Senior Project

A comparison of intraday solar power forecasting methods

Chayanont Potawananont ID 5930084921
Sararut Pranonsatid ID 5930515021

Advisor : Assist. Prof. Jitkomut Songsiri

Department of Electrical Engineering
Faculty of Engineering, Chulalongkorn University

Table of contents

01
Introduction

02
Methodology

03
Result & Discussion

04
Conclusion

01 Introduction

Why intraday forecasting is essential ?

The volatile nature of solar resource has posed the difficulties in grid management as solar penetration rate grows continuously.

An intraday forecasting task becomes important for

1. Power plant operation
2. Grid balancing
3. Real-time unit dispatching

Objective

- To study the relevant variables of intraday solar irradiance forecasting.

- To compare forecasting models including LR, MARS, ANN, SVR, RF in the aspects of
  - Forecasting performance
  - Computational complexity

02 Methodology

Notation:

- I: solar irradiance $[\mathrm{W/m^2}]$
- P: solar Power $[\mathrm{kW}]$
- θ: solar zenith angle
- $x(t)$: x at execution time
- $x(t+k)$: x in 30k-min ahead
- $x^{(\mathrm{d}-1)}(t)$: value of x in day $d-1$ at time t
- $\hat{x}(t)$: forecasted value of x at time t

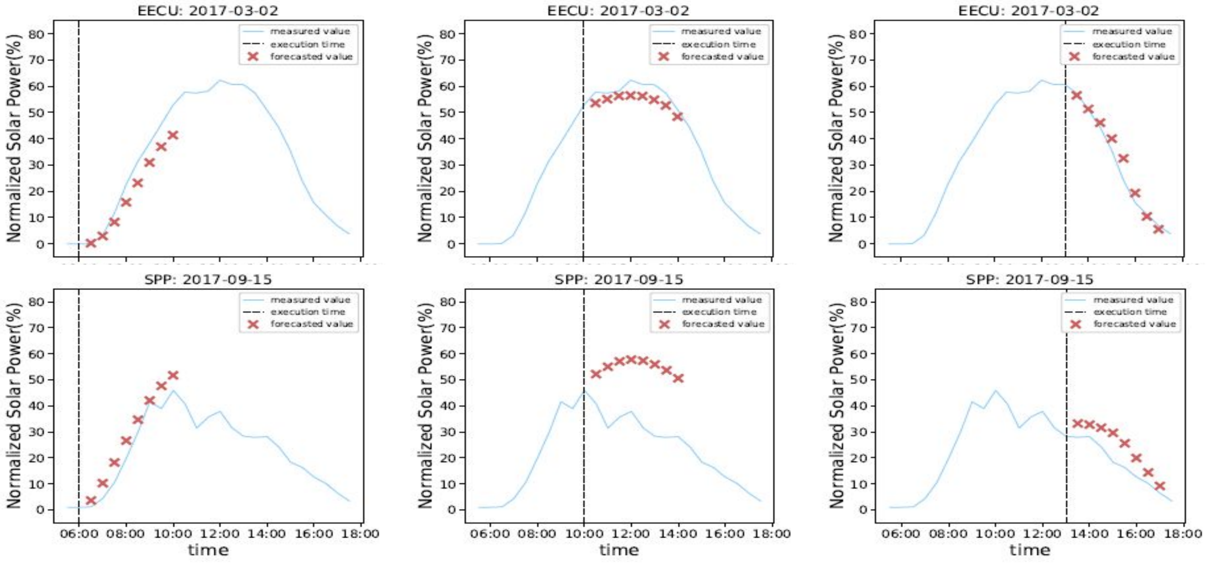
Forecasting Configuration

Goal: predict solar power with the horizon of 4 hours every 30 minutes

$$\hat{P}(t+1), \hat{P}(t+2), \ldots, \hat{P}(t+8)$$

Time of forecast values: 6:00 - 17:30: every 30 min

Execution time: 5:30 - 17:00



In clear sky days,
the solar power prediction is quite straightforward.

In general days,
the solar power prediction
Is a challenging problem especially at midday.

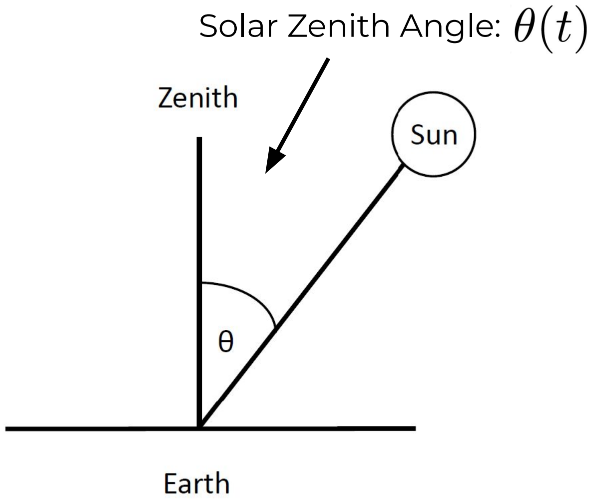
Clear sky models

Clear sky models estimate the terrestrial solar radiation under a cloudless sky as a function of the solar elevation angle and various atmospheric conditions.

