ABSTRACT

The increasing adaptation towards renewable energy (RE) technologies has brought the need for reliable RE forecasting systems. Such systems can bring significant environmental and economic benefits to both distribution network and end user levels in terms of efficient energy management. As more RE sources integrate with existing or future energy supply structures, several challenges arise in the management and deliverance of electricity due to its intermittent and unpredictable nature, especially for solar and wind. Hence, RE forecasting systems are necessary to enable more efficient energy production and usage as well as improved energy management.

Within this study, a comparative model analysis was conducted to evaluate the forecast performance of hourly solar irradiance at Melbourne Airport, Australia. These data for the location of interest were obtained from the Australian Bureau of Meteorology. Various models were considered such as: (i) Multi-Linear Regression (MLR), (ii) Support Vector Regression (SVR), (iii) Multi-Layer Perceptron (MLP), and (iv) Long Short-Term Memory (LSTM) network. During cross-validation, the SVR and LSTM models attained high predictive performances relative to the MLR model as a baseline when temporal structures (i.e. trends and seasonality) were retained in the dataset. Upon evaluating the model performance on unseen data, the LSTM model was the best performing model as it learnt to recognise temporal patterns throughout the dataset and incorporate it in its predictions. Despite the findings presented in this study, more emphasis should be placed on feature selection, model tuning and optimizing model design to achieve better performance.

INTRODUCTION

The purpose of this study is to determine the most accurate model in forecasting hourly solar irradiance using historical ground-measured weather and solar irradiance data. The site of interest is located at Melbourne Airport, Australia. Several methods were undertaken to address the study's objective that involves (i) attaining measured data from the Australian Bureau of Meteorology, (ii) pre-processing and transforming the datasets to ensure its compliance with the computational models, (iii) model selection, and (iv) performance analysis.

DATA SELECTION

The site of interest is located at Melbourne Airport, Australia (Latitude = -37.6655, Longitude = 144.8321). Both weather and solar irradiance datasets were attained from the Australian Bureau of Meteorology (Station Number: #086282). The specifications of these datasets are described below:

- Automatic Weather Station Data:
    - Source: Australian Bureau of Meteorology
    - Location: Melbourne Airport
    - Year Range: 1 January 1990 – 8 September 2015
    - Interval: Half-hourly
- One Minute Solar Data:
    - Source: Australian Bureau of Meteorology
    - Location: Melbourne Airport
    - Year Range: 1 January 1999 – 31 November 2017
    - Interval: One-minute

DATA PREPROCESSING

Using both weather and solar datasets, a series of data pre-processing steps were employed to construct a continuous climate dataset that spans from 1st January 2005 to 31st August 2015 at an hourly resolution.

- <u>DATA FILLING</u>

Below briefly describes of the various data filling methods used to fill in missing data. For a more detailed description of how these data filling methods were implemented for the weather, solar, and climate datasets, please refer to 'weather_data_preprocess.py', 'solar_data_preprocess.py', and 'combined_datasets.py' scripts, respectively

- → Data Filling Methods
    - Linear Interpolation: Fill in missing observations by the computed average of its neighbouring non-missing observations.
    - Proxy Estimation: Compute the missing observations using numerical models
    - Daily Average Profiling: Fill in missing observations by the computed average of observations taken at the same hour of the previous and subsequent day
    - Climatological Estimation: Fill in the missing observations by the computed average of observations taken at the same time across all other years in the dataset

- <u>DATA TRANSFORMATION</u>

As suggested by Dr. Brownlee, the time series dataset should be transformed as follows [1]:

- (1) Transform the time series data so that it is stationary
    - Time series data may inherit temporal structures (i.e. trends and seasonality) and may not comply with certain models due to varying mean and variance in the dataset. In this study, a lag difference approach is taken that subtracts the current dataset by its previous set of observations. The previous set of observation used for lag differences is then omitted from the current dataset set for further model training and analysis.
- (2) Transform the time series data into a supervised learning problem
    - Time series data can be reconstructed as a supervised learning dataset by assuming that the dependent variable measured at the prior time interval is a predictor for the dependent variable of the current time. In other words, the predictors for the dependent variable at the forecast interval are the current input features of the dataset and measured dependent variable.
- (3) Transform the observations to have a specific scale
    - It is of best practise to ensure that all features and target variables are on relatively similar scale to avoid overpowering of values at different orders of magnitude during computation and allow for fast convergence for gradient descent. For time series data, the data is scaled within [-1,1] range [2].

- <u>FEATURE SELECTION</u>

A backward elimination algorithm was employed to remove features of negligible significance towards predicting the target variable. Any features that attains a p-value greater than defined significance level of 0.05 were removed from the dataset. It was identified that wind-related features (i.e. wind direction and wind speed) and cloud heights were insignificant predictors in the predicted solar irradiance. Below present the climate dataset's features after feature selection.

