Interim Report: Short-term Photovoltaic Power Forecasting using an LSTM Neural Network

*Jasmine M. Goode**

University of New South Wales Canberra at the Australian Defence Force Academy

A variety of factors have led to the role of renewable energy becoming more prevalent within the grid. Solar is one of the most implementable and consequently most common sustainable power sources. A major issue that comes with the increase of renewable energy is its time and weather dependency; particularly that solar power is generated during the day, while peak domestic power usage occurs at night. Renewable energy generation tends to have significant fluctuation compared to traditional energy production means such as coal-power. A potential solution to this is power storage. Photovoltaic (PV) output prediction combined with time-domain modelling for domestic and industrial power use, and sophisticated energy storage could provide the solution to the supply and demand issue. This project aims to explore a variety of data inputs, pre-processing techniques and forecasting models and determine which combination results in the best performance. A Long Short Term Memory (LSTM) Neural Network (NN) is found to be an optimal model for Short Term PV Forecasting (STPF).

Contents

*PLTOFF, School of Engineering and Information Technology. ZEIT4500

Nomenclature

PV	Photovoltaic
$LSTM$	Long Short Term Memory
NN	Neural Network
$STPF$	Short Term Power Forecasting
API	Application Programming Interface
ML	Machine Learning

I. Introduction

A. Solar Power and the Grid

The climate crisis, energy security concerns[9], the limited supply of fossil fuels, and critical fuel supply chain pinch points are a few of the many factors leading to the rise of renewable energy[11] within the grid and as a source for off-grid generation. As the supply of fossil fuel begins to diminish, the requirement for power traditionally supplied via these means increases. It is expected that the current population will double by 2050, resulting in a proportional increase in electricity demands[19]. In terms of renewable sources, solar and wind presently appear to be the best candidates to fulfil current and future energy requirements; though significant research and development would be required to meet the requisite scale. Solar power generation is more implementable than wind power generation due to the significant size and cost of infrastructure associated with wind turbines. Additionally, solar PV power generation can be more consistent than wind power generation (depending on geographic location) and solar cells can often be placed in closer proximity to residential areas, reducing transmission losses.

One of the main issues arising from the reliance on solar energy is the temporal mismatch between energy supply (generation) and demand (use). Typically, domestic power use peaks during the early-mid morning period and evening; when solar energy production is at a minimum. An increase in residential power use when solar energy is not being produced (overnight) is expected to occur as the popularity of Electric Vehicles (EVs) grow. According to [23] the individual integration of PV systems and EVs can lead to grid instabilities. It is also noted that coordinated operation of PV systems and EVs could eliminate the issues caused by their integration into the grid. Understanding the patterns of load requirements due to EV charging, and forecasting PV outputs can allow the coordinated operation required for a stable and effective grid.

B. Forecasting

Forecasting is essential to ensuring adequate power supply due to renewable energy supply fluctuation [19]. Aligning expected demand with forecasted supply minimises losses and shortages during peak demand by allowing the effective optimisation of power storage devices.

PV cell output is a function the module dimensions, material, magnitude of incident radiation and time exposed to radiation [12, 11]. Dimensions and material are reliant on the design of the panel while radiation depends on geographical location and weather. Consequently. areas with strong solar irradiance provide an ideal environment for solar power generation. As solar irradiance is such a critical factor determining PV power output, a reliable prediction of solar irradiance is required for accurate PV power forecasting. Though solar irradiance is the strongest indicator of PV output, additional weather data can provide further insight. Ambient temperature, humidity, pressure, wind speed, and precipitation also bear a linear relationship to PV output with varying influence [28]. Weather forecasting has existed for an extended period of time and continues to become more sophisticated and accurate. The weather forecast data has a critical influence on PV output prediction and sources used in this project are considered to be reliable.

Though many forecasting models assume a linear relationship between influencing factors and PV output, this is inaccurate for PV module temperature. A study on the impact of PV module temperature on output finds that module temperature effects PV power output $20\% - 30\%$ more than ambient air temperature [15]. It was additionally found in [15] that above a certain temperature limit a reduction $0.35\% - 0.45\%$ in module efficiency is observed for every $1°C$ increase in module temperature. Module temperature tends to increase with ambient temperature, ambient temperature increase is correlated with increased PV output, and an increase in module temperature results in decreased module efficiency. Module surface cleanliness is another influence of PV power generation that is not easily modeled or accounted for. Factors such as these introduce a significant non-linearity to the system that needs to be accounted for in PV power forecasting.

It is essential that any PV power forecasting model accounts for conversion of solar energy into usable electricity [19]. If this conversion is not accounted for, the forecast received by the grid will not be accurate and could impact grid supply-demand management. The efficiency of module energy conversion is affected by a variety of factors that are relatively steady over time compared with the fluctuation of environmental factors. The exception to this is the partial shading of solar modules which results in a significant decrease in PV power output [19]. This shading may ~~predicted~~ by weather forecasting in the case of cloud coverage or may be caused by an object obstruction incident solar radiation unexpectedly.

