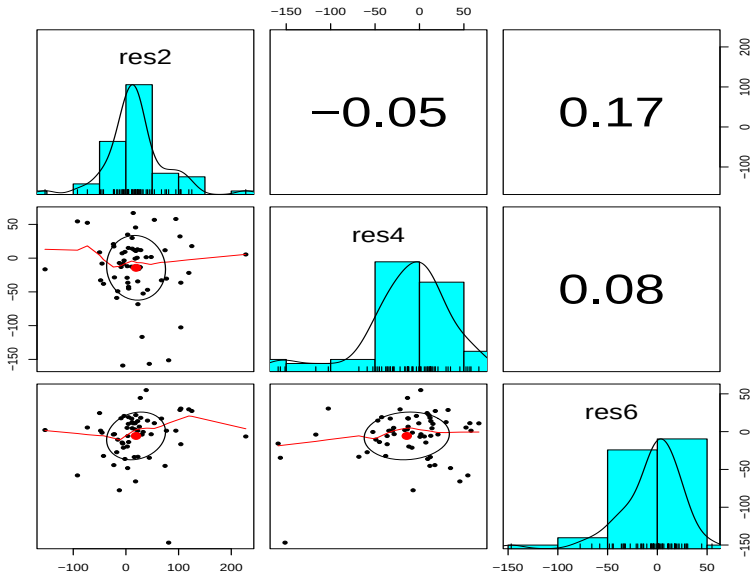


Figure 7.10: Histograms (diagonal), pairwise scatter plots (bottom left), and pairwise Kendall's rank correlation coefficients (top right) of the prediction errors for stations 2, 4 and 6, which are CUT, UNV and UPR.

Standardising forecasts

To evaluate the forecasts' accuracy, we standardised them by normalisation and then evaluated if their accuracy had changed. Standardisation creates consistency in the dataset by avoiding wider ranges of the predicted values from the actual values, though the decreases are small. For example, RMSE is extremely small (around 1%).

Table 7.5 compares the RMSE, MAE, Skill Score, DM, and GW scores, of the predicted values versus the standardised forecasts for the validation set UNV, UPR and MIN. The models with standardised predictions achieved better forecasting accuracy, as shown by the positive skill scores. The skill scores show that the predictions improved after standardisation.

Table 7.5: Standardised based models evaluation.

Model	MAE	RMSE	Skill Score
predicted4 (UNV)	34.365	49.682	
predicted4.adjusted	33.738	49.191	0.988
predicted6 (UPR)	21.435	31.851	
predicted6.adjusted	20.766	31.217	1.991
predicted7 (MIN)	39.962	56.482	
predicted7.adjusted	39.714	56.544	0.110

Combining forecasts

Combining forecasts takes advantage of diverse information, compensates for errors, reduces variability, and enhances predictive insights. This leads to improved accuracy and the ability to make more extreme predictions that reflect underlying trends or events better. However, it is important to note that forecast combination methods vary, and not all combinations result in improved accuracy or more extreme predictions. The success of forecast combinations depends on the quality and diversity of individual forecasts and the appropriateness of the combination method.

Whilst the forecast combination using MCQRNN was applied to all the forecasts from the five validation sets, we present the results from validation set 5 (UNV, UPR, MIN) here.

Figure 7.11 shows plots of combined forecasts with GHI.

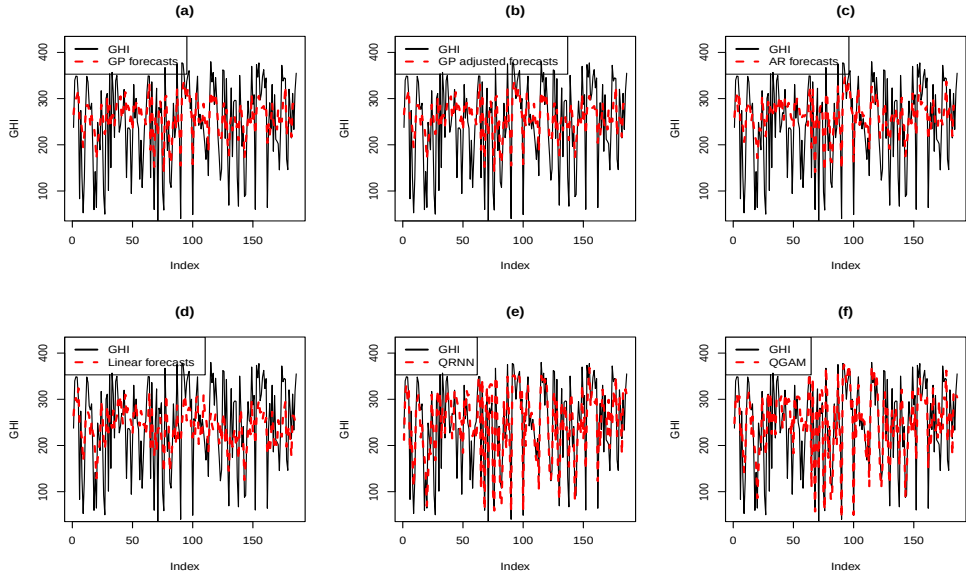


Figure 7.11: Plots of GHI with forecasts from (**a**) GP, (**b**) GP adjusted (**c**), AR (**d**) Linear models and combined forecasts using (**e**) MCQRNN and (**f**) QGAM.

Table 7.6 shows forecast evaluations of all the models based on the validation set 5, models GP and AR, Linear and Standardised GP. MCQRNN represents the forecasts that were combined using the QMCRNN model. It is noted from Table 7.6 that adjusting the forecasts through standardisation and combining forecasts significantly improves the forecast accuracy.

Table 7.6: Accuracy evaluations (Validation set (467)).

Model	RMSE	MAE
GP	67.01	52.99
Standardised GP	66.53	52.52
AR	67.70	52.70
Linear	74.53	61.49
MCQRNN(GPR+Linear+AR)	51.53	37.97
QGAM(GPR+Linear+AR)	52.39	38.85

To assess the forecasting results, we use two performance evaluation criteria (RMSE and MAE) with three forecasting horizons, 10-, 15-, and 20-day ahead forecasts. The results are presented in Table 7.7. Using the linear model, the smallest RMSEs for the 10-, 15-, and 20-day ahead forecasts were 62.53, 73.78 and 69.15, respectively. Based on the GP model, the smallest RMSEs were 50.80, 50.32 and 46.10, respectively. Similarly, using the GP-AR model, the RMSEs are 52.96, 51.09 and 45.79, respectively. Based on this validation set and using the RMSE, the results suggest that the GP model is the best for the 10- and 15-day ahead forecasts, while the GP-AR produces good forecasts for the 20-day ahead forecast horizon. It should be noted that this conclusion is based on one validation set. However, this confirms that the spatio-temporal GP regression model produces the most accurate forecasts as summarised in Tables 7.3, 7.4 and 7.6, respectively.

Table 7.7: Accuracy evaluations for three forecast horizons using the validation set UNV, UPR and MIN.