- → Climate Dataset (Total number of observations: 93481)
    - Input Features:
        - Dry Bulb Temperature (degrees C)
        - Wet Bulb Temperature (degrees C)
        - Dew Point Temperature (degrees C)
        - Relative Humidity (%)
        - Vapour Pressure (hPa)
        - Saturated Vapour Pressure (hPa)
        - Cloud Amount (of first group) in eighths
        - Cloud Amount (of second group) in eighths
        - Cloud Amount (of third group) in eighths
        - Cloud Amount (of fourth group) in eighths
        - Station Level Pressure (hPa)
        - Mean Global Horizontal Irradiance of prior hour (W/m-sq)
    - Dependent Variable:
        - Mean Global Horizontal Irradiance of current hour (W/m-sq)

MODEL SELECTION

- METHODOLOGY

Traditionally, datasets are randomly shuffled before being divided into training, cross-validation, and test sets. Both training and cross-validation sets are used for model training, k-fold cross validation, and model tuning whilst the test set are used for evaluating the model performance on unseen data. For time series datasets, the order of observations must be retained due to time dependencies, thus, requiring an alternative method for splitting the dataset.

A 'multiple train-test splits' method is taken whereby multiple models are trained on various training set sizes whilst evaluating the model performance on a fixed test size during cross-validation [3]. The average model performances is computed and used as a guideline for further optimization or considerations of other models. Generally, this method is a more robust approach in estimating the expected model performance at the cost of additional computational expense.

Throughout the analysis, both stationary and non-stationary transformed climate dataset (i.e. supervised learning dataset with applied feature scaling) are examined to determine whether temporal structures are influential on the models' predictive performance. Observations of the climate dataset up to 31 December 2014 are used for multiple-train test splits and model selection whilst the remaining dataset are used for model performance testings on unseen data.

- BASELINE MODELLING

A 'persistence' model was used as an initial baseline that assumes the predicted solar irradiance at time 't' is the same as the observed solar irradiance at prior time 't-1'. Using this model, the root mean squared error (RMSE) was computed with a value of 98.23 W/m-sq and 100.74 W/m-sq for stationary and non-stationary climate dataset, respectively. By using a multilinear regression (MLR) model, a better baseline performance was achieved with an RMSE value of 9.707 and 9.764 W/m-sq for the stationary and non-stationary climate dataset, respectively.

Clearly, the MLR model was a good starting point to improve upon before attempting to consider other models. A variety of temporal features were incorporated into the dataset in hopes of improving its forecast performance. These features include: (i) type of season, (ii) hour of day, (iii) month of year, and (iv) 'day or night' time, all of which are one hot encoded.

For the stationary dataset, these added features had marginally improved the MLR model's performances. It was suspected that the removal of temporal structures in the stationary dataset had removed the model's ability to distinguish any temporal effects on the predicted output. Conversely for the non-stationary dataset where temporal structure exists, the added features had improved the MLR model performance. While most temporal features had slightly enhanced the model's performance, the 'hour of day' feature stood out to be the most impactful. With the inclusion of all temporal features, the RMSE value had reduced to 8.457 W/m-sq.

Refer to Appendix A that presents the performance metric results of the MLR model's performance that was fitted to a non-stationary climate dataset with and without temporal features.

- <u>MODEL IMPLEMENTATION</u>

Aside from the MLR model, the support vector regression (SVR), multi-layer perceptron (MLP), and long short-term memory (LSTM) models were considered. These choices were based on several literature papers that focused on evaluating solar irradiance forecast performance using machine learning models. Although LSTM models have not been widely used, this study seeks to evaluate its forecast performance as it is a recurrent neural network specialised to 'memorise' long term patterns. It is noted to the readers that these models were not developed from scratch, but rather, were constructed using 'sklearn' and 'keras' imported packages for Python which provided the basic building blocks in the models' construction.

- <u>MODEL SELECTION</u>

Interestingly, all other models performed slightly worse than the MLR model when trained on the stationary climate dataset. The only comparable models with minimal reduced performance were both SVR and LSTM models which differed in RMSE values by 0.07 and 0.06 W/m-sq, respectively. Had these models been tuned, perhaps such improvements may be evident but may not be justified by its noticeable computational expenses.

As for the non-stationary climate dataset, all other models achieved slight improvements in forecast performance. With an initial configuration of 50 neurons in a single hidden layer with 'ReLU' activation function and 'ADAM' optimiser, the MLP model attained an RMSE value of 8.359 W/m-sq. When considering for a wider and deeper MLP network (i.e. single hidden layer with 100 units; double hidden layer with 50 units each), the RMSE value reduced to 8.329 and 8.252 W/m-sq, respectively. Despite this marginal drop in RMSE, this implies that there are more room for potential performance gains via model tuning and network design optimization.

