Recursive Estimation of Solar Forecasting at Chulalongkorn University

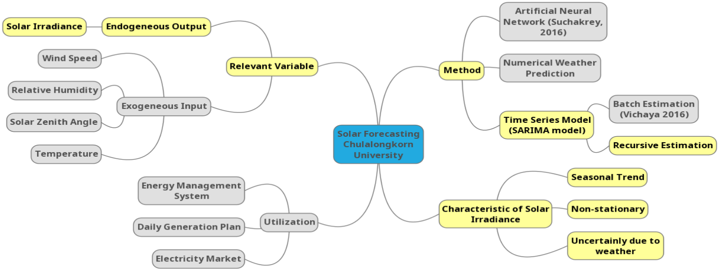
Tony Fang ID : 5730212721
Advisor : Assist. Prof. Jitkomut Songsiri

Department of Electrical Engineering
Chulalongkorn University

Outline

- Introduction
- Methodology / Preliminary Result
- Future Plan
- Conclusion

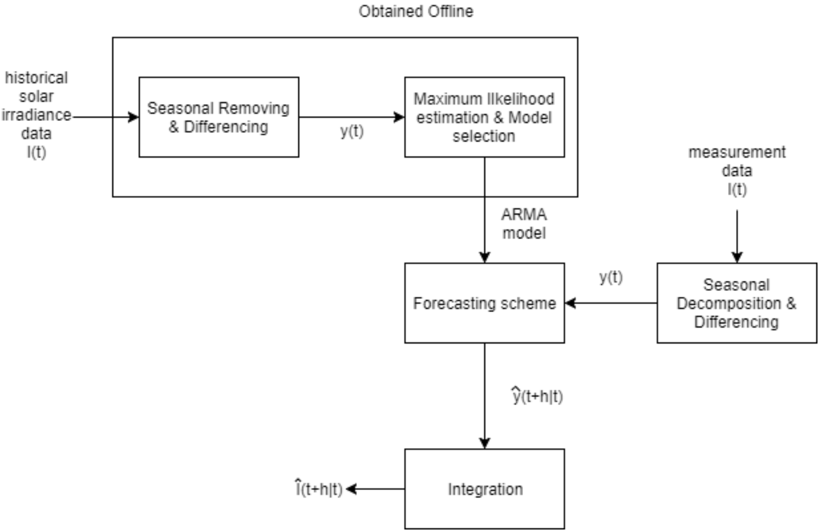
Introduction



[Vichaya, 2016] conclude that the other relevant variables is improve solar forecasting insignificantly. Thus, we focus only solar irradiance.

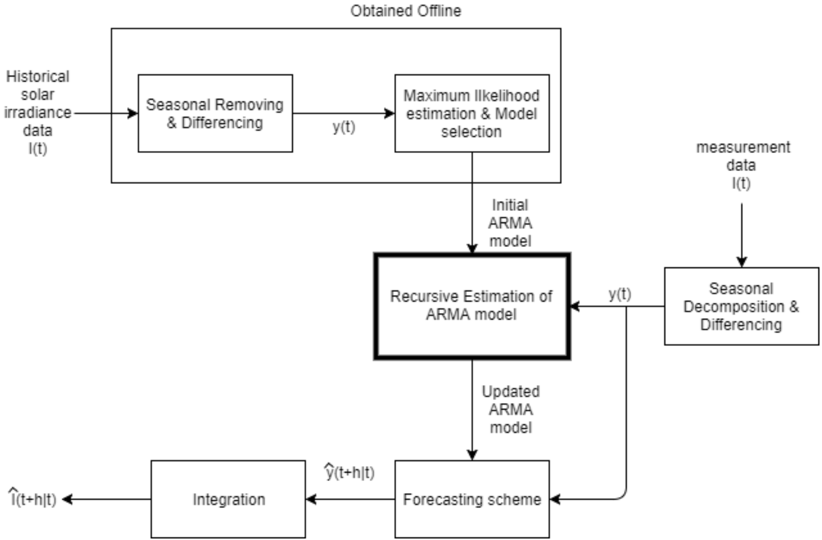
Batch Estimation

Batch estimtion is a model estimation which model parameters are not further updated.

