**QUEEN'S
UNIVERSITY
BELFAST**

Data Driven Solar Forecasting Model for Northern Ireland

Cowan, K., & Liu, X. A. (2021). Data Driven Solar Forecasting Model for Northern Ireland. In *2021 56th International Universities Power Engineering Conference: Proceedings: Powering Net Zero Emissions, UPEC 2021 - Proceedings* (2021 56th International Universities Power Engineering Conference: Powering Net Zero Emissions, UPEC 2021 - Proceedings). Institute of Electrical and Electronics Engineers Inc..
<https://doi.org/10.1109/UPEC50034.2021.9548211>

Published in:

2021 56th International Universities Power Engineering Conference: Proceedings

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

© 2021 IEEE.

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Data Driven Solar Forecasting Model for Northern Ireland

Kellie Cowan

*School of Electronics, Electrical Engineering
and Computer Science
Queen's University Belfast
Belfast, BT9 5AH, UK
kcowan05@qub.ac.uk*

Xueqin Liu

*School of Electronics, Electrical Engineering
and Computer Science
Queen's University Belfast
Belfast, BT9 5AH, UK
x.liu@qub.ac.uk*

Abstract—Solar forecasting is an increasing problem for power system operators due to its variability from day to day. This inconsistent power generation effects not only load forecasting, but can cause increased network stability problems due to changing system inertia. Ample research has been carried out surrounding wind generation, with several accurate forecast methodologies created and used across the world. However, as the solar industry is only now beginning to take off in Northern Ireland (NI), with its capacity increasing exponentially every year, forecasting methods have not yet been accurately implemented by the System Operator for Northern Ireland (SONI). This report documents the creation of a NI specific photovoltaic (PV) forecasting model, based on meteorological and historic PV generation data, from data collection to testing. It was found that partial least squares out-performed multiple linear regression and therefore was used as the forecast base. A large scale model was created, achieving an average RMSE of 11.9MW and MAE of 7.7MW. The forecast also improved accuracy with the addition of temperature, wind speed and cloud cover data. A small scale forecast was also built and up-scaled to its representative capacity on the NI system. The model was tested using NI System Demand data, achieving positive results.

Index Terms—power, solar forecasting, PV generation

I. INTRODUCTION

The electricity network is evolving. The network designed for power flow from central power stations to the consumer, is now experiencing power flows in the opposite direction - this is due to an increase in decentralised generation (DG) such as PV and wind turbines. As of July 2019, there is 1549.7 MW of DG connected in Northern Ireland at 33kV and below, with a further 249.17 MW committed to be connected [1]. Wind is the largest source totalling 1,157 MW, whilst PV is the second largest source at 17%, with a capacity of 258 MW.

The use of PV technology has rapidly increased across Northern Ireland, from 0 MW in 2010 to 258 MW of installed capacity in 2019. There are ten large scale solar farms in NI with their maximum export capacities (MECs) totalling 124 MW - approximately half of NI PV production. These sites range from 4.9 MW to 29.9 MW, mainly situated in the East of NI as the high number of wind turbines connected in the West presents difficulties for integrating solar. As the West tends to be a more rural network, the sophisticated power infrastructure and greater load in the east provides closer proximity to large

substations for PV connection, reducing connection costs and enabling larger MECs.

II. METEOROLOGICAL EFFECTS ON PV GENERATION

A. Solar Irradiance

Solar Irradiance (SI) data is often used in PV forecasting, as it provides an instantaneous measurement of the solar power over given area. Irradiance can be expressed and separated into different components depending on the collection plane, portion of sunlight which is measured, or the field of view of the measurement instrument. There are four parameters of SI:

- Direct/Beam Normal Irradiance (BNI) - The amount of irradiance received per unit area by a surface that is perpendicular (or normal) to the beam of the sun at its current position in the sky.
- Diffuse Horizontal irradiance (DHI) - The irradiance received by a horizontal surface which has been scattered or diffused by the atmosphere (ie no direct path from the sun).
- Global Horizontal Irradiance (GHI) - The amount of irradiance falling on a surface horizontal to the surface of the earth. GHI can be calculated from BNI and DHI using (1), where: θ_z = solar zenith angle:

$$GHI = DHI + BNI \cdot \cos(\theta_z) \quad (1)$$

- Direct/Beam Horizontal Irradiance (BHI) - DHI is often overlooked when investigating solar generation. It represents the amount of irradiance received per unit area by a surface that is horizontal to the beam of the sun at its current position in the sky.