The LSTM model was able to perform better than the MLP model with an RMSE value of 8.173 W/m-sq. The model configuration consisted of 50 memory units and using the 'tanh' activation function and 'adam' optimizer. Interestingly, the performance had worsened when the memory capacity was increased to 100 units which implies that there exist an optimal number of memory units below 100. Due to time constraints, alternative configurations or model tuning were not conducted.

The SVR model attained comparable performance to the LSTM model with an RMSE value of 8.165 W/m-sq. Unfortunately, due to time constraints and the author's limited knowledge on SVR models, none of the parameters had been tuned. Surprisingly, the computational expense was significantly less than the LSTM model which raises potential avenues for the SVR model to achieve better, if not, comparable performance to the LSTM with reduced computational time.

Overall, models trained on the non-stationary climate dataset were shown to outperform the models trained on the stationary climate dataset. It was suspected that the inclusion of temporal features in the dataset had enabled the model to better distinguish the trends and seasonality in the form of numerical indicators. The following models for evaluating its performance on unseen data are the MLR, SVR, and LSTM models. Note that the model parameter and configurations are retained and that these selected models would had been different if extensive model tuning were employed.

Refer to Appendix B that presents the performance metric results of the all models' performances that was trained on both stationary and non-stationary climate dataset with the inclusion of temporal features.

MODEL ANALYSIS

The selected models for analysis (i.e. MLR, SVM, and LSTM models) were trained on the non-stationary climate dataset for observations up to 31 December 2014 whilst the final performance test was conducted on the remaining observations from 1 January 2015 to 31 August 2015.

Interestingly, the RMSE values attained by the selected models in predicting on unseen data were much lower than that from cross-validation. This might be due to the significant size difference between the training-test split. Nevertheless, the obtained RMSE values were 8.268, 7.957, and 7.785 W/m-sq by the MLR, SVR, and LSTM model, respectively. Aside from the LSTM model, both MLR and SVR models performed poorly in predicting the absence of solar irradiance (i.e. night time) throughout the timeframe of the test set. This itself is suspected to contribute heavily towards a higher RMSE value. With the temporary exclusion of night time hours in the performance metric evaluation, the errors in the forecast performance during daytime hours became more pronounced. While the MLR model still retained poor forecast performance with RMSE value of 9.509 W/m-sq, both SVR and LSTM model were slightly better despite suffering from higher RMSE of 9.016 and 8.979 W/m-sq, respectively.

As shown in Figure 1, all models struggled to forecast hourly solar irradiance within reasonable accuracy to the observed solar irradiance, especially during partially cloudy where unexpected real-time fluctuations of solar irradiance are difficult to capture by the model in advance, and often times may over-predict on overcast days as well as under-predict during clear-sky conditions.

Overall, the LSTM model was determined to be the best model in terms of accurate forecasting of hourly solar irradiance at Melbourne Airport, Australia. However, a combination of model tuning, network configuration, and perhaps the inclusion of more features should be emphasized to potentially elevate its forecast performance.

Refer to Appendix C that presents the performance metric results of the all models' performances that was trained on the non-stationary climate dataset with the inclusion of temporal features.

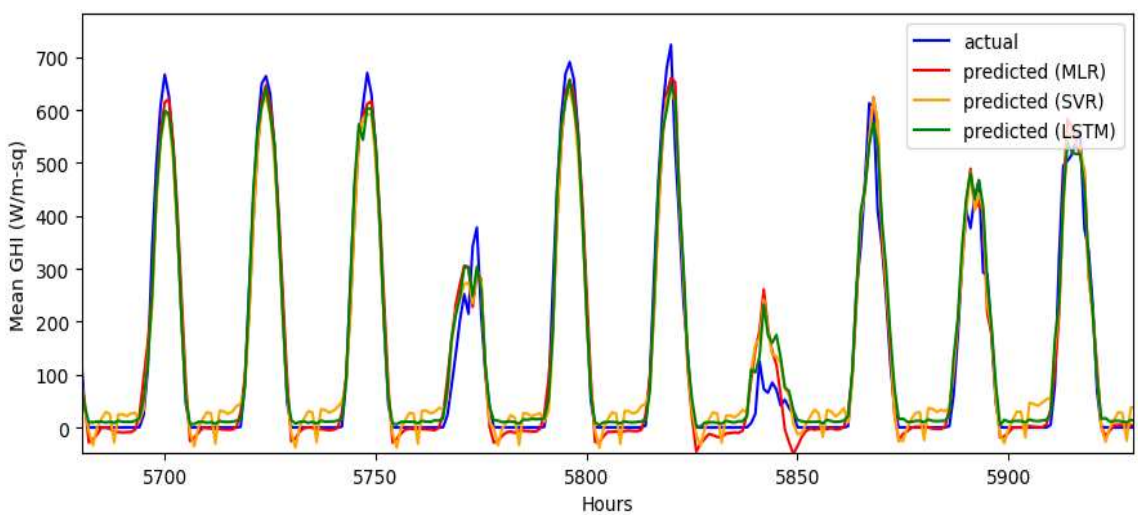

Figure 1: Actual and predicted hourly solar irradiance over the middle of July 2015