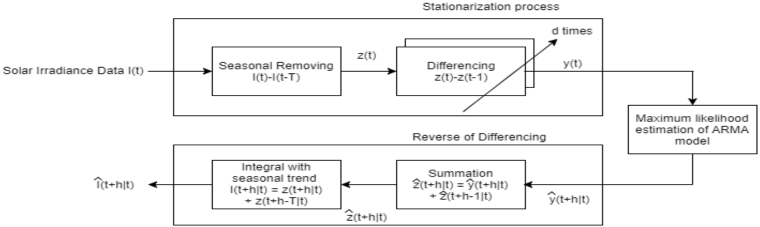


Recursive Estimation

Recursive estimation is a model estimation which model parameters are updated using new measurement data.



SARIMA model Estimation Flowchart



Seasonal AutoRegressive Integrated Moving Average (SARIMA) model defined by

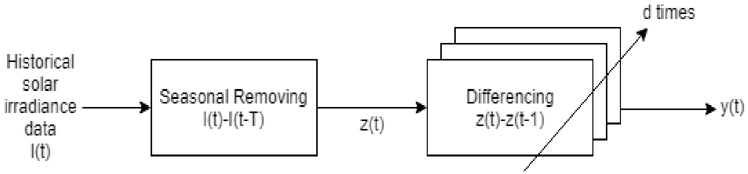
$$A(L)(1-L)^d(1-L^T)^D I(t) = C(L)v(t) \tag{1}$$

where

$$\begin{aligned} A(L) =& 1-(A_1L + A_2L^2 + \cdots + A_pL^p) \\ C(L) =& 1+(C_1L + C_2L^2 + \cdots + C_qL^q) \end{aligned} \tag{2}$$

L is a lag operator, T is a seasonal period, d is an integrated non-seasonal order, D is an integrated seasonal order

Stationarization



Seasonal Removing

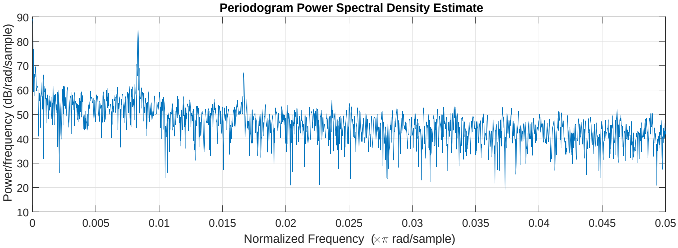
Finding seasonal period T from power spectral density by using `periodogram` in MATLAB.

Differencing

Apply Autocorrelation Function (ACF) test finding integrated order d.

Finding seasonal trend

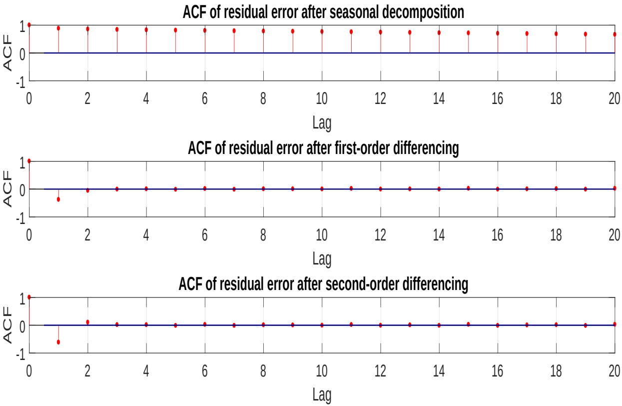
Using Fast Fourier Transform (FFT) to find the power spectral density (PSD). We choose only high-energy frequency and label them ω_i.



The PSD of this data set has 2 peaks at $\omega_1 = 0.0083\pi$ and $\omega_2 = 0.0166\pi$.

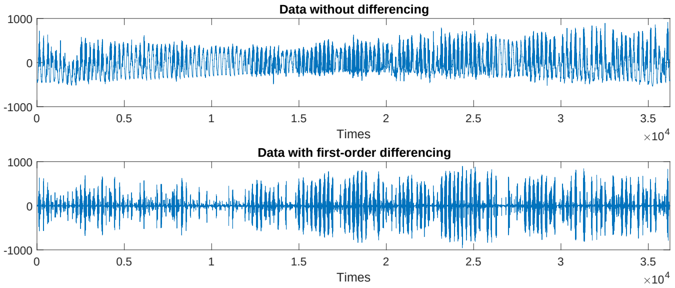
Differencing

Autocorrelation test of each data set.