Linear Model						
Forecast Horizon	**RMSE**			**MAE**		
	UNV	**UPR**	**MIN**	**UNV**	**UPR**	**MIN**
10-day	72.85	80.79	62.53	54.26	71.78	57.85
15-day	73.86	74.98	73.78	53.59	66.70	64 60
20-day	72.95	69.15	70.87	54.95	60.44	59.94
GP Model						
Forecast Horizon	**RMSE**			**MAE**		
	UNV	**UPR**	**MIN**	**UNV**	**UPR**	**MIN**
10-day	75.59	50.80	53.58	58.76	44.44	44.81
15-day	77.98	50.32	63.78	57.59	45.01	54.12
20-day	77.31	46.10	62.27	59.20	39.59	54.06
GP-AR Model						
Forecast Horizon	**RMSE**			**MAE**		
	UNV	**UPR**	**MIN**	**UNV**	**UPR**	**MIN**
10-day	77.00	52.96	58.17	59.66	45.88	45.11
15-day	76.67	51.09	68.20	58.26	44.97	55.44
20-day	76.98	45.79	67.03	59.25	38.31	56.51

7.6 Conclusion

This study investigated the spatio-temporal forecasting of global horizontal irradiance data from eight radiometric stations. The data are from SAURAN (`https://sauran.ac.za/`), accessed on 13 August 2022. This was done through a comparative analysis of two spatio-temporal predictive models, the Gaussian process regression and the autoregressive Gaus-

sian process regression. Variable selection was done using ElaticNet, a hybrid shrinkage method combining Lasso and ridge regressions. The results were compared to the benchmark model, the spatio-temporal linear model. The spatio-temporal GP outperformed the other two models, proving to be the most appropriate model for predicting GHI in South Africa. Most authors, see, for example, Chandiwana et al. (2021) and Sigauke et al. (2022) applied models based on single-site data sets for predicting solar power irradiation. Accurate forecasts of solar power from multi-sites are important to the system operator, as they facilitate large-scale integration of solar power onto the power grid. This modelling approach exploits production information of solar power from neighbouring radiometric stations.

Energy forecasting has many limitations, including coming up with accurate forecasts. In order to improve the forecasts, standardisation was done, and it proved that the models' forecasting accuracy improved. The forecasts from the individual models were combined using quantile regression neural networks and additive quantile regression models, significantly improving the forecast accuracy. As a result, there was a significant improvement in forecast accuracy. This modelling framework could be useful to power utility companies in making informed decisions when planning power grid management.

This study has several possible future research directions. The modelling done in the current study involved stations in geographically sparse areas. Future research could use more stations and cluster them before the spatio-temporal forecasting of solar power. An in-depth analysis of the extremal dependence of GHI in each cluster could help develop models with high predictive capabilities. Another interesting research could be using large data sets from stations at different periods and comparing them with those from gridded data sets.

7.A Appendix for this Chapter

Figure 7.12 shows the solar resource map for South Africa (SA). It shows the solar radiation each area (region) in SA receives. The largest amount of GHI is in the western part of South Africa.

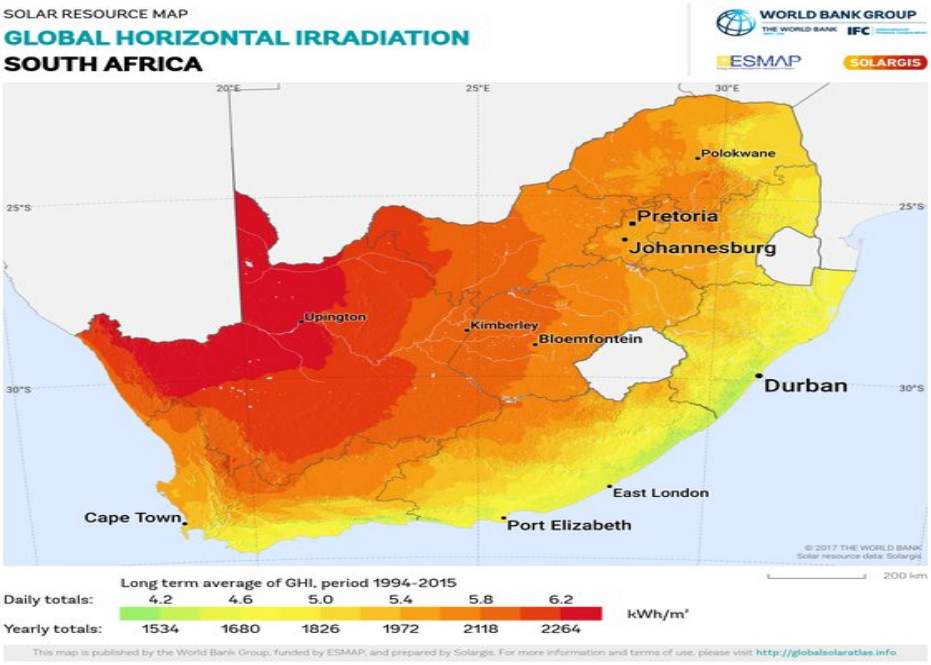


Figure 7.12: Solar heat map. Source: `https://solargis.com/maps-and-gis-data/download/south-africa`, accessed on 8 July 2022.

Chapter 8

Discussion of key findings and concluding remarks

8.1 Introduction

This study used various statistical tools to predict solar power using GPR. This Chapter summarises the research findings and recommendations based on the results produced. The results of this study were presented in Chapter 5, 6, and 7. Summaries of the chapters are given in the section that follows.

8.2 Summary of research findings

Chapter 5 looks at applying Gaussian process regression coupled with core vector regression. Core vector regression was used to solve the problem of finding an appropriate way of finding the appropriate covariance function for GPR. A core vector regression technique called the Minimum Enclosed Ball was used. The analysis was based on two radiometric stations, Stellenbosch University and the University of Venda. Variables were selected using Lasso via hierarchical interactions; pairwise interactions were used. This study used four kernel functions: Matern, rational quadratic, dot product and radial basis. The best kernel function was selected based on the minimum enclosed ball. The kernel function selected was

the rational quadratic used in GPR. Many other statistical tools were explored. Interaction effects were investigated as well. The results showed that interaction effects did not improve forecasting accuracy. A comparison analysis was done between models with and without interact effects and two benchmark models. The results showed that the GPR model without interactions produced superior results. The Chapter's results showed that the application of CVR improved the accuracy of the GPR model.

Chapter 6 looks at the application of GPR coupled with quantile regression analysis. Most researches were based on point estimation when performing GPR forecasting, so the use of quantile regression was introduced in order to broaden our prediction values. Various statistical methods were applied to make the methodology unique and superior. The first thing done was the selection of significant variables. Most researchers use one particular method for variable selection. For this research, various techniques were used, and one method which produced superior results was used for variable selection. The methods explored are Elasticnet, Mars, Boruta and Gradient Boosting Regression. Linear Quantile regression was then applied to these 5 to check for a method that performed better than the other models. The GPR was the best, so the variables selected were used for regression analysis. UPR dataset was then used to predict solar power based on GPR coupled with Quantile regression. BSTS was used as a benchmark model. Forecasts combination was done using the following quantile regression methods, partially linear additive quantile regression, QRNN and Linear Quantile regression. A comparison analysis was done based on the combined forecasts and the ordinary GPR results. Various methods were used to evaluate these techniques: Murphy diagrams and scoring rules. The results showed that the performance of the ordinary GPR outperformed the other models.

