

RESEARCH

Open Access



Energy forecasting based on predictive data mining techniques in smart energy grids

Ekanki Sharma

From The 7th DACH+ Conference on Energy Informatics
Oldenburg, Germany. 11-12 October 2018

Correspondence: Ekanki.Sharma@aau.at
Alpen-Adria-Universität Klagenfurt,
Universitätsstraße 65-67, Klagenfurt,
Austria

Abstract

Energy forecasting is a technique to predict future energy needs to achieve demand and supply equilibrium. In this paper we aim to assess the performance of a forecasting model which is a weather-free model created using a database containing relevant information about past produced power data and data mining techniques. The idea of using a weather-free data-driven model is first to alleviate the dependence on weather data which, in some scenarios is difficult to obtain and second to reduce the computational effort. In this work, we aim first to evaluate the interplay between anomaly detection techniques and forecasting model accuracy. Secondly we will determine out of the three defined performance metrics, which one is the best for this particular application.

Keywords: Energy forecasting, Weather-free data, Data-driven model, Anomaly detection

Motivation

The increasing penetration of renewable energy resources (RES) in the today's power system has made energy forecasting a popular theme. It is very important for grid operators and decision makers to know how much power RES will produce over next hours and days (Dobschinski et al. 2017). Along with this, predicting load demand and consumption plays a vital role in operation and planning of modern power system. Storage of electrical energy is necessary in the case when there is excess power production from the RES and less load demand. However, it cannot be massively stored as energy storage is expensive, requires high maintenance and have limited lifespans. Because of this, utilities have to balance supply and demand at every moment. These limitations lead to several interesting characteristics of energy forecasting, which includes data collection and the need for precise accuracy. Forecasting errors lead to unbalanced supply-demand, which adversely affects the operational cost, reliability and efficiency. Forecasting energy production and consumption are usually based upon meteorological data like solar irradiation, temperature and number of occupants and appliances respectively. However, there are some scenarios where the on-site

measurements for solar irradiation and other meteorological variables like temperature and humidity are unavailable and only the past power measurements are available. In such cases, data-driven models utilizing the available past power production data can be used. In this paper we aim to exploit the available past power data and to assess the performance of data-driven forecasting model in terms of accuracy by applying data pre-processing techniques. The idea is to choose an appropriate anomaly detection technique and data-driven methodology for energy production forecasting along with developing a unified model for long-term forecasting with step of short-term (hourly) accuracy.

Background

In the past decades, different approaches for forecasting energy production, distribution and consumption had been implemented. In the domain of energy consumption forecasting several techniques are used by researchers which includes traditional methods such as regression, time series, statistical methods along with soft computing techniques such as Artificial Neural Networks (ANNs), Support Vector Machine (SVM), fuzzy logic, and Grey prediction. A good overview of these techniques can be found in (Suganthi and Samuel 2012).

To perform predictions typically larger datasets in connection with deep learning are becoming common. In (Marino et al. 2016), authors implemented an energy load forecasting technique with long short term memory (LSTM). Two variants of LSTM are presented, standard LSTM and the LSTM-based Sequence-to-Sequence (S2S) architecture. Both variants are tested with one hour and one minute time step resolution data, the results indicate S2S worked well in both datasets.

To integrate RES in the power grid, forecasting photovoltaic (PV) yield is very important, as the output of PV systems is sensitive to weather conditions and to the varying strength of solar irradiance striking the PV surface throughout the day. The input variables and prediction horizon affect the accuracy of the prediction model. In general, the relevant variables which are available as inputs of the prediction model of solar power includes historical measurements of PV generation, historical measurements of explanatory variables like temperature, global irradiance, wind speed or cloud coverage (Wan et al. 2015). In the domain of energy production forecasting, there are several studies which reveal the potential of Artificial Intelligence (AI). Authors in (Saberian et al. 2014) implemented solar power modelling method using artificial neural networks (ANNs) which includes two neural network structures, namely, general regression neural network (GRNN) and feedforward back propagation (FFBP) to model a PV panel output power. They used meteorological data and estimated generated power to train the GRNN and FFBP. The results indicated higher accuracy when using FFBP. In (Khatib and Elmenreich 2015), authors proposed a generalized regression artificial neural network for predicting hourly solar radiation. In (Alanazi et al. 2017), authors implemented a nonlinear autoregressive neural network for prediction of irradiance. Authors in (Gandelli et al. 2014) implemented a new hybrid method PHANN (Physical hybrid artificial neural network), combining physical model with statistical model (neural network) and concluded that the PHANN method is more accurate than ANN. Authors in (Dolara et al. 2018), employed a similar method considering theoretical clear sky solar radiation model and stated the improved accuracy in case of hybrid