Differencing

Both d=1 and d=2 pass the autocorrelation test. Because of model complexity, we use d=1. Thus, we get the SARIMA$(p, 1, q)(0, 1, 1)_{240}$. After finding order d, these post-processing data are used for find the polynomial in ARMA model.



Estimation of ARMA model and Model selection

Maximum Likelihood Estimation

From [Hamilton,1994] Parameter θ in ARMA model can find from the maximum of the cost function in (3)

$$\log \mathcal{L}(y|\theta) = -\frac{N-p}{2}\log(2\pi) - \frac{N-p}{2}\log(\sigma^2) - \sum_{t=p+1}^{N}\frac{v(t)^2}{2\sigma^2} \quad (3)$$

Candidate Score

Both AIC and BIC are trade-off between goodness of fit and model complexity.

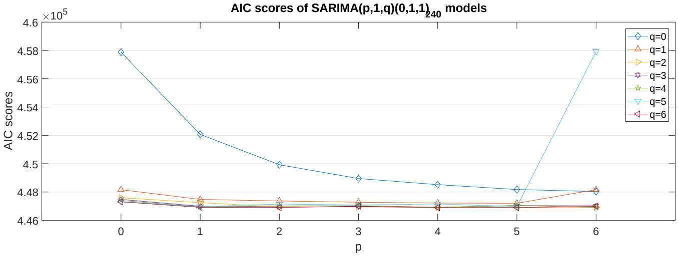
$$\mathrm{AIC} = -2\mathcal{L} + 2d \quad (4)$$

$$\mathrm{BIC} = -2\mathcal{L} + d\log(N) \quad (5)$$

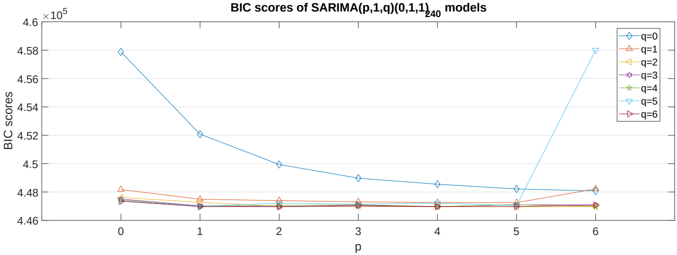
where $\mathcal{L}$ is a log-likelihood function, N is a number of data and d is a number of parameter in each models.

Maximum Likelihood Estimation and Model Validation

Finding the AIC score and BIC score after doing ML to find the candidate model.

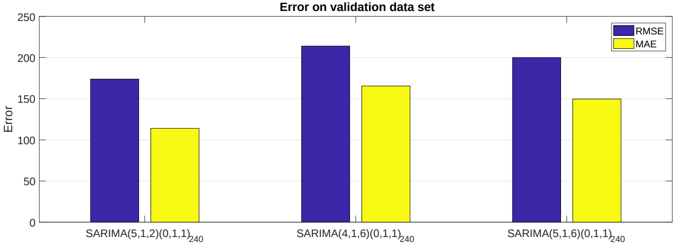


AIC choose SARIMA$(4,1,6)(0,1,1)_{240}$ and SARIMA$(5,1,6)(0,1,1)_{240}$.



BIC choose SARIMA$(5,1,2)(0,1,1)_{240}$.

After chose the candidate model, we find RMSE and MAE on validation data set of all models which we choose from AIC or BIC graph.



SARIMA$(5,1,2)(0,1,1)_{240}$ has less RMSE and MAE than others model.

Recursive Prediction Error Method

The aim of Recursive Prediction Error Method (RPEM) is to minimize the cost function which is defined as

$$V_t(\theta) = \frac{1}{2}\sum_{s=1}^{t} \varepsilon(s,\theta)^T \varepsilon(s,\theta) \tag{6}$$

[Young, 2011] assume that

- $\hat{\theta}(t-1)$ minimize $V_t(\theta)$
- minimum point of cost function $V_t(\theta)$ is close to $\hat{\theta}(t-1)$

