Comparative Study of Forecasting Algorithms for Energy Data

Master Thesis Presentation by

Nishat Fariha Rimi

20th May, 2019

Albert-Ludwigs-Universität Freiburg

Motivation

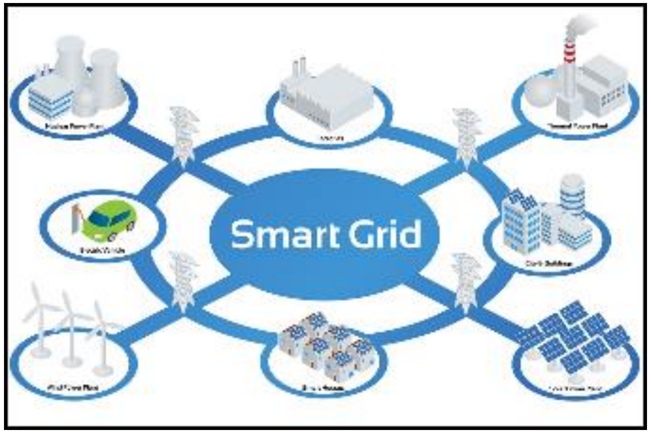
- Wind and solar energy varies
- What is produced must be used

Benefits of energy demand forecasting–

- Balances supply and demand
- Prevents energy waste
- Reduces operation cost

Comparative analysis of forecasting methods depending on

- Time scale
- Dataset type and sample size



Source: Science direct. smart grid and solar energy

Source: https://www.iass-potsdam.de

Outline

Overview

- Implemented forecasting methods
- Considered forecasting scenarios

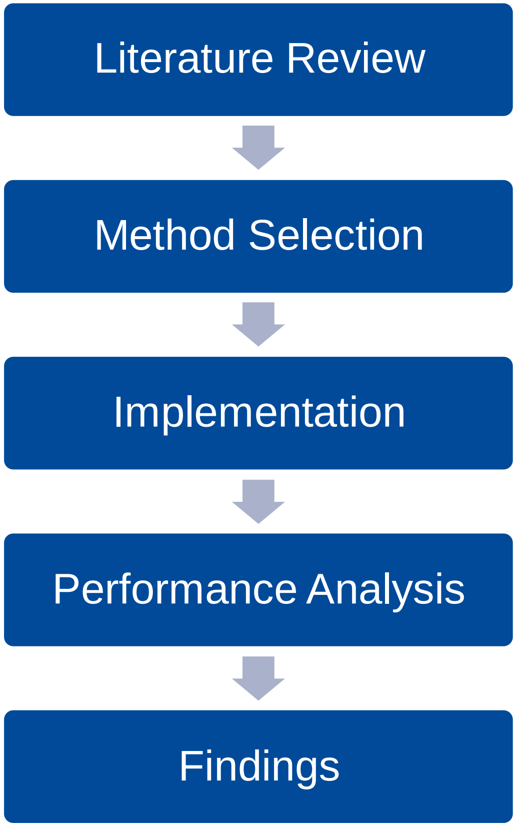
Methodology

- Methods
- Forecasting toolbox

Performance analysis

- Performance comparison

Conclusion and Future Work

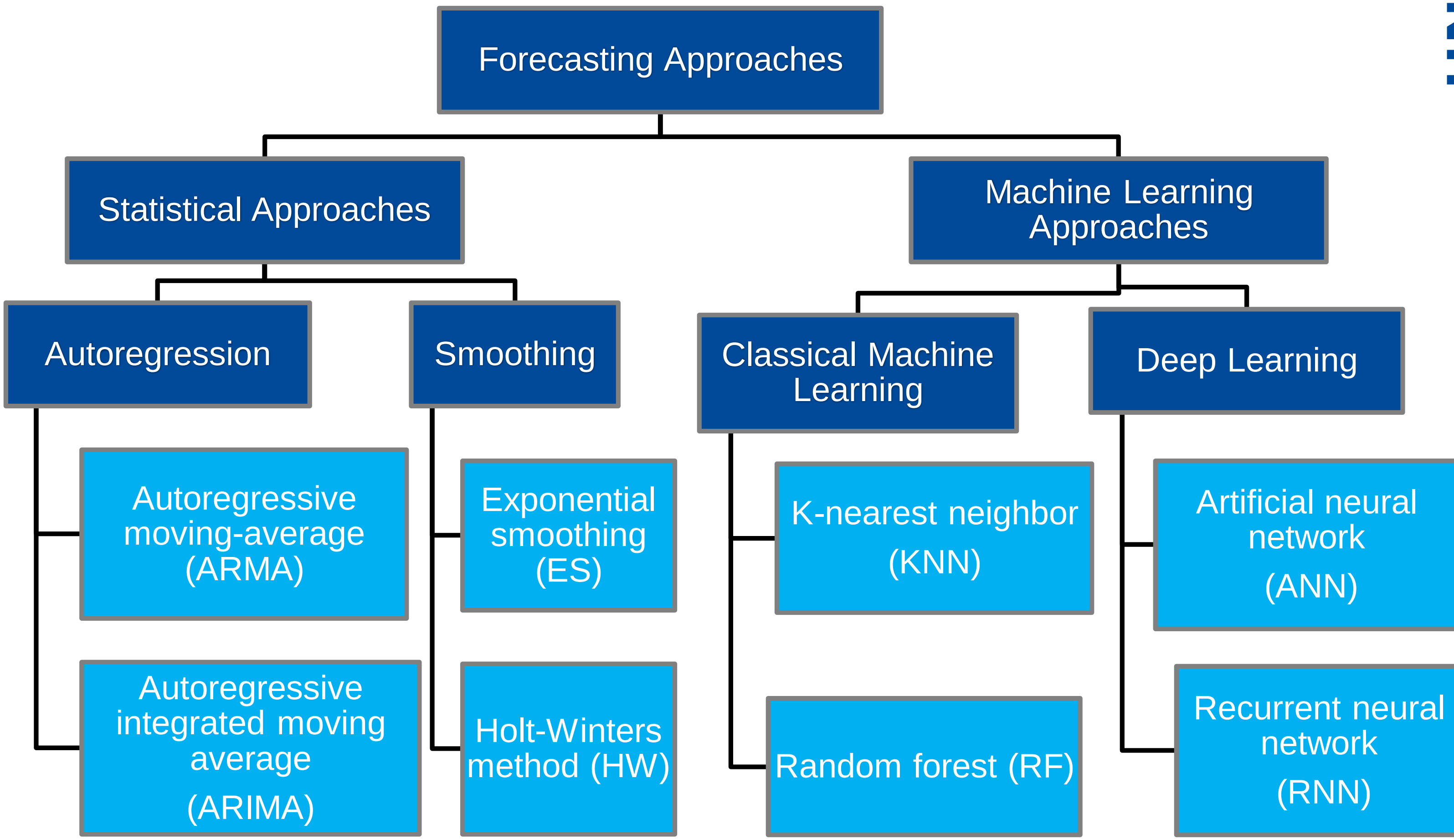


Goal

➢ **Create:** Structure of a forecasting toolbox

➢ **Compare:** Methods performance according to datasets and forecasting scenarios

Selected Forecasting Approaches



Lahouar and J. Ben Hadj Slama Energy Conversion and Management, vol. 103, pp. 1040–1051, 2015.
R. J. Hyndman and G. Athanasopoulos, Forecasting : Principles and Practice. OTexts, 2018.

Considered Scenarios

Performance comparison according to different forecasting aspects.

