Senior Project Proposal 2102499 Year 2016

Solar irradiance forecasting for Chulalongkorn University location using time series models

Vichaya Layanun ID 5630550721

Advisor: Assist. Prof. Jitkomut Songsiri

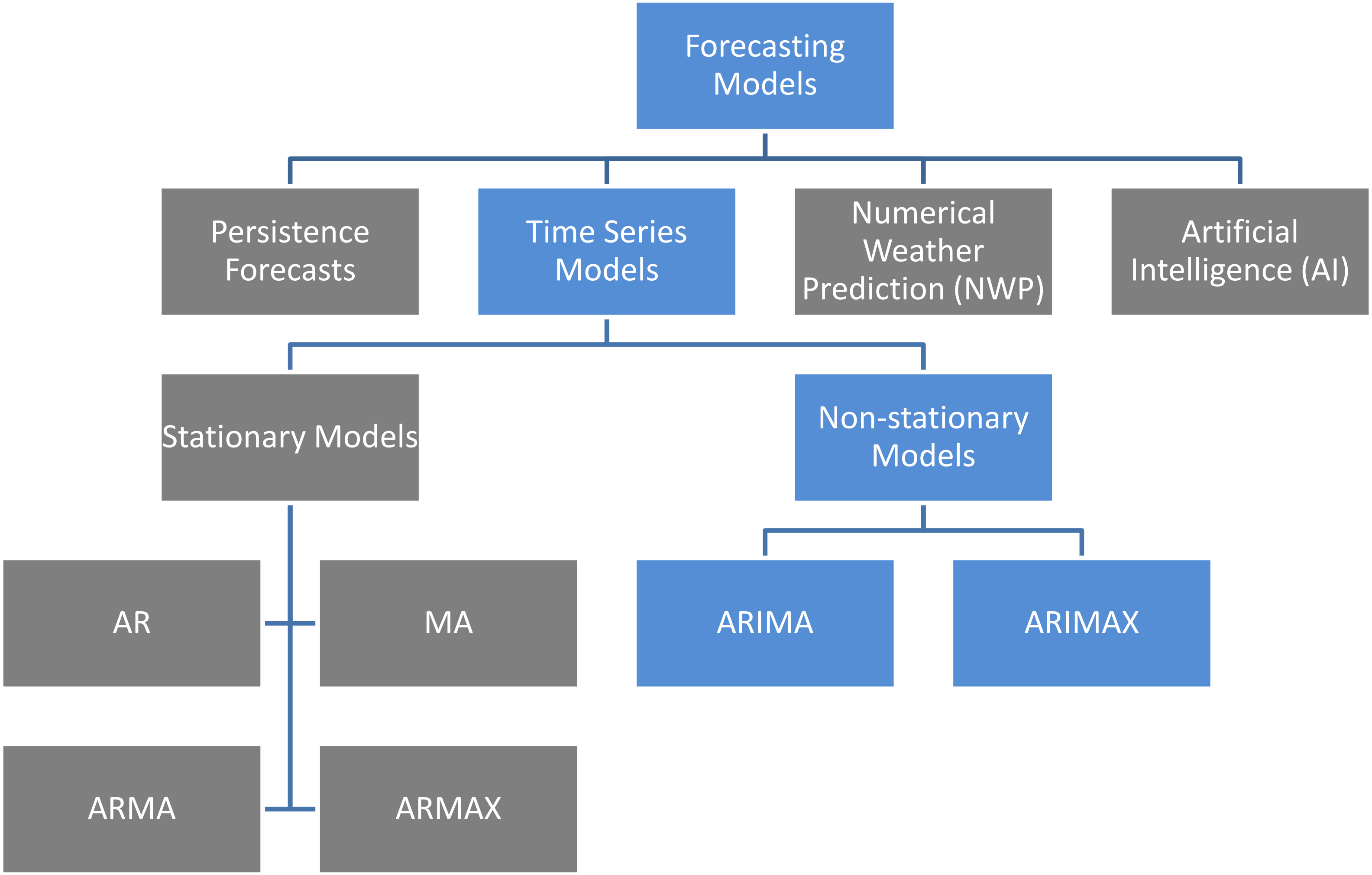
Department of Electrical Engineering

Faculty of Engineering, Chulalongkorn University

Why solar forecasting is important?

Because, **the unreliability of solar power generating** has made the entire electrical power generation difficult for power management.

Solar forecasting is a widely-common approach to deal with the problem. An improved accuracy in the forecast can provide a better management of electrical power production.



There are many forecasting models to predict the future solar irradiance. We focus on time series models.

Objective

- **To study the relevant variables** of solar irradiance forecasting.
- **To apply ARIMA** models, a **Seasonal ARIMA** models and **ARIMAX** models to forecast solar irradiance.
- **To validate results** of forecasting performance among the models using RMSE, MAE and a sample autocovarince function of the residual as model validation criterion.
- **To solve the practical issues** on data pre-processing.

Scope of work

- We focus on solar irradiance in the area of Chulalongkorn University location by using ARIMAX and seasonal ARIMAX models.
- The exogenous inputs consist of the local temperature, relative humidity, wind speed and air pressure will be included in the models.
- We will conduct experiments to verify our approach by using data obtained from Thai Meteorological Department (TMD).
- We will study the consequence of seasons of the year in the location.

Expected outcomes

- **Schemes for solving practical issues** on data pre-processing which are 1) missing data and 2) timely asynchronous data.
- **Comparison results of forecasting performance** among Persistence forecast, ARIMA models, a Seasonal ARIMA models and ARIMAX models using Root Mean Squared Error (RMSE), Mean Absolute Error as model validation criterion and a sample autocovariance function (ACV) of the residual.

Practical issues

- Missing data
- Asynchronous data

Relevant Variables

Global Horizontal Irradiance (GHI) is the considered variable which is the geometric sum of Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI).

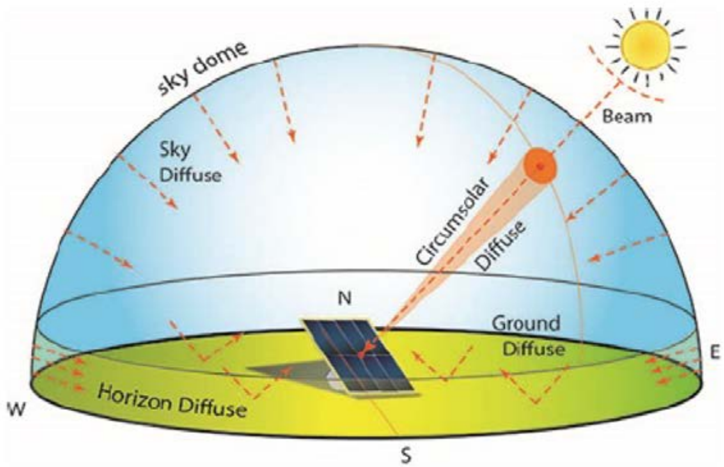


Figure 4: Solar irradiance component

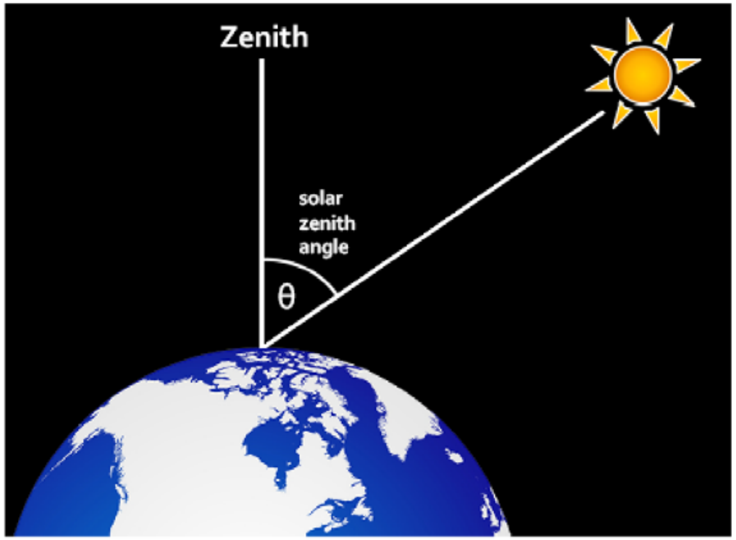
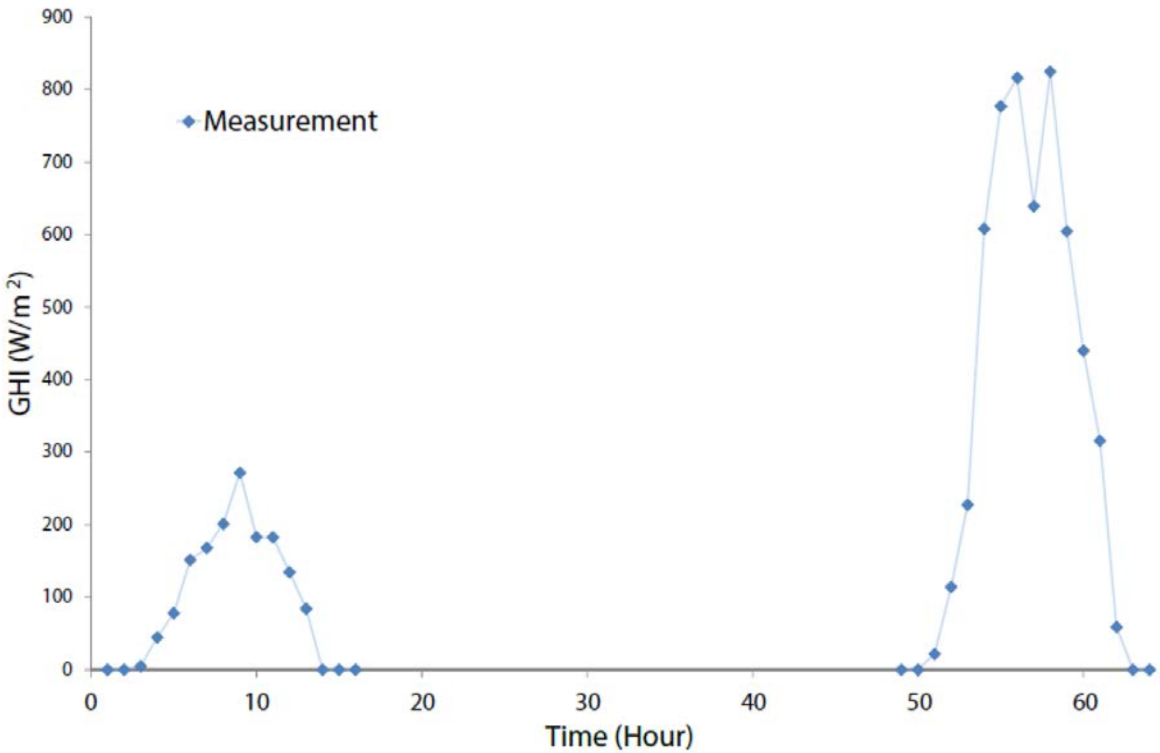


Figure 5: Solar zenith angle

This study predicted GHI and we will use the notation of $I(t)$ throughout this project.