CONCLUSION

Within this study, a comparative model analysis was conducted to determine the most accurate model in forecasting hourly solar irradiance using historical ground-measured weather and solar data. A variety of models were initially considered during cross-validation which includes multilinear regression (MLR), support vector regression (SVR), multi-layer perceptron (MLP), and long short-term memory (LSTM) model.

With the incorporation of temporal features, the retainment of temporal structure in the climate dataset (i.e. non-stationary dataset) had enabled these models to achieve higher forecast performance. Although the LSTM model was revealed to be the best predictive model in terms of accuracy, further improvements in forecast performance have yet to be made by placing more emphasis in model tuning, design configuration, and extensive feature selection.

Despite the relative ease of implementation and usage of MLR models, the machine learning models were able to achieve higher forecast accuracy as shown by its relatively low RMSE. Since forecast accuracy are highly crucial, machine learning models have a significant advantage over MLR models and should be considered in not just for solar irradiance forecast but for other forecast applications.

REFERENCES

[1]: https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-network-python/

[2]: https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-network-python/

[3]: https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/

APPENDIX A: Performance metric results of MLR model fitted to the non-stationary climate dataset with and without temporal features

	Case #1	Case #2	Case #3	Case #4	Case #5
MSE (W/m-sq)	95.35	94.429	92.519	71.701	71.528
RMSE (W/m-sq)	9.764	9.717	9.618	8.467	8.457
MAE (W/m-sq)	64.482	64.13	62.393	44.03	44.716
MBE (W/m-sq)	-0.722	-0.645	-0.655	-0.109	-0.076
R^2	0.872	0.874	0.879	0.927	0.928
Adjusted R^2	0.872	0.874	0.88	0.928	0.928

Note #1: Performance metrics include: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean bias error (MBE), R-squared, and adjusted R-squared.

Note #2:

- Case #1: MLR model fitted to climate dataset with no temporal features
- Case #2: MLR model fitted to climate dataset with ‘type of season’ features
- Case #3: Case #2 with ‘day/night time’ features
- Case #4: Case #3 with ‘hour of day’ features
- Case #5: Case #4 with ‘month of year’ features (i.e. all temporal features)

APPENDIX B: Performance metric results of all models fitted to the non-stationary and stationary climate dataset with temporal features

	MLR	SVR	MLP #1	MLP #2	MLP #3	LSTM #1	LSTM #2
MSE (W/m-sq)	71.528	66.676	69.996	69.436	68.156	66.825	68.341
RMSE (W/m-sq)	8.457	8.165	8.359	8.329	8.252	8.173	8.258
MAE (W/m-sq)	44.716	42.523	43.192	42.406	41.181	37.082	38.942
MBE (W/m-sq)	-0.076	-6.006	5.331	-0.308	-2.778	2.327	5.263
R^2	0.928	0.937	0.930	0.932	0.935	0.937	0.934

Note:

- MLR: Multilinear regression model
- SVR: Support vector regression model
- MLP #1: Multi-layer perceptron model
    - Single hidden layer, 50 hidden units, 50 epochs, 72 batch sizes
    - 'ReLU' activation, 'ADAM' optimizer
- MLP #2: Multi-layer perceptron model
    - Single hidden layer, 100 hidden units, 50 epochs, 72 batch sizes
    - 'ReLU' activation, 'ADAM' optimizer
- MLP #3: Multi-layer perceptron model
    - Two hidden layers, 50 hidden units each, 50 epochs, 72 batch sizes
    - 'ReLU' activation, 'ADAM' optimizer
- LSTM #1: Long short-term memory model
    - Single layer, 50 memory units, 50 epochs, 72 batch sizes
- LSTM #2: Long short-term memory model
    - Single layer, 100 memory units, 50 epochs, 72 batch sizes

APPENDIX C: Performance metric results of all models fitted to non-stationary climate dataset with temporal features

	MLR	SVR	LSTM
MSE (W/m-sq)	68.361	63.308	60.603
RMSE (W/m-sq)	8.268	7.957	7.785
MAE (W/m-sq)	43.284	42.146	33.369
MBE (W/m-sq)	-0.414	-7.269	-2.529
R^2	0.928	0.938	0.944

Note:

- MLR: Multilinear regression model
- SVR: Support vector regression model
- LSTM: Long short-term memory model
    - Single layer, 50 memory units, 50 epochs, 72 batch sizes