Forecasting horizon	Time scale	Considered time scale	Forecasting sample size
Very short-term	5 min- 1 h	1 h	1
Short-term	1 h- 24 h	1 d	24
Medium-term	24 h- weeks	1 w	24*7
Long-term	month-years	1 m 3 m	24*7*4 24*7*4*3

Training Sample Size: 50 | 100 | 200 | 400 | 800 | 1600 | 2200

Datasets: PV generation | Electrical load

P. Kuo, Energies, vol. 11, January, pp. 1–13, 2018
J. W. Taylor, International Journal of Forecasting, vol. 24, no. 4, pp. 645–658, 2008

Statistical Approaches

Depend on the past values of endogenous variable for forecasting

ARMA (*p,q*):

- Combination of $\mathrm{AR}(p)$ and $\mathrm{MA}(q)$ models for stationary time series

ARIMA (*p,d,q*):

- Transform the non-stationary data into stationary by differencing

ES (α):

- Assign exponentially decreasing weights for past observations

HW (*α, β, γ*):

- Design to capture trend and seasonality

C.-M. Lee *et al. Expert Systems with Applications*, vol. 38, pp. 5902–5911, 2011.
R. Weron, Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach. 2006

Machine Learning & Deep Learning Approaches

Exogenous information and endogenous variables used together

RF (n_tree, max_depth):

- Constructs multiple decision trees during training

KNN (*k*):

- Searches for a group of k samples nearest based on distance function

ANN (hidden_node, hidden layer, epoch):