Berger-Duffie model [Vio97]: $I_{\mathrm{clr}}(t) = 956.27\cos(\theta(t)))$

ASHRAE model [PC07]: $I_{\mathrm{clr}}(t) = Ke^{-B\sec(\theta(t))}$

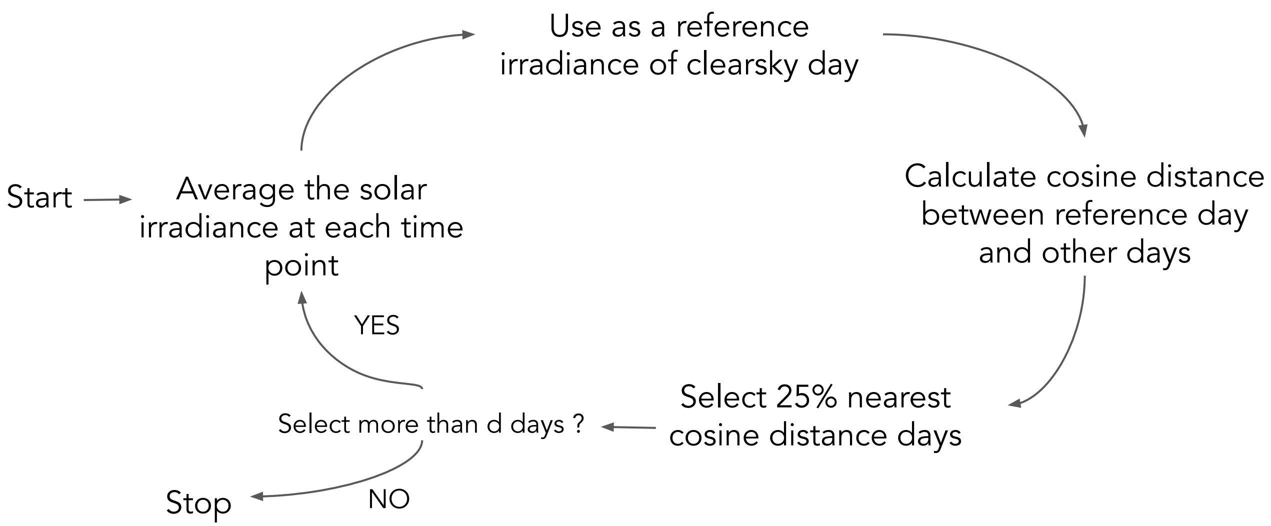
K, B are constant estimated by measurement data.

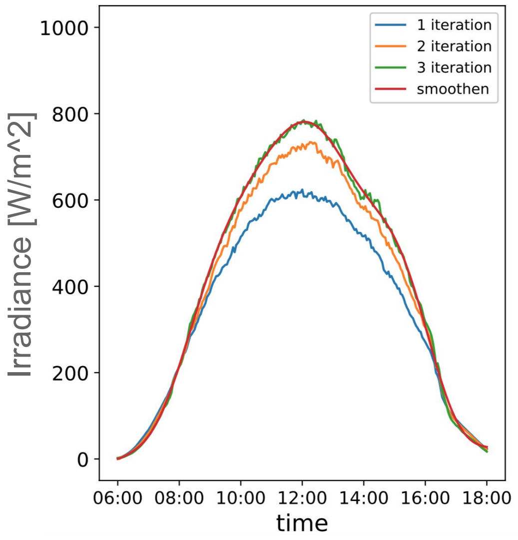


Clear sky detection algorithm

Concept: we will find the reference clearsky day then select the days that are similar to this day as a clearsky day. [we use cosine distance to measure similarity]

- Finding reference clearsky day



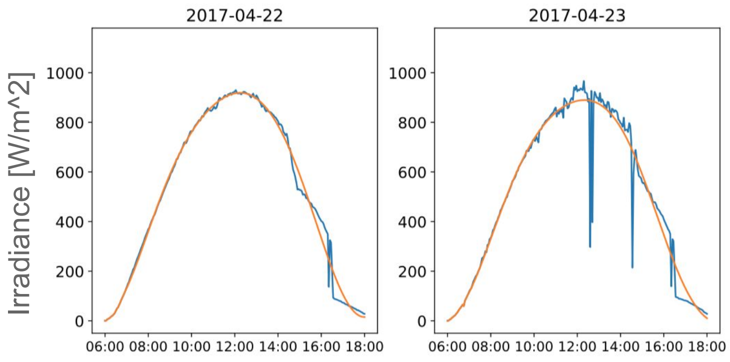


Example of reference clearsky days : EECU

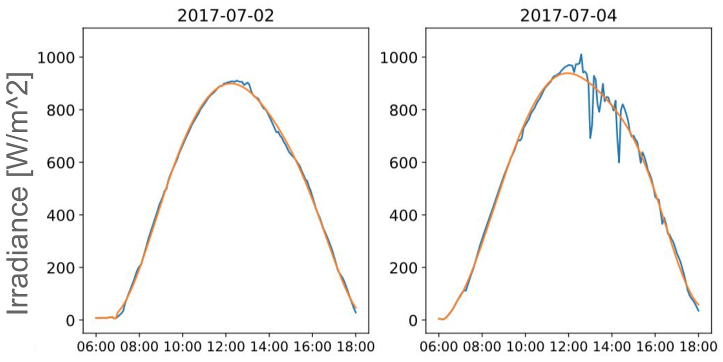
Clear sky detection algorithm

Concept : we will find the reference clearsky day then select the days that are similar to this day as a clearsky day. [we use cosine distance to measure similarity]

- Select clearsky days: set cosine distance threshold for selecting clearsky day
- Smoothen the data: apply bi-directional butterworth low-pass filter



Example of selected clearsky days : EECU

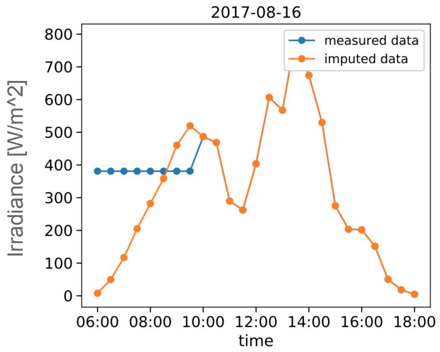


Example of selected clearsky days : SPP

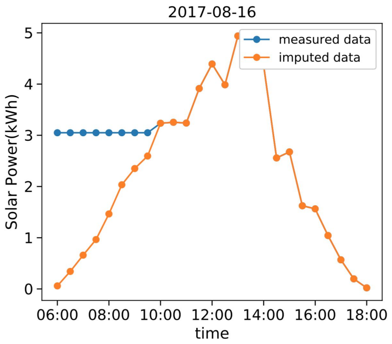
Data preprocessing: data verification

We found abnormal patterns in some irradiance and power data

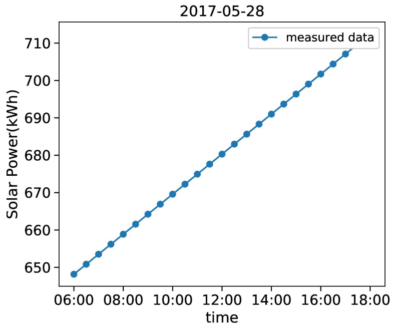
1. Data are held constant more than 2 hrs (but not entire day).
    - Replace with an (±15 days) average of solar irradiance at that time point
2. Data are held constant for entire day or have unnatural pattern.
    - Delete an information of that day