Broad categories of forecasting methods utilised in PV power forecasting include time series forecasting, statistical modelling, stochastic modelling and predictive Machine Learning (ML). Each of these categories include unique methods and a cost-benefit analysis is required to determine the appropriate forecasting method for the application.

II. Aim

The aim of this project is to produce a system that processes data to forecast PV power output. An extensive analysis of forecasting methods will result in the election of a high performing and implementable predictive model. This project will investigate new combinations of input data and processing algorithms, and assess their affect on short term forecasting performance. Additionally, this paper will utilise new training data, resulting in outputs unique to the current field of research.

III. Scope

There are two main components to the project. The first is obtaining and processing the input data. Application Programming Interface (API) connections are used to obtain forecasted solar irradiance and weather data to be used for solar forecasting. Future horizon PV output data is to be utilised as bench-marking data. A variety of data processing techniques are examined to ensure that data is prepared for forecasting and optimal performance can be achieved.

The second component of the project is the production of predictive outputs. A variety of forecasting methods including are explored. The chosen forecasting method, a LSTM NN is implemented, trained and tested with combinations of of inputs and pre-processing algorithms to produce an accurate short term PV power forecast.

IV. Literature Review

A. PV Power Output

PV power forecasting can be grouped into four time frames which each have real-world applications. Ultra short term systems forecast five minutes [27] to half an hour into the future. Ultra short term and short term forecasting can be used to monitor the health of PV systems and to optimise economic dispatch and managing grid supply and demand[3]. Short term PV forecasting can be defined as extending from half an hour to 24 hours into the future[9]. Medium-range PV power forecasting, up to 72 h into the future is primarily used as a planning reference for power plants. Long term PV forecasting, often predicts PV outputs weeks in advance. It is significantly more difficult than shorter term predictions and requires a broader range of data inputs. Long term forecast can be used make strategic decisions such as infrastructure planning and electricity supply determinations. This paper will focus on the implementation of short term PV power forecasting.

B. Time Series Forecasting and Stochastic Processes

Time series forecasting models tend to rely on the linear relationship between the forecasted output and another time series data set[10]. Though this may be an adequate assumption for some environmental factors affecting PV power forecasting, it is inaccurate for other factors. Non-linear regression models can be fitted to a wider range of data sets. Non-linear regression models can better describe relationships by increasing the order of the fitting polynomial (up to cubic or quadratic), modelling exponentially behaviour or describing behaviour with a piece-wise function[10]. Commonly used approaches for time series forecasting, the exponential smoothing and Auto-Regressive Integrated Moving Average (ARIMA), Grey-Markov model, and Kalman filter will be investigated for STPF.

<u>Exponential Smoothing</u>
Forecasting models using exponential smoothing operate on the principle that more recent observations will be better indicators of future behaviour[10]. Forecasting often includes single or double exponential smoothing. The performance of double smoothing applied to a PV system in [4] finds that the approach improved forecasting by 3% compared with the literature benchmark methods.

<u>ARIMA</u>
An auto-regression model forecasts using a linear combination of past values[10]. A moving average model uses a similar premise with past forecast errors. The ARIMA model is a combination of the auto-regression and moving average model, with an integrated term for differencing[22]. It was found in [19] that ARIMA model ~~provide~~ accurate PV power forecasts for ultra short to short forecasting ranges (minutes to tens of minute) and are less accurate for longer range forecasts.

<u>Grey-Markov</u>
The grey system is defined as when a system has partially known information, opposed to completely known information in a white system and completely unknown information in a black box system[2]. The widely used GM(1,1) model utilises a monotonically increasing sequence, differential equation and the Ordinary Least Squares (OLS) method to forecast values. A Markov chain is a stochastic model that predicts probability based on a previous state. The Grey-Markov chain produces a forecast based on only the most recent previous state[14]. Findings by [14] find that Grey-Markov chain forecasting is highly accurate for STPF.

<u>Kalman Filter</u>
Kalman filters utilise a prediction engine to forecast a Gaussian distributed result. This result is combined with a measurement result from sensor data, also of Gaussian distribution, to produce an adjusted predic-

tion. The mean and covariance of this adjusted prediction ~~is~~ used as feedback into the prediction engine, allowing the Kalman filter to adapt to a dynamic system [13]. Multiple approaches using Kalman filtering are investigated in [13], finding that the proposed approach is able to reliably forecast temporal and permanent drops in PV outputs due to faults.

C. Machine Learning

Artificial Intelligence (AI) is the process of problem solving utilising characteristics of human cognition [16]. ML is a subset of AI and an exceptionally broad field defined by the ability to learn without explicit programming. ML techniques allow for complex problems to be solved beyond the realm of traditional computation. Difficult and monotonous tasks are being increasingly taken on by ML algorithms that can be more efficient and sometimes identify patterns that may not initially be obvious to humans. Identification and prediction are two of the main tasks performed by ML algorithms. The two main forms of ML are supervised and unsupervised learning[5].

