algorithms

Article

Supplementary Material for the Paper "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: This paper provides supplementary material for the study discussed in the paper "Twenty-four-hour ahead probabilistic global horizontal irradiance forecasting using Gaussian process regression" which includes a detailed analysis of the benchmark models used in the study. The data used in the study is also included as supplemenray material.

check for updates

Citation: Chandiwana, E.; Sigauke, C.; Bere, A. . *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. Benchmark Models and Evaluation of Prediction Techniques

Two benchmark models, gradient boosting method (GBM) and support vector regression (SVR) were used as a basis of comparison to the Gaussian process regression (GPR) model.

Stochastic Gradient Boosting Model

We start with an analysis of the stochastic GBM which was done in two parts, one without interactions and the other with interactions. Since we are going to do a real-time analysis of our forecasts for the next 10 hours will make use of 150 observations because we found our period to be equal to 15 from the periodogram.

1.1. VEN Data with No Interactions

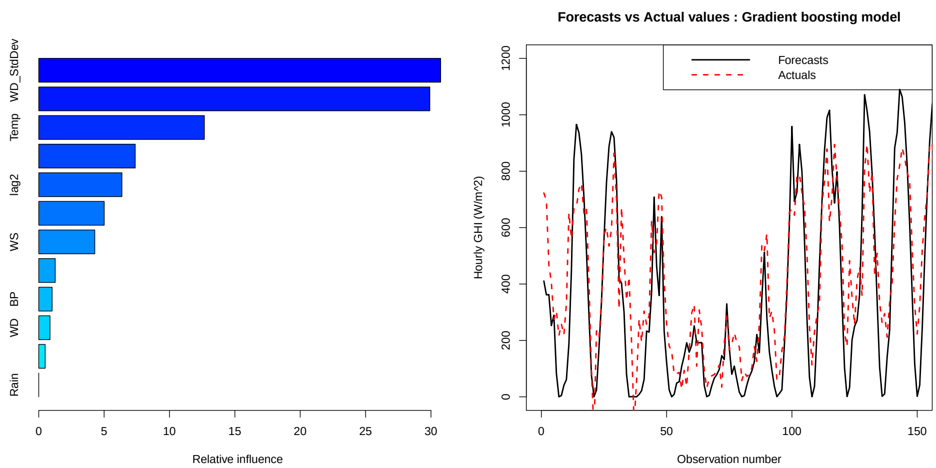

Figure 1. GBM **Left panel**: Relative influence of different variables. **Right panel**: GHI superimposed with predictions.

In Figure 1, the left panel shows the relative influence of different variables, while the right panel shows results of GHI superimposed with predictions from the GBM.

1.2. SUN Data

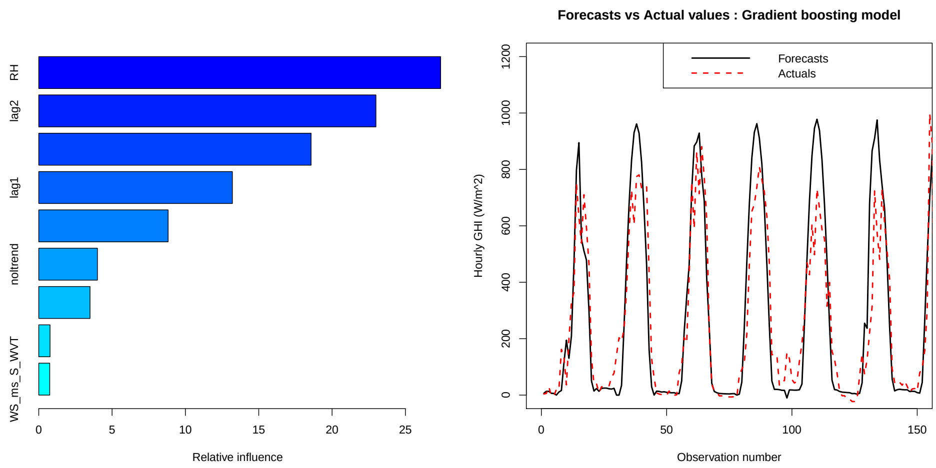

Figure 2. GBM Actual vs prediction plot.

GBM forecasting was also done SUN data without interactions. The forecasted solar power, GHI for the SUN data is shown in Figure 2 (to the right), the red plot shows the actual data, and the forecasted data is in black for the period 2019 to 2020 and we observe that the Gradient Boosting forecaster follow the actual demand close enough. We implemented the gradient boosting regression and checked the relative influence of different variables, the results are shown in Figure 2 (left panel).

