**Abstract**

There has been a recent push to integrate renewable sources of energy into the power grid, with Solar PV being one of the most promising among them. Among other things, effective integration involves the ability to predict generation so that other sources can be curtailed or increased accordingly to meet the power demand. Time series modeling is the preferred technique due to the complexity of the factors. This paper explores the effect of various factors like training data size and resolution, and time series model order on the prediction accuracy. We will also be implementing custom programs to estimate the model parameters, make predictions and evaluate these models.

**Introduction**

Solar PV

Due to the irreversible effects of Global Warming, there is a worldwide initiative to integrate renewable sources of energy into the power grid. Solar Photo-Voltaic (PV) generation is the most prominent among these that is being explored and is promising due to the lower upfront costs involved, ease of installation and adoption, and the passive nature of generation with minimal human oversight. However, one of the major drawbacks associated with it is the restriction of power generation to hours of daylight. The demands of the grid that are not met by solar will have to be balanced out by other conventional and renewable sources. Therefore, the effective integration of solar will require the ability to predict the amount and duration of solar PV power generation and using other sources to match the excess demand.

Time-series forecasting

A time-series is a set of data points that is spaced at equal intervals of time. In time-series forecasting, we use time-series data to create system models which use past time instants as input and generate predictions of future time instants as outputs. This model is then used to make predictions about future instants.

Time-series forecasting is a very important technique that is used in a wide array of fields like Economics, Statistics, Weather-Forecasting and Signal Processing. It is generally adopted when the system is too complex to be represented as a mathematical model of various other inputs resulting in an output. This may be because of the complex relationships between the various inputs affecting the output, or the lack of input data that is spaced at the same time intervals as the output training data.

In these cases, a time-series model is used, wherein the output prediction of a random variable at any time-instant is a linear result of the random variable at past time instants. The output will not depend on other inputs to the system, unlike other conventional system models.

Time-series forecasting for Solar PV Generation Prediction

The power generated by a Solar PV cell is dependent on various factors like Incident Solar Radiation Intensity, the ambient temperature as well as the meteorological conditions. All these factors are strongly correlated, which makes it very difficult to characterize the power generated as a function of these parameters. Therefore, we choose to use a time-series model where the solar power generation is the only variable considered and the power generation at any time instant is a function of the power generated at past instants.

Autoregressive-moving-average (ARMA) model

“A stationary process is a stochastic process whose unconditional joint probability distribution does not change when shifted in time. It’s parameters such as mean and variance do not change over time. Such a process can be described using an ARMA model, consisting of two polynomials. One of these is for the autoregression (AR) and the second is for the moving average (MA). The AR part involves regressing the variable on its own lagged values. The MA part involves modeling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past. The model is usually referred to as the ARMA (p,q) model where p is the order of the AR part and q is the order of the MA part.”

“ARMA models can be estimated using the Box-Jenkins method.”

By studying the past Solar Generation data, we can see that the process is weakly stationary with AR(3) and MA(0). Therefore, an ARMA(3,0) model can be used to represent the power generation timeseries and make predictions with.