1. Unsupervised Learning

Unsupervised machine learning is commonly referred to as clustering [7]. Supervised and unsupervised machine learning are differentiated by the training data. Clustering has no capability of cross-validating outputs with expected results. This form of machine learning is adaptable to a range of data types and its implementation should be based on the features of the specific data set as no effective broadly applied methods have been established [7].

Predictive Clustering

Clustering may only be used to predict future behavior or outcomes for a particular clusters, hence it is an inadequate model for PV power forecasting [1].

2. Supervised Learning

The addition of training data characterises supervised machine learning [7]. The most widely known example of supervised learning is the artificial neural network, though there are many other algorithms within the field that perform similar functions.

Decision Tree Learning

Decision trees use rule-based predictive learning, where the model will start at the 'root' and branch out [1]. Outcomes are represented as leaves or nodes and branches are representations of observations. The M5' decision tree model in [30] was found to produce reliable results for day-ahead forecasting.

Support Vector Machine Algorithm

The Support Vector Machine (SVM) algorithm separates data via a hyper plane. Data separation acts as a form of classification which is then used by the model to forecast predictions. In [24] the SVM algorithm is found to produce accurate STPFs for some weather conditions and unreliable for other weather conditions. It was also established that for an integrated model combining the SVM with a statistical model, short term forecasting accuracy is improved.

D. Neural Networks

Artificial Neural Networks (ANNs) are based of the structure of neurons in the human brain [25]. Typically, ANNs will contain three layers, the input layer hidden neural layer and output layer. ANNs utilise these neurons and the weights between their connections and backward propagation (BP) to form predictions. BP is the process of adjusting the neuron weights based on the difference between expected output and predicted output (error). Perceptrons, feed-forward networks, multi-layer perceptrons, and Convolutional Neural Networks (CNNs) are most effective at classifying problems; while Radial Basis Networks (RBNs), Recurrent Neural Networks (RNNs) and LSTM NNs can be accurate time series forecasting tools.

1. Radial Basis Network

RBNs are characterised by their activation function, the radial function and its unique storage of classes [25]. A study by [29] on the RBN implementations of varying complexities found that RBN forecasting was relatively accurate for STPF.

2. Recurrent Neural Network

Recurrent Neural Networks (RNNs) have a time-delayed input and associated memory that distinguishes then from other architectures [25]. Though this architecture produces accurate PV power forecasts in [9], it is outperformed by the LSTM NN.

3. Long-Short Term Memory Neural Network

LSTM NNs extend the principle of memory utilised in RNNs by adding a memory cells for long term storage [25]. The added complexity of LSTM NN is often worth the improved performance due to the memory cell overcoming the vanishing gradient shortfall of RNNs. The classical architecture of an LSTM NN in Figure 1 is not significantly more complex than the structure of a generalised NN. The connection between input layer and hidden layer, hidden layer and output layer and within the hidden layers allow an forecasting method based on the principle of BP.

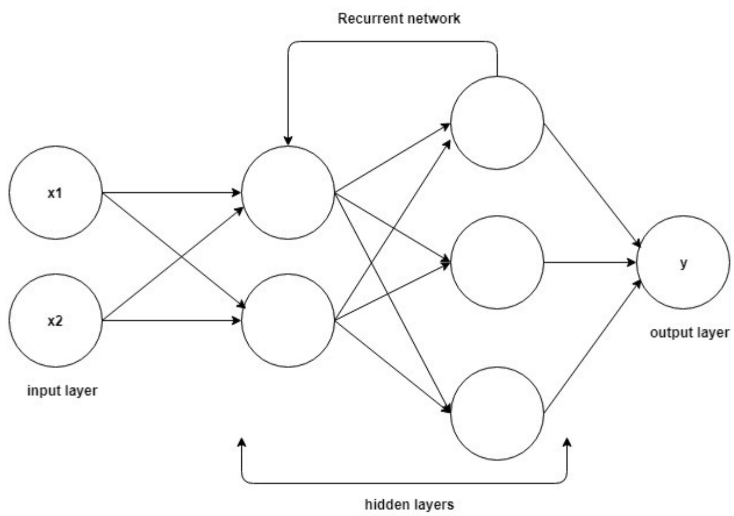

Figure 1. Generalised Architecture of LSTM NN from Medium [17]

LSTM NNs are an exceptionally common model for STPF, which can be attributed to their memory characteristics. The LSTM used in [9] is found to produce a highly accurate forecast, outperforming the RNN, generalised RNN, and extreme learning machine. The superior forecasting accuracy of the LSTM NN is reflected in [18], though it is also noted that the execution time required for the LSTM NN was long.