1.3. SUN Data with Interactions

SUN data with interactions were analysed using GBM, Figure 3 shows the results of SUN data using the GBM with interactions.

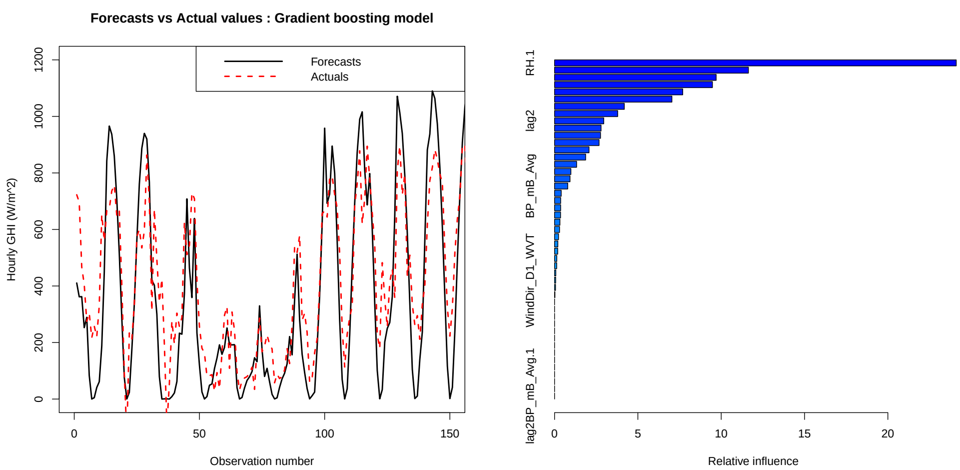

Figure 3. GBM with interactions-SUN.

1.4. Support Vector Regression

The analysis that follows is based on the Support vector regression, which was also done in 2 parts that is one with interactions included and the other without.

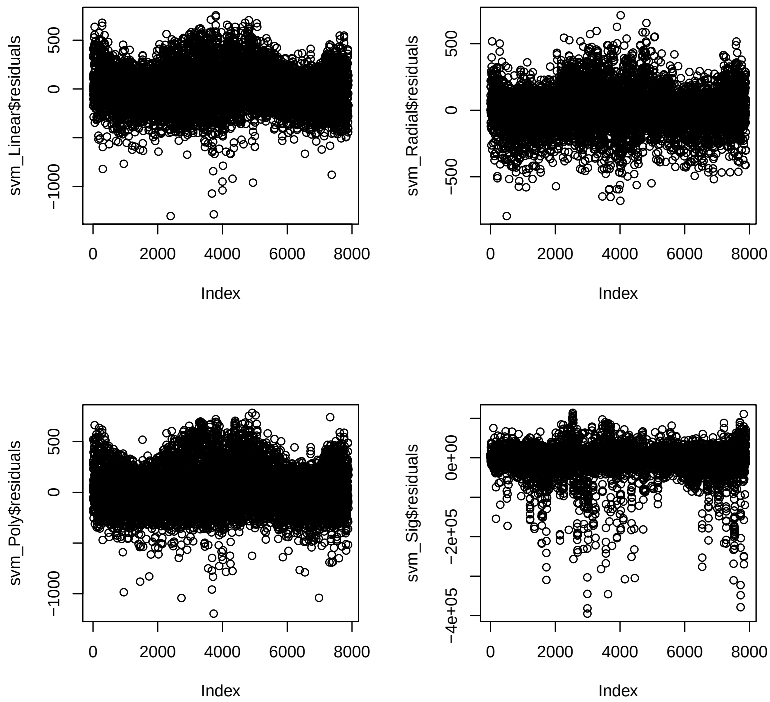

Figure 4. Residuals plots of different kernels for the UNV data.

1.5. VEN Data

Figure 4 shows residuals plots of different kernels for the VEN data, the residuals are reasonably well spread above and below with slightly less variance there and some points clustering about the point 0. In these plots, each point shows the prediction made by the model on the y-axis, and the accuracy of the prediction is shown on the x-axis. They appear symmetrically distributed and we cannot identify any clear patterns, hence the models are good, we just have to select the best one. The best model is the one with a radial kernel, selection was based on RMSE.