- Allows data signals to process the output in one way

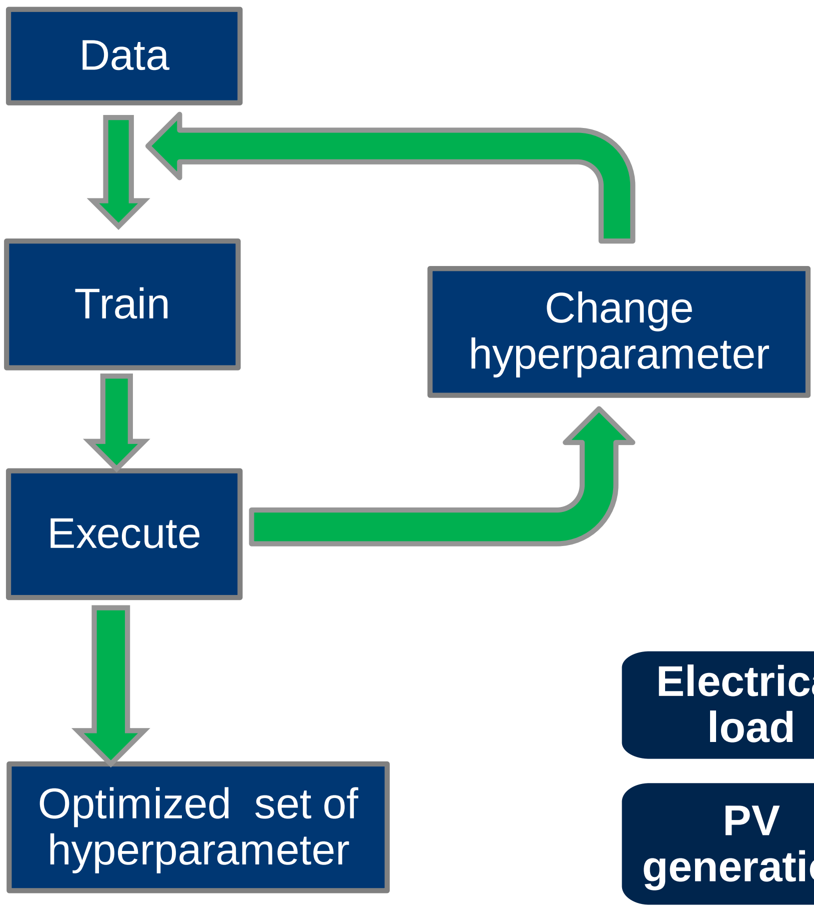
RNN (hidden_node, hidden layer, epoch):

- Use internal state (memory) to process sequences of inputs

C. Xia *et al.* International Journal of Electrical Power & Energy Systems, vol. 32, pp. 743–750, 2010.
M. Thanh Noi *et al.* Sensors, vol. 18, no. 1, 2018.

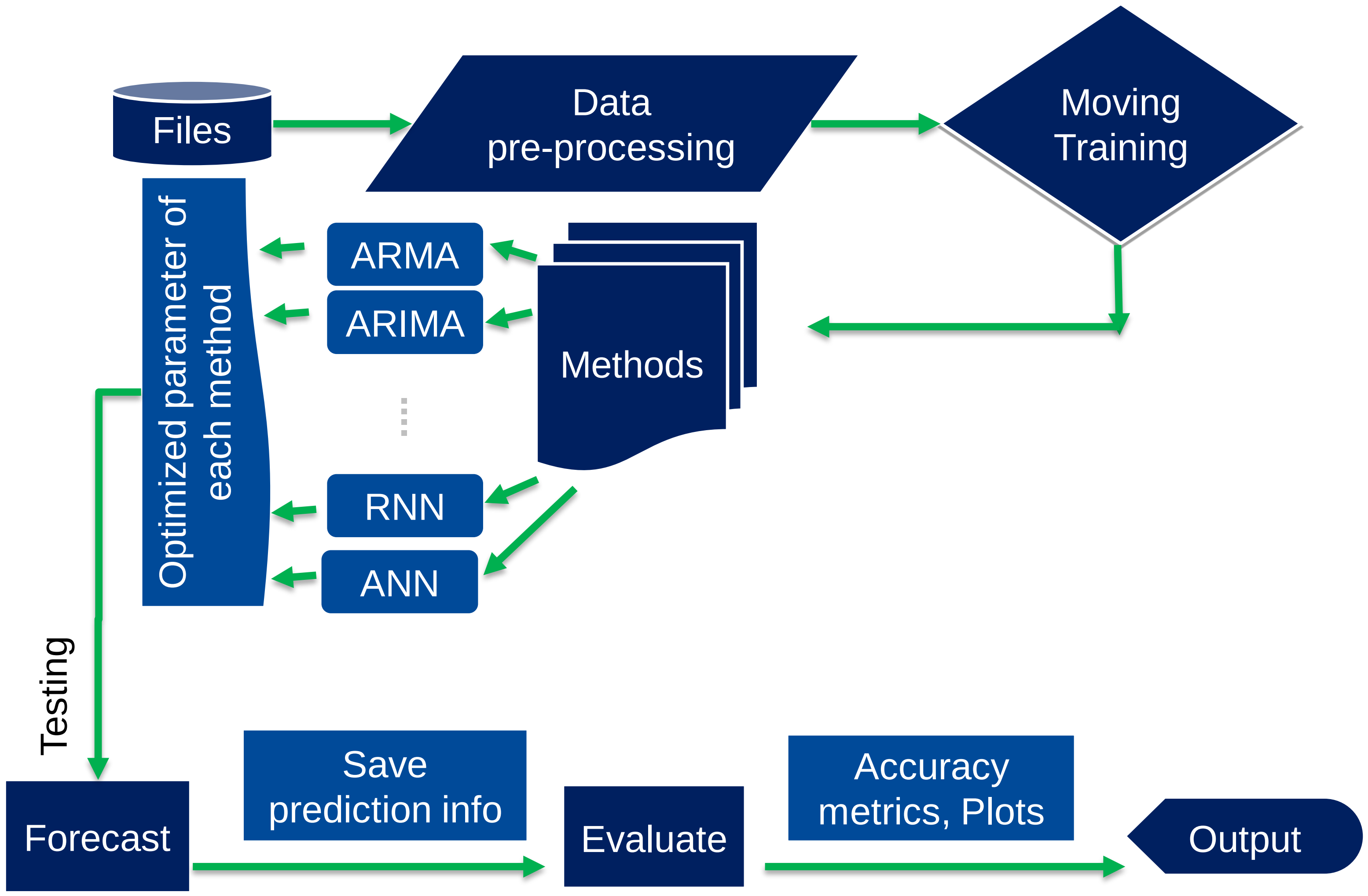
Methodology
Parameter Optimization

Optimization of hyperparameter:



	Endogenous	Exogenous
Electrical load	Electrical load time series	Hour of the day
PV generation	PV generation	Diffused Solar radiation

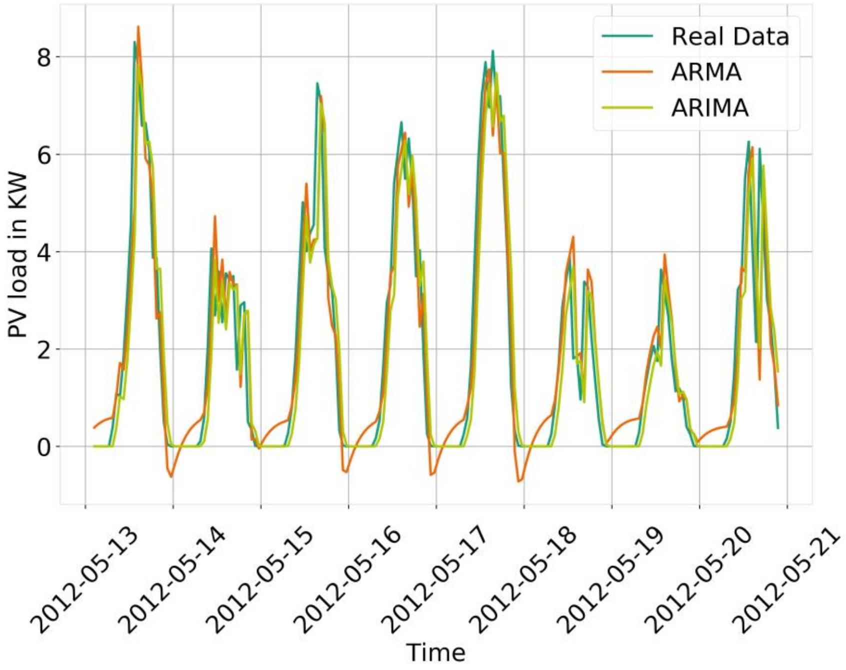
Methodology
Forecasting Toolbox

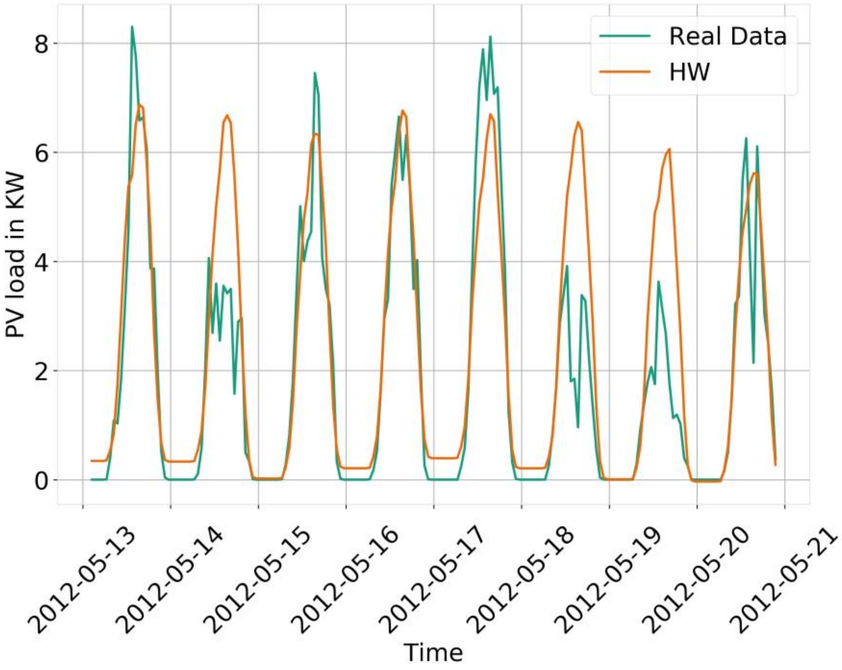


Performance Analysis

Statistical Approaches

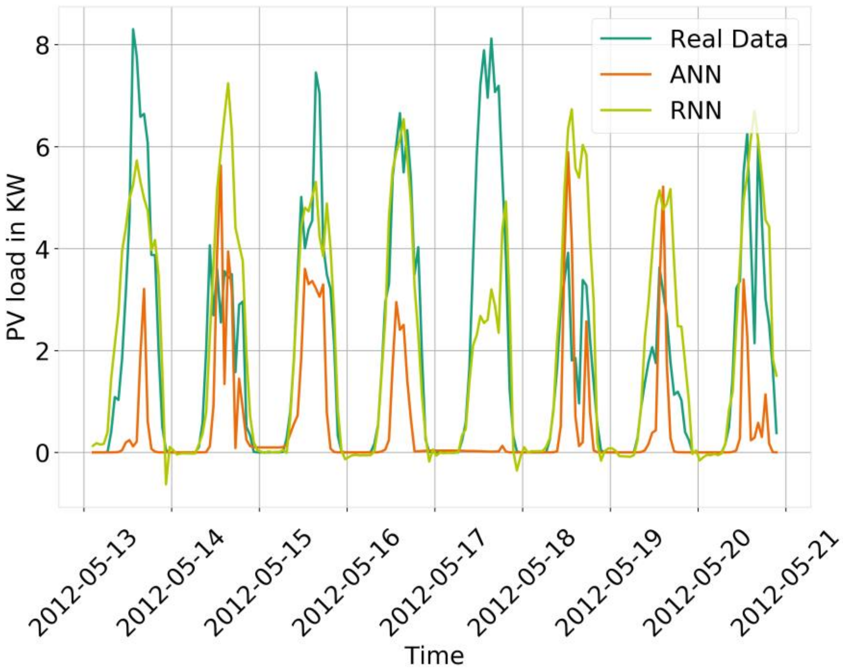
- Days Plot – PV Generation
- Forecasting sample size- 1
- Training sample size- 2200

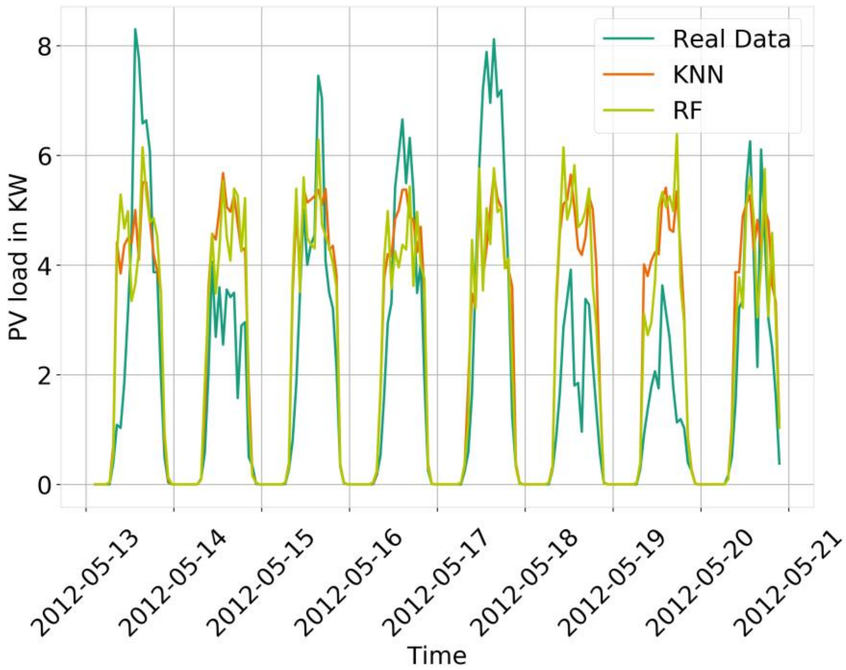




Performance Analysis
Machine Learning Approaches

- Days Plot – PV Generation
- Forecasting sample size- 1
- Training sample size- 2200