Chapter 7 looked at the Gaussian process of regression coupled with spatial-temporal regression. Most research has been done on single-site analysis, but this study shifted to multi-site forecasting. Multi-site analysis was explored to incorporate the dependencies be-

tween different weather stations. Variables were selected using elasticnet. The dataset used in this study was obtained from eight radiometric stations from different geographical areas. The training set was based on five stations, especially those far apart. The other three were set aside as the test set. Spatial analysis was done using the GP regression, Autoregressive, and Linear Spatial models. A comparison between the predictive models and the Spatial GPR model proved to be the best. In order to improve the models' prediction accuracy, standardisation was done. The results showed that the accuracy of the models improved.

8.3 Modelling discussions and summary of key findings

8.3.1 Solar power forecasting via GPR

Various ideas were applied in this research to predict solar power. GPR predictions were compared with several other models' prediction accuracy. The Benchmark models have been explored and used before to predict solar power. The benchmark models used are SVM, GBR, BSTS and Spatial Linear. These comparisons proved that GPR produced superior results and was effective and efficient.

8.3.2 Solar power forecasting via GPR coupled with core vector regression

Particular attention was given to kernel function selection when CVR was used. Covariance functions are important components of GPR, which must be selected appropriately. The effectiveness of the kernel selection method used was even checked by comparing the results of the ordinary GPR and the GPR coupled with the core vector. The results showed that the cooperation of CVR improved model accuracy, and the application of MEB was superior.

8.3.3 Solar power forecasting via GPR based on quantile regression

Quantile regression was coupled with GPR to improve the robustness of the model. It was used to explore investigations from point estimation to interval estimation. This was done to come up with more improved predicted values of GHI. The results of this research showed that quantile regression improved the forecasting accuracy of the models.

8.3.4 Solar power forecasting via GPR based on spatio-temporal regression

GPR was coupled with spatial analysis to produce a model capturing spatial dependencies. The distance between weather stations might cause differences in the values recorded for GHI. There was a need to include different weather stations to capture dependencies. This improved the robustness of the model since we shifted from just selecting a dataset from one station and ignoring the rest. Using spatio-temporal GPR proved very effective based on the results it produced. There were superior.

8.4 Future research studies

Other possible areas can be explored in the future. Highlights of some of them follow:

Future research can consider other statistical tools to find an appropriate kernel function for GPR modelling. These include methods such as combining kernel functions to produce a hybrid kernel. Another possible area that can be explored is the area of Spatial Gaussian process modelling. The Spatial GPR was made robust, but other methods of investigating the dependencies can be explored. Archimedean copulas can be explored. This method can be used to find the relationship that exists between stations.

8.5 Conclusion

The study results provide answers to three questions stated in Chapter 1. The research questions and how we responded to them follow. The first question is, Does the Core vector regression provide a better way of selecting a kernel function for short-term Global Horizontal Irradiance forecasting? To address this question, various approaches were considered. The results proved that Corevector regression provided a better way of selecting kernels. The second question is, Does Quantile regression and GPR combination improve GHI predictions? The combination of GPR and quantile regression produced better results. This can be seen from the results produced in Chapter 6 and 7, the combined forecasts reduced the error produced. The third question is, Does capturing spatial dependencies improve GHI predictions? This question was addressed in Chapter 7. The results showed that applying GPR and spatial regression produced results that reduced prediction errors. The major predictors of solar power are Air Temperature, Relative Humidity, Wind Speed, Wind Direction and Barometer Pressure.

Having addressed the aim and objectives and answered the research questions. The results of this study have predicted solar power using GPR, which proved to be more accurate than the other models used. Hence the results can be used to advise policymakers to use them to develop future solar power predictions. Policymakers can implement measures promoting the use of GHI-predicted values. This will benefit the nation in areas of public health by reducing carbon emissions in the air and reducing associated costs. Various economic costs and the problem of electricity shortage are also reduced.

Appendix A1: Publications

Research articles (title pages and abstracts only) from this thesis are given in the next three pages.

1. Chandiwana, E., **Sigauke, C.** and Bere, A. (2021). Twenty-four-hour ahead Probabilistic Global Horizontal Irradiance Forecasting Using Gaussian Process Regression, Algorithms, vol. 14, no. 177. https://doi.org/10.3390/a14060177

2. Chandiwana, E., **Sigauke, C.** and Bere, A. (2021). Robust Modelling Framework for Short-Term Forecasting of Global Horizontal Irradiance. Journal of Statistics Applications & Probability, Vol. 12, No. 3 (Sep. 2023), PP:1203-1221. The paper is now available online on https://doi.org/10.18576/jsap/120327

3. **Sigauke, C.**, Chandiwana, E. and Bere, A. (2023). Spatio-temporal forecasting of global horizontal irradiance using Bayesian inference, Applied Sciences, vol. 13, 201, pp. 1–24. https://doi.org//10.3390/app13010201

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

Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120327

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

Received: 2 Feb. 2023, Revised: 22 Apr. 2023, Accepted: 28 Apr. 2023
Published online: 1 Sep. 2023

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]

* Corresponding author e-mail: caston.sigauke@univen.ac.za

applied sciences

MDPI

Article

Spatio-Temporal Forecasting of Global Horizontal Irradiance Using Bayesian Inference

Caston Sigauke †, **Edina Chandiwana** *,† and **Alphonce Bere**

Department of Mathematical and Computational Sciences, University of Venda, Private Bag X5050, Thohoyandou 0950, South Africa
* Correspondence: 18021412@mvula.univen.ac.za; Tel.: +27-15-962-8135
† These authors contributed equally to this work.

Abstract: Accurate global horizontal irradiance (GHI) forecasting promotes power grid stability. Most of the research on solar irradiance forecasting has been based on a single-site analysis. It is crucial to explore multisite modeling to capture variations in weather conditions between various sites, thereby producing a more robust model. In this research, we propose the use of spatial regression coupled with Gaussian Process Regression (GP Spatial) and the GP Autoregressive Spatial model (GP-AR Spatial) for the prediction of GHI using data from seven radiometric stations from South Africa and one from Namibia. The results of the proposed methods were compared with a benchmark model, the Linear Spatial Temporal Regression (LSTR) model. Five validation sets each comprised of three stations were chosen. For each validation set, the remaining five stations were used for training. Based on root mean square error, the GP model gave the most accurate forecasts across the validation sets. These results were confirmed by the statistical significance tests using the Giacommini–White test. In terms of coverage probability, there was a 100% coverage on three validation sets and the other two had 97% and 99%. The GP model dominated the other two models. One of the study's contributions is using standardized forecasts and including a nonlinear trend covariate, which improved the accuracy of the forecasts. The forecasts were combined using a monotone composite quantile regression neural network and a quantile generalized additive model. This modeling framework could be useful to power utility companies in making informed decisions when planning power grid management, including large-scale solar power integration onto the power grid.