4. Hybrid Models

Two or more forecasting methods may be used in conjunction to improve forecasting accuracy. Hybrid model of the same forecasting method category may be used, though combining vastly different methods tends to yield better performance. Hybrid model forecasting can be simple and implementable, depending on the involved forecasting methods and expected performance output.

Combined Grey Model and Back-Propagation Neural Network

A study conducted by [26] implemented a hybrid Grey-BP NN forecasting model and analysed its performance compared to the individual Grey forecasting method. The time series Grey model outlined above produces a prediction which is used to train the BP NN. It was found that the implementation of the hybrid model markedly reduces the forecasting error, resulting in an efficient model that is fairly accurate for STPF.

E. Processing Algorithms

Data pre-processing prior to forecasting has a major impact on performance and efficiency [8]. Data cleaning, transformation, and reduction are often employed to improve the quality of data [6] used as inputs to forecasting models.

1. Data Cleaning

Data cleaning involves performing processes that ensure data is clear and consistent, improving the accuracy of forecasts. Identifying and removing outlier data, interpolating missing data and managing noisy data is often performed in the data cleaning stage of pre-processing [6]. For the purpose of PV forecasting removing outlier data will be imperative to ensuring accuracy, particularly if clustering is also a step in data prepossessing. Noisy data may be addressed by binning (for sorted data), linear regression, and clustering for grouping.

2. Data Transformation

Normalisation, attribute selection, discretisation, feature selection, and concept hierarchy generation are data transformation methods that may be used to ensure that the format of the data is appropriate for forecasting[6]. Data transformation methods most applicable to PV power forecasting are normalisation and feature selection [20]. Data normalisation ensures accurate scaling [6] where input variables have different units. Feature engineering extracts key data features improving training data quality and or some cases improving forecasting stability [20].

3. Data Reduction

Reducing the features of input data for forecasting models decreases the computational cost of the forecast and reduces over-fitting due to the weight of irrelevant data [6]. Dimensionality and numerosity reduction methods decrease the number of features used and reduce data size respectively. The data transformation method of feature engineering can also be applied to reduce data size, improving computational efficiency in addition to the benefits outlined above [20]. Due to the requirement for multiple inputs to establish accurate PV power forecasts, data reduction is a particularly useful pre-processing technique.

F. Input Data

PV systems have a heavy reliance on weather data, hence the range of data and how it is processed contributes significantly to the accuracy of the forecast [31]. Logically, the more data that can be provided to a model, the more accurate the model will be as there is more information to inform the prediction. Though an increase in the amount of data and data diversification can lead to improved forecasts accuracy, this is not true for all forecast models [9]. Some forecasting models can only utilise single data types, and an excess of data often leads diminishing returns at a certain threshold for forecasting models that can handle multiple inputs and large data inputs. Consequently, the choice of data used for forecasting is a parameter that will have a great affect on forecasting performance.

Data selection to produce an accurate forecasting system can be an iterative. After data has been selected it is processed and used in forecasting. Through pre-processing or once the forecast has been made it may be established that data is inadequate in some capacity. If this is the case, input data should be reassessed.

A linear relationship between inputs and forecasted outputs is often assumed in forecasting models. An assessment of data types based on correlation is a common method. The correlation between the input data and forecast output is a measure of the extent to which the input informs the output, though the effect is not always causal. The Pearson correlation coefficient is a particularly common measurement relating input and and output, used by [9] and [20]. The relationship between PV power generation and solar irradiance was found to be the strongest of the inputs investigated according to the correlation coefficient in [9]. It was also found in this study that temperature, relative humidity and wind speed are strong indicators of PV power output; while pressure and precipitate water are not strongly correlated. It is similarly found by [20] that solar irradiance is the strongest indicator, though it is followed by humidity, temperature then wind speed. Both sources agree that pressure and rainfall are not good predictors of PV power outputs.

V. Approach

Figure 2 illustrates the proposed system architecture modelled off the approach used in [28]. Weather forecast data may include temperature, wind, humidity, precipitation and pressure. This forecast data and pre-processed solar irradiance data form the synthetic weather predictors. Historical weather time series data which may also include temperature, wind, humidity, precipitation and pressure. Inputs to the NN are grouped into synthetic weather predictors, historical PV power and historical weather predictors. The diversity of input into the NN is projected to improve the performance of the NN by allowing the NN to make correlations that may not be expected.

