Robert Ulbricht, Ulrike Fischer, Wolfgang Lehner, Hilko Donker

# Optimized renewable energy forecasting in local distribution networks

**Erstveröffentlichung in / First published in:**

*EDBT/ICDT '13: Joint 2013 EDBT/ICDT Conferences*, Genoa 18.03. – 22.03.2013. ACM Digital Library, S. 262-266. ISBN 978-1-4503-1599-9

DOI: https://doi.org/10.1145/2457317.2457360

Final edited form was published in "EDBT/ICDT '13: Joint 2013 EDBT/ICDT Conferences. Genoa 2013", S. 262-266, ISBN 978-1-4503-1599-9
https://doi.org/10.1145/2457317.2457360

# Optimized Renewable Energy Forecasting in Local Distribution Networks

Robert Ulbricht[2], Ulrike Fischer[1], Wolfgang Lehner[1], Hilko Donker[2]

[1] Technische Universität Dresden, Database Technology Group, Dresden, Germany
[2] Robotron Datenbank-Software GmbH, Dresden, Germany
[1] {first name.lastname}@tu-dresden.de
[2] {first name.lastname}@robotron.de

## ABSTRACT

The integration of renewable energy sources (RES) into local energy distribution networks becomes increasingly important. Renewable energy highly depends on weather conditions, making it difficult to maintain stability in such networks. To still enable efficient planning and balancing, forecasts of energy supply are essential. However, typical distribution networks contain a variety of heterogeneous RES installations (e.g. wind, solar, water), each providing different characteristics and weather dependencies. Additionally, advanced meters, which allow the communication of finalgranular production curves to the network operator, are not available at all RES sites. Despite these heterogeneities and missing measurements, reliable forecasts over the whole local distribution network have to be provided. This poses high challenges on choosing the right input parameters, statistical models and forecasting granularity (e.g. single RES installations vs. aggregated data). In this paper, we will discuss such problems in energy supply forecasting using a real-world scenario. Subsequently, we introduce our idea of a generalized optimization approach that determines the best forecasting strategy for a given scenario and sketch research challenges we are planning to investigate in future work.

## 1. INTRODUCTION

We can notice the strong influence of constantly increasing capacities of renewable energy sources (RES) due to excessive funding policies (e.g. [4], [9]) and industrial promotion, making conventional power plants becoming less attractive to private investors. Renewable energy is characterized by a decentralized allocation and fluctuating output, thus making it difficult to maintain stability in power networks where energy supply and demand must be balanced carefully. To prevent collapsing grids in the near future, and to battle the negative influences of the growing economies' ever increasing energy consumption on greenhouse gas emissions, new technical concepts are introduced. This includes intelligent and networked measurement devices (the so called smart meters), efficient storage systems or the integration of mobile consumption points.



However, even the smartest meter technology can only provide a recorded view on past situations, but the key to balance an energy distribution network successfully is to plan actions foresightfully by predicting as many of the most influencing (correlated) parameters for operations as possible. For energy demand, very good forecast results can already be achieved based on historical data, user experience and the application of simple mathematical functions [10]. In contrast, the supply side is much more challenging as it heavily depends on weather conditions like wind speed and direction, global heating or cloud coverage. This adds a new dimension to energy forecasting and is the main reason why in the past years a lot of research has been conducted to improve prediction quality.

Typical RES allocations in local or regional networks contain a variety of different RES installations, including solar panels, wind parks or water power plants. Such installations range from large parks to single installations deployed in private households. Many meters installed at smaller RES sites still do not have a communication module. Load curves are not available and production data has to be retrieved manually in several periods, usually once or twice a year. Only a few large installations provide automatically measurements to the network operator in regular time intervals (usually 15 minutes). Additionally, private households increasingly auto-consume produced energy (prosumers) and provide only surpluses to the energy grid, leading to a mixed load curve containing energy demand and supply. Moreover, private storage capacities, such as night storage heating or electricity car batteries, allow the storage of energy for later use.