Keywords: autoregressive; Gaussian process; global horizontal irradiance; spatial analysis

Citation: Sigauke, C.; Chandiwana, E.; Bere, A. Spatio-Temporal Forecasting of Global Horizontal Irradiance Using Bayesian Inference. *Appl. Sci.* **2023**, *13*, 201. https://doi.org/10.3390/app13010201

Academic Editor: Constantinos A. Balaras

Received: 18 November 2022
Revised: 13 December 2022
Accepted: 16 December 2022
Published: 23 December 2022

1. Introduction

1.1. Overview

The use of energy sources that are clean and renewable has been on the rise, hence the need to efficiently manage the power grid. This increase has led to the need to come up with predictions of the available energy sources to ease the management of the power grid. Power grid planning is a challenging process, as it requires efficient and accurate power predictions inputs. In this research, we focused on probabilistic, short-term forecasting to efficiently model GHI. Probability forecasting is an emerging research area that handles the uncertainty of weather variables, which produces more comprehensive results [1]. This probability approach was combined with spatial analysis to assess the relatedness of meteorological stations. Spatial analysis is a very important aspect in regression, and in this study it was used to explore spatial dependence between various meteorological stations.

1.2. Survey of Related Literature

Some authors have applied various methods incorporating spatial analysis. Andre et al. [2] used a spatial–temporal model for short-term solar irradiation forecasting. They used a spatial–temporal vector autoregressive model, which was designed in such a way that

References

1. Abdessalem, A. B., Dervilis, N., Wagg, D. and Worden, K. (2018), 'Model selection and parameter estimation in structural dynamics using approximate bayesian computation', *Mechanical Systems and Signal Processing* **99**, 306–325.

2. Adedeji, P. A., Akinlabi, S., Madushele, N. and Olatunji, O. (2022), 'Hybrid adaptive neuro-fuzzy inference system (anfis) for a multi-campus university energy consumption forecast', *International Journal of Ambient Energy* **43**(1), 1685–1694.

3. Agoua, X., Girard, R. and Kariniotakis, G. (2018), 'Probabilistic models for spatio-temporal photovoltaic power forecasting', *IEEE Transactions on Sustainable Energy* **10**(2), 780–789.

4. Al-Shedivat, M., Wilson, A. G., Saatchi, Y., Hu, Z. and Xing, E. (2017), 'Learning scalable deep kernels with recurrent structure', *The Journal of Machine Learning Research* **18**(1), 2850–2886.

5. Amarasinghe, P., Abeygunawardana, N., Jayasekara, T., Edirisinghe, E. and Abeyunawardane, S. (2020), 'Ensemble models for solar power forecasting—a weather classification approach', *AIMS Energy* **8**(2), 252–271.

6. André, M., Dabo-Niang, S., Soubdhan, T. and Ould-Baba, H. (2016), 'Predictive spatio-temporal model for spatially sparse global solar radiation data', *Energy* **111**, 599–608.

7. Ardia, D. and Hoogerheide, L. (2010), 'Efficient bayesian estimation and combination of garch-type models', *Rethinking Risk Measurement and Reporting: Examples and Applications from Finance* **2**.

8. Armstrong, J. (2001), 'Combining forecasts', *Principles of forecasting: A handbook for researchers and practitioners* pp. 417–439.

9. Aslam, M., Lee, J., Kim, H., Lee, S. and Hong, S. (2020), 'Deep learning models for long-term solar radiation forecasting considering microgrid installation: A comparative study', *Energies* **13**(1), 147.

10. Bacher, P., Madsen, H. and Nielsen, H. A. (2009), 'Online short-term solar power forecasting', *Solar energy* **83**(10), 1772–1783.

11. Badoiu, M. and Clarkson, K. L. (2003), Smaller core-sets for balls, *in* 'SODA', Vol. 3, pp. 801–802.

12. Banerjee, S., Gelfand, A., Finley, A. and Sang, H. (2008), 'Gaussian predictive process models for large spatial data sets', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(4), 825–848.

13. Barlett, P. and Downs, T. (1992), 'Using random weights to train multilayer networks of hard-limiting units', *IEEE Transactions on Neural Networks* **3**(2), 202–210.

14. Batchelor, R. and Dua, P. (1995), 'Forecaster diversity and the benefits of combining forecasts', *Management Science* **41**(1), 68–75.

15. Bates, J. M. and Granger, C. W. (1969), 'The combination of forecasts', *Journal of the Operational Research Society* **20**(4), 451–468.

16. Bayarri, M. and Berger, J. O. (2004), 'The interplay of bayesian and frequentist analysis', *Statistical Science* **19**(1), 58 – 80.

17. Bien, J., Taylor, J. and Tibshirani, R. (2013), 'A lasso for hierarchical interactions', *Annals of statistics* **41**(3), 1111.

18. Breiman, L. (1997), Arcing the edge, Technical report, Technical Report 486, Statistics Department, University of California at ....

19. Brodersen, K., Gallusser, F., Koehler, J., Remy, N. and Scott, S. (2015), 'Inferring causal impact using bayesian structural time-series models', *The Annals of Applied Statistics* **9**(1), 247–274.

20. Cao, D., Zhao, J., Hu, W., Zhang, Y., Liao, Q., Chen, Z. and Blaabjerg, F. (2021), 'Robust deep gaussian process-based probabilistic electrical load forecasting against anomalous events', *IEEE Transactions on Industrial Informatics* .

21. Cervone, G., Clemente-Harding, L., Alessandrini, S. and L., D. M. (2017), 'Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble', *Renewable Energy* **108**, 274–286.

22. Chandiwana, E., Sigauke, C. and Bere, A. (2021), 'Twenty-four-hour ahead probabilistic global horizontal irradiance forecasting using gaussian process regression', *Algorithms* **14**(6), 177.

23. Chen, N., Qian, Z., Nabney, I. T. and Meng, X. (2013), 'Short-term wind power forecasting using gaussian processes', *In Twenty-third international joint conference on artificial intelligence.* .

24. Chikobvu, D. and Sigauke, C. (2013), 'Modelling influence of temperature on daily peak electricity demand in south africa', *Journal of Energy in Southern Africa* **24**(4), 63–70.

25. Christensen, O. and Waagepetersen, R. (2002), 'Bayesian prediction of spatial count data using generalized linear mixed models', *Biometrics* **58**(2), 280–286.

26. Clemen, R. (1989), 'Combining forecasts: A review and annotated bibliography', *International journal of forecasting* **5**(4), 559–583.

27. Clements, M. (2004), 'Evaluating the bank of england density forecasts of inflation', *The Economic Journal* **114**(498), 844–866.

28. Comber, A., Harris, P. and Brunsdon, C. (2022), 'Spatially varying coefficient regression with gam gaussian process splines: Gam (e)-on', *AGILE: GIScience Series* **3**, 31.

29. Crane, D. B. and Crotty, J. (1967), 'A two-stage forecasting model: Exponential smoothing and multiple regression', *Management Science* **13**(8), B–501.

30. Dahl, A. and Bonilla, E. V. (2019), 'Grouped gaussian processes for solar power prediction', *Machine Learning* **108**(8), 1287–1306.

31. Dairi, A., Harrou, F., Sun, Y. and Khadraoui, S. (2020), 'Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach', *Applied Sciences* **10**(23), 8400.

32. Daniel, L., Sigauke, C., Chibaya, C. and Mbuvha, R. (2020), 'Short-term wind speed forecasting using statistical and machine learning methods', *Algorithms* **13**(6), 132.

33. Daniely, A., Frostig, R. and Singer, Y. (2016), 'Toward deeper understanding of neural networks: The power of initialization and a dual view on expressivity', *Advances in neural information processing systems* **29**.