GHI is the most common SI metric used in solar forecasting, as it provides the best representation of the overall SI of an area (derived from BNI and DHI). GHI is very predictable, reaching its peak when the sun is highest in the sky and following a yearly trend - low SI throughout Winter months and high throughout Summer months, particularly May and June. As NI is a relatively small, it could be considered to use one GHI reading to represent the whole country. However, Fig. 1 illustrates clearly just how variable solar irradiance can

be between the towns and cities of NI. The three PV farms plotted are located approximately 25 miles from each other and experience completely different GHI profiles over the same 24 hour period. Therefore a PV forecasting model requires solar irradiance data for individual PV sites.

Fig. 1. GHI Daily Profiles for Three NI PV Farms

B. Cloud Cover

Cloud cover is one of the most important weather observations for solar forecasting. The okta scale of measurement is used, where cloud cover is classified as a value ranging from 0, if a completely clear sky, to 8, if completely overcast. A value of 9 indicates the sky has been obscured from view usually due to heavy fog/snow. Increasing cloud cover can rapidly change the output generation of PV by obstructing the SI reaching the panels, reflecting it back into the atmosphere and reducing BNI by up to 100% [3] for days of 8 okta cover. Figure 2 illustrates the effect of atmospheric conditions, including cloud cover, on the SI which reaches the ground. Clear Sky GHI represents the incident SI under perfect conditions with no obstructions, whilst the actual GHI displays the recorded SI which reached the ground on the 14th June 2019. As increased cloud cover reduces the SI reaching the PV panels, this will have the knock on effect of reducing the generation. Figure 3 demonstrates the output of PV Farm 3 on three separate days, each with varying levels of cloud cover. To ensure a fair test, all three days were from the same month and have similar clear sky GHI curves. The clear sky day had an average cloud cover of 1.89, the slightly cloudy day 5.33, and the overcast day an average of 7.78. The resultant plot highlights how days with higher levels of cloud cover tend to generate less power than those with clear skies, with an extreme peak difference of 28MW between the clear and overcast curves.

C. Temperature, Windspeed and Humidity

When an increase in temperature causes a solar panel to heat up past its ambient temperature, the efficiency of the panel can decrease due to a reduction in the band gap of the semiconductor. As a result, heat can severely reduce the solar panel's production of power. The most common polycrystalline PV cells have an average temperature coefficient of -0.50% [4], ie, for one degree temperature rise, the solar generation output

Fig. 2. Clear Sky GHI vs Actual GHI

Fig. 3. GHI Daily Profiles for Three PV Farms in NI

falls by 0.5%. An increase in wind therefore improves the efficiency of the panels, by reducing the panel temperature and providing ventilation to remove hot air. However, humidity can decrease efficiency of PV panels by causing a collection of water vapour on panel surface, refracting sunlight away from the panel [5]. As Northern Ireland has a relatively cool and non-humid climate, PV is able to generate at its optimum efficiency.

III. THE DATA DRIVEN PV FORECASTING MODEL

The main objective of this report is to develop forecasting methods for PV and demonstrate model performance for NI. The model produces accurate results when using four types of solar irradiance data as the input - GHI, BHI, DHI, BNI (Note Beam Radiation is often called Direct Radiation).

A. Data Collection

The following data was collected to commence a data driven, statistical model for NI:

- Data for several large scale solar farms across NI (one year, fifteen minute resolution) totalling 114MW of PV Capacity in NI. Data for twelve small scale sites ranging from 8kW to 50kW across NI (two months, half hourly resolution).
- Copernicus Atmosphere Monitoring Service (CAMS) provided all types of Solar Irradiance data (Global, Direct, Diffuse, Beam) for both actual weather conditions

and clear sky conditions. (One year, fifteen minute intervals.)