method. It can be inferred that combining and implementing several techniques like statistical and physical could improve the performance. Based on the reviewed papers, we assume that it is possible to improve the accuracy by applying advanced approaches like soft computing techniques which outperforms naive methods coming from statistical theory. However, complex models such as deep learning models do have a limitation in terms of interpretability.

Nevertheless, it is possible to improve the accuracy by applying data pre-processing techniques (anomaly detection) i.e. feature selection and outlier rejection. Before applying any forecasting model these two important issues should be considered (Saleh et al. 2016), as both have a direct impact on the forecasting model performance. In large datasets, there is often the case where we have many ineffective features and a feature selection process could minimize the considered features to effective ones. This process can improve the model performance and provide faster decisions. The authors of (Saleh et al. 2016) implemented a data mining-based load forecasting strategy and divided the whole process in two parts data pre-processing and load estimation. The data pre-processing step performed outlier rejection to eliminate the bad data using a distance-based outlier rejection and feature selection using genetic algorithm. However, authors have clearly mentioned that the outliers are rejected based on a global view, where extreme values are considered as outliers. It is noteworthy that these values do represent real measured value and in some circumstances extreme values may indicate sudden events. Secondly to construct the case study they used historical electricity load dataset. They did not explore the validity of their model on the real time data set which may pose additional challenges.

In energy and power applications, anomaly detection emerges as an important aspect in fields like electric load forecasting (Chen et al. 2014; Chakhchoukh et al. 2011), energy production forecasting etc. In (Luo et al. 2018) authors implemented a model-based anomaly detection method for very short-term load forecasting. The method includes two components, an underlying model i.e. dynamic regression model (DRM) and an adaptive anomaly threshold. Some of the recent work on anomaly detection is presented in (Table 1).

Data-driven modelling (DDM) is emerging as another important aspect in forecasting energy production problem. The output power produced by PV is highly correlated

Table 1 State-of-the-art: Anomaly detection or outlier rejection

Author	Problem Targeted	Method applied	Contribution & Perspective
Panapakidis (Panapakidis et al. 2018)	Missing data treatment	Data processing (Clustering phase + completion phase) – Clustering phase is unsupervised machine learning tool K-means – Completion phase filling technique application	– Proposed a new methodology for data filling – Applicable for both complete and partial absence of data – Presents a novel methodology for missing and incomplete data completion – Methodology is not dependent on data size, data resolution and amount of missing data – Incomplete data artificially completed with data entries of high similarity
Daliento (Daliento et al. 2017)	Monitoring & diagnosis of faults in single and multiple PV array strings	Monitoring and diagnosis techniques based data mining i.e. decision tree method, K-nearest neighbour and SVM	– Reviewed methods for fault detection – Presented reliability issues

with the weather conditions. Hence, they are usually considered as an important parameter in training the prediction algorithm. However, in cases when the weather data is unavailable, it is interesting to use data-driven models using only past PV output production data. DDM is based on analysing the data about a system, in particular finding connection between the system state variables (input, internal and output variables) without explicit knowledge of physical behaviour of the system. Authors in (Ordiano et al. 2017) implemented a simple weather-free data-driven models by considering only the past generated power and time of the day as an input. (Table 2) presents short review on work done in the domain of energy production forecasting.

(Filik et al. 2011) proposed a novel unified model for short, medium and long-term for hourly electric energy demand forecasting. The authors compared the accuracy of analytically developed model with three different ANN architectures and achieved highest accuracy with time delay back propagation ANN architecture.

