Very Short Term Forecasting for Solar Farms

John Boland
University of South Australia

July 2019

Credentials

- Professor of Environmental Mathematics
- Director, Centre for Industrial and Applied Mathematics, UniSA
- Past Member, Task 46, International Energy Agency, Solar Forecasting
- Member, Task 16, IEA, High Penetration PV
- Head, Urban Microclimates Project, Cooperative Research Centre for Low Carbon Living

The Problem

Any one step ahead statistical forecasting method can be encapsulated by the structure

$$Y_t = f(S; Y_{t-1}, \ldots, Y_{t-p}; X_{i,t-1}, \ldots, X_{i,t-q}) + Z_t$$

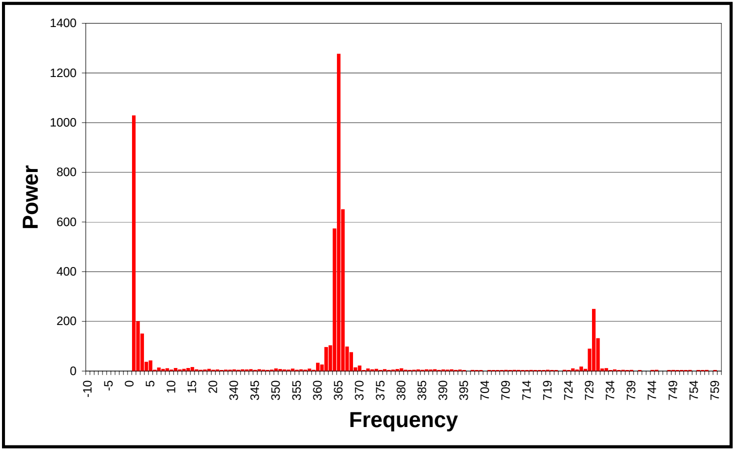
Knowledge of the statistical qualities of Z_t is necessary in order to construct the error bounds of the forecast. In this formulation, it is hoped, and sometimes assumed that Z_t is independent and identically distributed (i.i.d.).

The Situation with Solar Data

For this data, and I assume other short time scales at least,

- Independent - NO. Z_t Uncorrelated but dependent. Note that correlation is only a linear property, so higher moments can be, and are, correlated.
- Identically distributed - NO.

Power Spectrum



Seasonality

- The first step is to identify and model the seasonality. We have identified several significant cycles using spectral analysis. Fourier series will be used in this step.

$$
\begin{aligned}
S_t &= \alpha_0 + \alpha_1 \times \cos\frac{2\pi t}{T} + \beta_1 \times \sin\frac{2\pi t}{T} + \\
&\alpha_2 \times \cos\frac{4\pi t}{T} + \beta_2 \times \sin\frac{4\pi t}{T} + \\
&\sum_{i=3}^{11}\sum_{n=1}^{3}\sum_{m=-1}^{1}[\alpha_i \times \cos\frac{2\pi(365n+m)t}{T} + \\
&\beta_i \times \sin\frac{2\pi(365n+m)t}{T}]
\end{aligned}
$$

here S_t is seasonal component.

UniSA

Clear Sky Index

- A majority of researchers instead use multiplicative deseasoning, employing a clear sky model.
- There are a multitude of such models, so which do you use?
- Once you have decided which one, you divide the Global Horizontal Irradiation (GHI) by the clear sky model to get a clear sky index (CSI).
- See the next slide for problems with the CSI.

Maximum value of Unity?