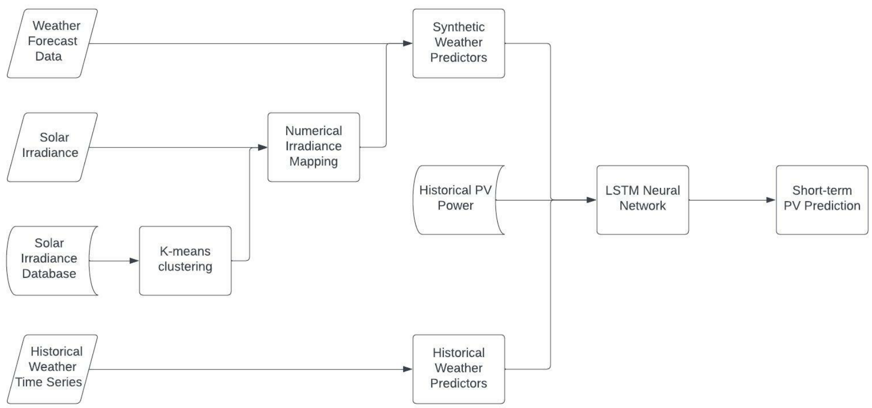

Figure 2. Proposed System Architecture

The project approach ~~consisted of~~ four major components:

- obtaining the input data
- utilising processing algorithms to prepare data
- developing the NN
- training the NN

Prior to the outlined approach being undertaken, the literature review was completed. This research evaluated potential data inputs, pre-processing algorithms and forecasting models. The analysis found that a range of data inputs would be tested with priority placed on weather inputs that have strong correlation with PV output. The primary pre-processing algorithm would be K-means clustering due to its simple implementation, ability to process large data sets and adaptability to new data. This algorithm is intended to be used for solar irradiance clustering though will also be tested with other input data to determine which combination provides the most accurate PV power prediction. Statistical and machine learning methods were explored for PV forecasting and it was determined that supervised machine learning would likely provide the most accurate forecast. More specifically, a LSTM NN would be most appropriate for due to its accuracy, ability to handle a broad range of data inputs, and ability to conduct continuous forecasts, essential to the short term PV power output application.

Synthetic predictor input data for the geographic location of Mount Majura Solar Farm in Canberra will be used for this project. The forecasted PV power output from the NN will be compared to the forecasts produced by *Forcast.Solar* and *Solcast* for this location. Additionally, inquiries will be made to the Mount Majura Solar Farm to determine if measured data may be used to compare to the forecasts produced by the LSTM NN.

A. Input Data

Data to be used in the project was obtained from various web sources via APIs. Two sources have been used to obtain data to date. Current solar irradiance data has been acquired through *Solcast*. This data is to be processed by the K-means clustering algorithm to determine a sky type. *Forcast.Solar* is the source for current solar cell output and horizon data; solar cell output forecasting three days into the future. The *Forcast.Solar* forecast output is the bench-marking data for the NN implemented by this project. The historical solar cell output data is compared to the NN output to inform and train the NN. Additional potential data inputs including historical solar irradiance, temperature, wind and presage biology measurements are noted in the future work section of this report.

B. Processing Algorithms

Presently K-means clustering is the only processing algorithm that has been tested. The K-means clustering algorithm sorts the data into groups depending on characteristics of the data [7]. For solar irradiance data this results in sky types being identified and grouped prior to NN input. Clustering is predicted to improve the computational efficiency of the NN significantly.

K-means clustering processing is to be performed on additional data types in future work. Various combinations of these processing algorithms are to be tested with the NN implementation and are predicted to further improve performance. Clustering is immensely common in PV power forecasting to classify weather types, used in [9] and [20]. K-means is the most popular clustering application for solar irradiance data as irradiance is such a strong indicator of PV output and K-means clustering is an efficient method of classifying sky types. Classifications of weather types often include sunny, cloudy, partially cloudy, and rainy [20].

K-means clustering on a the range of data inputs used for the LSTM NN will be performed and performance tested. Prior to k-means clustering processing, the minimum covariant determinant method will be used to remove outliers, improving the performance of the clustering algorithm and presumably the accuracy of the LSTM forecast. Principle Component Analysis (PCA) [21] may also be explored as a means of reducing dimensionality, improving the computational efficiency of the forecast.

INSERT EQUATIONS

C. Neural Network Architecture

The architecture of the NN will be based on LSTM NN architectures found to have a high level of performance for STPF [9]. It is presently expected that the LSTM NN will have two layers, the first layer performing a memory function and the second layer synchronously updating the NN output.

D. Neural Network Training

There are two distinct tasks related to training the NN:

- Providing feedback to the NN
- Measuring performance

The NN is provided feedback by the historical PV output provided by *Forecast.Solar*. PV forecast bench-marking data for the NN will be provided by *Forecast.Solar* and *Solcast*.

Error equations:
- MAPE
- RMSE
- NRMSE

Compare to other forecasts:
- Forecast.Solar
- solcast

Compare to real world measurements:
- data obtained from Mount Majura Solar Farm

VI. Results

A. Neural Network Performance

A simple NN was used to develop an understanding of NN architecture and implementation. It demonstrated the relationship between the weightings and bias vectors in the network and the consequent output. Additionally, this NN, in Figure 3 was utilised to explore the impact of the NN learning rate and how this can lead to the data being over-fit or under-fit.

