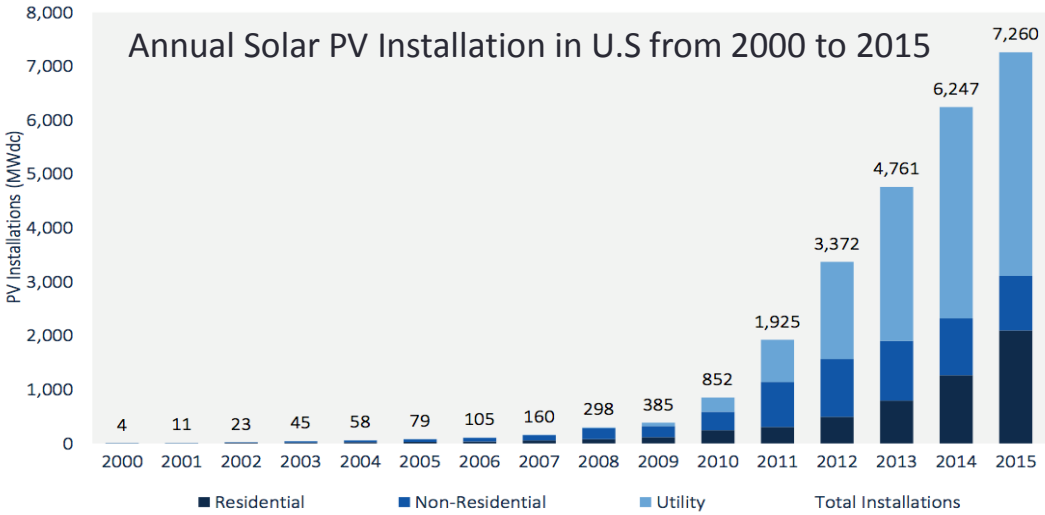
Investigating the Impact of Time Series Stationarization on Day-Ahead Solar Forecasting

Mohana Alanazi & Amin Khodaei
University of Denver
USA

Shay Bahramirad & Aleksi Paaso
ComEd
USA

Solar Installation

- At the end of 2015, solar installation has increased 16% over 2014 and reached up to 7,260 MW - 8.5 times of total installation in 2010.
- The residential PV installation represented around 27.5% (2 GW) of the total installed PV in 2015, which had grown 66% over 2014.
- As the federal Investment Tax Credit (ITC) was extended through 2021, a total of 70 GW of the new PV installation is expected to be added to the U.S. electric grid by the end of 2020.



Solar Energy Industries Association, "US Solar Market Insight: 2015 Year in Review," *September 2016*

Solar Forecasting

- Forecasting aims to estimate the variability and to reduce the uncertainty of the solar output so the grid operator will be able to efficiently accommodate its generation.
  - Variability refers to the intermittent (i.e., not always available) and fluctuating (i.e., constantly changing from seconds to minutes to hours) nature of renewable generation
  - Uncertainty represents the inability to predict in advance the timing and the magnitude of the generation variability

- Accurate forecasting helps alleviate many of the short-term operation (economic dispatch, power flow, etc.), mid-term maintenance, and long-term planning (solar investment) challenges.

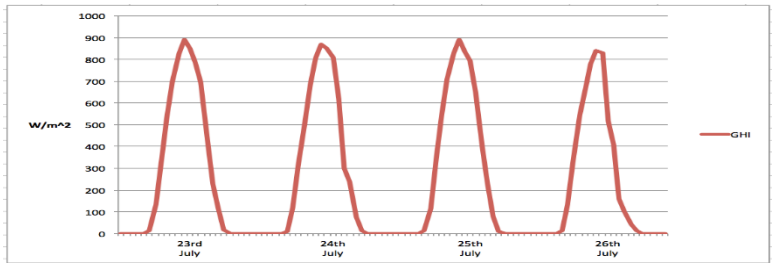
Forecasting Techniques

- Forecasting methods can be categorized into three different methods: Physical, Statistical, and Hybrid.