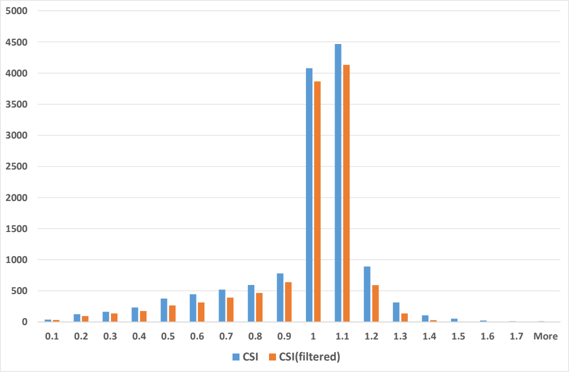


Figure : CSI values for Los Vegas

Horns

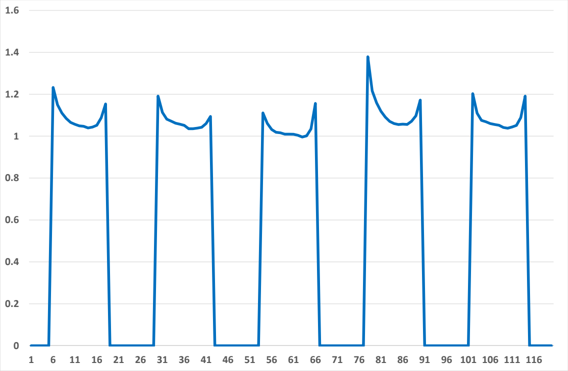
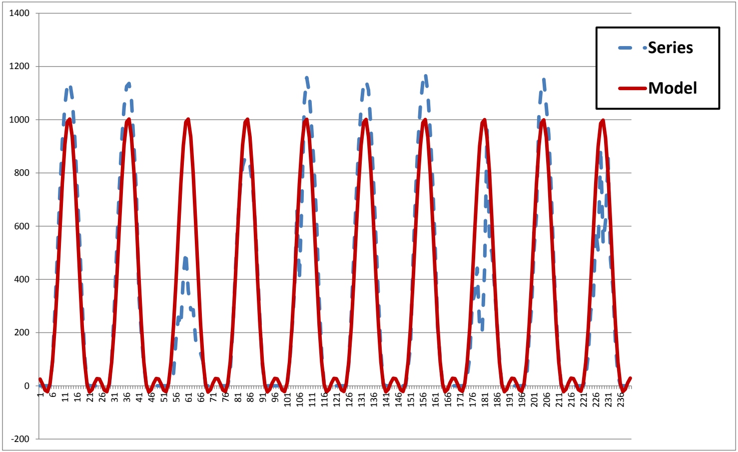
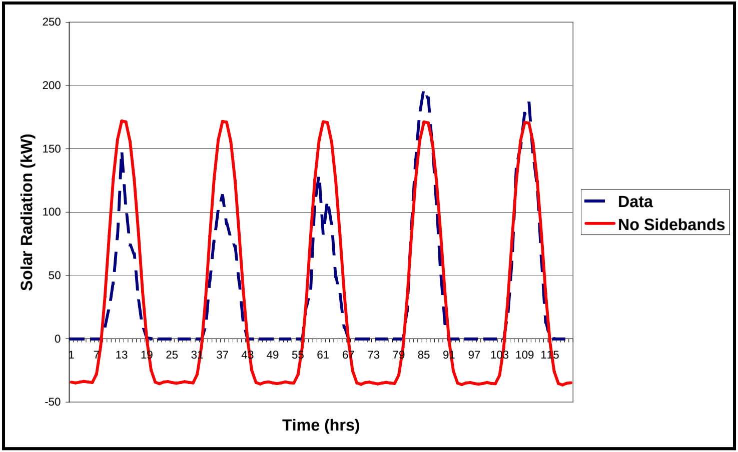


Figure : Five days of CSI for Los Vegas

For ten days Fourier series and data



Effect of ignoring sidebands - winter



The Forecast Model

- The standard time series model (Box-Jenkins) is based on finding any deterministic components like trend or seasonality and then removing them.
- Next, one tries to find any remaining structure, the most basic type being autoregressive.
- Thus, we take the original series $Y(t)$ and subtract the seasonality $S(t)$ to form the residual series $X(t) = Y(t) - S(t)$.
- Then we try to find some relation between $X(t)$ and $X(t-1), X(t-2), \ldots$.

Solar Farms in Australia

- As of March 2019, Australia had over 12,000 MW of installed PV, with 4,000 MW being commissioned in the previous 12 months.
- The largest solar farm is 220 MW (Bungala in SA) and there are twelve over 50 MW.
- ARENA grant for **Solar Power Ensemble Forecaster** - $1.25 million over 18 months.
- Boland, Kay (UNSW), Huang, West, Knight (CSIRO).
- Solar Farms - Manildra, Clare, Darling Downs, Gannawarra, Parkes.