Missing data

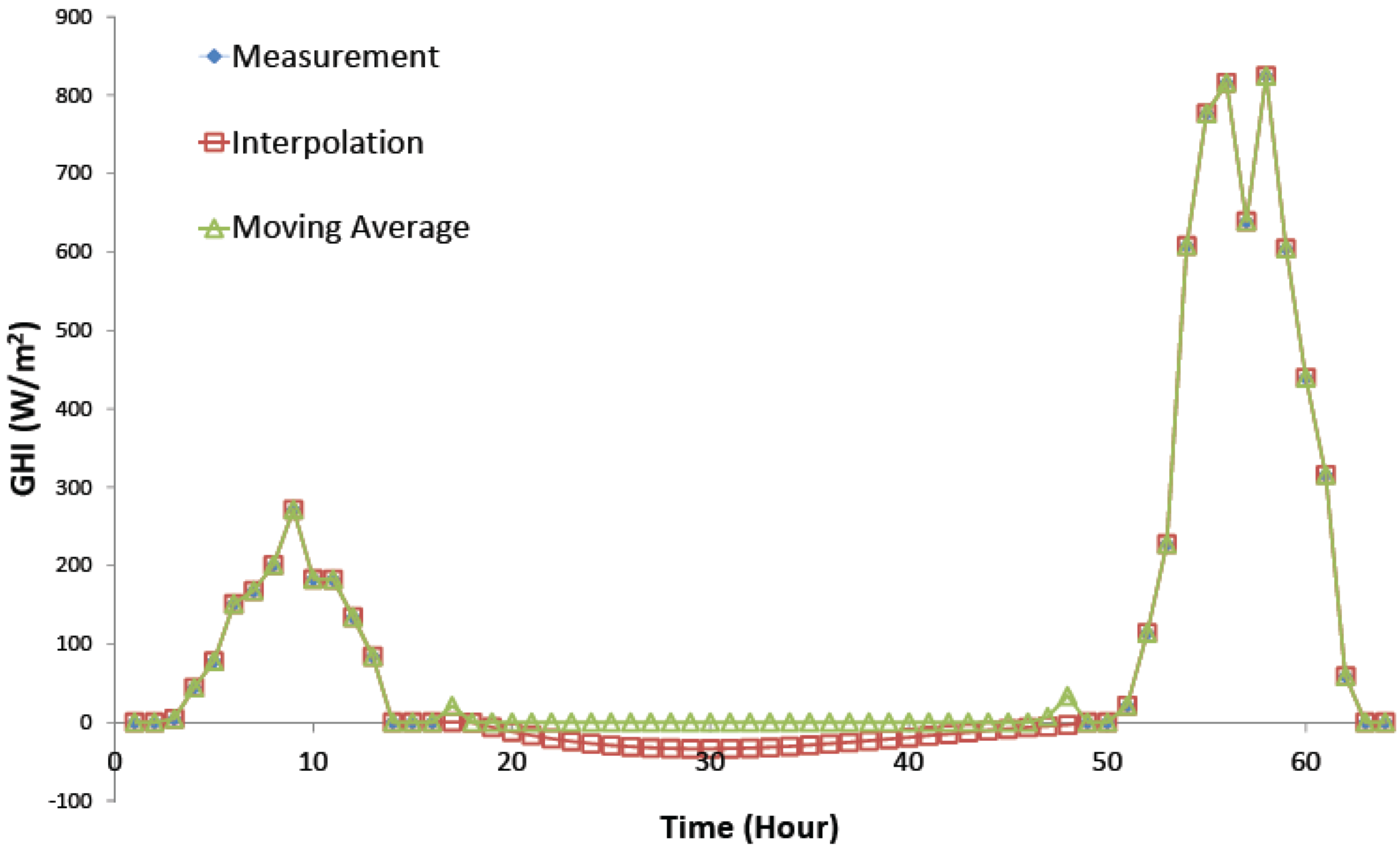
Missing data



Data missing usually holds for several consecutive days.

Estimating parameters in a time series model definitely requires a complete historical data set.

Missing-data imputation using typical methods



Moving average (MA) and linear interpolation that exploits the variable dynamic **cannot perform well** in this case as the imputed value is a linear combination of nearby available values.

Concept of the proposed method for missing-data

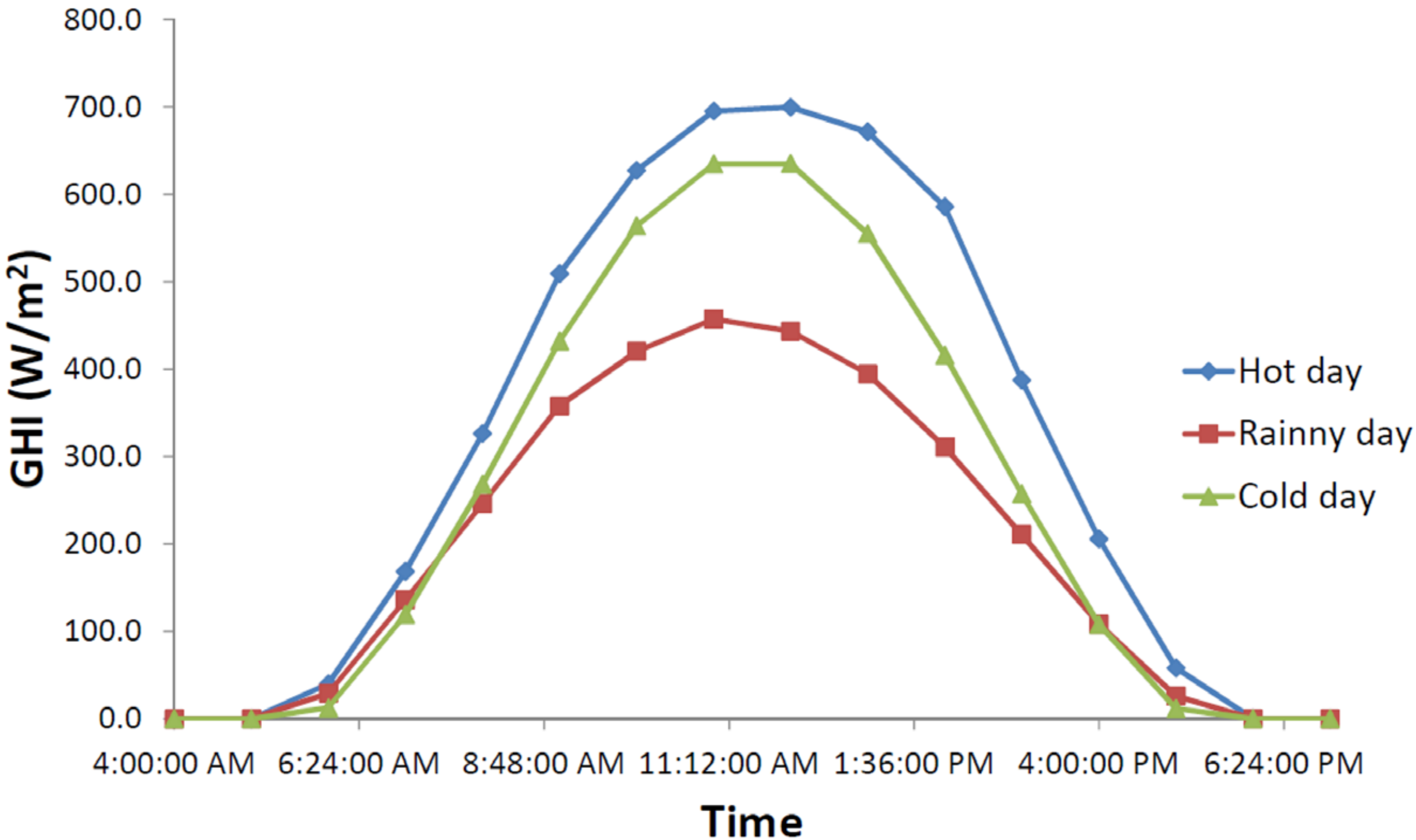
Date/Time	08-09	09-10	10-11	11-12	12-13	13-14	14-15
7-Apr	284.8	596.0	771.0	801.5	850.9	807.1	673.7
8-Apr	338.0	589.3	761.6	806.3	834.4	807.4	699.5
9-Apr	NA	NA	NA	NA	NA	NA	NA
10-Apr	NA	NA	NA	NA	NA	NA	NA
11-Apr	NA	NA	NA	NA	NA	NA	NA
12-Apr	NA	NA	NA	NA	NA	NA	NA
13-Apr	NA	NA	NA	NA	NA	NA	NA
14-Apr	NA	NA	NA	NA	NA	NA	NA
15-Apr	NA	NA	NA	NA	NA	NA	NA
16-Apr	NA	NA	NA	NA	NA	NA	NA
17-Apr	NA	NA	NA	NA	NA	NA	NA
18-Apr	NA	NA	NA	NA	NA	NA	NA
19-Apr	NA	NA	NA	NA	NA	NA	NA
20-Apr	NA	NA	NA	NA	NA	NA	NA
21-Apr	NA	NA	NA	NA	NA	NA	NA
22-Apr	383.1	594.1	763.4	861.6	881.9	852.2	711.2
23-Apr	278.9	404.3	577.7	677.9	656.7	668.2	711.6
Mean	**269.0**	**428.2**	**555.6**	**598.9**	**622.3**	**591.3**	**482.5**

One obvious choice is to fill the missing values with the mean.