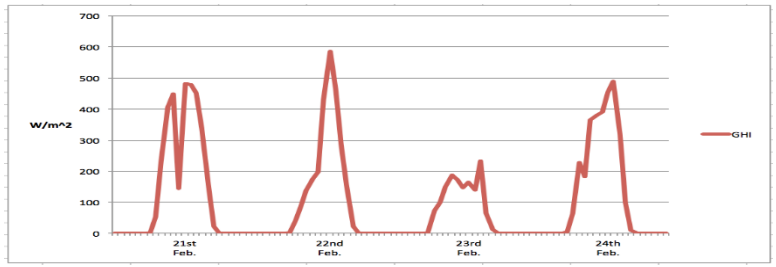
- Physical Method
  - Utilize physical weather data such as temperature, clouds, and pressure:
    - Numerical Weather Prediction (NWP)
    - Satellite and Cloud Imagery Model
- Statistical Method
  - Mathematical model, uses historical data
  - Persistence model
  - Time series models : AR, MA, ARMA
- Hybrid Method
  - Combination of physical and statistical

Forecasting Challenges

(1) The time series of solar irradiance is unpredictable, caused mainly by weather changes and partial/full cloud cover. As a result, the solar time series is considered non-stationary → difficult to predict the trending term in solar irradiance time series



Solar irradiance in sunny days



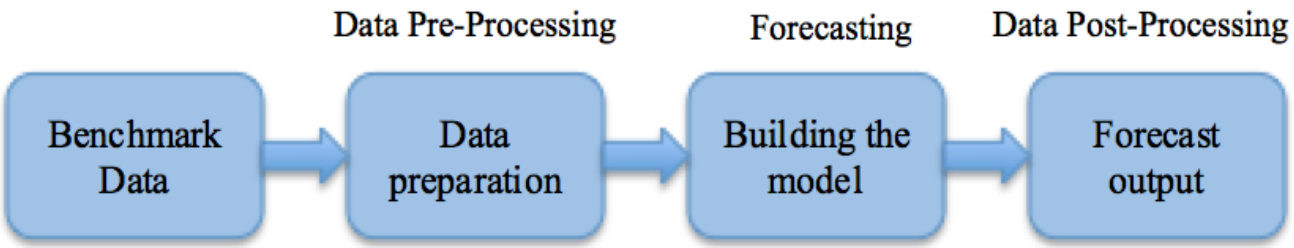
Solar irradiance in cloudy days

(2) The changes with respect to daytime hours (i.e. sunrise and sunset) impose limitations in forecasting. The duration of daytime hours changes throughout the year → impacts the patterns of the solar irradiance time series and thus increases the forecast error

Day	Sunrise	Sunset	Daytime hours
22-Jan	7:15	17:06	9 h, 51 min
18-Jul	5:46	20:24	14 h, 38 min
3-Nov	6:30	16:54	10 h, 24 min

Proposed Forecasting Model

- The model includes three stages:



- Data pre-processing
- Data processing (Forecasting)
- Data post-processing

→ Instead of working on non-stationary data we work on stationary data

Proposed Forecasting Model

- Data pre-processing:

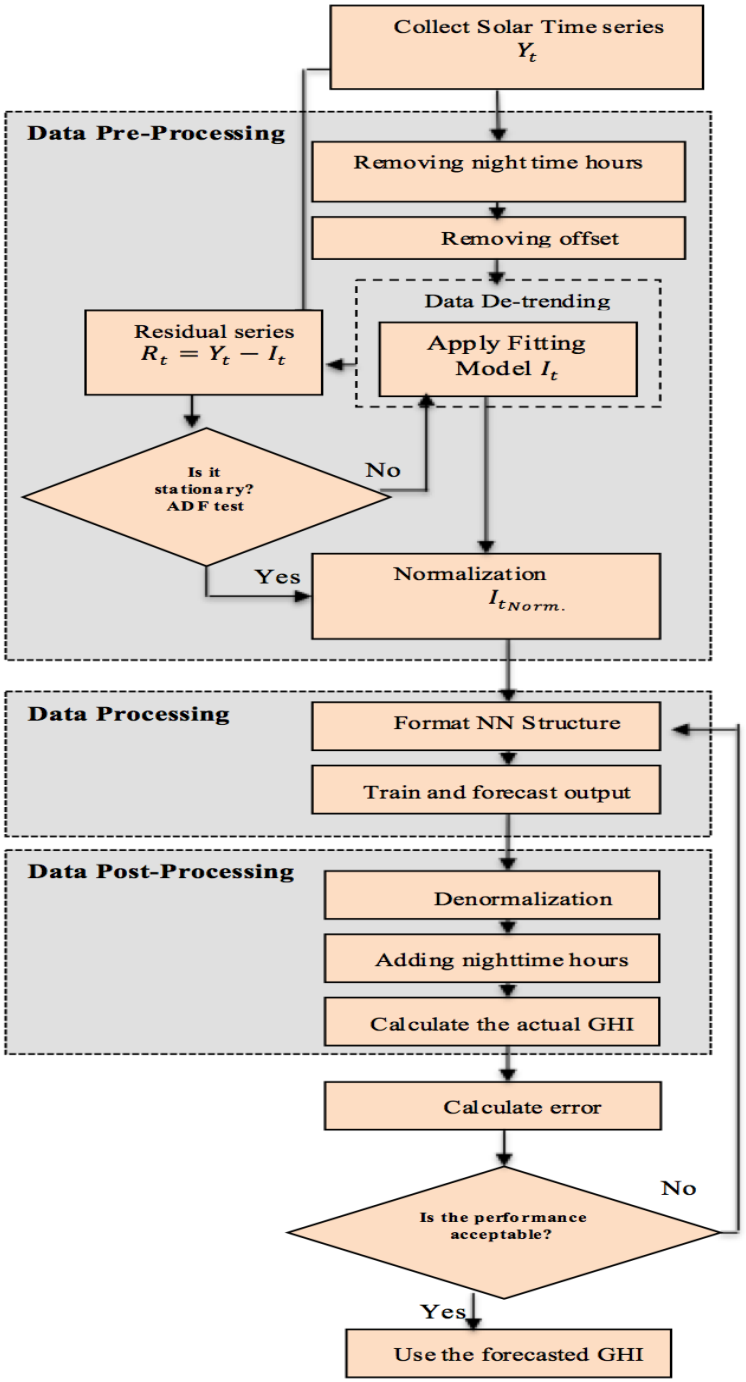
removing offset, removing zero solar radiation nighttime, Data detrending and normalization.

- Data processing (Forecasting):

model construction, training, forecasting

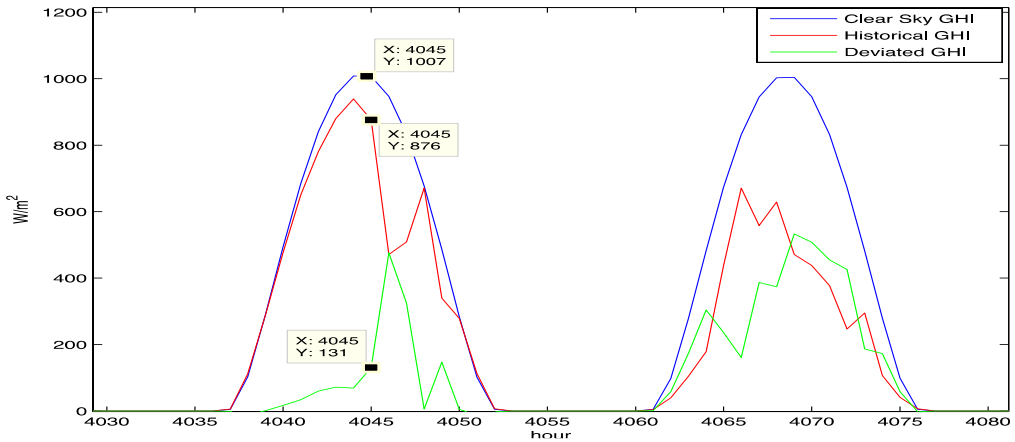
- Data post-processing:

denormalization, adding solar radiation nighttime hours, calculate the actual GHI.