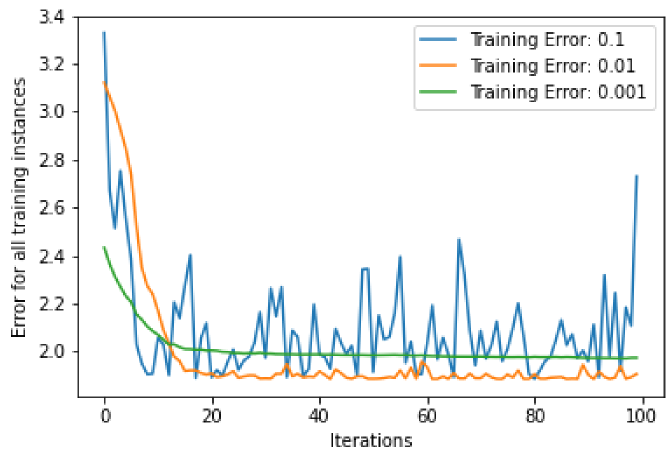

Figure 3. Error Occurrence for a Range of Training Errors

A training error of 0.1 results in a significant fluctuation of errors throughout the duration of iterations, indicating that the data is under-fit. For a training error of 0.001, the frequency of errors over the iteration range is relatively stable and reaches a steady state, indicating that the data is over-fit. The ideal learning rate for this simple NN is 0.01 as it avoids the inaccurate predictions associated with under-fit data and the lack of adaptability to new data that occurs when data is over-fit.

Visual tools such as this plot will be a vital component of gauging NN performance and optimisation once the NN has been established.

B. Neural Network Bench-marking Data

Multiple data sets are to be used to ratify the performance of the NN, as outlined in the approach. Figure 4 contains the PV output forecasted by *Forecast.Solar* for the current day and a three day horizon. The purpose of this project is to conduct STPF so the data forecasted up to 24 hours into the future will be relevant for bench-marking.

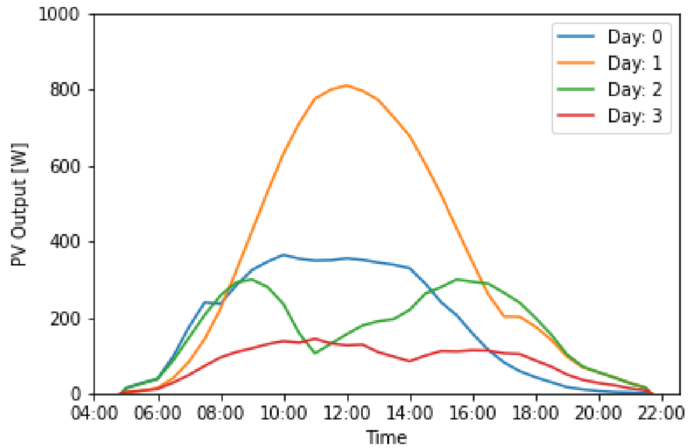

Figure 4. ***Forcast.Solar* PV Output Horizon Forecast**

Bench-marking forecast data will be compared visually through time series and with various statistical measures to evaluate performance.

VII. Future Work

A. Expanding Input Data

Input data is presently limited to current solar irradiance. In the future this input data is to be expanded to include temperature, wind, and historical solar irradiance. It is expected that additional input data will improve accuracy of the NN predictions

Additional, less conventional data inputs may also be explored to determine if a relationship between this data and solar cell output may be established. Presage biology, a study of the link between bio-indicators and weather prediction, may be considered.

B. Optimising Data Processing Algorithms for Neural Network Input

Presently some data processing algorithms have been explored in isolation from one another and the implementation of the neural network. Future work will include combining various processing algorithms and observing the resultant performance of the NN.

C. Training Neural Network

The architecture of the NN has not yet been fully established. Subsequent to architecture establishment and implementation, testing of the NN will be conducted. Testing will consist of alternating data inputs, interchanging data processing algorithms and adjusting NN architecture and comparing predictive outputs to bench-marking data.

Performance will be measured via three methods. A time-series comparison between forecasting output and bench-marking data provides a visual representation of performance compared to established data. Error autocorrelation and cross-correlation provide a measure of error correlation in time and correlation between the input and prediction error respectively.

VIII. Conclusion

PV power forecasting is becoming an increasingly important mechanism in managing the grid as the proportion of renewable energy increases. Short term PV power forecasting and modelling energy use patterns can resolve the time disparity between energy production and utilisation. A range of input data, pre-processing algorithms and short term forecasting methods have been investigated. This report has established that an LSTM NN should produce an accurate STPF; particularly when conscientious data selection and fitting data pre-processing algorithms are implemented.

References

[1] Seth Adler. *Predictive analytics defined, explained and simplified.* May 2019. URL: https://www.intelligentautomation.network/decision-ai/articles/a-basic-guide-to-predictive-analytics.