Concept of the proposed method for missing-data

Date/Time	08-09	09-10	10-11	11-12	12-13	13-14	14-15
7-Apr	284.8	596.0	771.0	801.5	850.9	807.1	673.7
8-Apr	338.0	589.3	761.6	806.3	834.4	807.4	699.5
9-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
10-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
11-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
12-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
13-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
14-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
15-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
16-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
17-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
18-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
19-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
20-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
21-Apr	269.0	428.2	555.6	598.9	622.3	591.3	482.5
22-Apr	383.1	594.1	763.4	861.6	881.9	852.2	711.2
23-Apr	278.9	404.3	577.7	677.9	656.7	668.2	711.6
Mean	**269.0**	**428.2**	**555.6**	**598.9**	**622.3**	**591.3**	**482.5**

One obvious choice is to fill the missing values with the mean **(over yearly data)**.

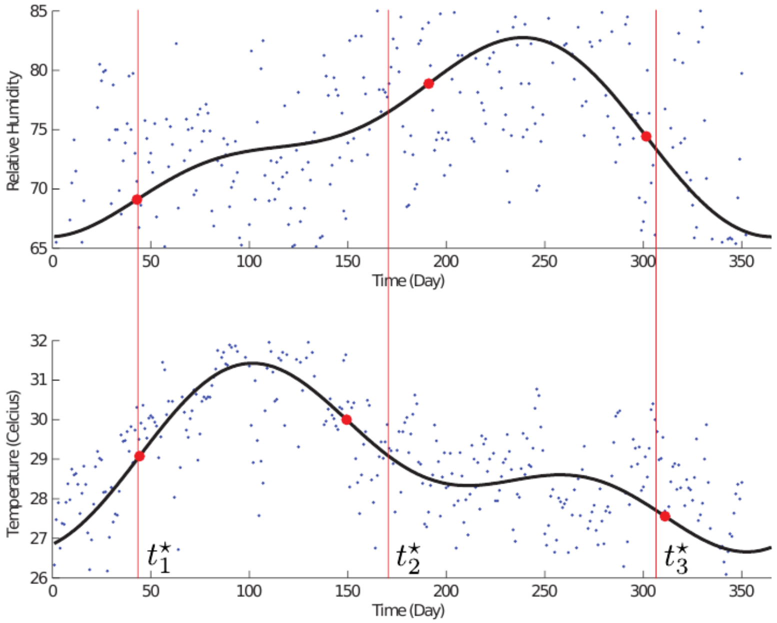


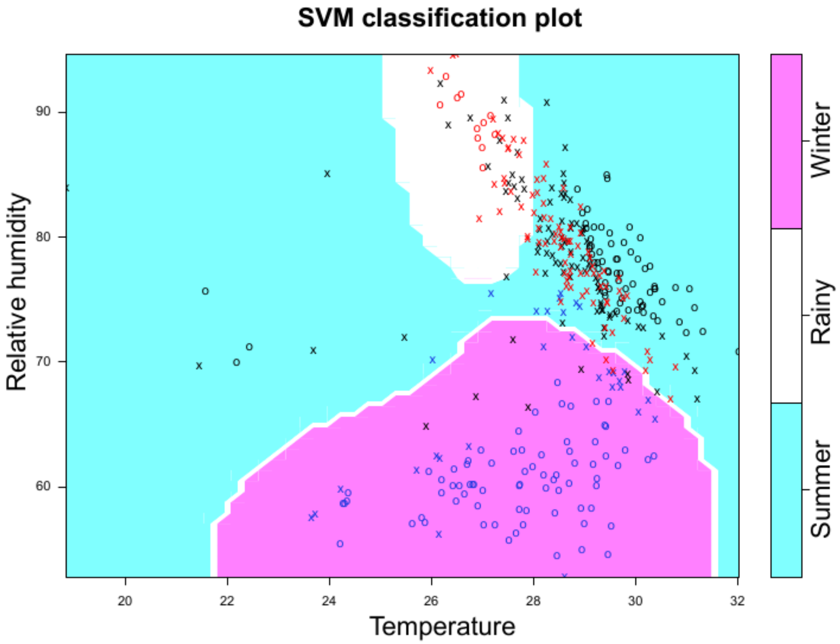
- A mean over yearly data is seem to be not a good representative as an imputed value due to different weather types of each day.
- The idea is that **the mean** should be the averaged irradiance over the values **from the same weather type**.

The proposed method for missing values

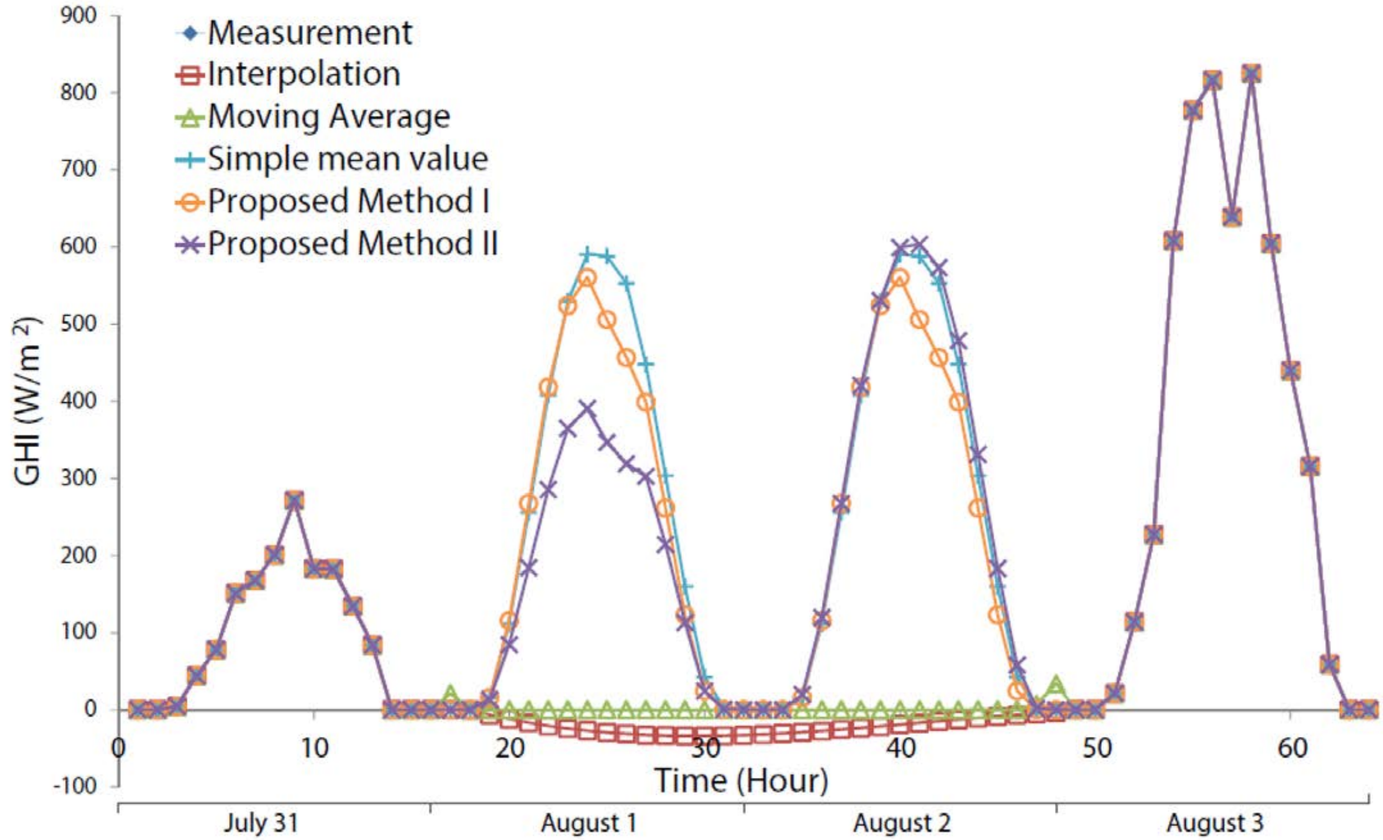
The required weather classicationvconsists of two steps:

1. a seasonal segmentation based on detecting changes of monotonic properties of temperature and humidity time series
2. a nonlinear support vector machine (SVM) that uses weather labels from the previous seasonal segmentation.





Missing-data imputation



- The second and third cycles are imputed.
- Both imputed cycles using simple mean value (over yearly data) are the same.
- Proposed method imputes both cycles differently according to different weather types of each day.