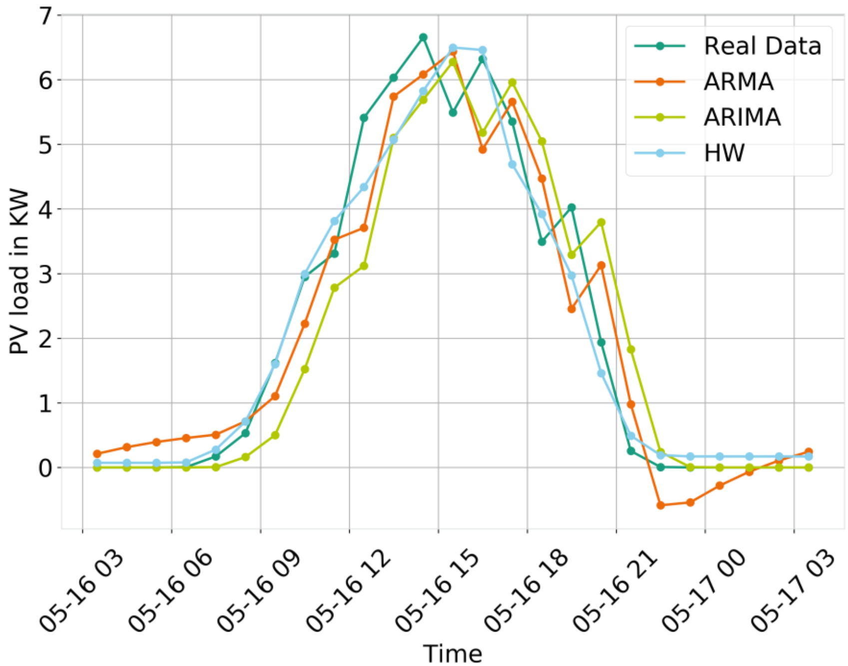


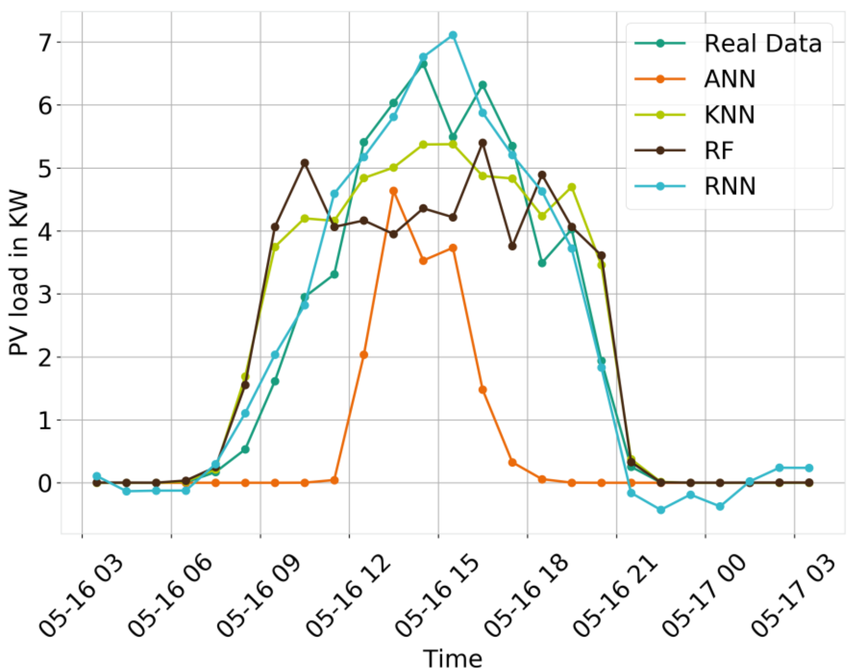


Performance Analysis

All methods

- 1 Day Plot – PV Generation
- **Statistical methods** and **Machine learning methods**
- Forecasting sample size- 1
- Training sample size- 2200





Learning Time Comparison

Mean **training time** in **seconds**

PV generation: (daily forecasting for 100 training sample size)

HW	KNN	ES	ARMA	ARIMA	RF	ANN	RNN
0.004	0.005	0.006	0.009	0.11	0.29	8.33	32.983

Electrical load: (daily forecasting and 100 training sample size)

KNN	ES	HW	ARMA	ARIMA	RF	ANN	RNN
0.0084	0.0079	0.25	0.034	0.106	0.35	8.41	48.27