Finally, the final step of the update formulas are shown in

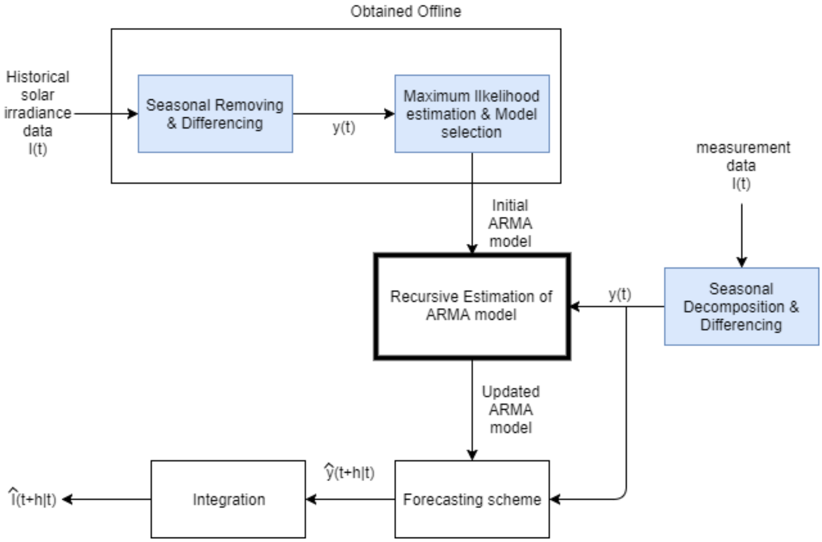
$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)K(t)\varepsilon(t) \tag{7}$$

$$g(t) = P(t)K(t)[1 + K^T P(t-1)K(t)]^{-1} \tag{8}$$

$$P(t) = P(t-1) - P(t-1)K^T(t)g(t) \tag{9}$$

where $\varepsilon(t) = y(t) - \hat{y}(t|t-1)$, $K(t) = -\nabla\varepsilon(t)$.
This method is used in ARMA model.

Conclusion


Obtained Offline
Historical solar irradiance data I(t)
Seasonal Removing & Differencing
y(t)
Maximum llkelihood estimation & Model selection
measurement data I(t)
Initial ARMA model
Recursive Estimation of ARMA model
y(t)
Seasonal Decomposition & Differencing
Updated ARMA model
Î(t+h|t)
Integration
ŷ(t+h|t)
Forecasting scheme

Q&A

Back up

Maximum Likelihood Estimation

Maximum Likelihood Estimation is a method to find parameter θ in ARMA model.

$$\hat{\theta}_{ML} = \underset{\theta}{\text{maximize}}\, f(y(1), y(2), \cdots, y(N)|\theta) \tag{10}$$

where $\theta = \begin{bmatrix} A_1 & A_2 & \cdots & A_p & C_1 & C_2 & \cdots & C_q & \sigma^2 \end{bmatrix}^T$, $f(y(1), y(2), \cdots, y(N)|\theta)$ is likelihood function. $v(t)$ in ARMA model can write into

$$\begin{aligned} v(t) = y(t) - (A_1 y(t-1) + A_2 y(t-2) + \cdots + A_p y(t-p)) - \\ (C_1 v(t-1) + C_2 v(t-2) + \cdots + C_q v(t-q)) \end{aligned} \tag{11}$$

From [1], if $y(t)$ has real value from 1 to p and $v(t) = 0$ since $t = p, p-1, \cdots, p-q+1$, so that $y(t)$ also has normal distribution.

$$\mathcal{L}(y|\theta) = f(y(p+1), y(p+2), \cdots, y(N)|y(1), y(2), \cdots, y(p), \theta) \tag{12}$$

Thus, we start at $t = p + 1$. At the same time, the conditional likelihood function is change to (12). Problem in (10) subject to root of polynomial

$$\begin{aligned} A(z^{-1}) =& 1 - (A_1 z^{-1} + A_2 z^{-2} + \cdots + A_p z^{-p}) \\ C(z^{-1}) =& 1 + (C_1 z^{-1} + C_2 z^{-2} + \cdots + C_q z^{-q}) \end{aligned} \tag{13}$$

lie inside the unit circle.