We assess the imputation methods by presumably **deleting the recorded values** from the data sets, and then **we evaluate how well the deleted values are predicted** in a yearly basis

Evaluation Measures

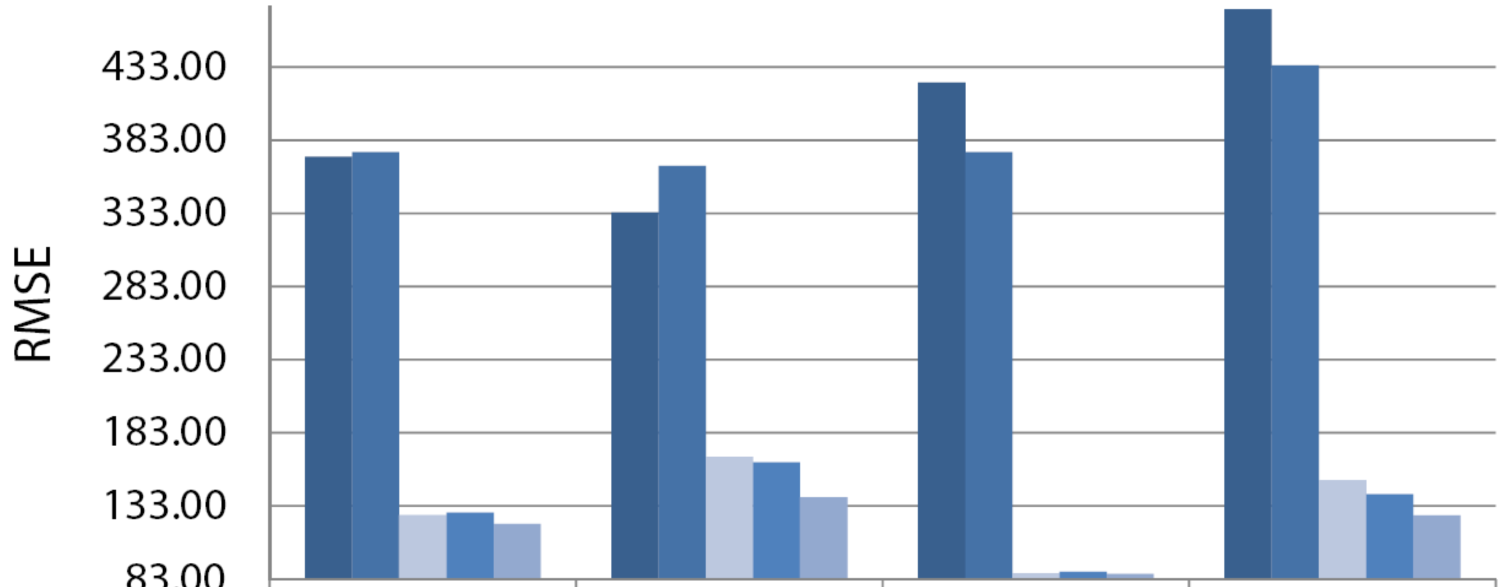
Common evaluation measures are **Root Mean Squared Error (RMSE)** and **Mean Absolute Error (MAE)** which are used to validate a forecasting method. RMSE and MAE can be defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(I(t) - \hat{I}\,(t)\right)^2}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|I(t) - \hat{I}\,(t)\right|$$

where N is the length of the time horizon. **Desired value of RMSE and MAE is minimized**.

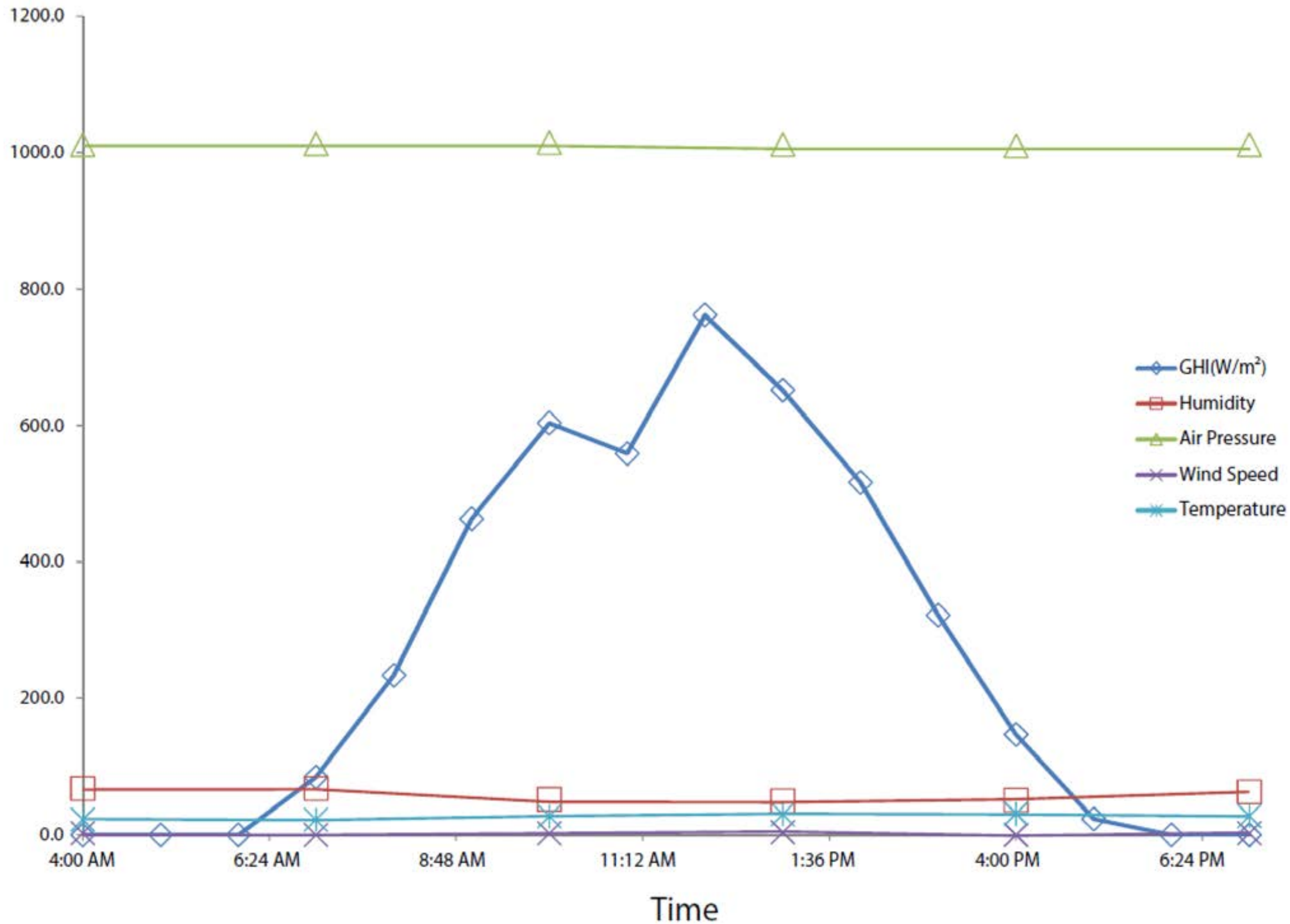
Imputation error



	2011	2012	2013	2014
Interpolation	371.79	333.61	422.41	472.42
Moving Average	374.72	365.64	374.90	434.10
Mean in one year	126.81	166.62	87.00	150.69
Proposed Method I	128.46	162.87	87.98	141.08
Proposed Method II	120.77	138.94	86.70	126.53

Asynchronous data

Asynchronous sampled data

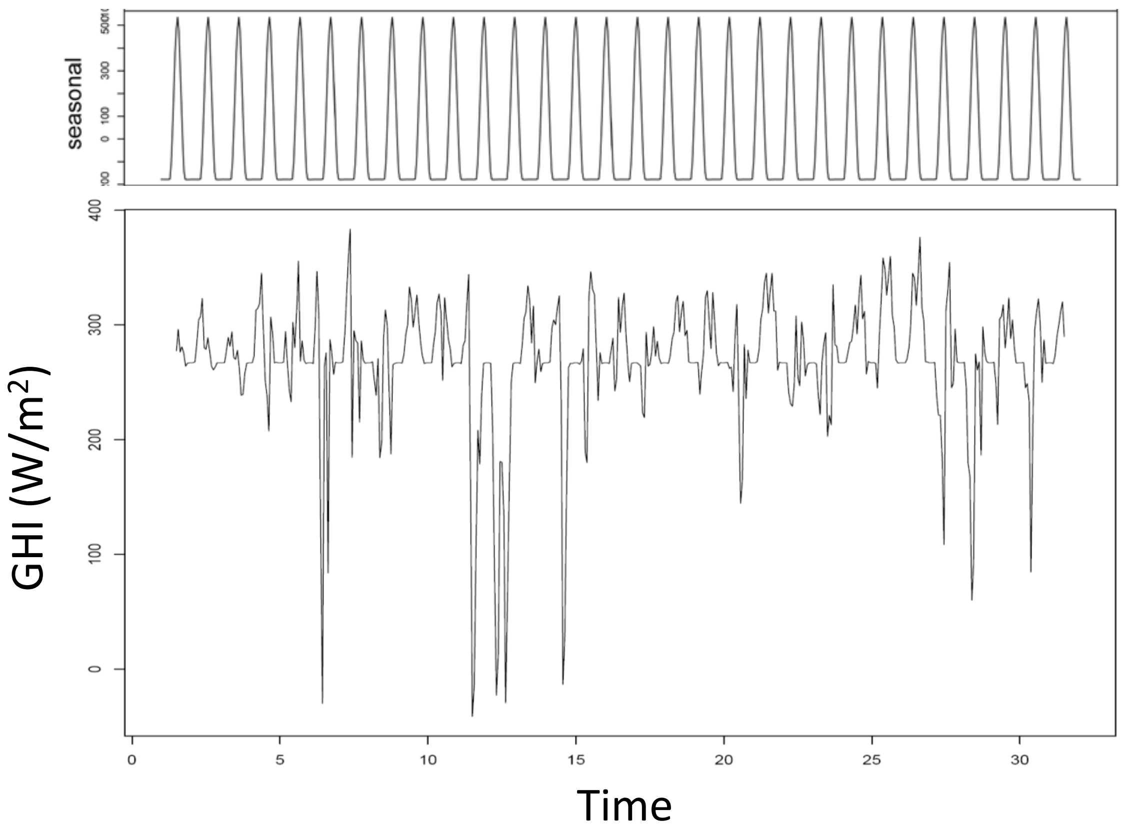


- GHI data were recorded hourly.
- Other meteorological variables were recorded 3-hourly.
- Cubic Spline Interpolation is used to impute exogenous variables to be at hourly sampled rate.

Time Series Models

- Models without removing seasonal trend
- Models with removing seasonal trend
    - Seasonal ARIMA models
    - Fitting of seasonal trends

GHI data has a seasonal trend



GHI after removing seasonal trend in January 2014 in Bangkok.