34. DeAar (2022), 'De aar solar power-ariel photos', `https://deaarsolar.co.za/image-gallery/aerial-photographs-de-aar-solar-power/`. Accessed: 2022-05-24.

35. Devaine, M., Gaillard, P., Goude, Y. and Stoltz, G. (2013), 'Forecasting electricity consumption by aggregating specialized experts', *Machine Learning* **90**(2), 231–260.

36. Diebold, F. and Lopez, J. (1996), '8 forecast evaluation and combination', *Handbook of statistics* **14**, 241–268.

37. Diebold, F. and Mariano, R. (2002), 'Comparing predictive accuracy', *Journal of Business and economic statistics* **20**(1), 134–144.

38. Dragomiretskiy, K. and Zosso, D. (2013), 'Variational mode decomposition', *IEEE transactions on signal processing* **62**(3), 531–544.

39. Drucker, H., Burges, C., Kaufman, L., Smola, A., Vapnik, V. et al. (1997), 'Support vector regression machines', *Advances in neural information processing systems* **9**, 155–161.

40. Durbin, J. and Koopman, S. (2002), 'A simple and efficient simulation smoother for state space time series analysis', *Biometrika* **89**(3), 603–616.

41. Ehm, W., Gneiting, T., Jordan, A. and Krüger, F. (2016), 'Of quantiles and expectiles: consistent scoring functions, choquet representations and forecast rankings', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**(3), 505–562.

42. Elamin, N. (2018), Quantile regression model for peak load demand forecasting with approximation by triangular distribution to avoid blackouts, *in* 'Quantile regression model for peak load demand forecasting with approximation by triangular distribution to avoid blackouts: Elamin, Niematallah'.

43. energy economics, D. and statistics (2021), 'The south african energy sector report', https://www.energy.gov.za/files/media/explained/2021-South-African-Energy-Sector-Report.pdf. Accessed: 23-02-2023.

44. Energycentre, C. R. e. c. (2022), 'Statistics of utility-scale power generation in south africa in 2021', https://www.csir.co.za/sites/default/files/Documents/20220503-Statistics%20of%20power%20in%20SA%20H2-2021-CSIR-%5BFINAL%5D%20%281%29.pdf. Accessed: 2023-02-2023.

45. Erickson, C., Ankenman, B. and Sanchez, S. (2018), 'Comparison of gaussian process modeling software', *European Journal of Operational Research* **266**(1), 179–192.

46. Eschenbach, A., Yepes, G., Tenllado, C., Gomez-Perez, J., Pinuel, L., Zarzalejo, L. and Wilbert, S. (2020), 'Spatio-temporal resolution of irradiance samples in machine learning approaches for irradiance forecasting', *IEEE Access* **8**, 51518–51531.

47. Fasiolo, M., Wood, S., Zaffran, M., Nedellec, R. and Goude, Y. (2020), 'qgam: Bayesian non-parametric quantile regression modelling in r', *arXiv preprint arXiv:2007.03303* .

48. Ferkous, K., Chellali, F., Kouzou, A. and Bekkar, B. (2021), 'Wavelet-gaussian process regression model for forecasting daily solar radiation in the saharan climate', *Clean Energy* **5**(2), 316–328.

49. Fernandez, C., Ley, E. and Steel, M. (2001), 'Model uncertainty in cross-country growth regressions', *Journal of applied Econometrics* **16**(5), 563–576.

50. Friedman, J. (1991), 'Multivariate adaptive regression splines', *The annals of statistics* **19**(1), 1–67.

51. Friedman, J. (1999), 'Stochastic gradient boosting. department of statistics'.

52. Friedman, J. H. and Fisher, N. (1999), 'Bump hunting in high-dimensional data', *Statistics and computing* **9**(2), 123–143.

53. Gaillard, P., Goude, Y., Plagne, L., Dubois, T., Thieurmel, B. and Gaillard, M. P. (2021), 'Package 'opera'', *Rcpp, LinkingTo and Rcpp, RcppEigen Imports and Rdpack, RdMacros* .

54. Gal, Y. and Ghahramani, Z. (2016), Dropout as a bayesian approximation: Representing model uncertainty in deep learning, *in* 'international conference on machine learning', PMLR, pp. 1050–1059.

55. Gal, Y., Islam, R. and Ghahramani, Z. (2017), Deep bayesian active learning with image data, *in* 'International conference on machine learning', PMLR, pp. 1183–1192.

56. Gbémou, S., Eynard, J., Thil, S. and Grieu, S. (2022), Comparison between time- and observation-based gaussian process regression models for global horizontal irradiance forecasting, *in* 'Solar', Vol. 2, MDPI, pp. 445–468.

57. Gbemou, S., Eynard, J., Thil, S., Guillot, E. and Grieu, S. (2021), 'A comparative study of machine learning-based methods for global horizontal irradiance forecasting', *Energies* **14**(11), 1–24.

58. Gbémou, S., Tolba, H., Thil, S. and Grieu, S. (2019), Global horizontal irradiance forecasting using online sparse gaussian process regression based on quasiperiodic kernels, *in* '2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)', IEEE, pp. 1–6.

59. Gensler, A., Henze, J., Sick, B. and Raabe, N. (2016), Deep learning for solar power forecasting—an approach using autoencoder and lstm neural networks, *in* '2016 IEEE international conference on systems, man, and cybernetics (SMC)', IEEE, pp. 002858–002865.

60. George, E. I. and McCulloch, R. (1997), 'Approaches for bayesian variable selection', *Statistica sinica* pp. 339–373.

61. Gill, J., Rubiera, J., Martin, C., Cacic, I., Mylne, K., Dehui, C., Jiafeng, G., Xu, T., Yamaguchi, M. and Foamouhoue, K. (2022), 'Guidelines on communicating for uncertainty', `https://public.wmo.int/en/`. Accessed: 2022-07-06.

62. Gneiting, T. and Raftery, A. (2007), 'Strictly proper scoring rules, prediction, and estimation', *Journal of the American statistical Association* **102**(477), 359–378.

63. Govender, P., Brooks, M. and Matthews, A. P. (2018), 'Cluster analysis for classification and forecasting of solar irradiance in durban, south africa', *Journal of Energy in Southern Africa* **29**(2), 51–62.

64. Government, S. (2022), 'President cyril ramaphosa: Address to the nation on energy crisis', `https://www.gov.za/speeches/president-cyril-ramaphosa-address-nation-energy-crisis-25-jul-2022-0000`. Accessed: 23-02-2023.

65. Graham, J. (2009), 'Missing data analysis: Making it work in the real world', *Annual review of psychology* **60**, 549–576.

66. Granger, C. W. (1989), 'Invited review combining forecasts—twenty years later', *Journal of forecasting* **8**(3), 167–173.

67. Hafner, D., Tran, D., Irpan, A., Lillicrap, T. and Davidson, J. (2018), 'Reliable uncertainty estimates in deep neural networks using noise contrastive priors', *stat* **1050**, 24.

68. Hair, J., Hult, G., Ringle, C. M., Sarstedt, M., Danks, N., Ray, S., Hair, J. F., Hult, G., Ringle, C., Sarstedt, M. et al. (2021), 'An introduction to structural equation modeling', *Partial least squares structural equation modeling (PLS-SEM) using R: a workbook* pp. 1–29.