Despite all these heterogeneities as well as missing and hidden measurements in existing regional networks, reliable supply forecasts over the whole network have to be provided. To cope with all these challenges a distribution system operator (DSO), who is responsible for balancing energy supply and demand, can carry out different kinds of strategies. For example, a straight forward approach might be to aggregate available supply curves, extrapolate and build a single sta-

Final edited form was published in "EDBT/ICDT '13: Joint 2013 EDBT/ICDT Conferences. Genoa 2013", S. 262-266, ISBN 978-1-4503-1599-9
https://doi.org/10.1145/2457317.2457360

tistical model on the aggregated curve. Alternatively, statistical models might be built for existing individual curves, allowing customized approaches for certain types of RES or prosumers. Finally, clustering approaches, as already applied in energy demand forecasting [7], can identify clusters of supply patterns and compute a common forecast for each cluster.

In this paper, we introduce the challenge of forecasting renewable energy resources in local distribution networks. Using a real-world scenario, we first describe existing components and problems in a local distribution network (Section 2). Then, in Section 3, we present the status of our work on a generalized optimization approach for determining the best forecasting strategy in such networks. We conclude with open challenges that we are planning to investigate in future work (Section 4).

## 2. REAL-WORLD SCENARIO

To demonstrate the problem introduced in Section 1 we use a real-world dataset. The scenario represents all renewable energy installations of an exemplary town and it's neighboring villages located in central Germany. For our work, we require (1) the installations master data like the type of energy production, location and installed power, (2) every suppliers yearly output and (3) load curves of all available measured devices. In addition, numerical weather forecasts (4) are needed in order to apply regression-based forecast models, as this is the common energy supply prediction approach [5].

According to the German energy domain's transparency rules, master data (1) and output (2) of renewable energy sources has to be published periodically by the corresponding transmission network operator[1] and is available to the public. RES load curves (3) were provided by the local DSO and weather data (4) could be obtained from meteorological services[2].

| Symbol | Type | Number | Power |
|---|---|---|---|
| SOL | Solar load profile | 10 | 3625 kW |
| SOT | Solar non-profiled | 82 | 850 kW |
| WAP | Water load profile | 1 | 120 kW |

**Table 1: Scenario structure**

The combination of these four elements results in a scenario structured as follows: 10 photovoltaic (PV) installations equipped with load profile meters (SOL) with a total of 3,625 kW installed capacity, 82 PV-installations without profile meters (SOT) adding up to 850 kW and one run-of-the-river power plant (WAP) with a capacity of 120 kW (compare Table 1). Further, we have one centrally located weather station providing global radiation and outside temperature as time series with hourly resolution. We consider this to be a representative constellation for such local distribution networks, as renewable energy in settled areas is mainly provided by a couple of powerful free standing PV-installations combined with more numerous but smaller roof-top ones. Although the scenario's network has a total load capacity of 19,700 kW we have to bear in mind that consumption peaks will usually take place during the winter season, where PV output is naturally lower. Instead, as peak PV production values are obtained on clear-sky summer days coinciding with lower energy demand (what happens especially on non-working days), the share of uncertain RES supply on the network's distributed energy reaches up to 30% for such moments. This increases the risk of unstable supply situations, so careful balancing is needed.

As weather influences strongly vary over increasing distances, an RES's geographic location is the most important context information in order to determine the expected reliability of weather forecasts at a given point. By using their coordinates we can visualize the allocation of RES within the distribution network as a grid-like structure. This is shown in Figure 1, displaying all unmeasured (SOP, red square) and measured PV-installations (SOL, blue rhombus), complemented by the single water power plant (WAP, grey cross) and the weather station (green triangle).