We imply that GHI can be described as ARMA models containing s season:

$$A(q^{-1})y(t) = \mathrm{s(t)} + \alpha + C(q^{-1})v(t)$$

Models without removing seasonal trend

Persistence Forecasts

Persistence forecast as a baseline prediction method is used for a comparison to more advanced methods. There are many equations of persistence forecast in accordance with each study.

$$\hat{I}(t+h) = I(t)$$

Persistence model supposes that **solar irradiance at time $t+h$ can be predicted by it value at time t.**

ARIMAX description

An autoregressive integrated moving average model with an exogenous input (ARIMAX) is employed to predict the future solar irradiance. The model ARIMAX(*p,d,q*) is defined by

$$A(q^{-1})(1-q^{-1})^d y(t) = B(q^{-1})u(t) + \ C(q^{-1})v(t)$$

where

- $A(q^{-1}) = I - (a_1 q^{-1} + a_2 q^{-2} + \cdots + a_p q^{-p})$, Autoregressive term
- $B(q^{-1}) = B_1 q^{-1} + B_2 q^{-2} + \cdots + B_m q^{-p}$, Exogenous term
- $C(q^{-1}) = I + c_1 q^{-1} + c_2 q^{-2} + \cdots + c_q q^{-p}$, Moving Average term
- $(1-q^{-1})^d$, Integrated term

Model Estimation and Selection

Model Estimation

- The Maximum likelihood method (ML) is applied to estimate the parameters of models.
- The ML estimation is nonlinear optimization problem.

Model selection

- The AIC is defined as

$$\mathrm{AIC} = 2L + 2d$$

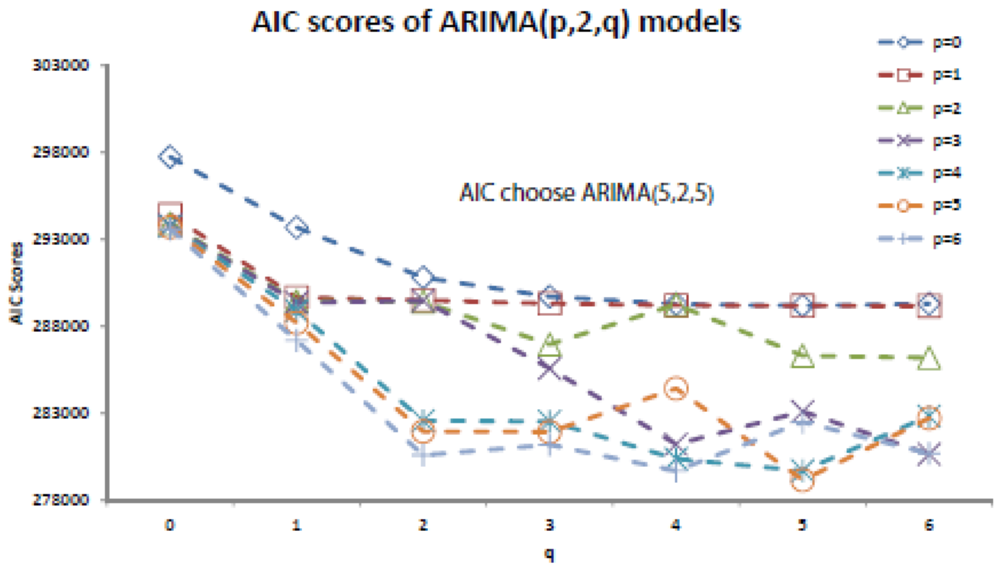
where L is the loglikelihood function and d is the number of effective parameters.

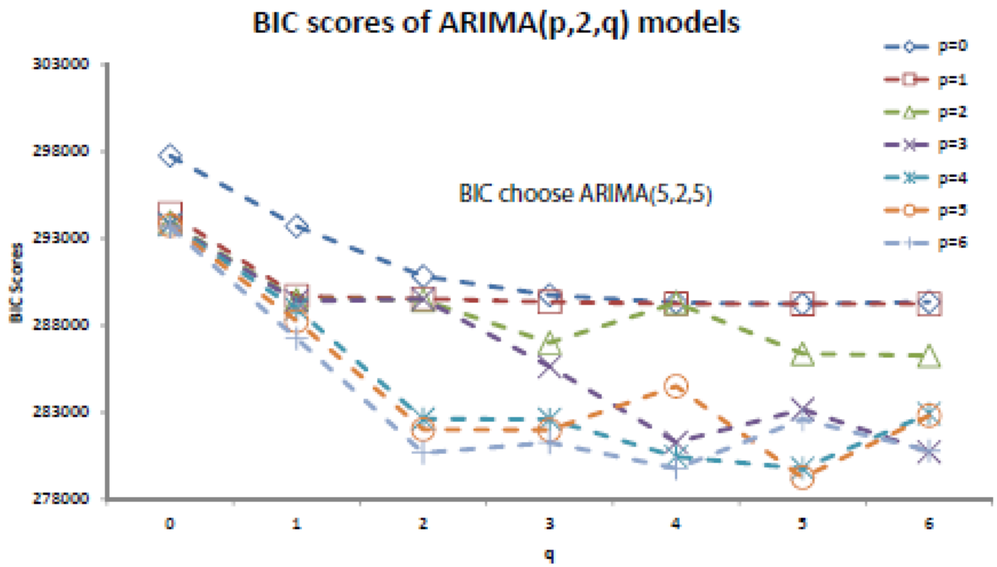
- The BIC is defined as

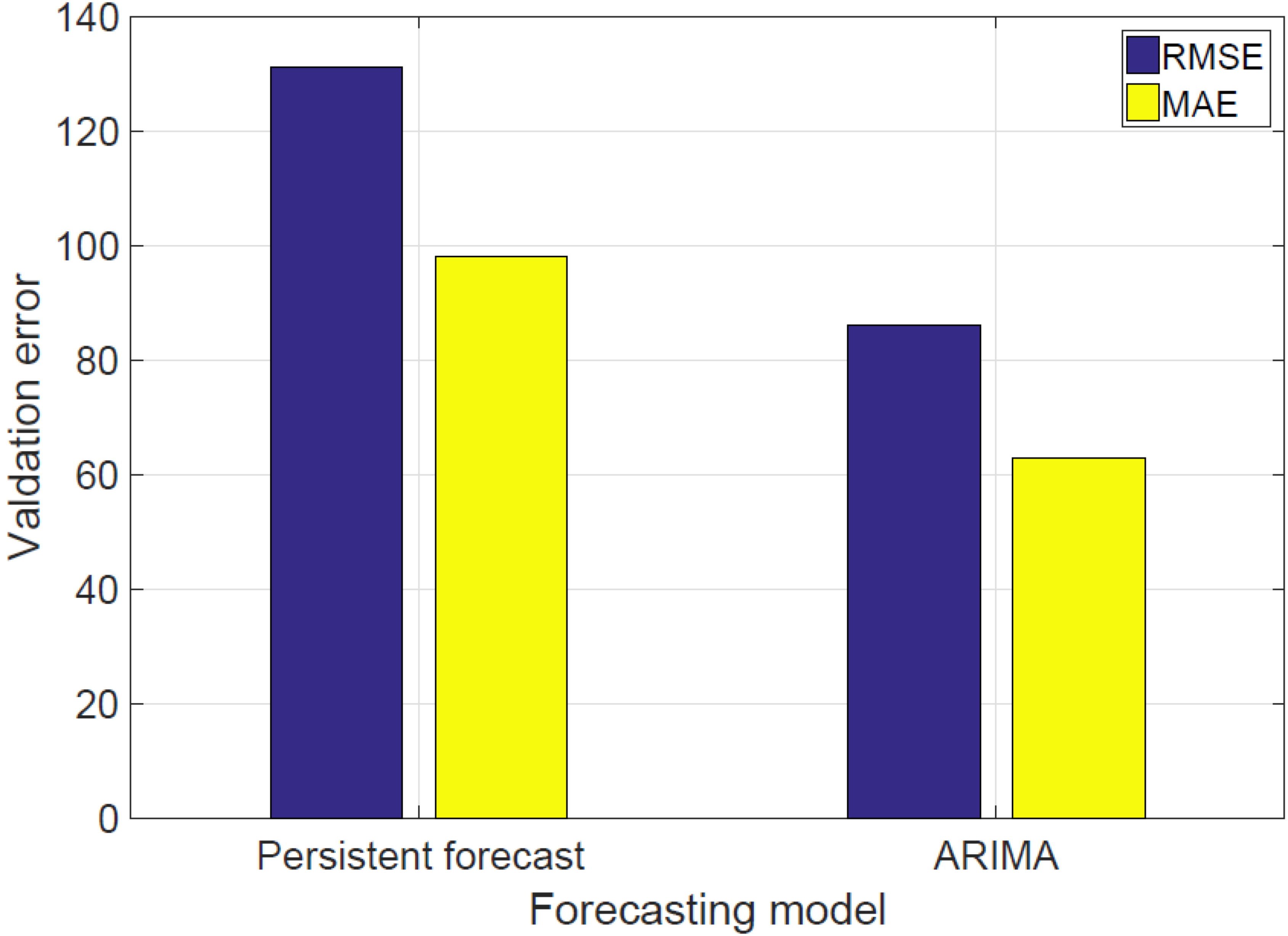
$$\mathrm{BIC} = 2L + d\log N$$

where N is the number of sample.