Example of constant irradiance : EECU

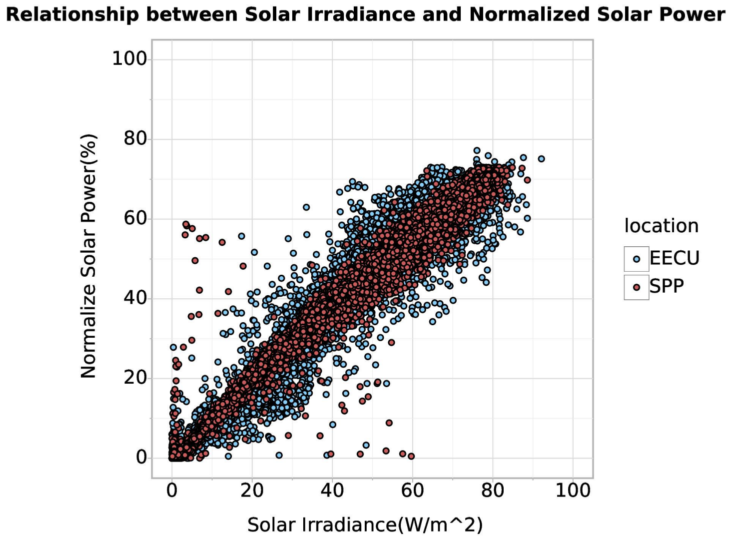


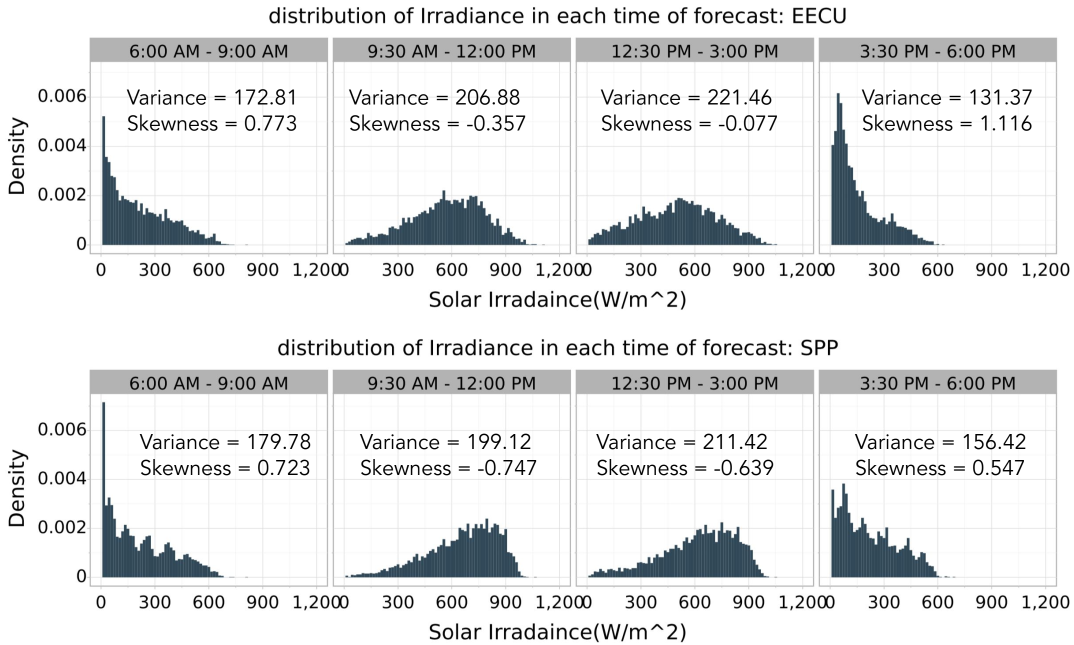
Example of constant power : EECU



Example of unnatural pattern in power : SPP

Data pre-analysis





The Irradiance and power at SPP have stronger relationship than at EECU

The distribution can be separated into 3 groups: morning/midday/evening

The SPP have lower variance/skewness than EECU

Forecasting technique

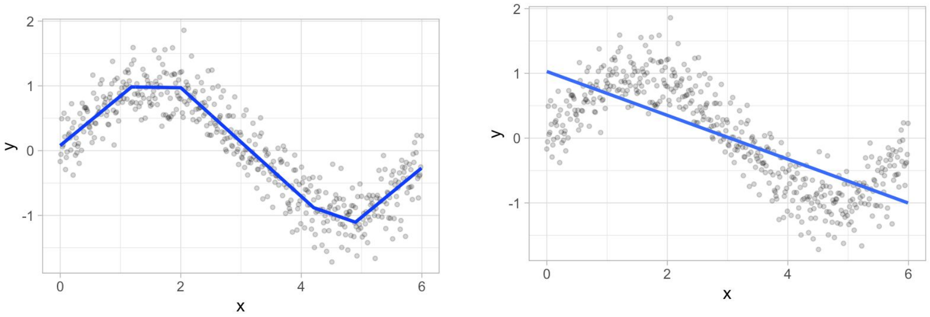
- Baseline models
  - Linear regression (LR)
  - Multivariate adaptive regression splines (MARS)
  - Artificial neural network (ANN)