The National Electricity Market (NEM)

- Every 5 minutes, generators supply a bid stack, with the amount of energy they can provide in the next 5 minutes at each of 10 price bands, from -\$1,000 to \$14,000.
- The Australian Energy Market Operator (AEMO) then runs a linear program for each region of the NEM to determine how far up the bid stack they have to go to satisfy their forecasted net load.
- This determines the 5 minute price for all energy, and the mean of 6 five minute prices gives the spot price for the half hour.

The Project

- One big problem for AEMO is that there wind and solar forecast tool is based on a numerical weather prediction model (NWP) from Europe, originally developed for the day ahead market, so it does not work very well at short time scales.
- They asked the Australian Renewable Energy Agency (ARENA) to fund projects to improve the 5 minute solar and wind farm forecasts with a view to the farms being self-forecasting.
- Since the farms we are working with are recent additions, we are using Broken Hill Solar Farm 5 minute data to practise with.

Difference of Solar Farm Output from Solar Energy Profile - Summer

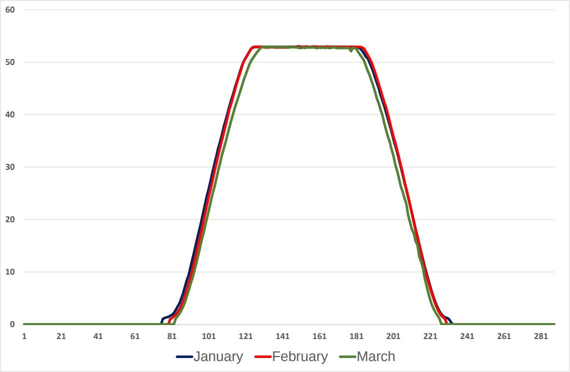


Figure : January-March

Difference of Solar Farm Output from Solar Energy Profile

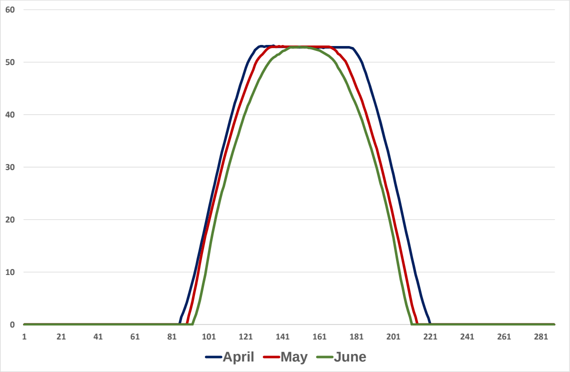
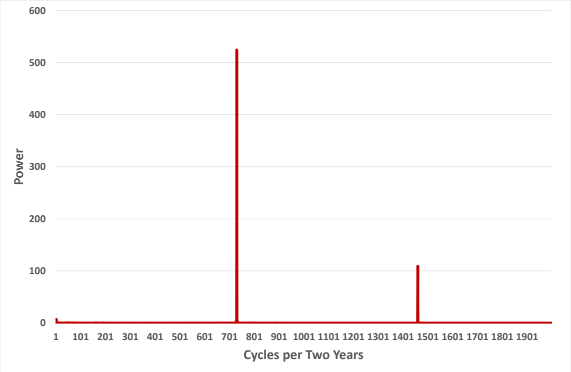


Figure : April-June

Power Spectrum



Comparison of the Fourier series modelling of solar versus solar farm

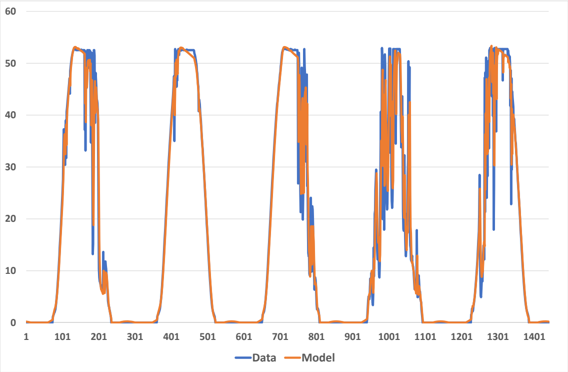
The variance of the series explained by the inclusion of frequency ω is equal to $(\alpha_\omega^2 + \beta_\omega^2)/2$ in percentage terms.