Models without removing seasonal trend







140
120
100
80
60
40
20
0
Valdation error
RMSE
MAE
Persistent forecast
ARIMA
Forecasting model

Models with removing seasonal trend

Seasonal ARIMA models

Seasonal ARIMA model

In this study, **we used a seasonal ARIMA models to remove a seasonal trend.** The seasonal term and constant are removed by using this transformation. This method is called a Seasonal ARIMA $(p, d, q)(P, D, Q)_T$ models which can be defined as

$$\tilde{\boldsymbol{A}}(\boldsymbol{q}^{-\boldsymbol{T}})A(q^{-1})(\boldsymbol{1}-\boldsymbol{q}^{-\boldsymbol{T}})^{\boldsymbol{D}}(1-q^{-1})^{d}y(t)=\tilde{\boldsymbol{C}}(\boldsymbol{q}^{-\boldsymbol{T}})C(q^{-1})v(t)$$

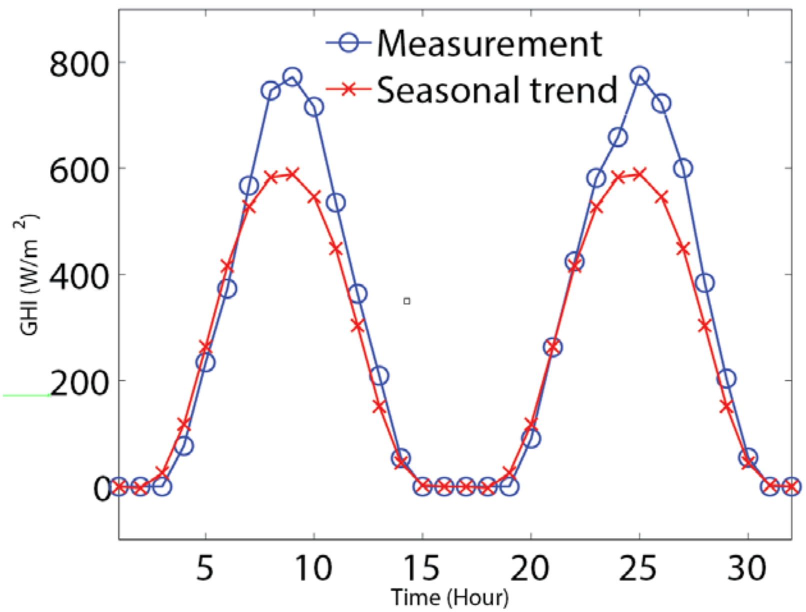
where

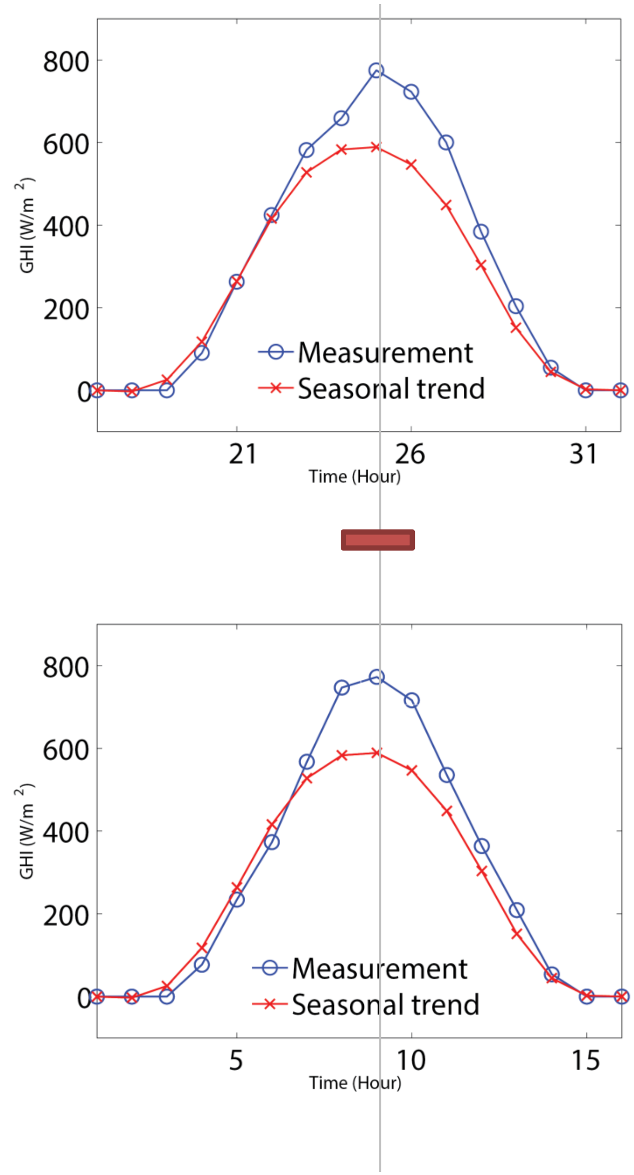
$$\tilde{A}(q^{-1})=I-(\tilde{a}_1 q^{-T}+\tilde{a}_2 q^{-2T}+\cdots+\tilde{a}_p q^{-pT})$$
$$\tilde{C}(q^{-1})=I+\tilde{c}_1 q^{-T}+\tilde{c}_2 q^{-2T}+\cdots+\tilde{c}_q q^{-qT}$$

T is a seasonal period (giving $T=16$ as daily cycle) and D is integrated seasonal order.

Seasonal ARIMA model

$$(1 - q^{-T})y(t) = \mathrm{y(t)} - \mathrm{y(t-T)}$$



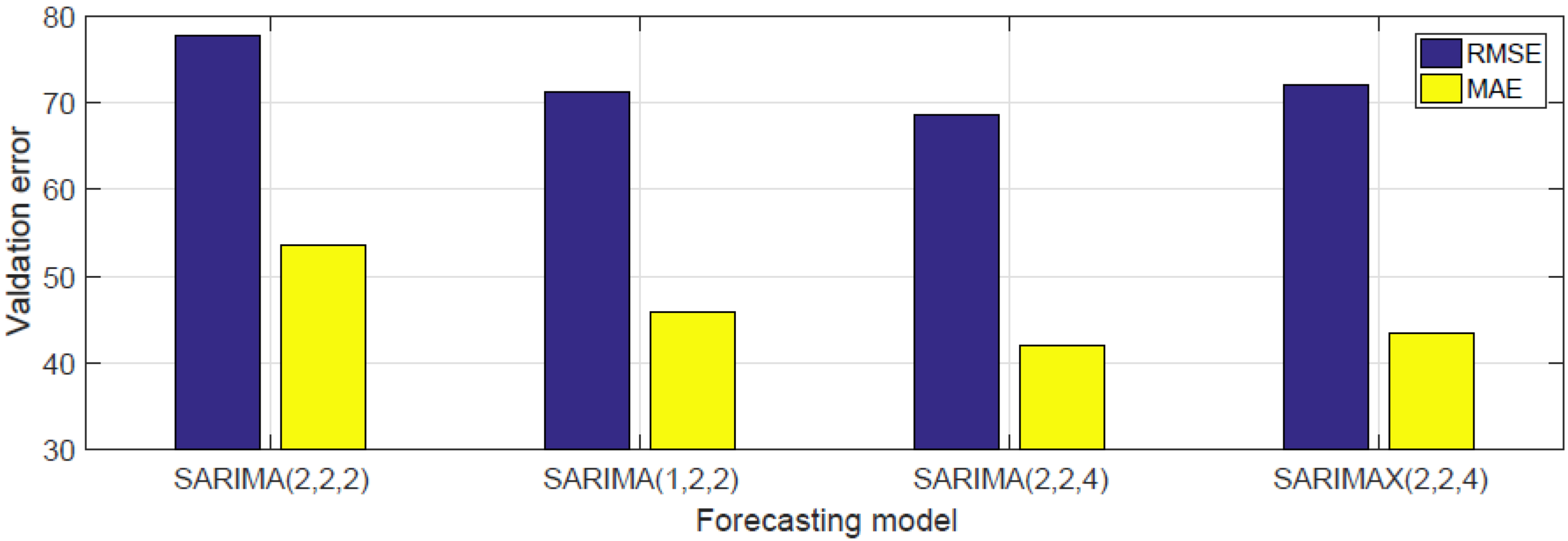


$\mathrm{y(t)}$

$\mathrm{y}(\mathrm{t} - T)$

Assuming the seasonal trend is known.

Seasonal ARIMA models



- Exogenous terms consist of temperature, relative humidity, air pressure and wind speed.
- Exogenous terms of ARIMAX or SARIMAX marginally affect the forecasting performance.

- To apply Seasonal ARIMA models, a cycle time (T) is fixed considered by priority information.
- Seasonal trend may consist of many cycles.
- The idea is to fit seasonal trend, and then subtract it from data.

Fitting of seasonal trends

Fitting of seasonal trend

We imply that GHI can be described as ARMA models containing s season:

$$A(q^{-1})y(t) = \mathrm{s(t)} + \alpha + C(q^{-1})v(t)$$

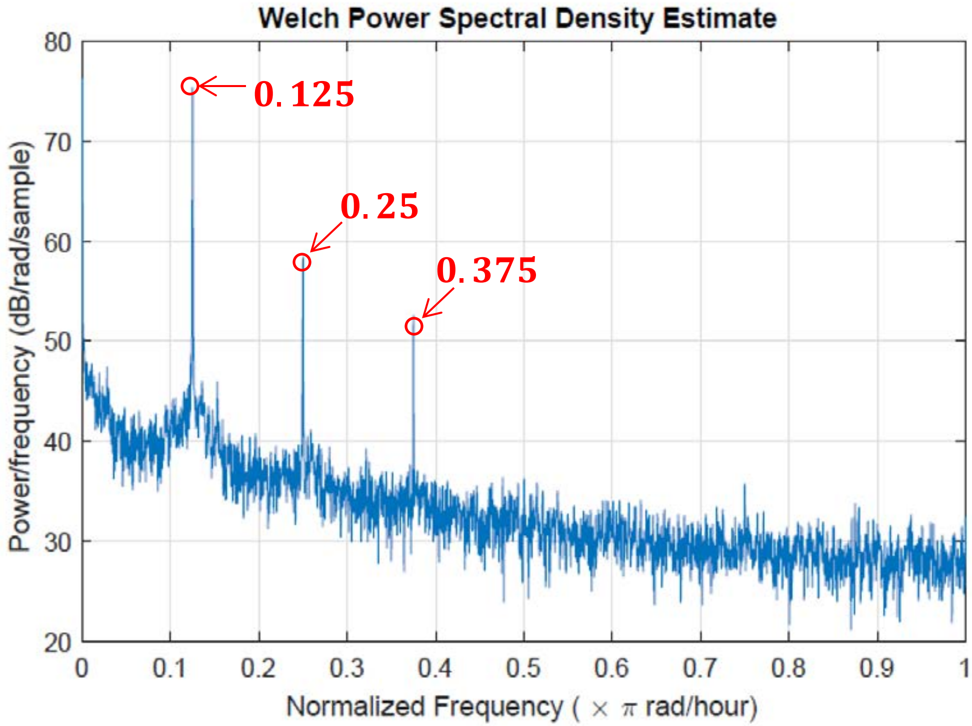
The seasonal term can be expressed as sine wave of certain frequencies f_i. It can be written as

$$s(t) = \sum_{i=1}^{M} \sigma_i \sin \omega_i t + \beta_i \cos \omega_i t$$

where $i = 1,2,3, \dots, t = 0,1,2, \dots, N$,
σ_i is the coefficient of sine component of each frequency ω_i,
β_i is the coefficient of cosine component of each frequency ω_i