- Proposed models
  - Support vector regression (SVR)
  - Random Forest (RF)

Baseline model: Linear Regression & MARS

MARS:

MARS capture the nonlinear relationships in the data by assessing cutpoints (knots) using expansions in piecewise linear basis functions [FHT01]



Example of MARS & Linear regression

Target:

- $\hat{I}(t+1), \hat{I}(t+2), \ldots, \hat{I}(t+8)$

Input:

- $I(t), I(t-1), \ldots, I(t-7)$
- $I^{(\mathrm{d}-1)}(t+k), k=1,2,\ldots,8$
- $I_{\mathrm{clr}}(t+k), k=1,2,\ldots,8$

Baseline model: Artificial Neural Network (ANN)

Artificial neural network model proposed from SGRU has the following architecture

1. Input layer: consist of
   - $I(t)$
   - $I_{\mathrm{ema}}^{(\mathrm{d}-1)}(t+k), k=1,2,...,8$
   - $P(t)$
   - $T(t)$
2. Hidden layer: fully connected 5 hidden layers, each hidden layer consist of 128 neurons
3. Output layer:
   - $\hat{I}(t+1), \hat{I}(t+2), \ldots, \hat{I}(t+8)$

Proposed model: Random Forest & SVR

Random Forest (RF)

Random forest is an ensemble regressor that consists of many regression trees.[FHT01]

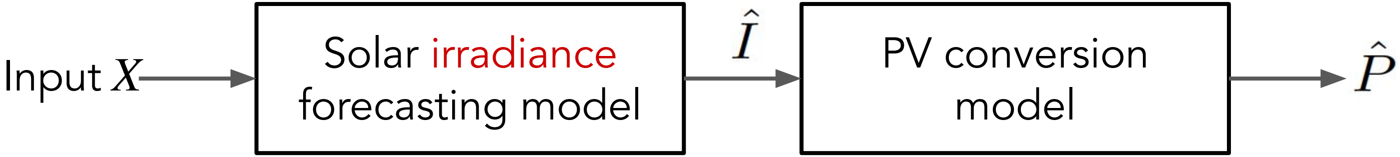
Support Vector Regression

SVR is an regression technique based on the concepts of support vector. The idea is to find the function that minimize the ε-loss function and at the same time, as flat as possible [SS04]

This 2 approaches are split and responsible for providing forecasting values at morning, midday and evening.

Forecasting approaches

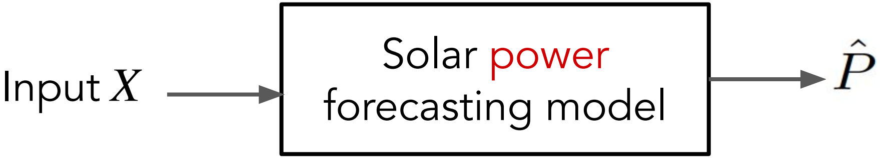
Indirect approach



- Linear regression
- MARS
- SVR
- Random forest
- ANN

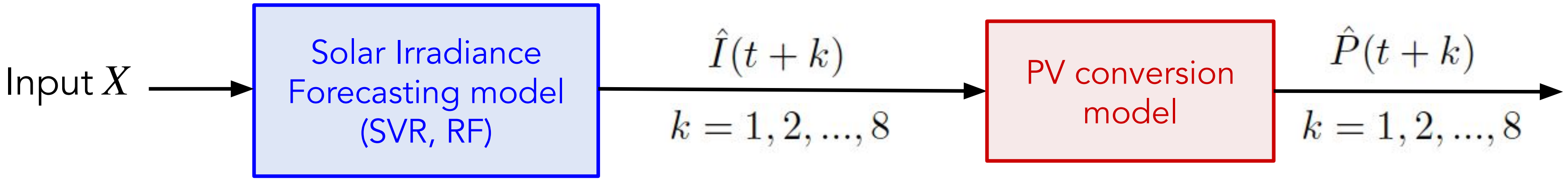
Polynomial regression

Direct approach



- Linear regression
- MARS
- SVR
- Random forest
- ANN

Indirect approach



Morning model:

Input

- $I(t)$
- $I^{(\mathrm{d}-1)}(t+k)$
- $\cos(\theta(t+k))$
- $I_{\mathrm{clr}}(t+k)$

Midday model:

Input

- $I(t), I(t-1), ..., I(t-7)$
- $I_{\mathrm{ema}}(t)$
- $I^{(\mathrm{d}-1)}(t+k)$
- $\cos(\theta(t+k))$
- $I_{\mathrm{clr}}(t+k)$

Evening model:

Input

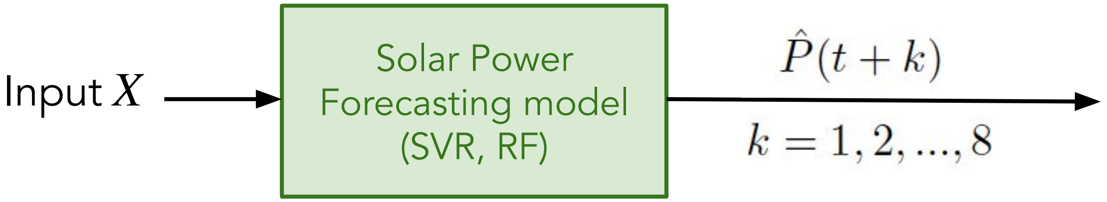
- $I(t)$
- $I^{(\mathrm{d}-1)}(t+k)$
- $\cos(\theta(t+k))$
- $I_{\mathrm{clr}}(t+k)$

Polynomial model:

$$P(I) = a_1 I + a_2 I^2 + a_3 I^3$$

When a_1, a_2, a_3 are constant estimated by measurement data.