Methods

Material

Energy forecasting algorithms are trained and tested on energy consumption and production datasets. These datasets contain energy readings from the smart meter and power output produced by PV. The forecasting approaches which are present in the literature usually utilize proprietary data. Instead, we will use freely available benchmark data for testing future energy forecast models which makes the comparison between approaches easier to understand. In this study we intend to use the Open Power System Data (OPSD)(open-power-system-data.org) and the Australian Solar home electricity dataset provided by Ausgrid (aus-grid.com.au).

Model evaluation

We first split the data into training and testing datasets and then run the machine learning algorithm on the training dataset to generate the prediction model. Then we use the test dataset to evaluate the model. To avoid underfitting and overfitting cross validation will be performed. In order to evaluate the performance of the forecasting algorithm, various performance metrics are available in the literature. These standardized performance measures or metrics helps in providing forecast evaluations and benchmarking (Pelland

Table 2 State-of-the-art: Data-driven modelling

Author	Forecast model	Forecast horizon	Performance metrics & forecast error measurement	Contribution & Perspective
Liu et al. (Liu et al. 2015)	– Back propagation based ANN	24 h ahead	MAPE = 7.65%	– Aerosol index parameter used as input, which resulted in slightly improved accuracy
Ramsami, Oree (Ramsami and Oree 2015)	– Stepwise Regression – GRNN – FFNN – MLR	24 h ahead	RMSE = 2.74%	– Stepwise regression: select I/P variables highly correlated to PV power O/P – GRNN,FFNN,MLR and their hybrid were applied on the I/P – Hybrid model showed slight improvement
Ordiano et al. (Ordiano et al. 2017)	– ANN6 – ANN10 (Data-driven)	24 h ahead	MAE = 6.64---7.25% RMSE = 12.47---13.3% $r(\rho, \bar{\rho}) = 85.81---87.65\%$	– Accuracy of model strictly related to accuracy of historical database (Achieved reasonable accuracy)

et al. 2013). This includes Pearson correlation (ρ), mean bias error (MBE, or bias), mean square error (MSE) and root mean square error (RMSE), mean absolute error (MAE) and standard deviation (SDE).

1. Pearson correlation is the coefficient that measures the correlation between actual and forecasted value defined below in (1)

$$\rho = \frac{\text{cov}(\rho, \bar{p})}{\sigma_{\rho} \sigma_{\bar{p}}}$$

2. The metric RMSE introduced by (Zhang et al. 2015) provides a global error measure throughout the entire forecasting period, given by (2)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_{pred} - p_{meas})^2}$$

3. Metric MAPE assess uniform prediction errors given by (3)

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{p_{pred} - p_{meas}}{p_0} \right|$$