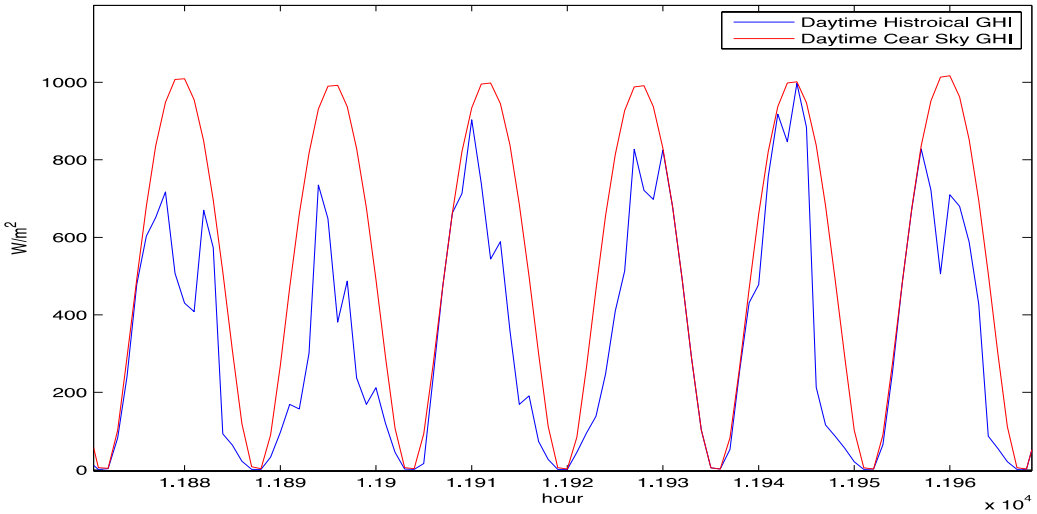


Data Pre-Processing

- Removing offset
- Removing nighttime
- Data detrending
- Normalization



GHI deviation between clear sky GHI and historical GHI



Removing Nighttime hours

Data Pre-Processing – cont'd

- A fitting model (here a high order polynomial) is used to detrend the solar time series.

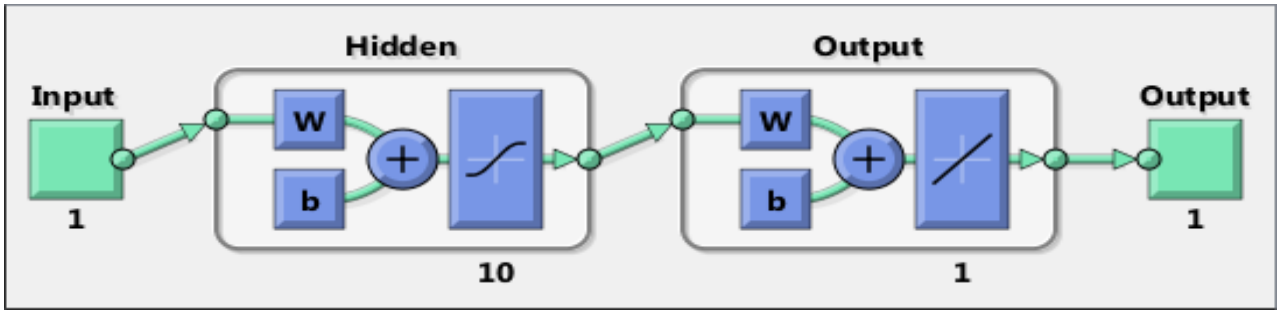
$$I_t = a_0 + \quad a_1 h + \; a_2 h^2 + \cdots + a_n h^n$$

$$h \notin nighttime\ hours$$

- The Augmented Dickey–Fuller (ADF) test examines whether there is a unit root in the time series. If there is a unit root, which means that the test result is above the critical value, the null hypothesis should be accepted and the time series is not stationary, otherwise the null hypothesis is rejected and the time series is stationary.

Data Processing

- Introduce the pre-processed data to the forecasting model, construct the model, train and forecast



Data Post-Processing

- Denormalization, adding solar radiation nighttime hours, calculate the actual GHI.

- The forecasting accuracy is checked using a variety of error measures
  - We use mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{N} * \sum_{t=1}^{N} \left| \frac{GHI(t)_{actual} - GHI(t)_{forcast}}{GHI(t)_{actual}} \right| * 100$$

Case Study - non-stationary data

- The hourly average GHI data is tested, using ADF, for checking stationarity without data detrending.