Direct approach



Morning model:

Input

- $I(t)$
- $P^{(\mathrm{d}-1)}(t+k)$
- $P(t)$
- $I_{\mathrm{clr}}(t+k)$
- $\cos(\theta(t+k))$

Midday model:

Input

- $I(t), I(t-1), ..., I(t-7)$
- $P(t), P(t-1), ..., P(t-7)$
- $P_{\mathrm{ema}}(t)$
- $I_{\mathrm{clr}}(t+k)$
- $P^{(\mathrm{d}-1)}(t+k)$
- $\cos(\theta(t+k))$

Evening model:

Input

- $I(t)$
- $P^{(\mathrm{d}-1)}(t+k)$
- $P(t)$
- $I_{\mathrm{clr}}(t+k)$
- $\cos(\theta(t+k))$

03 Result & Discussion

Performance comparison between direct & indirect approach

- NRMSE versus prediction horizon at EECU/SPP

  Represent the forecast accuracy in each prediction horizon

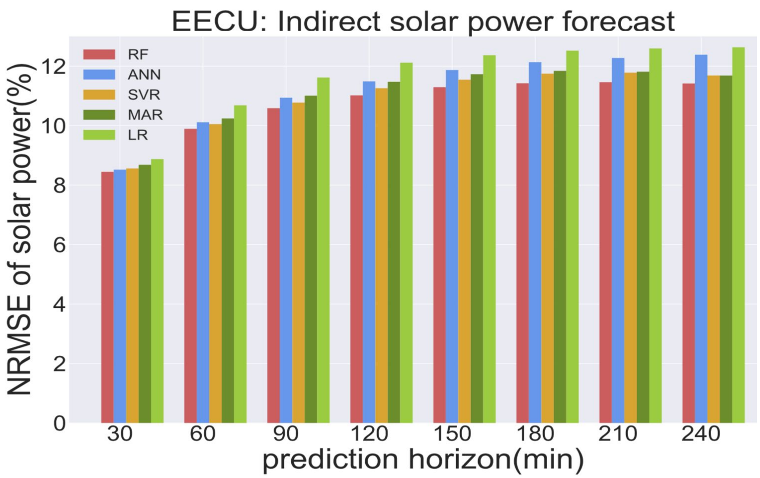
- NRMSE versus time of forecasted values at EECU/SPP

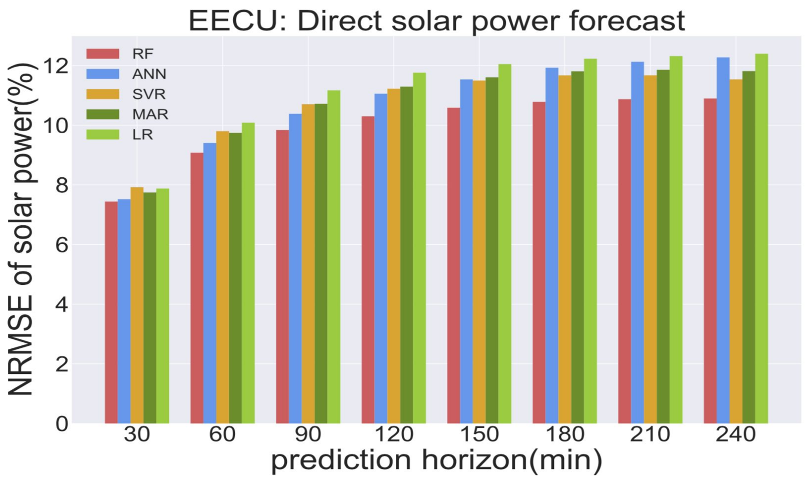
  Represent the forecast accuracy in each time point

- NMBE versus time of forecasted values at EECU/SPP

  Show that the model overestimate/underestimate solar the power which may lead to over operating cost and lacking power in power distribution system respectively

Performance comparison between direct & indirect approach
- NRMSE versus prediction horizon at EECU



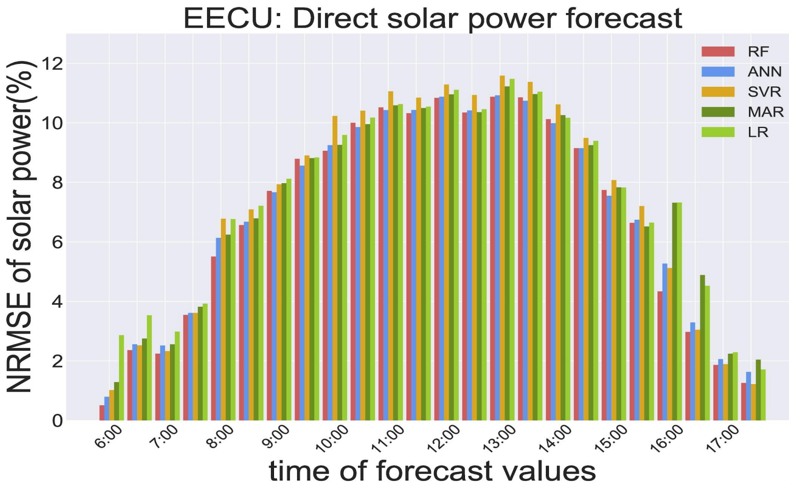


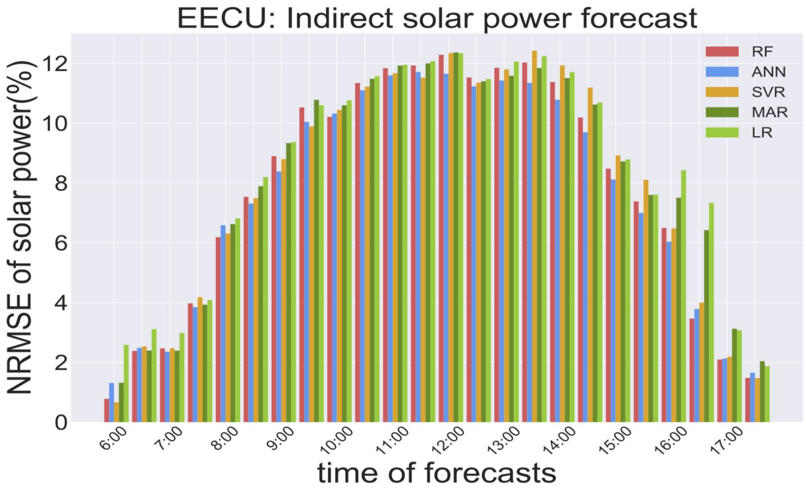
- Direct approach: random forest achieved the best performance. (7.44% in 30-min ahead)
- Linear regression which is the simplest model have the worst performance.
- The proposed models achieved the better performance than the baseline and showed the competing result with the former literature.