69. Hamelijnck, O., Wilkinson, W., Loppi, N., Solin, A. and Damoulas, T. (2021), 'Spatio-temporal variational gaussian processes', *Advances in Neural Information Processing Systems* **34**, 23621–23633.

70. Harvey, A. (1990), 'Forecasting, structural time series models and the kalman filter'.

71. Hastie, T., Qian, J. and Tay, K. (2021), 'An introduction to glmnet', *CRAN R Repositary* .

72. Hoeting, J., Madigan, D., Raftery, A. and Volinsky, C. (1999), 'Bayesian model averaging: a tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors', *Statistical science* **14**(4), 382–417.

73. Hong, T. and Shahidehpour, M. (2015), 'Load forecasting case study', *EISPC, US Department of Energy* .

74. Hong, T., Wilson, J. and Xie, J. (2013), 'Long term probabilistic load forecasting and normalization with hourly information', *IEEE Transactions on Smart Grid* **5**(1), 456–462.

75. IEA. (2023), 'Data and statics, iea', `https://www.iea.org/data-and-statistics`. Accessed: 02-02-23.

76. Isaksson, E. and Karpe, C. (2018), 'Solar power forecasting with machine learning techniques'.

77. Ismail, A., Ramirez-Iniguez, R., Asif, M., Munir, A. and F., M.-S. (2015), 'Progress of solar photovoltaic in asean countries: A review', *Renewable and Sustainable Energy Reviews* **48**, 399–412.

78. Jäkel, F., Schölkopf, B. and Wichmann, F. (2007), 'A tutorial on kernel methods for categorization', *Journal of Mathematical Psychology* **51**(6), 343–358.

79. Jang, H. S., Bae, K. Y., Park, H.-S. and Sung, D. K. (2016), 'Solar power prediction based on satellite images and support vector machine', *IEEE Transactions on Sustainable Energy* **7**(3), 1255–1263.

80. Jordan, A., Krüger, F. and Lerch, S. (2017), 'Evaluating probabilistic forecasts with scoringrules', *arXiv preprint arXiv:1709.04743* .

81. Juban, R., Ohlsson, H., Maasoumy, M., Poirier, L. and Kolter, J. Z. (2016), 'A multiple quantile regression approach to the wind, solar, and price tracks of gefcom2014', *International Journal of Forecasting* **32**(3), 1094–1102.

82. Kalman, R. (1960), 'A new approach to linear filtering and prediction problems'.

83. Kim, B. and Suh, D. (2020), 'A hybrid spatio-temporal prediction model for solar photovoltaic generation using numerical weather data and satellite images', *Remote Sensing* **12**(22), 3706.

84. Koenker, R. (2005), *Quantile regression*, Vol. 38, Cambridge university press.

85. Koenker, R. (2008), 'Censored quantile regression redux', *Journal of Statistical Software* **27**, 1–25.

86. Koenker, R. and Bassett, G. (1978), 'Regression quantiles', *Econometrica: journal of the Econometric Society* pp. 33–50.

87. Krige, D. (1976), 'Some basic considerations in the application of geostatistics to the valuation of ore in south african gold mines', *Journal of the Southern African Institute of Mining and Metallurgy* **76**(9), 383–391.

88. Larson, D. P., Nonnenmacher, L. and Coimbra, C. (2016), 'Day-ahead forecasting of solar power output from photovoltaic plants in the american southwest', *Renewable Energy* **91**, 11–20.

89. Lawin, A. E., Niyongendako, M. and Manirakiza, C. (2019), 'Solar irradiance and temperature variability and projected trends analysis in burundi', *Climate* **7**(6), 83.

90. Lee, K. J. and Carlin, J. (2010), 'Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation', *American journal of epidemiology* **171**(5), 624–632.

91. Leholo, S., Owolawi, P. and Akindeji, K. (2019), Solar energy potential forecasting and optimization using artificial neural network: South africa case study, *in* '2019 Amity International Conference on Artificial Intelligence (AICAI)', IEEE, pp. 533–536.

92. Leith, D., Heidl, M. and Ringwood, J. (2004), Gaussian process prior models for electrical load forecasting, *in* '2004 International Conference on Probabilistic Methods Applied to Power Systems', IEEE, pp. 112–117.

93. Ley, E. and Steel, M. (2009), 'On the effect of prior assumptions in bayesian model averaging with applications to growth regression', *Journal of applied econometrics* **24**(4), 651–674.

94. Lim, M. and Hastie, T. (2015), 'Learning interactions via hierarchical group-lasso regularization', *Journal of Computational and Graphical Statistics* **24**(3), 627–654.

95. Little, R. (1988), 'Missing-data adjustments in large surveys', *Journal of Business & Economic Statistics* **6**(3), 287–296.

96. Liu, B. and Jordan, R. (1960), 'The interrelationship and characteristic distribution of direct, diffuse and total solar radiation', *Solar energy* **4**(3), 1–19.

97. Liu, B., Nowotarski, J., Hong, T. and Weron, R. (2015), 'Probabilistic load forecasting via quantile regression averaging on sister forecasts', *IEEE Transactions on Smart Grid* **8**(2), 730–737.

98. Liu, Y., Qin, H., Zhang, Z., Pei, S., Wang, C., Yu, X., Jiang, Z. and Zhou, J. (2019), 'Ensemble spatiotemporal forecasting of solar irradiation using variational bayesian convolutional gate recurrent unit network', *Applied Energy* **253**, 113596.

99. Lourenço, J. and Santos, P. (2010), 'Short term load forecasting using gaussian process models', *Proceedings of Instituto de Engenharia de Sistemas e Computadores de Coimbra* .

100. Lubbe, F., Maritz, J. and Harms, T. (2020), 'Evaluating the potential of gaussian process regression for solar radiation forecasting: A case study', *Energies* **13**(20), 5509.

101. Luttinen, J. and Ilin, A. (2009), 'Variational gaussian-process factor analysis for modeling spatio-temporal data', *Advances in neural information processing systems* **22**.

102. Ma, T., Barajas-Solano, D., Tipireddy, R. and Tartakovsky, A. (2020), 'Physics-informed gaussian process regression for probabilistic states estimation and forecasting in power grids', *arXiv preprint arXiv:2010.04591* .

103. Mabasa, B., Lysko, M., Tazvinga, H., Mulaudzi, S. T., Zwane, N. and Moloi, S. (2020), 'The ångström–prescott regression coefficients for six climatic zones in south africa', *Energies* **13**(20), 5418.

104. MacKay, D. and Neal, R. M. (1997), 'Near shannon limit performance of low density parity check codes', *Electronics letters* **33**(6), 457–458.

105. Maritz, J., Lubbe, F. and Lagrange, L. (2018), 'A practical guide to gaussian process regression for energy measurement and verification within the bayesian framework', *Energies* **11**(4), 935.

106. Marlin, J., Diez, R. and Insua, D. (2003), 'Bayesian methods in plant conservation biology', *Biological Conservation* **113**(3), 379–387.

107. Martins, F., DaCunha, J. and Serra, F. (2018), 'Secondary data in research–uses and opportunities', *PODIUM sport, leisure and tourism review* **7**(3).