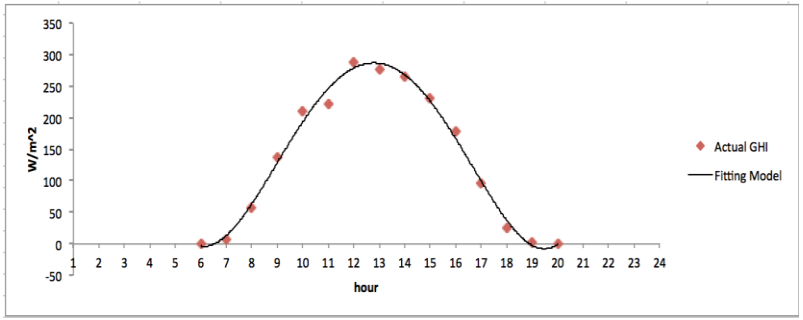
Statistical Power	Significance level	Test result	Critical value
0.47	0.05	-0.465	-1.9567

- The test result is above the critical value, which indicates that there is a unit root and null hypothesis should be accepted, i.e., non-stationary
- This data is directly fed to the neural network forecasting tool to forecast GHI values

Weather Condition	Day	MAPE (%)
Partly Cloudy	March 8th	2.624
Partly Cloudy	May 5th	2.670
Sunny	August 13	1.117

Case Study - stationary data

- First, the hourly actual and clear sky irradiance monthly average are calculated to fit the high order polynomial.
- The data undergoes the first two steps in the pre-process which are removing the nighttime hours and removing the offset. The resultant data is fitted.



Actual hourly average and the fitting model series

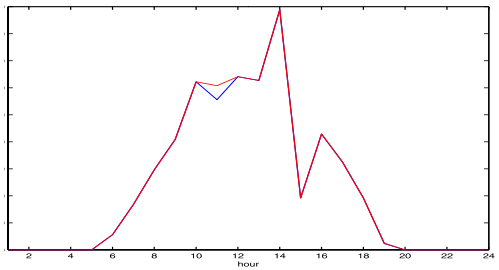
- The ADF test is applied to the residual to check if the series is stationary or not.

Statistical Power	Significance level	Test result	Critical value
0.001	0.05	-5.12	-1.957

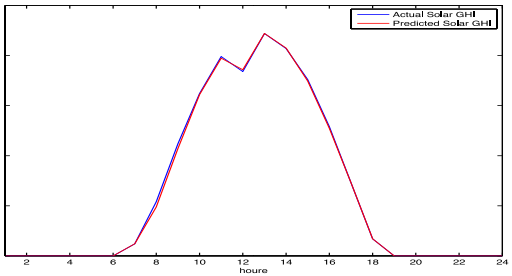
Case Study - stationary data – cont'd

- The test result is below the critical value and that means there is no unit root in the data tested, so the null hypothesis should be rejected and the time series is stationary.
- The statistical power shows the probability that the detrended time series has unit root.

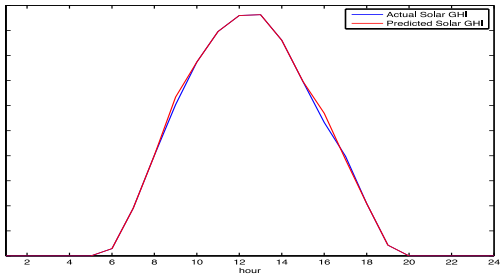
Weather Condition	Day	MAPE (%)
Partly Cloudy	March 8th	0.8107
Partly Cloudy	May 5th	0.7998
Sunny	August 13	0.6385



March 8th 2010 (partly cloudy)



May 5th 2010 (partly cloudy)



August 13th 2010 (sunny)

Conclusion

- The model converts the solar irradiance time series to a stationary time series and apply ADF test to check the stationarity.
- The application of stationary data set could reduce the MAPE by as much as 42% in sunny days and 70% in cloudy days.
- The new model performs well under different weather conditions, including cloudy conditions.
- The new model reduces the size of the time series to almost half by removing nighttime values, which accelerates the time needed for simulations.

Future Work

- This work can be extended to include hybrid forecasting methods that could potentially outperform single methods.
- The model can be applied to different time horizons, week-ahead, day-ahead, hour-ahead, etc.
- Geographical Information System (GIS) will be used to improve the forecasts

Thank you

amin.khodaei@du.edu