(1-hour ahead) Direct approach random forest : 9.08% , Direct approach SVR : 9.80%
[VKSB16] SARIMA : 8.12% , [BMP13] (SARIMA-SVR) 9.40%

Performance comparison between direct & indirect approach

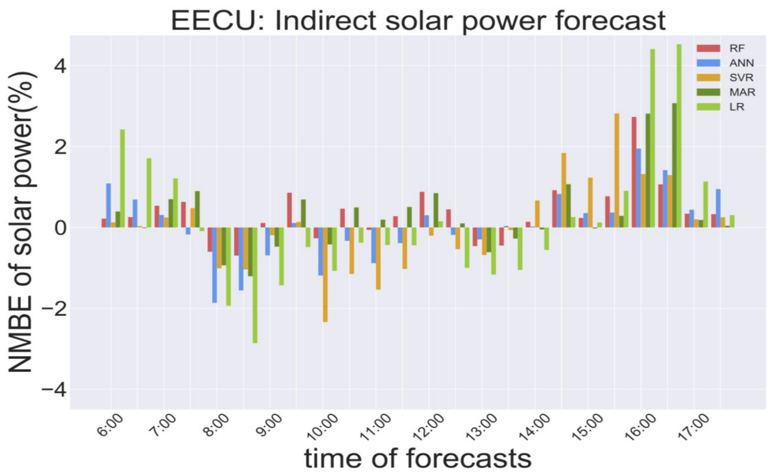
- NRMSE versus time of forecasted values at EECU

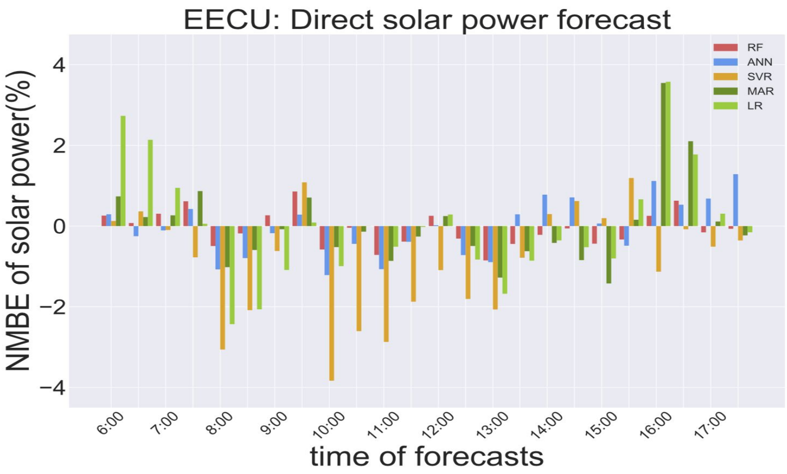




- NRMSE in the midday is usually higher than in the morning/evening
- The worst performance appear at 1:00-2:00 p.m. in both direct & indirect approach.
  - Direct approach RF: 10.12%
  - Direct approach SVR: 10.61%
- Linear regression and MARS reported high error at 4:00-5:00 p.m. due to rapid drop of solar irradiance

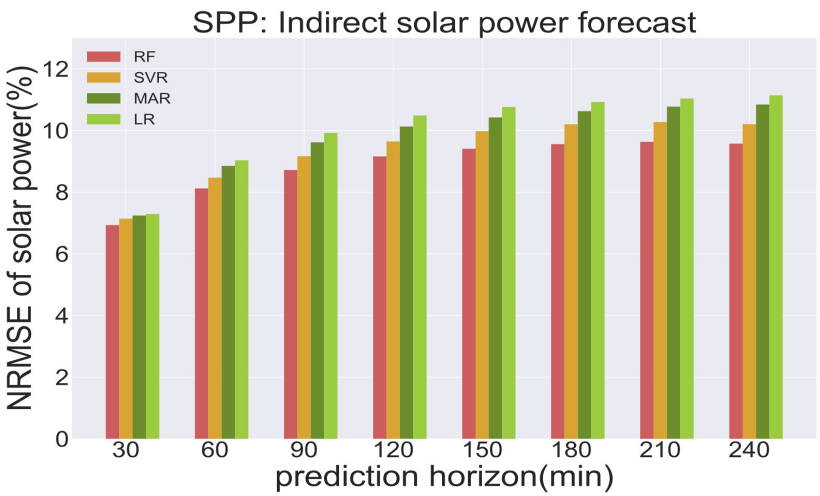
Performance comparison between direct & indirect approach
- NMBE versus time of forecasted at EECU