Type	1/year	Beat	1/day	Beat	2/day
Solar	7.37	3.72	56.67	3.34	11.05
Solar Farm	1.03	0.58	76.82	0.18	15.98

Model for the deseasoned data

- Form $R_t = S_t = F_t$, where S_t is the solar farm output in MWh, and F_t is the Fourier series representation of the climate.
- Check the sample autocorrelation function (SACF) and sample partial autocorrelation function (SPACF) to see what form an $ARMA(p, q)$ should be used.
- There are two things to note here:
  - The Fourier series and $ARMA(p, q)$ models are estimated on a training set and then tested on a year not in the training set.
  - Many people at this stage use much more esoteric means for modelling, like ANN or other machine learning techniques - I even have a PhD student about to try deep learning. But, no one has been able to beat these simple methods as yet.

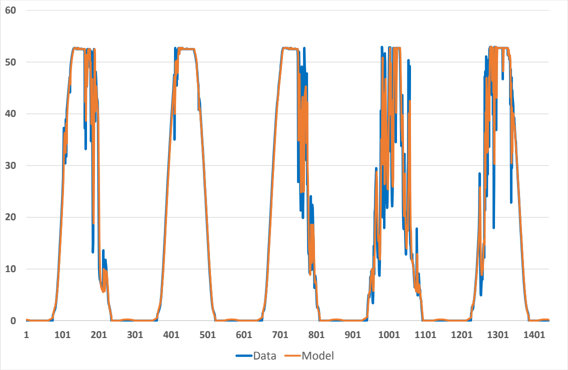
Forecast with $AR(5)$



Alteration

- I noticed that there is the falling away from the capacity after a few time steps.
- When the output reaches capacity, it tends to stay there for at least a few time steps. This is reasonable since clear or nearly clear skies will ensure that the output stays at capacity.
- To make use of this characteristic for forecasting I constructed a rule.
- If $S_t = S_{t-1} \approx C$, then $\hat{S}_{t+1} = S_t$, where C is the capacity of the farm.

Forecast with $AR(5)$ plus working with capacity



Definitions of all the measures

Normalised Mean Bias Error

$$NMBE = \frac{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\bar{y}} = 0.09\%. \tag{1}$$

Normalised Mean Absolute Error

$$NMAE = \frac{\frac{1}{n}\sum_{i=1}^{n}(|\hat{y}_i - y_i|)}{\bar{y}} = 6.5\%. \tag{2}$$

where $\hat{y}_i$ are predicted values, y_i are measured values and $\bar{y}$ is the average of measured values. For comparison, the lowest NMAE we have found for a similar type of site is approximately 10-12%.

UniSA

Normalised Root Mean Squared Error

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}}{\bar{y}} = 13.3\%. \tag{3}$$

There is an associated measure, called the Skill Score (SS). It is defined as

$$SS = 1 - \frac{RMSE_{Forecast}}{RMSE_{Reference}} = 16.3\%. \tag{4}$$

where $RMSE_{Reference}$ is the root mean squared error of some reference or benchmark model. We use persistence, $\hat{S}_{t+1} = S_t$.

Conclusions and further work

- Note that I also tried the idea of a clear sky output index, using a rolling window of thirty days to calculate the maximum at each five minute interval over those days. It seemed to capture this phenomenon, but the overall model did not work as well as the simpler approach.
- This analysis was performed on data from the Broken Hill Solar Farm, which has been operating for over three years.
- The five solar farms we are working with have been operational for less than a year.
- The proposal is to use a scaled version of the Fourier series model to suit the capacity and then use an adaptive autoregressive model for the deseasoned series.

Continued

- The intention is to then blend this approach with the use of sky camera and satellite based forecasts.
- We will also try other approaches such as deep learning.
- We will add probabilistic forecasting methods.
- John Boland and Adrian Grantham (2018) Nonparametric Conditional Heteroscedastic Hourly Probabilistic Forecasting of Solar Radiation, *J Multidisciplinary Journal*, pp. 174-191; *doi* : 10.3390/*j*1010016

UniSA