- Ceda Web Processing Service was used to access Met Office weather data containing the temperature, wind speed and cloud cover of several sites.

B. Linear Regression Model

Regression analysis, a statistical technique used to estimate the relationship among variables, has been used to create the a PV model. Linear regression models the relationship between a dependent variable and one or more explanatory variables using a linear approach [7]. Multiple linear regression (MLR) has been used, which is the most common form of regression analysis utilising more than one explanatory variable. MLR can provide three key insights into PV generation - identify if there is correlation between the variables, quantify the impact on the dependent variable when there is change in the explanatory variables, and forecast future trends and values. The mathematical equation for MLR is as follows:

$$\gamma_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i \quad (2)$$

$i = 1, \dots, n$

C. Partial Least Squares Regression Model

An alternative PV model was based on the method of Partial least squares (PLS), an extension of MLR, used to construct predictive models when there are many variables with a large amount of correlation, or even collinearity, among them [8]. PLS excels in predicting responses rather than understanding the underlying relationship between the variables. Although both PLS and MLR can be used with many variables, MLR suffers from the phenomenon of “over-fitting”, where the model will begin to fit the sampled data perfectly but fail to predict new data accurately. PLS aims to extract latent variables, accounting for as much of the variation as possible whilst modelling the responses well [9]. The underlying model of multivariate PLS is as follows:

$$\begin{aligned} X &= TP^T + E \\ Y &= UQ^T + F \end{aligned} \quad (3)$$

D. Model Comparison

A years worth of SI data was collected for each large scale PV site in NI (114 MW), based on their latitude and longitude, which was used to create a PLS and LR model for each PV farm. These individual forecasts were then combined to create the total large scale forecast.

- *Short Term Forecasting* - The models were constructed using four months of generation data, and tested on the following month, illustrated in Fig. 4. The PLS method had a greater accuracy than LR as the prediction range increased and therefore should be favoured for short term forecasting.
- *The Effect of Overnight Data* - The model was simulated both including and excluding night time data achieving similar percentage errors of +7.76% and +7.63% respectively. (Night time data should be removed to prevent interference with error metrics.)

Fig. 4. Plot of Forecasts Using PLS and MLR Against Real Generation Data

IV. MODEL ACCURACY

PV forecasting is becoming increasingly important for System Operators in order to maximise the use of solar resources on the network and minimise generation uncertainties, therefore it is important that the model used has a high level of accuracy.

A. Error Metrics

To precisely measure the error level of this forecast, the output will be compared to real generation data using various statistical metrics discussed below. In each equation, p_i represents actual solar power generation, $\bar{p}_i$ is the corresponding forecasted generation, and N is the number of points estimated in the forecasting period.

The **Root Mean Squared Error (RMSE)** is often used in solar forecasting as presented in Eq. (4). RMSE is scale dependent therefore is more often used to compare accuracy between models, and not between datasets. Other metrics should be used alongside RMSE.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\bar{p}_i - p_i)^2} \quad (4)$$

The **Mean Absolute Error (MAE)** presents an average of each individual error in the forecast shown in Eq. (5). Unlike RMSE, MAE is not as sensitive to extreme forecasting events.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\bar{p}_i - p_i| \quad (5)$$

The **Mean Bias Error (MBE)** is used to indicate the tendency of the PV model to under or over-forecast, presented in Eq. (6). A negative MBE highlights that predictions are mostly smaller than the real generation values, and a positive MBE represents higher.

$$MBE = \frac{1}{N} \sum_{i=1}^N (\bar{p}_i - p_i) \quad (6)$$

B. Short Term Forecasting

The PLS forecast was built using ten months of PV generation data (November to July) and tested across the full month of August, illustrated in Fig. 5. Note, night-time data has been removed. The resultant error metrics are given in Table 1. The forecast illustrated a slight positive bias, mainly influenced due to days with a sudden extreme decrease in generation. High Okta days can be unpredictable but could be accounted for if the System Operator knew of incoming major storms. Overall, the forecast achieves an accurate result considering the impact of unpredictable weather patterns and cloud cover.