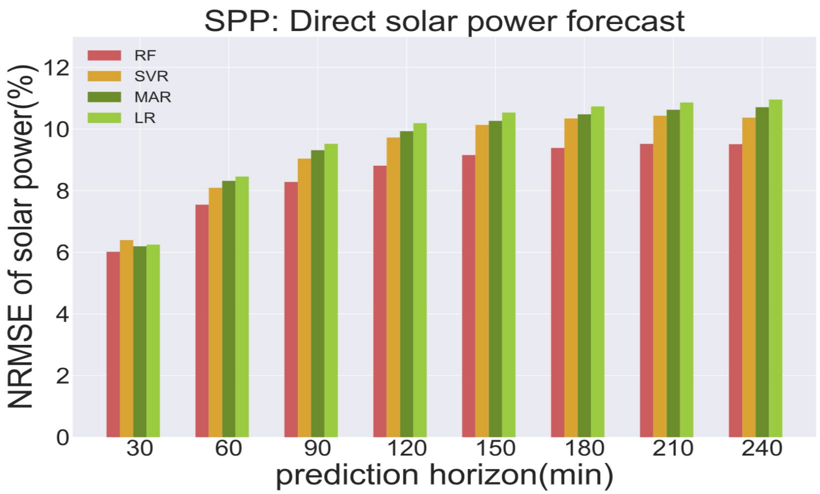




- Direct approach: random forest achieved the best performance.
- Direct approach: SVR is usually underestimate the solar power.
  This may lead to over operating cost.
- LR & MARS tend to overestimate the solar power at 4:00 - 4:30 p.m.
  This may lead to lacking power in power distribution system.

Performance comparison between direct & indirect approach
- NRMSE versus prediction horizon at SPP

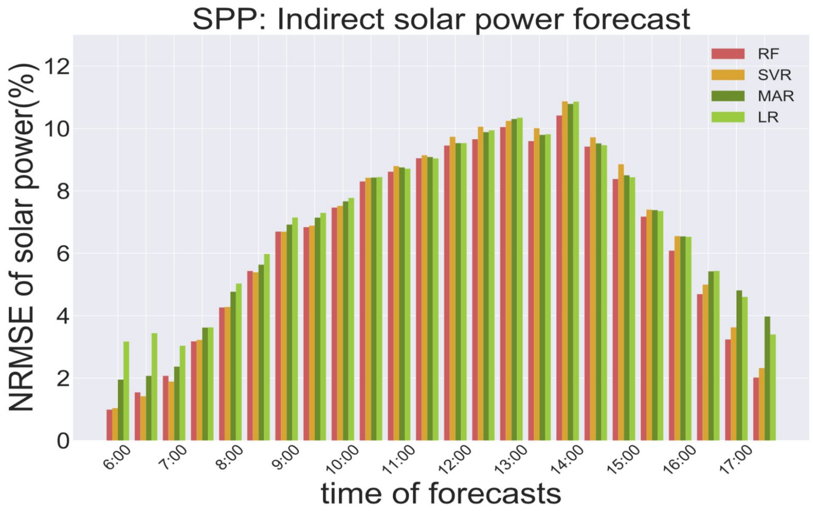


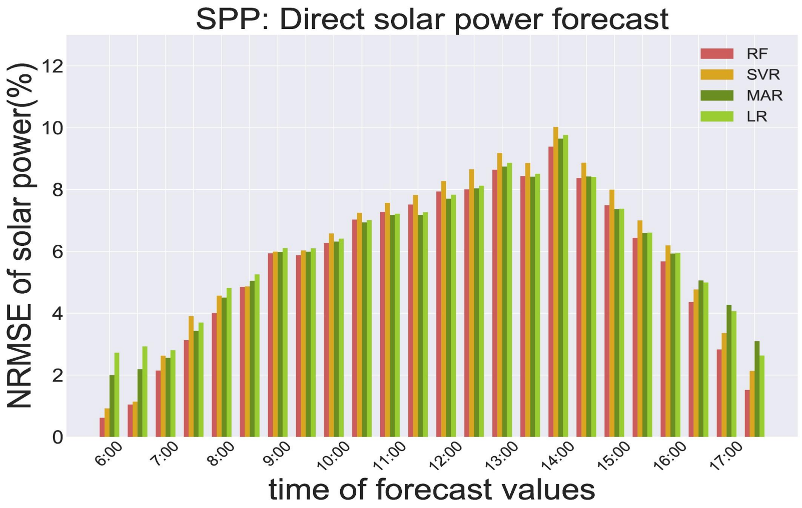


- Direct approach random forest still achieved the best performance in every prediction horizon.
  30 min-ahead: 6.02% 240 min-ahead: 9.51%
- The result showed the competing result with the former literature.

(1-hour ahead) Direct approach random forest : 7.55% Direct approach SVR : 8.09%
[VKSB16] SARIMA : 8.12% [BMP13] (SARIMA-SVR) 9.40%

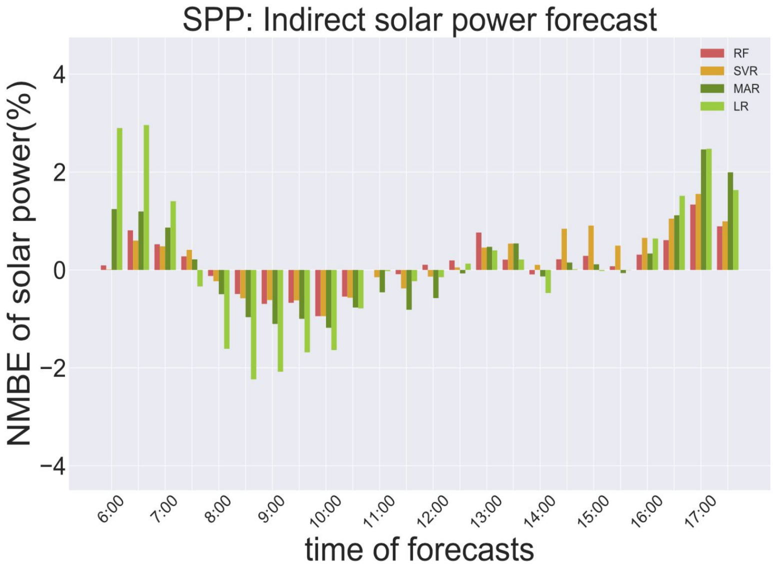
Performance comparison between direct & indirect approach
- NRMSE versus time of forecasted values at SPP