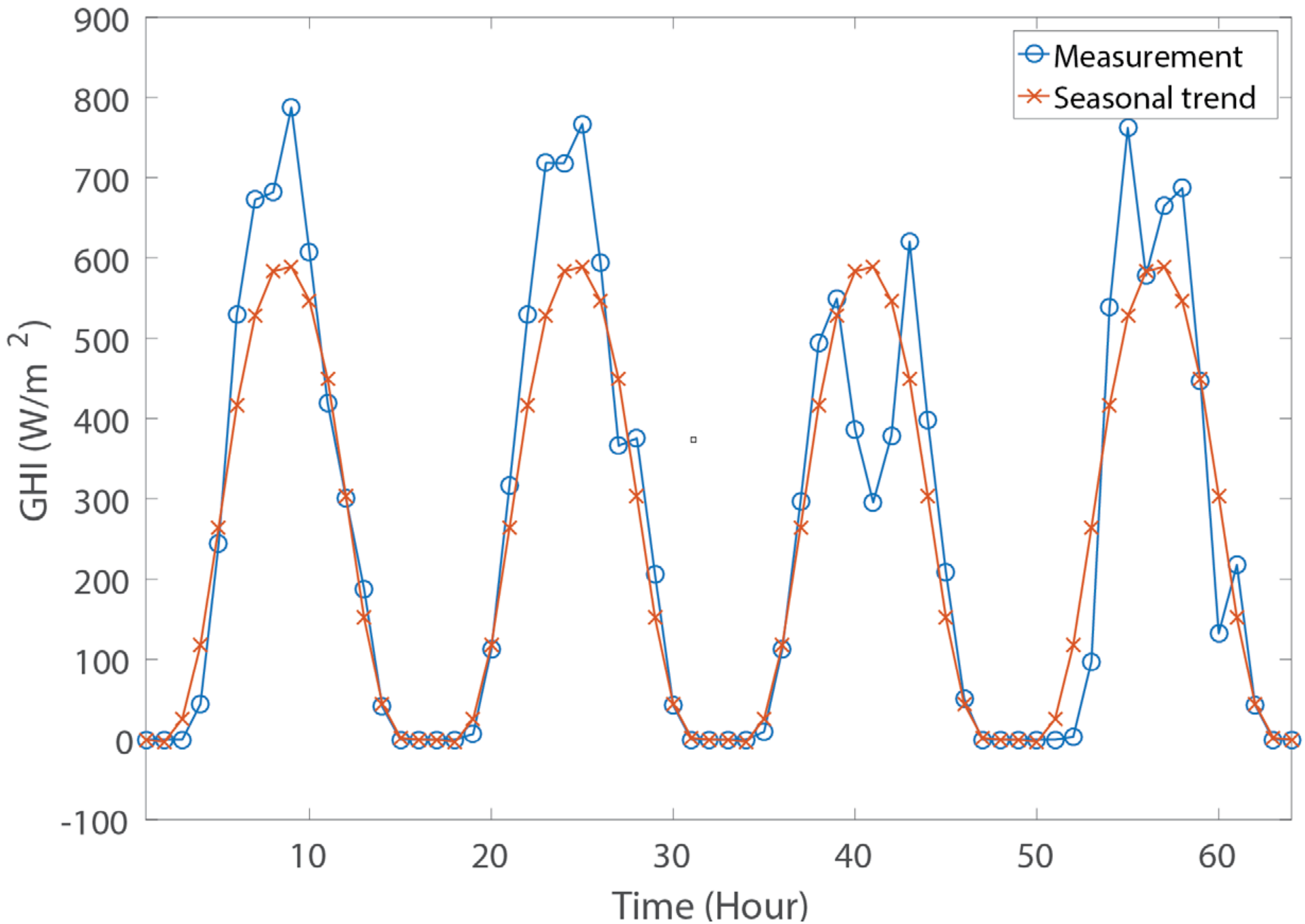
Power spectral analysis

- Power spectral analysis is performed to find a dominant frequency of GHI data.

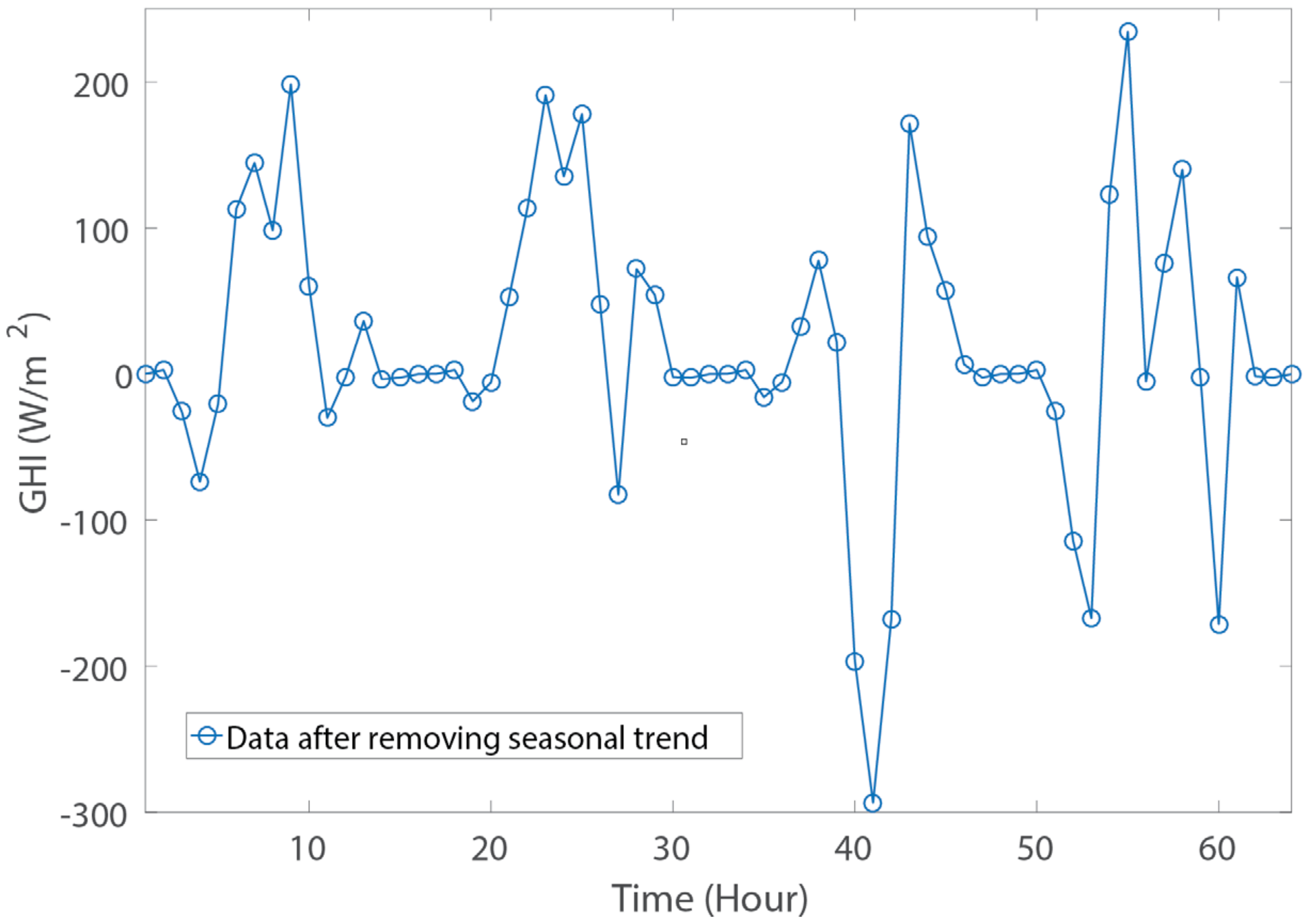


- **The signal repeats in every 16 hours, 8 hours, and 5.3 hours.**
- Other unknown parameters are computed using method of least square after frequencies ω_i are known.