[2] Mehdi Askari and Hadi Askari. "Time Series Grey System Prediction-based Models". In: *Trends in Applied Research* 6.11 (2011).

[3] Biaowei Chen et al. "Very-short-term power prediction for PV power plants using a simple and effective RCC-LSTM model based on short term multivariate historical datasets". In: *Electronics (Switzerland)* 9.2 (Feb. 2020). ISSN: 20799292. DOI: 10.3390/electronics9020289.

[4] Pasquale De Falco, Luigi Pio Di Noia, and Renato Rizzo. "Exponential Smoothing Model for Photovoltaic Power Forecasting". In: *Proceedings of 2021 9th International Conference on Modern Power Systems, MPS 2021.* Institute of Electrical and Electronics Engineers Inc., June 2021. ISBN: 9781665433815. DOI: 10.1109/MPS52805.2021.9492538.

[5] *Difference between Artificial intelligence and Machine learning.* URL: https://www.javatpoint.com/difference-between-artificial-intelligence-and-machine-learning.

[6] Yulia Gavrilova and Olga Bolgurtseva. *What Is Data Preprocessing in ML?* Sept. 2020. URL: https://serokell.io/blog/data-preprocessing.

[7] Robert Gentleman, Kurt Hornik, and Giovanni Parmigiani. *Bioconductor Case Studies.* New York: Springer Science+Business Media, 2008.

[8] Samadrita Ghosh. *A Comprehensive Guide to Data Preprocessing.* Dec. 2021. URL: https://neptune.ai/blog/data-preprocessing-guide.

[9] Mohammad Safayet Hossain and Hisham Mahmood. "Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast". In: *IEEE Access* 8 (2020), pp. 172524–172533. ISSN: 21693536. DOI: 10.1109/ACCESS.2020.3024901.

[10] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice.* 2nd ed. Melbourne, Australia: OTexts, 2018. URL: https://otexts.com/fpp2/.

[11] Robina Iram et al. "Prioritization of renewable solar energy to prevent energy insecurity: An integrated role". In: *Singapore Economic Review* 66.2 (Mar. 2021), pp. 391–412. ISSN: 02175908. DOI: 10.1142/S021759082043002X. URL: https://web-p-ebscohost-com.wwwproxy1.library.unsw.edu.au/ehost/pdfviewer/pdfviewer?vid=0&sid=1bf2d712-4936-4c8d-995f-6a3cfd5186b5%40redis.

[12] Anne Labouret and Michel Villoz. *Solar Photovoltaic Energy.* Institution of Engineering and Technology (The IET), 2010. URL: https://app-knovel-com.wwwproxy1.library.unsw.edu.au/kn/resources/kpSPE00006/toc?issue_id=kpSPE00006.

[13] Hyun Yong Lee et al. "Kalman Filter-based Adaptive Forecasting of PV Power Output". In: *International Conference on ICT Convergence.* Vol. 2020-October. IEEE Computer Society, Oct. 2020, pp. 1342–1347. ISBN: 9781728167589. DOI: 10.1109/ICTC49870.2020.9289532.

[14] Ying Zi Li, Lin He, and Ru Qing Nie. "Short-term forecast of power generation for grid-connected photovoltaic system based on advanced Grey-Markov chain". In: *2009 International Conference on Energy and Environment Technology, ICEET 2009.* Vol. 2. Beijing, 2009, pp. 275–278. ISBN: 9780769538198. DOI: 10.1109/ICEET.2009.305. URL: https://ieeexplore-ieee-org.wwwproxy1.library.unsw.edu.au/stamp/stamp.jsp?tp=&arnumber=5367173.

[15] Shourui Liu and Mo Dong. "Quantitative research on impact of ambient temperature and module temperature on short-term photovoltaic power forecasting". In: *2016 International Conference on Smart Grid and Clean Energy Technologies, ICSGCE 2016*. Institute of Electrical and Electronics Engineers Inc., Mar. 2017, pp. 262–266. ISBN: 9781467389044. DOI: 10.1109/ICSGCE.2016.7876066. URL: https://ieeexplore-ieee-org.wwwproxy1.library.unsw.edu.au/stamp/stamp.jsp?tp=&arnumber=7876066.

[16] Carolyn Mair et al. "An investigation of machine learning based prediction systems". In: *The Journal of Systems and Software* 53 (1999). URL: http://dec.bmth.ac.uk/ESERG..

[17] Cory Maklin. *LSTM Recurrent Neural Network Keras Example*. June 2019.

[18] S. Hamid Oudjana, A. Hellal, and I. Hadj Mahamed. "Short term photovoltaic power generation forecasting using neural network". In: *2012 11th International Conference on Environment and Electrical Engineering, EEEIC 2012 - Conference Proceedings*. 2012, pp. 706–711. ISBN: 9781457718281. DOI: 10.1109/EEEIC.2012.6221469. URL: https://ieeexplore-ieee-org.wwwproxy1.library.unsw.edu.au/stamp/stamp.jsp?tp=&arnumber=6221469.