108. Masike, K. and Vermeulen, C. (2022), 'The time-varying elasticity of south african electricity demand', *Energy* **238**, 121984.

109. Mason, L., Baxter, J., Bartlett, P. and Frean, M. (1999), 'Boosting algorithms as gradient descent', *Advances in neural information processing systems* **12**.

110. Matérn, B. (2013), *Spatial variation*, Vol. 36, Springer Science & Business Media.

111. Mendonça, D., Pires, P. S., Pinheiro, A., Gonçalves, M. E. and Mussetta, M. L. S. (2020), 'Multiple site intraday solar irradiance forecasting by machine learning algorithms: Mggp and mlp neural networks', *Energies* **13**(11), 3005.

112. Menyah, K. and Wolde-Rufael, Y. (2010), 'Energy consumption, pollutant emissions and economic growth in south africa', *Energy economics* **32**(6), 1374–1382.

113. Milligan, M., Schwartz, M. and Wan, Y. (2003), Statistical wind power forecasting for us wind farms, Technical report, National Renewable Energy Lab., Golden, CO (US).

114. Ministry of forestry, f. and the environment (2016), 'The south african energy sector report', `https://www.dffe.gov.za/mediarelease/southafricasignsparisagreementonclimate`. Accessed: 20-02-2023.

115. Ministry of minerals, r. and energy (2019), 'Intergrated resource plan 2019', `https://www.energy.gov.za/IRP/2019/IRP-2019.pdf`. Accessed: 23-02-2023.

116. Mitchell, T. and Beauchamp, J. (1988), 'Bayesian variable selection in linear regression', *Journal of the american statistical association* **83**(404), 1023–1032.

117. Mitrentsis, G. and Lens, H. (2022), 'An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting', *Applied Energy* **309**, 118473.

118. Mohammed, A. A., Yaqub, W. and Aung, Z. (2017), Probabilistic forecasting of solar power: An ensemble learning approach, *in* 'International Conference on Intelligent Decision Technologies', Springer, pp. 449–458.

119. Mosetlhe, T., Yusuff, A. and Hamam, Y. (2018), 'Investigating seasonal wind energy potential in vredendal, south africa', *Journal of Energy in Southern Africa* **29**(2), 77–83.

120. Mpfumali, P., Sigauke, C., Bere, A. and Mulaudzi, S. (2019), 'Day ahead hourly global horizontal irradiance forecasting—application to south african data', *Energies* **12**(18), 3569.

121. Mutavhatsindi, T., Sigauke, C. and Mbuvha, R. (2020*a*), 'Forecasting hourly global horizontal solar irradiance in south africa using machine learning models', *IEEE Access* **8**, 198872–198885.

122. Mutavhatsindi, T., Sigauke, C. and Mbuvha, R. (2020*b*), 'Forecasting hourly global horizontal solar irradiance in south africa using machine learning models', *IEEE Access* **8**, 198872–198885.

123. Najibi, F., Apostolopoulou, D. and Alonso, E. (2021*a*), Clustering sensitivity analysis for gaussian process regression based solar output forecast, *in* '2021 IEEE Madrid PowerTech', IEEE, pp. 1–6.

124. Najibi, F., Apostolopoulou, D. and Alonso, E. (2021*b*), 'Enhanced performance gaussian process regression for probabilistic short-term solar output forecast', *International Journal of Electrical Power & Energy Systems* **130**, 106916.

125. Nations., U. (2021), 'Undp, sustainable development goal 7: Affordable and clean energy', `https://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-7-affordable-and-clean-energy.html/`. Accessed: 2021-12-06.

126. Neal, R. (1996), 'Priors for infinite networks', *Bayesian learning for neural networks* pp. 29–53.

127. Nelder, J. and Wedderburn, R. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society: Series A (General)* **135**(3), 370–384.

128. NERSA (2016), 'Monitoring of renewable energy performance 2016', `https://www.nersa.org.za/ReadNews.aspx?NewsID=325`. Accessed: 2022-05-24.

129. Novak, R., Xiao, L., Lee, J., Bahri, Y., Yang, G., Hron, J., Abolafia, D., Pennington, J. and Sohl-Dickstein, J. (2018), 'Bayesian deep convolutional networks with many channels are gaussian processes', *arXiv preprint arXiv:1810.05148* .

130. Nowotarski, J., Raviv, E., Trück, S. and Weron, R. (2014), 'An empirical comparison of alternative schemes for combining electricity spot price forecasts', *Energy Economics* **46**, 395–412.

131. Palm, N., Landerer, M. and Palm, H. (2022), 'Gaussian process regression based multi-objective bayesian optimization for power system design', *Sustainability* **14**(19), 12777.

132. Pattanaik, D., Mishra, S., Khuntia, G., Dash, R. and Swain, S. (2020), 'An innovative learning approach for solar power forecasting using genetic algorithm and artificial neural network', *Open Engineering* **10**(1), 630–641.

133. Peters, J., Janzing, D. and Schölkopf, B. (2017), *Elements of causal inference: foundations and learning algorithms*, The MIT Press.

134. Petris, G., Petrone, S. and Campagnoli, P. (2009), Dynamic linear models, *in* 'Dynamic Linear Models with R', Springer, pp. 31–84.

135. Raftery, A. E., Madigan, D. and Hoeting, J. (1997), 'Bayesian model averaging for linear regression models', *Journal of the American Statistical Association* **92**(437), 179–191.

136. Ranganai, E. and Sigauke, C. (2020), 'Capturing long-range dependence and harmonic phenomena in 24-hour solar irradiance forecasting: A quantile regression robustification via forecasts combination approach', *IEEE Access* **8**, 172204–172218.

137. Raza, M., Nadarajah, M. and Ekanayake, C. (2016), 'On recent advances in pv output power forecast', *Solar Energy* **136**, 125–144.

138. Resources and Energy Department, S. (2016), 'Renewable energy: Solar energy', `http://www.energy.gov.za/files/renewables_frame.html`. Accessed: 2022-05-24.

139. Roman, I., Santana, R., Mendiburu, A. and Lozano, J. A. (2021), 'Evolving gaussian process kernels from elementary mathematical expressions for time series extrapolation', *Neurocomputing* **462**, 426–439.

140. Rubin, D. and Schenker, N. (1986), 'Multiple imputation for interval estimation from simple random samples with ignorable nonresponse', *Journal of the American statistical Association* **81**(394), 366–374.

141. Sahu, S. (2022), *Bayesian modeling of spatio-temporal data with R*, Chapman and Hall/CRC.

142. Sahu, S., Bakar, K., Zhan, J., Campbell, J. and Yanai, R. (2020), Spatio-temporal bayesian modeling of precipitation using rain gauge data from the hubbard brook experimental forest, new hampshire, usa, American Statistical Association.

143. Schafer, J. and Olsen, M. K. (1998), 'Multiple imputation for multivariate missing-data problems: A data analyst's perspective', *Multivariate behavioral research* **33**(4), 545–571.

144. Scott, S. and Varian, H. (2014), 'Predicting the present with bayesian structural time series', *International Journal of Mathematical Modelling and Numerical Optimisation* **5**(1-2), 4–23.

145. Scott, S. and Varian, H. (2015), Bayesian variable selection for nowcasting economic time series, *in* 'Economic analysis of the digital economy', University of Chicago Press, pp. 119–135.