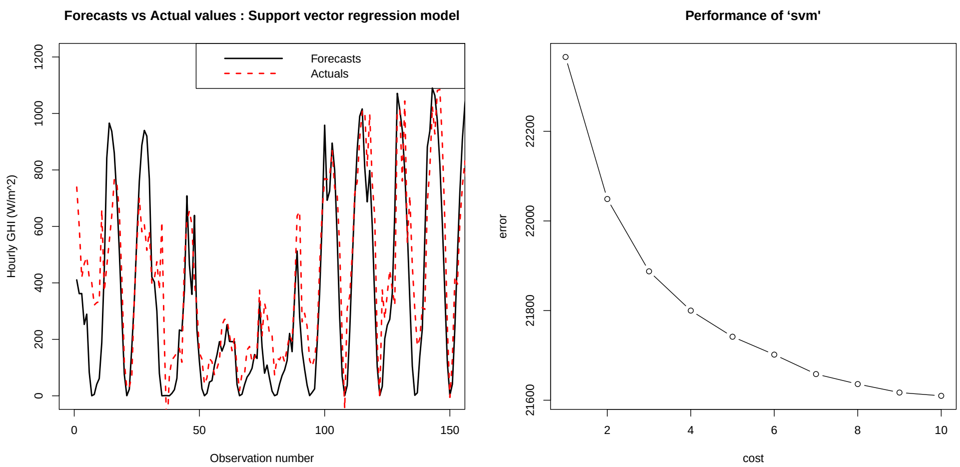

Figure 5. Plot of observations vs forecasts for UNV data using SVM.

The plot on the left panel in Figure 5 shows observations against the predicted UNV data using SVM. The black line represents the actual observations whereas the red line predicted values. It appears SVM fits well the data since the values are closer to the actual. On the right panel the plot shows the cost function for the UNV data which measures the performance of a machine learning model, in this case, we applied SVM. The goal of the function is to come up with values of the model parameters that give smaller values of a cost function that is minimizing errors. The graph shows that the error is varying with the index of complexity when the error is high, complexity is high and when it's low the complexity reaches the minimum, the cost is the value of complexity.

1.6. UNV Data Analysis with Interactions

SVM analysis was done on UNV data with interaction and the results are shown on Figure 6.

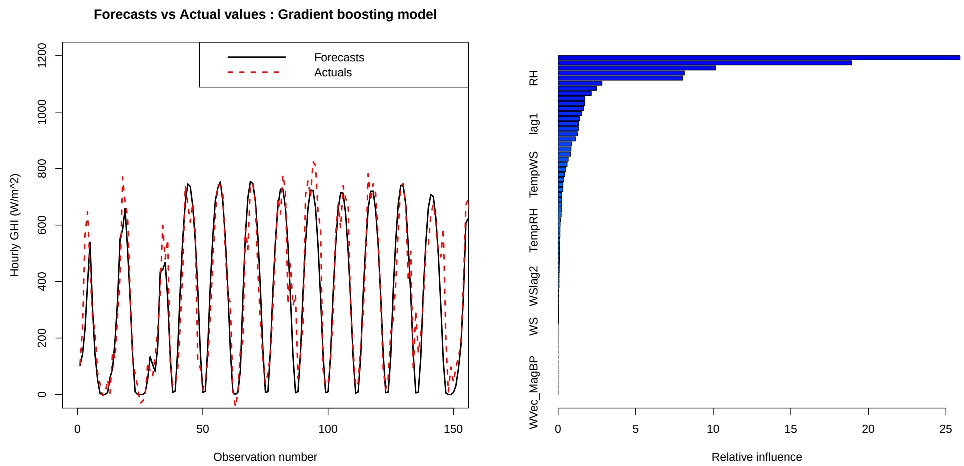

Figure 6. UNV Actual vs prediction plot.

1.7. SUN data

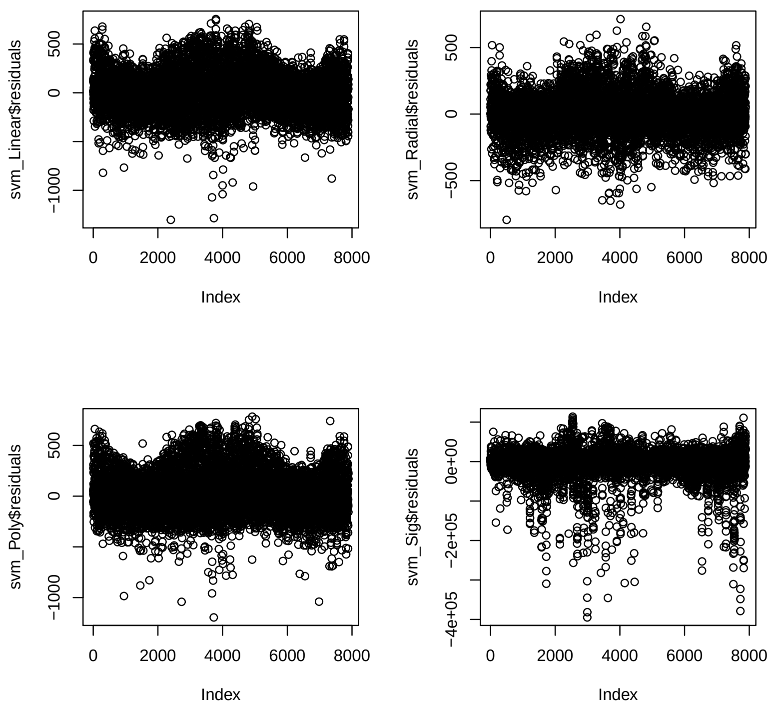

Figure 7. Residuals Plot for the SVM for SUN data.