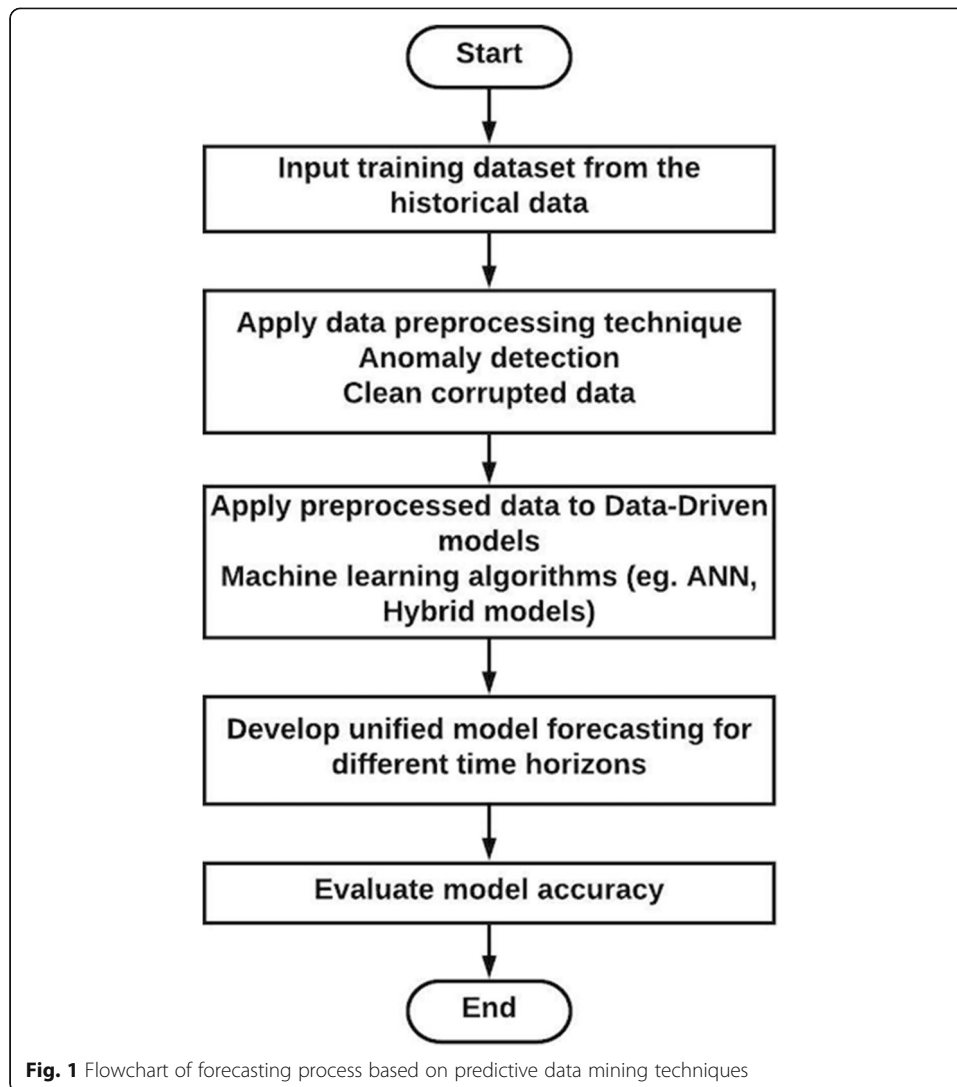
Where p_{meas} represents actual solar power generation at i_{th} time step, p_{pred} is the corresponding solar power generation estimated by forecasting model, N is the number of points estimated in the forecasting period. This metric is useful for evaluating the overall performance of the forecasts, especially when extreme events are a concern.

Methodology

Figure 1 presents the flowchart of the proposed forecasting process. To address the research questions, we first propose to conduct a case study that aims to benchmark the anomaly detection method and evaluate the link between forecasting accuracy and anomaly detection method.

In this work we plan to include three steps:

- In the first step, data pre-processing techniques are applied to perform anomaly detection and outlier rejection. Three machine learning based approaches are considered:
 - Density-based anomaly detection
 - Clustering-based anomaly detection
 - Support vector machine-based anomaly detection
- In the second step, the pre-processed data with chosen anomaly detection technique obtained from the first step is used to train the data-driven model based on predictive data mining techniques. The outcome of investigation from these steps will explore the interplay between anomaly detection technique and forecasting model accuracy.
- The third step involves developing a unified model which forecasts accurately for different time horizons i.e. short-term, medium-term and long-term forecasting.



Conclusion

Intelligent decision making is important to provide an unprecedented flexibility in the energy management for the future power system. This requires accurate forecasts of future energy production and demand/consumption.

In this paper we first discussed the terminology of energy forecasting and its classification based on different time horizons followed by a detailed state-of-the-art which revealed that applying advanced soft computing approaches will likely outperform the statistical methods. However, complex methods have a limitation in terms of interpretability.

We proposed to apply data pre-processing technique along with data-driven forecasting model which can possibly improve the accuracy even when using partial information i.e. past power data.

Acknowledgements

The author would like to thank Prof. Dr. Clemens Van Dinther for his valuable feedback and comments during the shepherding process.

Funding

Publication of this article was sponsored by funds of the Smart Grids research group.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

About this supplement

This article has been published as part of *Energy Informatics* Volume 1 Supplement 1, 2018: Proceedings of the 7th DACH+ Conference on Energy Informatics. The full contents of the supplement are available online at <https://energyinformatics.springeropen.com/articles/supplements/volume-1-supplement-1>.