- Training time increases gradually with the increase of training sample sizes

PC configuration: Windows 10 computer, with 4 Cores, 8GB of ram, and with 3.4 GHz clock speed.

Predicting Time Comparison

Mean **predicting time** in **seconds**

PV generation: (monthly forecasting with 2200 training sample size)

ARMA	ARIMA	ES	HW	KNN	RF	RNN	ANN
0.004	0.018	0.011	0.027	0.038	0.048	0.05	0.06

Electrical load : (monthly forecasting with 2200 training sample size)

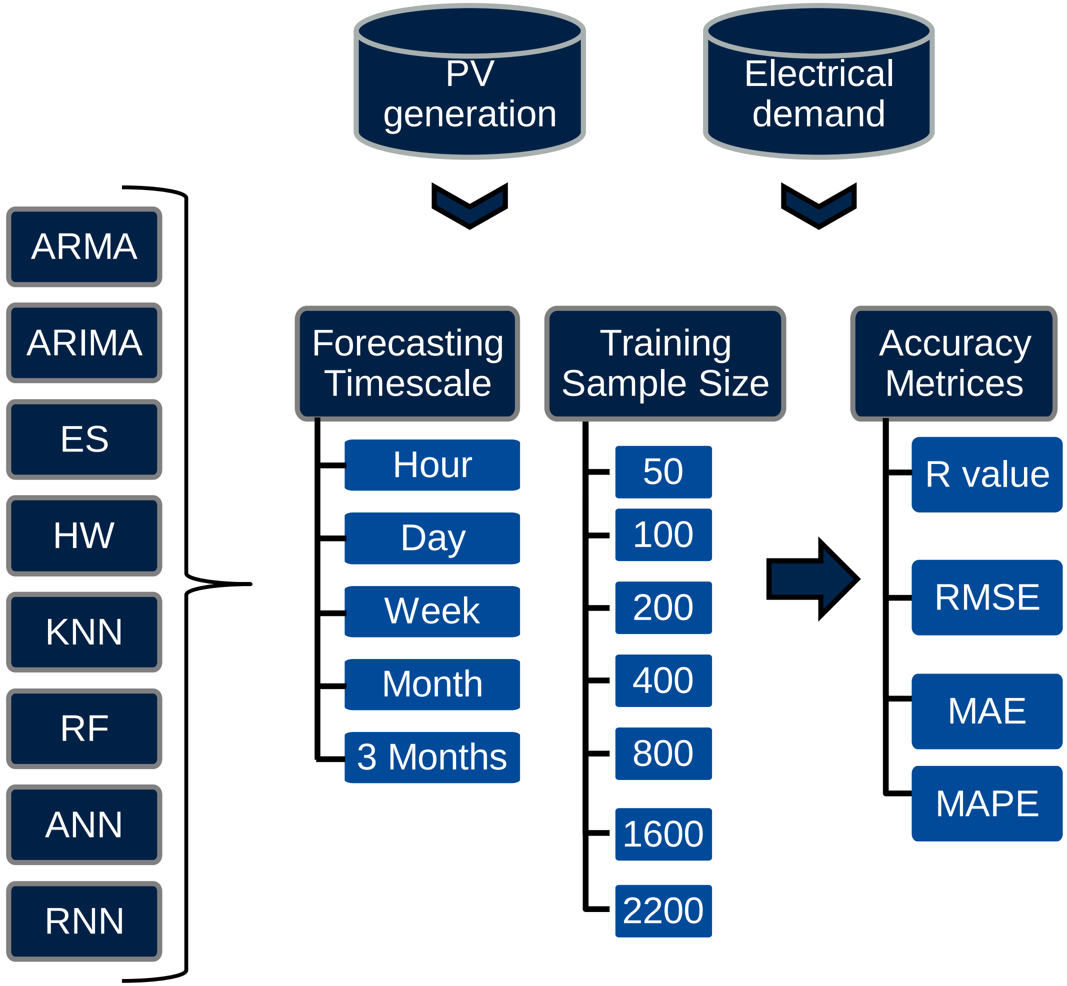
ARMA	ARIMA	KNN	ES	HW	RF	ANN	RNN
0.013	0.02	0.021	0.023	0.031	0.048	0.06	0.98