Figure 7 shows residuals plots of different kernels for the SUN data, the residuals are reasonably well spread above and below with slightly less variance there. They are pretty symmetrically distributed and we cannot identify any clear patterns, hence we can conclude that the 4 models are clearly good, we just have to select the best one. The best-selected model is the one with a radial kernel, selection was based on RMSE.

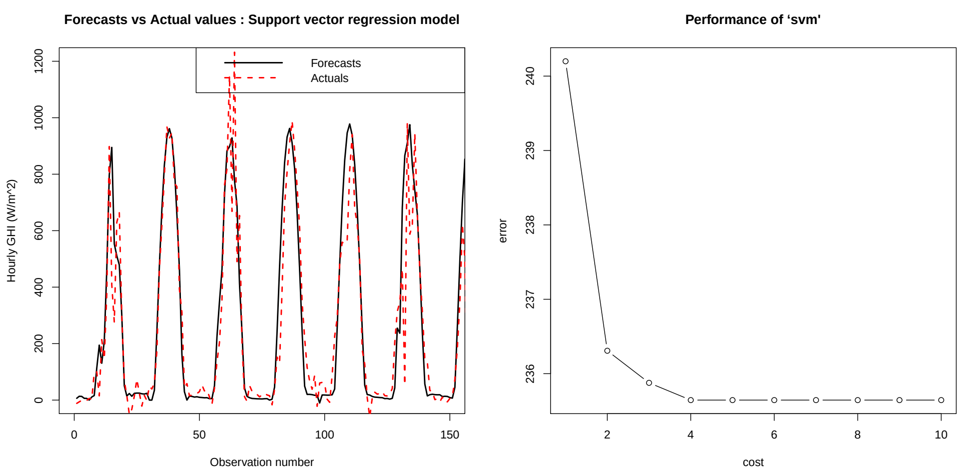

Figure 8. GBM Actual vs prediction plot.

The results of the predicted values of SUN data against actual are shown in the plot in Figure 8 (left panel), the forecasts were done using SVM. The black line represents the actual observations whereas the red line the predicted values. SVM appears to fit well the data.

On the right, Figure 8 shows a plot reflecting cost function which measures machine learning model performance for the SUN dataset. The goal of the function is to come up with values of the model parameters that give smaller values of a cost function that minimises errors. As we can see from the graph, the error is varying with the index of complexity, when the error is high, complexity is also high, and when it's low the complexity reaches the minimum.

1.8. SUN Data with Interactions

The results in Figure 9 show the SVM analysis of SUN data taking into consideration interactions.

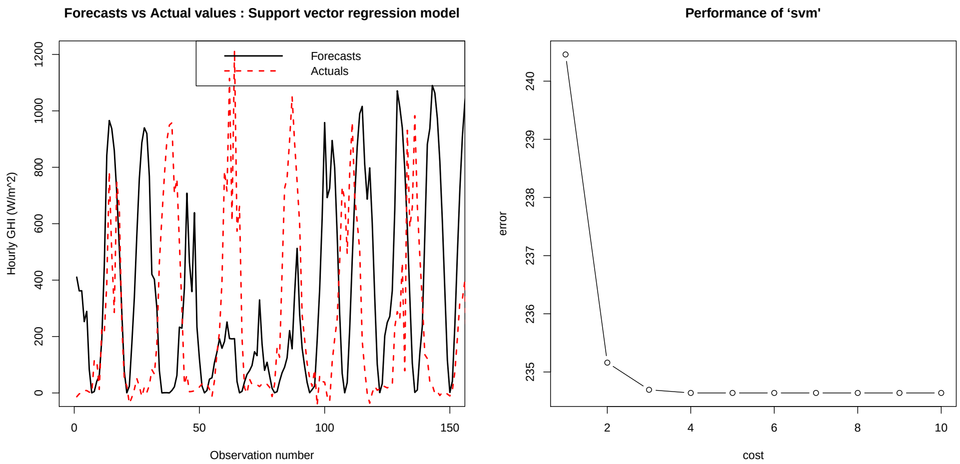

Figure 9. Performance of SVM.