[19] Marius Paulescu et al. "Weather Modeling and Forecasting of PV Systems Operation". In: *Green Energy and Technology* 103 (2013). ISSN: 18653537. DOI: 10.1007/978-1-4471-4649-0. URL: https://link-springer-com.wwwproxy1.library.unsw.edu.au/content/pdf/10.1007/978-1-4471-4649-0.pdf.

[20] Quoc-Thang Phan et al. "A Novel Forecasting Model for Solar Power Generation by a Deep Learning Framework with Data Preprocessing and Postprocessing". In: *2022 IEEE/IAS 58th Industrial and Commercial Power Systems Technical Conference (I&CPS)*. IEEE, May 2022, pp. 1–11. ISBN: 978-1-6654-0918-6. DOI: 10.1109/ICPS54075.2022.9773862. URL: https://ieeexplore.ieee.org/document/9773862/.

[21] Bertholddlausen Sabineekrolak-Schwerdt and Matthiassböhmer Editors. *Studies in Classifi cation, Data Analysis, and Knowledge Organization Data Science, Learning by Latent Structures, and Knowledge Discovery*. Tech. rep. London: Springer Heidelberg, 2015. URL: http://www.springer.com/series/1564.

[22] Shalin Shah. *Comparison of Stochastic Forecasting Models*. Tech. rep. URL: https://github.com/shah314/forecasting.

[23] Ahmad Tavakoli et al. "Impacts of grid integration of solar PV and electric vehicle on grid stability, power quality and energy economics: a review". In: *IET Energy Systems Integration* 2.3 (Sept. 2020), pp. 243–260. URL: https://digital-library.theiet.org/content/journals/10.1049/iet-esi.2019.0047.

[24] Lu Tianqui et al. "2019 IEEE PES International Conference on Innovative Smart Grid Technologies Asia (ISGT 2019) : May 21-24, 2019, Chengdu, China". In: *IEEE PES Innovative Smart Grid Technologies Asia*. 2019. ISBN: 9781728135205. URL: https://ieeexplore-ieee-org.wwwproxy1.library.unsw.edu.au/stamp/stamp.jsp?tp=&arnumber=8881582.

[25] Pavan Vadapalli. *7 Types of Neural Networks in Artificial Intelligence Explained*. Dec. 2020. URL: https://www.upgrad.com/blog/types-of-neural-networks/.

[26] Shouxiang Wang et al. "Photovoltaic system power forecasting based on combined grey model and BP neural network". In: *2011 International Conference on Electrical and Control Engineering, ICECE 2011 - Proceedings*. 2011, pp. 4623–4626. ISBN: 9781424481637. DOI: 10.1109/ICECENG.2011.6057634. URL: https://ieeexplore-ieee-org.wwwproxy1.library.unsw.edu.au/stamp/stamp.jsp?tp=&arnumber=6057634.

[27] Weiqing Wang et al. *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*. IEEE, 2019. ISBN: 9781728149301.

[28] Bo Xiao et al. "A Statistical Photovoltaic Power Forecast Model (SPF) based on Historical Power and Weather Data". In: *Conference Record of the IEEE Photovoltaic Specialists Conference*. Institute of Electrical and Electronics Engineers Inc., June 2021, pp. 26–28. ISBN: 9781665419222. DOI: 10.1109/PVSC43889.2021.9518549. URL: https://ieeexplore-ieee-org.wwwproxy1.library.unsw.edu.au/stamp/stamp.jsp?tp=&arnumber=9518549.

[29] Zhenlei Xu et al. “A short-term photovoltaic power forecasting model based on a radial basis function neural network and similar days”. In: *IOP Conference Series: Earth and Environmental Science* 227 (Mar. 2019), p. 22032. DOI: 10.1088/1755-1315/227/2/022032. URL: https://doi.org/10.1088/1755-1315/227/2/022032.

[30] Xiaoning Zhang, Xu Zhang, and Fang Fang. *Power Forecasting of Solar Photovoltaic Power Systems Based on Similar Day and M5' Model Trees.* Tech. rep. Beijing: North China Electric Power University, 2017. URL: https://ieeexplore-ieee-org.wwwproxy1.library.unsw.edu.au/stamp/stamp.jsp?tp=&arnumber=8028828.

[31] You Jing Zhong and Yuan Kang Wu. “Short-term solar power forecasts considering various weather variables”. In: *Proceedings - 2020 International Symposium on Computer, Consumer and Control, IS3C 2020.* Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 432–435. ISBN: 9781728193625. DOI: 10.1109/IS3C50286.2020.00117. URL: https://ieeexplore-ieee-org.wwwproxy1.library.unsw.edu.au/stamp/stamp.jsp?tp=&arnumber=9394178.