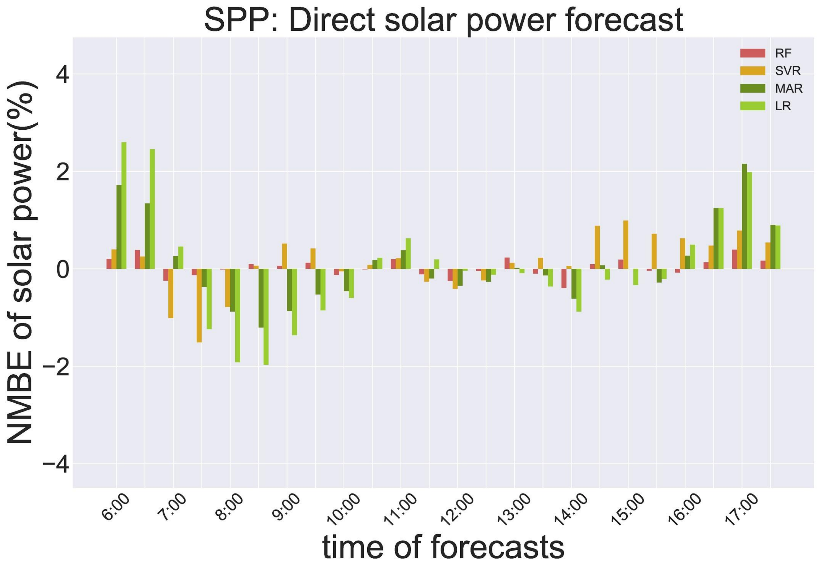




- NRMSE in the midday is usually higher than in the morning/evening
- The worst performance appear at 2.00 p.m. in both direct & indirect approach.
- The results showed the same pattern with EECU.

Performance comparison between direct & indirect approach
- NMBE versus time of forecasted values at SPP





- In the evening, we usually overestimate the solar power
- Overall error at SPP is lower than at EECU in accordance with the result from data pre-analysis.

Computational complexity

Main complexities consist of 3 parts

1. Feature computation
   - Calculate feature,e.g. exponential moving average of solar irradiance
2. Model training
   - Fitting model, e.g. Least square (linear regression), Binary splitting (random forest)
   - The most computational complex part but, can be an offline task
3. Prediction
   - Calculate prediction values
   - The most concerned part in real-time prediction (especially for intraday forecast)

Computational complexity

Hyper-parameter:

- p: number of feature
- M: number of basis (MARS)
- S: number of support vector (SVR)
- n_{tree}: number of tree in forest (RF)
- d: tree depth (RF)
- q: number of hidden layer (ANN)
- t: number of neuron in outer layer (ANN)
- r: number of neuron in inner layer (ANN)

Method	Prediction
Linear regression	$O(p)$
MARS	$O(Mp)$
SVR	$O(Sp)$
RF	$O(n_{\text{tree}}d)$
ANN	$O(pr + r(q-1) + rt)$

Reference: [RSO14], [Fri93], [CM16] , [Joh99]

Computational complexity

Complexity in prediction process (for hyper-parameter in this application):

LR < MARS < SVR < RF < ANN

But not significantly different in this application (in terms of computation time)

04

Conclusion

Conclusion: Forecasting approaches

- The result shows that the direct approach yielded better performance than the indirect approach because the error from irradiance prediction combining with the error from conversion model are greater than the error from direct power prediction.

- The forecast accuracy at SPP is better than EECU in every prediction horizon due to the different data characteristic. (e.g. variance, skewness)

Conclusion: Comparing to the literature

- The best model in every prediction horizon in terms of forecast accuracy is achieved by the direct approach: random forest model

NRMSE (%)	30-min ahead	1-hour ahead	2-hour ahead
SPP	6.02	7.55	8.81
EECU	7.44	9.08	10.30

- Compare to the literature,
  [VKSB16] (SARIMA) achieved 8.12% NRMSE for 1-hour ahead
  [BMP13] (SARIMA-SVR) achieved 9.40% NRMSE for 1-hour ahead
  [XCS12] (SVR) achieved 9.34% NRMSE for 2-hour ahead

05

Q&A

Thank you

Reference (1)

[Vio97] B. Viorel. Verification of some very simple clear and cloudy sky models to evaluate global solar irradiance.Solar Energy, 61(4):251–264, 1997.

[PC07] A. Pansak and S. Chumnong. An assessment of the ashrae clear sky model for irradiance prediction in thailand.Asian J. Energy Environ, 8(02):523–532, 2007.

[FHT01] J. Friedman, T. Hastie, and R. Tibshirani. The elements of statistical learning, volume 1. Springer series in statistics New York, 2001.

[SS04] Smola, A.J., Schölkopf, B. A tutorial on support vector regression. Statistics and Computing volume 14. 199–222, 2004.

Reference (2)

[RSO14] L. Roi, S. Shwartz, and S. Ohad. On the computational efficiency of training neural networks. In Z. Ghahramani,M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 27, pages 855–863. Curran Associates, Inc., 2014.

[Fri93] J. Friedman. Fast MARS. Technical report, Technical Report, 1993

[XCS12] R. Xu, H. Chen, and X. Sun. Short-term photovoltaic power forecasting with weighted support vector machine. In 2012 IEEE International Conference on Automation and Logistics, pages 248–253. IEEE, 2012.

Reference (3)

[CM16] G. Callejas and A. Miguel. The effects of model and data complexity on predictions from species distributions models. Ecological Modelling, 326:4–12, 2016

[Joh99] P. John. Fast training of support vector machines using sequential minimal optimization. advances in kernel methods—support vector learning (pp. 185–208). AJ, MIT Press, Cambridge, MA, 1999.

[BMP13] M. Bouzerdoum, A. Mellit, and A.M. Pavan. A hybrid model SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. Solar Energy, 98:226–235, 2013.

Reference (4)

[VKSB16] S. Vagropoulos, G. Chouliarasand . Kardakosand, C. Simoglou, and A. Bakirtzis. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. In 2016 IEEE International Energy Conference (ENERGYCON), pages 1–6. IEEE, 2016.