Authors' contributions

ES analysed related work, identified open issues, and developed a research proposal related to her PhD project. The author read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The author declares that she has no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published: 10 October 2018

References

- Alanazi M, Alanazi A, Khodaei A (2017) Long-term solar generation forecasting. 2016 IEEE/PES Trans Distrib Conf Expo (T & D): 1–13
- Chakhchoukh Y, Panciatici P, Mili L (2011) Electric load forecasting based on statistical robust methods. *IEEE Trans Power Syst* 26:982–991
- Chen X, Kang C, Tong X, Xia Q, Yang J (2014) Improving the accuracy of bus load forecasting by two-stage bad data identification method. *IEEE Trans Power Syst* 29:1634–1641
- Daliento S, Chouder A, Guerriero P, Pavan AM, Mellit A, Moeini R, Tricoli P (2017) Monitoring, diagnosis and power forecasting for photovoltaic fields: a review. *Int J Photoenergy*. 1–13
- Dobschinski J, Bessa R, Du P, Gleiser K, Haupt SE, Lange M, Mhrlen C, Nakafuji D, Rodriguez M (2017) Uncertainty forecasting in a nutshell: prediction models designed to prevent significant errors. *IEEE Power Energ Mag* 15:40–49
- Dolara A, Grimaccia F, Leva S, Mussetta M, Ogliaari E (2018) A physical hybrid artificial neural network for short term forecasting of PV plant power output. *Energies* 8:1138–1153
- Filik UB, Gerek ON, Kurban M (2011) Hourly forecasting of long term electric energy demand using novel mathematical models and neural networks. *Innovative computing. Inf Control* 7:115–118
- Gandelli A, Grimaccia F, Leva S, Mussetta M, Ogliaari E (2014) Hybrid model analysis and validation for PV energy production forecasting. 2014 Int Joint Conf Neural Netw (IJCNN):1957–1962
- Khatib T, Elmenreich W (2015) A model for hourly solar radiation data generation from daily solar radiation data using a generalized regression artificial neural network. *Int J Photoenergy*:1–13
- Liu J, Fang W, Zhang X, Yang C (2015) An improved photovoltaic power forecasting model with the assistance of aerosol index data. *IEEE Trans Sustainable Energy* 6:434–442
- Luo L, Hong T, Yue M (2018) Real-time anomaly detection for very short-term load forecasting. *J Mod Power Syst Clean Energy* 6:235–243
- Marino DL, Amarasinghe K, Manic M (2016) Building energy load forecasting using deep neural networks. *IECON 2016-42nd Ann Conf IEEE Ind Electron Soc*:7046–7051. [abs/1610.09460](https://doi.org/10.1109/IECON2016.7801946):1-6.
- Ordiano JAG, Waczowicz S, Reischl M, Mikut R, Hagenmeyer V (2017) Photovoltaic power forecasting using simple data-driven models without weather data. *Comput Sci Res Dev* 32:237–246
- Panapakidis IP, Bouhouras AS, Christoforidis GC (2018) A missing data treatment method for photovoltaic installations. 2018 IEEE Int Energy Conf (ENERGYCON):1–6
- Pelland S, Remund J, Kleissl J, Oozeki T (2013) Brabandere KD (2013) Photovoltaic and solar forecasting: State of the art. *Tech Rep IEA PVPS:T14–T01*. 1–36
- Ramsami P, Oree V (2015) A hybrid method for forecasting the energy output of photovoltaic systems. *Energy Convers Manag* 95:406–413
- Saberian A, Hizam H, Razid MAM, Kadir MZAA, Mirzaei M (2014) Modelling and prediction of photovoltaic power output using artificial neural networks. *Int J Photoenergy* 14:1–10
- Saleh AI, Rabie AH, Abo-Al-Ez KM (2016) A data mining based load forecasting strategy for smart electrical grids. *Adv Eng Inform* 30:422–448
- Suganthi L, Samuel AA (2012) Energy models for demand forecasting a review. *Renew Sust Energ Rev* 16:1223–1240
- Wan C, Zhao J, Song Y, Xu Z, Lin J, Hu Z (2015) Photovoltaic and solar power forecasting for smart grid energy management. *CSEE J Power Energy Syst* 1:38–46
- Zhang J, Florita A, Hodge B, Lu S, Hamann HF, Banunarayan V, Brockway AM (2015) A suite of metrics for assessing the performance of solar power forecasting. *Sol Energy* 111:157–175