146. Sharma, V., Ordiano, J., Mikut, R. and Cali, U. (2021), 'Probabilistic solar power forecasting: Long short-term memory network vs simpler approaches', *arXiv preprint arXiv:2101.08236* .

147. Sigauke, C. (2017), 'Forecasting medium-term electricity demand in a south african electric power supply system', *Journal of Energy in Southern Africa* **28**(4), 54–67.

148. Sigauke, C., Ravele, T. and Jhamba, L. (2022), 'Extremal dependence modelling of global horizontal irradiance with temperature and humidity: An application using south african data', *Energies* **15**(16), 5965.

149. Słoński, M. (2011), 'Bayesian neural networks and gaussian processes in identification of concrete properties', *Computer Assisted Methods in Engineering and Science* **18**(4), 291–302.

150. Smart-Energy (2022), 'Alternative sustainable energy solutions', `https://smartenergy-sa.co.za/about/`. Accessed: 2022-05-24.

151. Stein, M. (1999), *Interpolation of spatial data: some theory for kriging*, Springer Science & Business Media.

152. Sun, K., Li, K., Zhang, Z., Liang, Y., Liu, Z. and Lee, W. (2021), 'An integration scheme of renewable energies, hydrogen plant, and logistics center in the suburban power grid', *IEEE Transactions on Industry Applications* **58**(2), 2771–2779.

153. Tangwe, S., Simon, M. and Mhundwa, R. (2018), 'The performance of split and integrated types air-source heat pump water heaters in south africa', *Journal of Energy in Southern Africa* **29**(2), 12–20.

154. Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

155. Timmermann, A. (2006), 'Forecast combinations', *Handbook of economic forecasting* **1**, 135–196.

156. Todescato, M., Carron, A., Carli, R., Pillonetto, G. and Schenato, L. (2020), 'Efficient spatio-temporal gaussian regression via kalman filtering', *Automatica* **118**, 109032.

157. Tolba, H., Dkhili, N., Nou, J., Eynard, J., Thil, S. and Grieu, S. (2019), 'Ghi forecasting using gaussian process regression: kernel study', *IFAC-PapersOnLine* **52**(4), 455–460.

158. Tomizawa, Y. and Yoshida, I. (2022), 'Benchmarking of gaussian process regression with multiple random fields for spatial variability estimation', *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* **8**(4), 04022052.

159. Trapero, J. (2016), 'Calculation of solar irradiation prediction intervals combining volatility and kernel density estimates', *Energy* **114**, 266–274.

160. Tsang, I., Kwok, J., Cheung, P. and Cristianini, N. (2005), 'Core vector machines: Fast svm training on very large data sets.', *Journal of Machine Learning Research* **6**(4).

161. Tsymbalov, E., Makarychev, S., Shapeev, A. and Panov, M. (2019), 'Deeper connections between neural networks and gaussian processes speed-up active learning', *arXiv preprint arXiv:1902.10350* .

162. Uniejewski, B. and Weron, R. (2021), 'Regularized quantile regression averaging for probabilistic electricity price forecasting', *Energy Economics* **95**, 105121.

163. Ver Hoef, J. and Cressie, N. (2020), Spatial statistics: analysis of field experiments, *in* 'Design and analysis of ecological experiments', Chapman and Hall/CRC, pp. 319–341.

164. Verrelst, J., Rivera, J., Moreno, J. and Camps-Valls, G. (2013), 'Gaussian processes uncertainty estimates in experimental sentinel-2 lai and leaf chlorophyll content retrieval', *ISPRS journal of photogrammetry and remote sensing* **86**, 157–167.

165. Wang, B., Wang, Q., Wei, Y. and Li, Z. (2018), 'Role of renewable energy in china's energy security and climate change mitigation: An index decomposition analysis', *Renewable and sustainable energy reviews* **90**, 187–194.

166. Wang, F., Chen, P., Zhen, Z., Yin, R., Cao, C., Zhang, Y. and Duić, N. (2022), 'Dynamic spatio-temporal correlation and hierarchical directed graph structure based ultra-short-term wind farm cluster power forecasting method', *Applied Energy* **323**, 119579.

167. Wang, Z., Lin, X., Tong, N., Li, Z., Sun, S. and Liu, C. (2020), 'Optimal planning of a 100% renewable energy island supply system based on the integration of a concentrating solar power plant and desalination units', *International Journal of Electrical Power & Energy Systems* **117**, 105707.

168. Wilkins, A. S. (2018), 'To lag or not to lag?: Re-evaluating the use of lagged dependent variables in regression analysis', *Political Science Research and Methods* **6**(2), 393–411.

169. Williams, C. (1996), 'Computing with infinite networks', *Advances in neural information processing systems* **9**.

170. Williams, C. and Rasmussen, C. (2006), *Gaussian processes for machine learning*, Vol. 2, MIT press Cambridge, MA.

171. Wilson, A. (2014), Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes, PhD thesis, University of Cambridge Cambridge, UK.

172. WorldBank (2022), 'South africa: Country climate and development report', `https://www.worldbank.org/en/news/infographic/2022/11/01/south-africa-country-climate-and-development-report`. Accessed: 20-02-2023.

173. Xue, R., Wang, S., Long, W., Gao, G., Liu, D. and Zhang, R. (2021), 'Uncovering ghg emission characteristics of industrial parks in central china via emission inventory and cluster analysis', *Energy Policy* **151**, 112191.

174. Yang, D. (2018), 'Ultra-fast preselection in lasso-type spatio-temporal solar forecasting problems', *Solar Energy* **176**, 788–796.

175. Yang, D., Gu, C., Dong, Z., Jirutitijaroen, P., Chen, N. and Walsh, W. (2013), 'Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging', *Renewable Energy* **60**, 235–245.

176. Yildirim, E. (2008), 'Two algorithms for the minimum enclosing ball problem', *SIAM Journal on Optimization* **19**(3), 1368–1391.

177. Yu, B., Cunningham, J., Santhanam, G., Ryu, S., Shenoy, K. and Sahani, M. (2008), 'Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity', *Advances in neural information processing systems* **21**.

178. Zhandire, E. (2017), 'Predicting clear-sky global horizontal irradiance at eight locations in south africa using four models', *Journal of Energy in Southern Africa* **28**(4), 77–86.

179. Zhang, H. (2003), 'Optimal interpolation and the appropriateness of cross-validating variogram in spatial generalized linear mixed models', *Journal of Computational and Graphical Statistics* **12**(3), 698–713.

180. Zhang, J., Ju, Y., Mu, B., Zhong, R. and Chen, T. (2023), 'An efficient implementation for spatial–temporal gaussian process regression and its applications', *Automatica* **147**, 110679.

181. Zhang, W., Quan, H. and Srinivasan, D. (2018), 'An improved quantile regression neural network for probabilistic load forecasting', *IEEE Transactions on Smart Grid* **10**(4), 4425–4434.

182. Zhang, Z., Wang, C., Peng, X., Qin, H., Hao, L. and Wang, H. (2021), 'Solar radiation intensity probabilistic forecasting based on k-means time series clustering and gaussian process regression', *IEEE Access* **9**, 89079–89092.

183. Zou, H. and Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the royal statistical society: series B (statistical methodology)* **67**(2), 301–320.