Fig. 5. Predicted versus Actual Aggregated Large Scale Solar Generation in a Summer Month

TABLE I
ERROR METRICS

Metric	August Forecast
RMS (MW)	11.91
MAE (MW)	7.73
MBE	1.76

C. Addition of Meteorological Data

The model was expanded to include both generation and weather data, including solar irradiance, temperature, wind speed and cloud cover. Figure 6 illustrates both the original and new PLS forecasts in comparison to the real generation output on the 2nd of August 2019. It is evident that whilst both forecasts under perform, the orange plot, representing the new model, outputs a more accurate prediction in comparison to the original forecast which inputs only historic generation and solar irradiance data. Therefore reinforcing that temperature, wind speed and cloud cover each influence the output of solar generation.

V. UPSCALING THE MODEL

The forecasting model accounts for 114MW of the total 258MW installed PV capacity in NI (the data used for this work was collected in 2019), therefore, it is essential to employ an upscaling technique. Upscaling techniques can simplify calculations and decrease the need for historic generation data

Fig. 6. Forecast Comparison For The 2nd August

of all photovoltaic systems installed in a given area. The chosen subset must have an output which is representative of the behaviour of all PV systems in NI. Research indicates that large scale ground mounted PV is more efficient than roof top small scale, due to cooling or direction of panel placement therefore it is necessary to have separate forecasts for both small scale and large scale PV, upscaled to their representative capacities on the system.

- 1) **Large Scale Forecast:** A scaling factor of 1.088 was chosen and applied to the existing 114MW combined large scale forecast, in order to represent the entire 124MW large scale PV installed in NI.
- 2) **Small Scale Forecast:** The small scale PV sample consisted of 11 sites, totalling 252kW, with a scaling factor of 531.75 applied to represent the small scale total of 134MW. SI data collected for each location, with sites spread out across NI to provide a good geographical representation.

VI. SYSTEM OPERATOR APPLICATIONS

The PV model can not only be used to forecast both large and small scale solar generation, but could become a useful tool for the System Operator when load forecasting. Typically the NI system demand follows a predictable daily pattern, from the sudden rise in the morning time to the ‘tea time peak’ in the evening. The NI system also tends to follow a weekly pattern, requiring a higher demand during the working week and less at the weekends. Domestic PV can cause the appearance of a ‘dip’ in the system demand during the strongest hours of sunlight and high solar irradiance (typically between 12pm to 5pm), even though energy consumption across the province does not actually reduce. This drop in system demand is due to localised small scale PV supplying a portion of, or in some cases all, the energy to the installation to which they are connected. As a result, less power is supplied across the electricity network. Figure 7 displays the load curve of two consecutive Tuesdays, the 14th of May (blue) and the 21st May 2019 (green). The 21st of May was a much sunnier day, with a higher peak SI in comparison to the 14th of May. Whilst both days still had minimal wind generation,

the solar output on the 21st obscured the load curve, causing the appearance of a ‘load drop’ between the peak sunlight hours. The accuracy of the created PV forecasting model can be tested using the above dates. In theory, the output of the small scale PV forecast (representing the hidden domestic PV) plus the system demand on the 14th of May, should equal the system demand on the 21st of May (where there was minimal PV generation). However, as there is not enough data available to extrapolate the small scale forecast to May, the large scale forecast has been used, with a scaling factor of 0.7. This value was calculated, representing the average difference between upscaled small and large scale generation. The resultant plot is represented in Fig. 7 by the yellow trace.