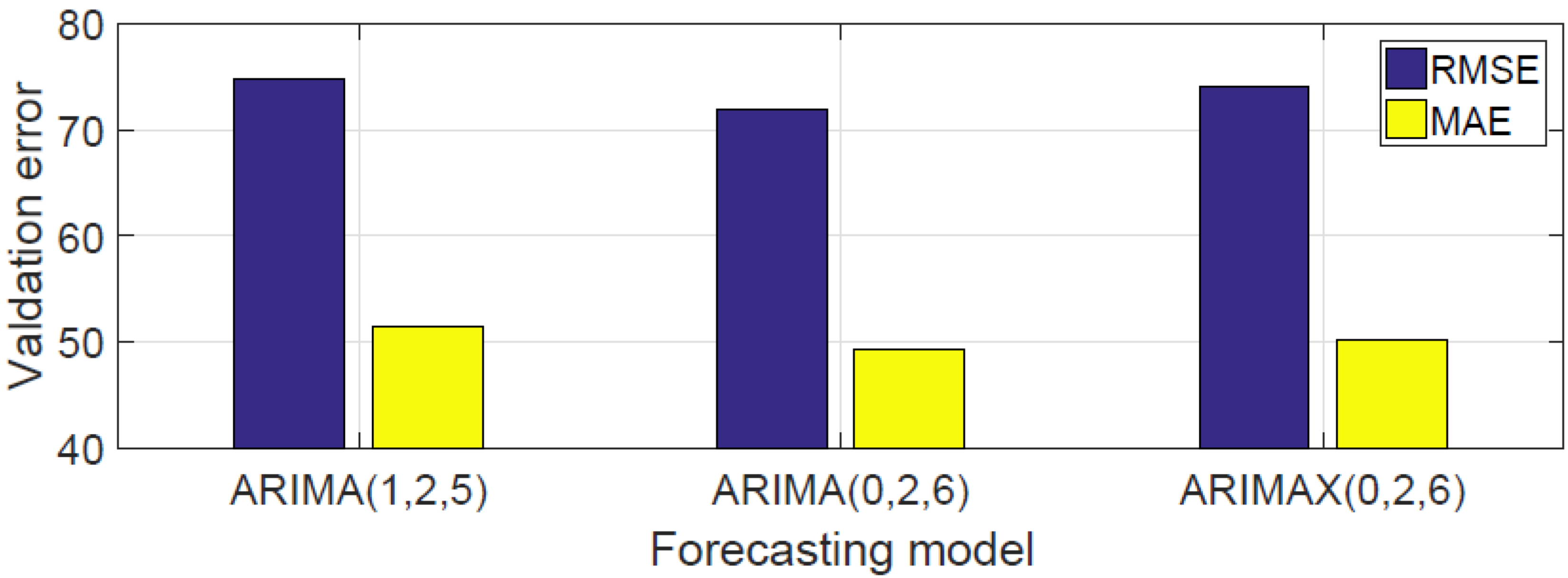
Plot of fitted seasonal model



Plot of data after removing seasonal trend



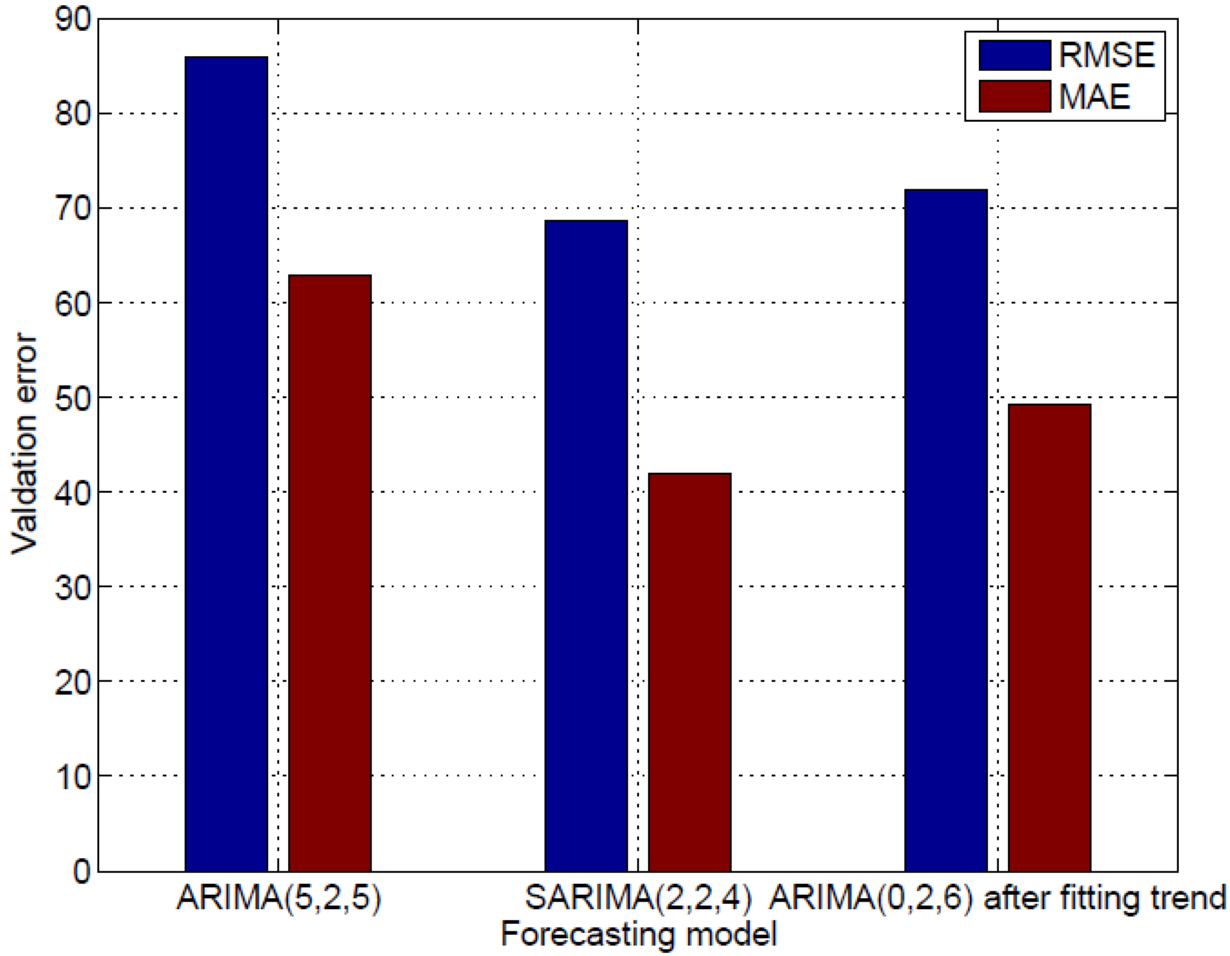
Fitting of seasonal trend



- Exogenous terms consist of temperature, relative humidity, air pressure and wind speed.
- Exogenous terms of ARIMAX or SARIMAX marginally affect the forecasting performance.

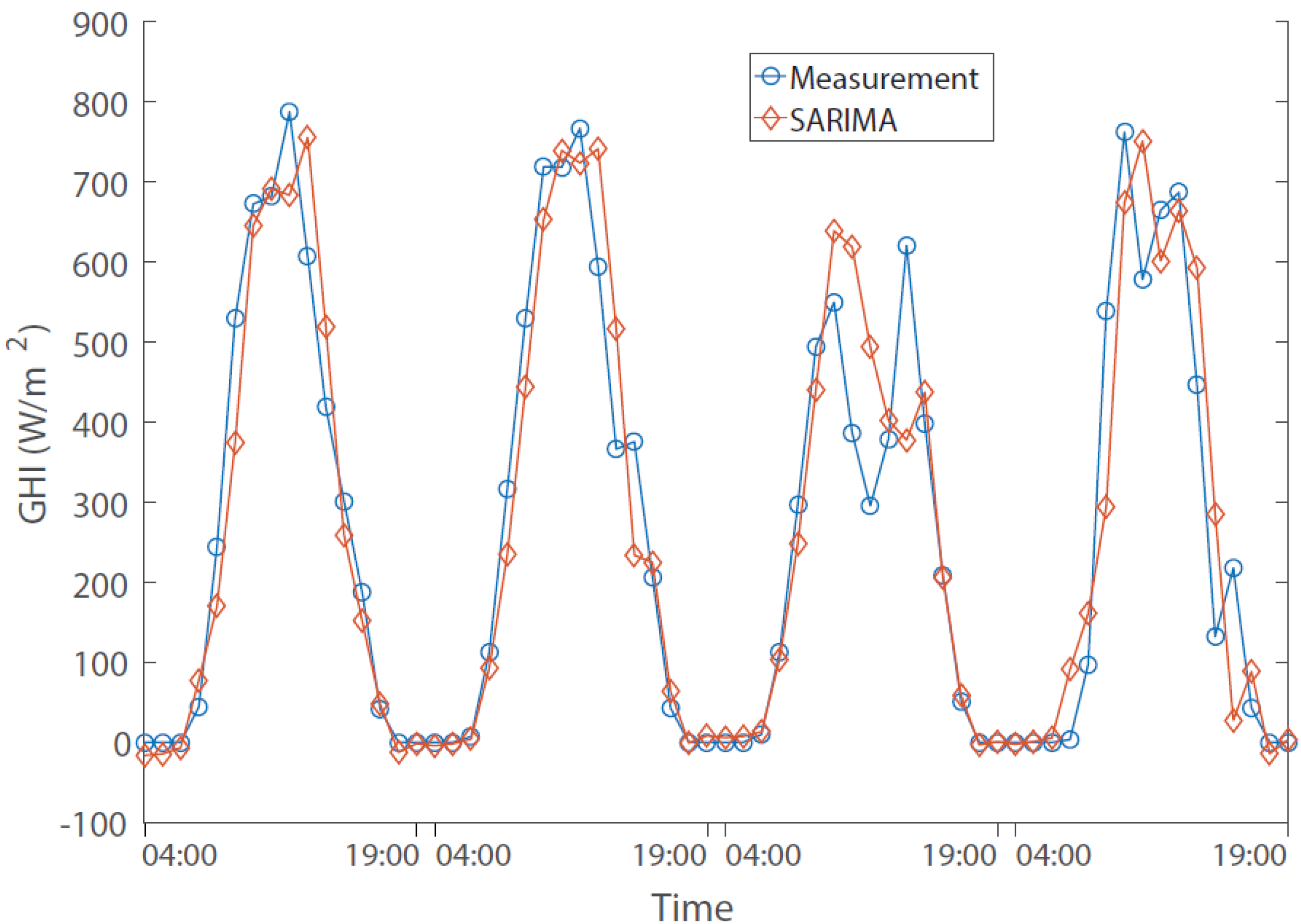
Conclusions

- Historical data contain missing data which are recommended to be imputed using our proposed method.
- Seasonal effect has to be considered in solar forecasting using time series model.
- Seasonal removal improves the forecast.
- Exogenous terms have marginal effect in forecasting solar irradiance.



we have an recommendation to use SARIMA(2,2,4) as the forecasting model. The equation of this model is

$$(1 + 0.40q^{-1} - 0.58q^{-2})I(t) = (1 - 0.94q^{-16})(1 - 0.83q^{-1} - 1.11q^{-2} + 0.83q^{-3} + 0.12q^{-4})v(t)$$



Holistically, the forecast is statistically accurate except when the measured GHI fluctuated sharply.

Q&A