- predicting time has increased gradually with the increase of prediction horizons

PC configuration: Windows 10 computer, with 4 Cores, 8GB of ram, and with 3.4 GHz clock speed.

Performance Analysis

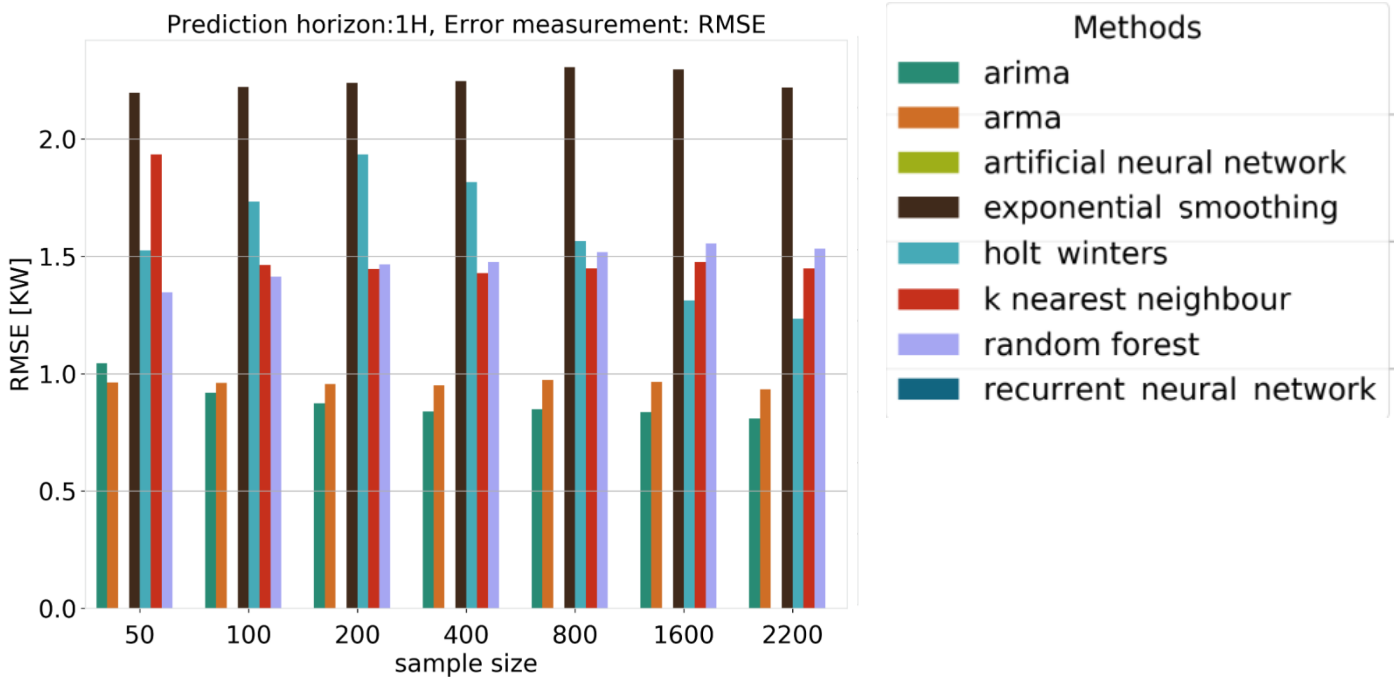
Indicators



RMSE Comparison for PV Generation

Hourly

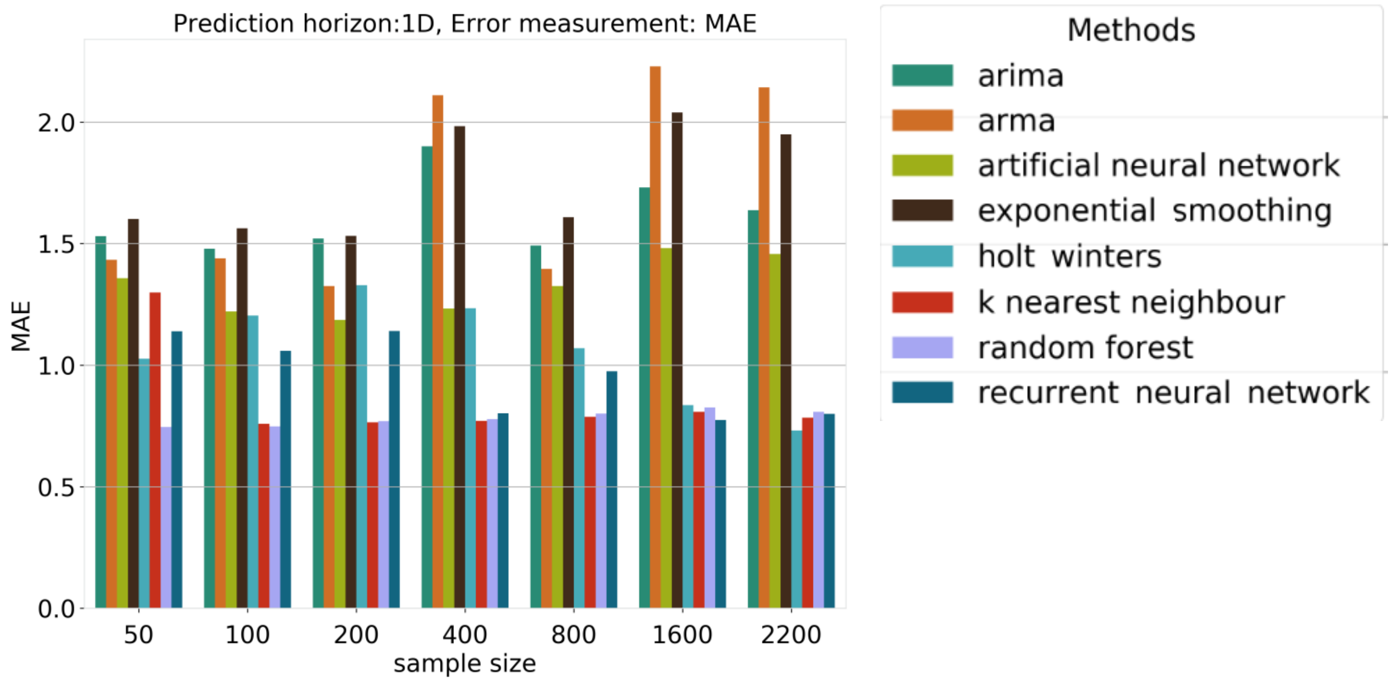
- ✓ ARIMA, ARMA
- − ES



MAE Comparison for PV Generation

Daily

- ✓ RF, KNN
- ✓ HW, RNN (1600, 2200)
- – ES, ARMA



RMSE Comparison Summary

	Photovoltaic Dataset	Electrical Load Dataset
Hourly	✓ ARIMA, ARMA – ES	✓ ARIMA, ARMA – HW
Daily	✓ RF, KNN ✓ HW, RNN (1600, 2200) – ARMA, RNN (small sample)	✓ RF, KNN, ✓ RNN, HW (>400) – ARMA, ES
Weekly	✓ RF, KNN, ✓ HW (1600,2200),RNN (2200) – RNN (small sample)	✓ Similar to Daily – Similar to Daily
Monthly	✓ RF, KNN, ✓ RNN (1600,2200) – HW	✓ RF, KNN,RNN ✓ HW (>200) – ARMA, ES

MAE Comparison Summary

	Photovoltaic Dataset	Electrical Load Dataset
Hourly	✓ ARIMA, ARMA − ES	✓ ARIMA, ARMA − HW
Daily	✓ RF, KNN, ✓ HW,RNN (1600, 2200) − ES, ARMA	✓ RF, KNN, ✓ RNN,HW (>400) − ARMA, ES
Weekly	✓ Similar to Daily − Similar to Daily	✓ Similar to Daily − Similar to Daily
Monthly	✓ RF, KNN ✓ RNN (sample size > 800) − HW	✓ RF, KNN,RNN ✓ HW (>200) − ARMA, ES

Conclusion

Comparative analysis of eight forecasting methods –

- ❑ Prediction horizon
- ❑ Training sample size
- ❑ PV generation and electrical load

Hourly

- ❑ ARMA and ARIMA – optimum choice
- ❑ Computation time of ARMA < Computation time of ARIMA

Daily, weekly and monthly

- ❑ RF and KNN
- ❑ Computation time of KNN < Computation time of RF

HW and RNN - large dependency on sample size

Future Work

Adding other forecasting approaches e.g.

- Support vector regression
- Gaussian process regression etc.

Optimizing parameter with dynamic optimization function

- Genetic algorithm

Training with –

- Datasets of shorter time interval like 15 or 30 minutes
- More datasets / applications

Thank You