Autocorrelation Function

Autocorrelation function is defined by

$$\mathrm{ACF} = \frac{R(\tau)}{R(0)} \tag{14}$$

where the sample autocovariance function is defined by

$$R(\tau) = \frac{1}{N}\sum_{t=\tau}^{N} y(t)y(t-\tau) \tag{15}$$

In ARMA model, ACF of the ARMA(p,q) is shown in 16

$$R(\tau)-(A_1R(\tau-1)+A_2R(\tau-2)+\cdots+A_pR(\tau-p)) = 0, 0 \geq \max(p, q+1) \tag{16}$$

with initial condition

$$R(\tau) - \sum_{i=1}^{p} A_i R(\tau-i) = \sigma^2 \sum_{i=\tau}^{q} H(i-\tau), 0 \leq \tau \leq \max(p, q+1) \tag{17}$$

where $H(z^{-1}) = \frac{y(z^{-1})}{v(z^{-1})} = \frac{C(z^{-1})}{A(z^{-1})}$

Prediction Error Method

After finding the model, we can find the optimal prediction from the ARMA model by using Prediction Error Method (PEM)

$$\hat{y}(t|t-1) = (1 - C^{-1}(L)A(L))y(t) \tag{18a}$$

$$e(t) = C^{-1}(L)A(L)y(t) \tag{18b}$$

We can find the estimated ARMA(p, q) model

$$\begin{aligned}
\hat{y}(t|t-1) =& (C(L) - 1)e(t) - (A(L) - 1)y(t) \\
\hat{y}(t|t-1) =& (C_1 L + C_2 L^2 + \cdots + C_q L^q)e(t) + \\
& (A_1 L + A_2 L^2 + \cdots + A_p L^p)y(t)
\end{aligned} \tag{19}$$

where $e(t) = y(t) - \hat{y}(t|t-1)$

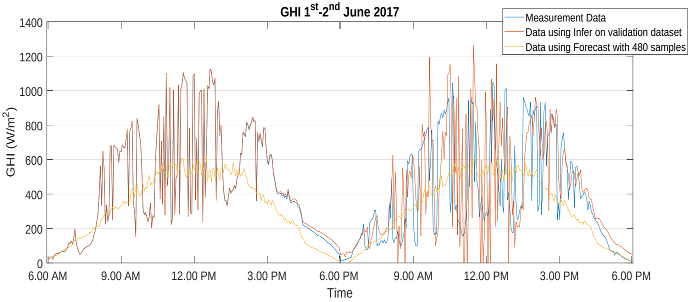
Then we can find the estimate ARIMA model in 20.

$$\hat{z}(t|t-1) = \hat{y}(t|t-1) + \hat{z}(t-1|t-1)$$
$$\hat{z}(t|t-1) = \hat{y}(t|t-1) + \sum_{k=1}^{\infty} \hat{y}(t-k|t-1) \tag{20}$$

Moreover, we can find the estimated SARIMA model in 21.

$$\hat{I}(t|t-1) = \hat{z}(t|t-1) + \hat{I}(t-T|t-1)$$
$$\hat{I}(t|t-1) = \hat{z}(t|t-1) + \sum_{k=1}^{\infty} \hat{z}(t-kT|t-1) \tag{21}$$

Forecasting Result



where the yellow line is described in (24)

We can find the forecast model by using Prediction Error Method (PEM) for an ARMA model.

$$\hat{y}(t+h|t) = (1 - 1.84L + 0.843L^2)e(t+h)$$
$$(1 + 1.26L - 0.21L^2 - 0.03L^3 - 0.02L^4 - 0.03L^5)\hat{y}(t+h|t) \tag{22}$$

We can find the forecast of ARIMA model by summation

$$\hat{z}(t+h|t) = (1 - 1.84L + 0.843L^2)e(t+h)$$
$$(1 + 1.26L - 0.21L^2 - 0.03L^3 - 0.02L^4 - 0.03L^5)\hat{y}(t+h|t) +$$
$$\sum_{k=0}^{\infty} \hat{y}(t-k+h-1|t) \tag{23}$$

And also find the forecast of SARIMA model by summation

$$\hat{I}(t+h|t) = \hat{z}(t+h|t) + \sum_{k=1}^{\infty} \hat{z}(t-240k+h|t) \tag{24}$$

Reference

J. D. Hamilton.
Time Series Analysis.
Princeton University Press, 1994.

V. Layanun and J. Songsiri.
Solar irradiance forecasting for chulalongkorn university location using time series models.
http://jitkomut.eng.chula.ac.th/group/vichaya_report.pdf, 2016.

Peter C. Young.
Recursive Estimation and Time Series Analysis.
Springer, 2011.