Fig. 7. NI Load Curve For Consecutive Tuesdays, Including PV Forecast On The 14th

It is clear that the forecast can accurately predict the domestic PV generation, as the yellow and green traces are almost identical. On previous testing, the forecast had a slightly positive mean bias error. This becomes evident between the hours of 6pm and 8pm, where the yellow trace is slightly higher than the system demand on the 21st May. As the model has been proven to produce an accurate domestic PV forecast, it could be utilised to improve the System Operator’s load forecasting process by unmasking the load which is supplied by PV, particularly during the hours of 12-5pm. Typically, extreme differences in SI across consecutive days creates a sizeable difference between the forecasted and actual system load, this can cause frequency, voltage and stability issues. Even a slight increase in load forecasting accuracy could provide not only increased stability across the system, but an economic benefit to the System Operator, allowing for optimal dispatch of generation and ultimately reducing the price of electricity for the customer.

VII. CONCLUSION

Historic PV generation for several large scale PV sites and 12 small scale sites in NI was collected. The monthly and daily weather patterns were then investigated, to include cloud cover, temperature and solar irradiance, and compared at different PV sites across NI - concluding that the model would require separate localised weather data for each PV site.

Statistical models were explored, including PLS and MLR, demonstrating PLS to produce the most robust results over a short term forecast. The resultant PV forecasting model, specific to Northern Ireland, has been created using MATLAB, and testing carried out to reveal a slight positive mean bias, RMSE of 11.91MW and MAE of 7.73MW. On the addition of meteorological data as inputs to the model, the forecast results had a visible improvement in accuracy. A second forecast was then created, to represent the small scale PV in Northern Ireland, and both models upscaled to their representative capacity on the system. Testing was then carried out to confirm the accuracy of the forecast, using demand curve data from SONI. This found the model to be quite accurate in forecasting the hidden domestic PV on the load curve. In conclusion, the two PV forecasts produce a rather accurate result, with some days having an MAE of only 4.3MW. Considering the uncertainty introduced by the continuously changing clouds, curtailment skewing historic generation data and the general unavailability of small scale data due to data protection, the model performs very well. The model could potentially improve the operation of the power system, and even improve load forecasting by unmasking hidden domestic PV generation.

ACKNOWLEDGMENT

I wish to thank SONI Ltd, with special thanks to Bryan Murphy for providing solar generation data. This work is part of SPIRE2 project supported by the European Union’s INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB). The views and opinions expressed in this paper do not necessarily reflect those of the European Commission or the SEUPB.

REFERENCES

- [1] M. Robinson, "Connected DG On The NI Network", NIE Networks, Fortwilliam House, November 2020.
- [2] "Electricity Consumption and Renewable Generation Statistics — Department for the Economy", Economy, 2019. [Online]. Available: <https://www.economy-ni.gov.uk/articles/electricity-consumption-and-renewable-generation-statistics>. [Accessed: 11- Nov- 2019].
- [3] "Solar Radiation Basics", Energy.gov, 2020. [Online]. Available: <https://www.energy.gov/eere/solar/articles/solar-radiation-basics>. [Accessed: 15- Mar- 2020].
- [4] "The Impact of Temperature on Solar Panels - TheGreenAge", TheGreenAge, 2020. [Online]. Available: <https://www.thegreenage.co.uk/article/the-impact-of-temperature-on-solar-panels/>. [Accessed: 15- Mar- 2020].
- [5] "Do Solar Panels Work Less Efficiently at Certain Temperatures? — Solar.com", Solar.com, 2020. [Online]. Available: <https://www.solar.com/learn/do-solar-panels-work-less-efficiently-at-certain-temperatures/>. [Accessed: 16- Mar- 2020].
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [7] A. Rencher and W. Christensen, "Chapter 10, Multivariate regression – Section 10.1, Introduction", Methods of multivariate analysis. Hoboken, N.J.: Wiley, 2013. ISBN 9781118391679.
- [8] "Partial Least Squares (PLS)", Statsoft.com, 2020. [Online]. Available: <http://www.statsoft.com/Textbook/Partial-Least-Squares>. [Accessed: 15- Jan- 2020].
- [9] "An Introduction To Partial Least Squares Regression", Stats.idre.ucla.edu, 2020. [Online]. Available: <https://stats.idre.ucla.edu/wp-content/uploads/2016/02/pls.pdf>. [Accessed: 13- Feb- 2020].