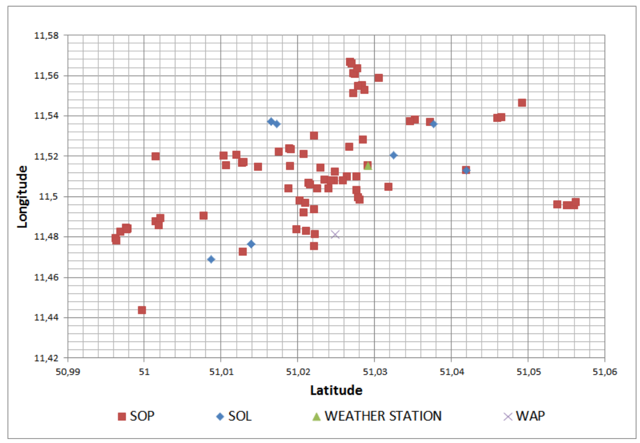


**Figure 1: Local RES distribution**

The presented scenario covers only 23 km$^2$ and contains just ten measured PV installations. A big city's or even a regional distribution network will be much more challenging, containing much more measured as well as unmeasured RES installations. Further, the regional networks are of special interest, as they are usually composed of numerous towns and villages connected by vast rural areas. Besides PV-based energy supply, such datasets will offer additional types of fluctuating RES like windmills, wave energy or solar-thermal power, which requires e.g. the coupling of wind data for geographical areas.

The DSO is interested in the overall (aggregated) energy supply forecast. Due to the fact that many PV installations are unmeasured, we cannot simply aggregate over all load profiles and compute the forecast. We have to cope with unmeasured installations by using reference profiles or extrapolation. Additionally, we have several possibilities to choose and compute statistical forecast models, i.e., single RES installations vs. aggregated load curves. The chosen aggregation level and forecasting technique highly influence the forecast accuracy and are the main challenges we want to discuss in this paper.

[1]http://www.50hertz.com/de/163.htm
[2]http://wetterstationen.meteomedia.de

Final edited form was published in "EDBT/ICDT '13: Joint 2013 EDBT/ICDT Conferences. Genoa 2013", S. 262-266, ISBN 978-1-4503-1599-9
https://doi.org/10.1145/2457317.2457360

# 3. OPTIMIZATION PROBLEM

We model the calculation of predicted RES production in local distribution networks as an optimization problem, defining input and output values and describing our optimization approach.

*Input.*
As input values, we consider sets of heterogeneous RES (e.g. solar, wind, water) with different context parameters, such as total capacity, maintenance times, and physical characteristics. The latter might be, if available, an PV-installations' inclination angle or the production year of the solar panels in use, since it is known that their output decreases over age. For each single RES installation the historical yearly amount of produced energy is available, while historical load profiles in hourly or smaller granularity are only available for a subset of the installations. The number of historical values of the total produced energy as well as the load profiles varies and depends on the beginning of the operation of the RES installation. Finally, different weather stations might be available within the local distribution network, which provide weather condition measurements and forecasts in possibly varying granularities. Depending on the types of RES, the energy output will be influenced by values such as outside temperature, cloud coverage, global radiation, wind speed and -direction or even the rainfall close to the RES location and in nearby areas if connected by rivers.

*Output.*
On the output side, we want to obtain an aggregated load curve for next days' expected supply of renewable energy in the whole network with a resolution of 15min. This time series consists of combined forecasts for all available types of RES included in our scenario, thus concerning both measured and unmeasured installations. Our overall optimization goal is the minimization of the forecast error, i.e. the deviation of the forecasted load curve from the real aggregated one.

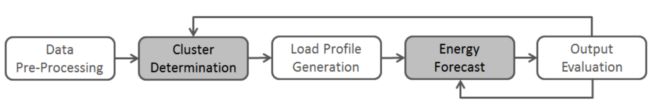


**Figure 2: Iterative output optimization**

*Optimization Approach.*
Our general approach tries to find suitable clusters of RES installations, containing measured and unmeasured installations, and computes supply forecasts for each individual cluster. In detail, our approach can be described as an iterative process shown in Figure 2. First, *pre-processing* of raw input data is needed in order to guarantee suitable data quality. This may include the normalization of non-equidistant source time series or the generation of substitution values for any missing but required information. Second, $N$ *clusters* are determined, e.g. by using an installation's context parameters or applying pattern recognition techniques in order to identify similar load curves. Based on this classification, we generate a *load profile* for each cluster $CLC_{1..N}$. This might be done by aggregating historical load curves and by extrapolating the aggregated curve to integrate unmeasured installations and to obtain a cluster's total load curve. Next step is the *energy forecasting*, which generates predicted time series $FLC_{1..N}$ for each cluster by applying forecasting techniques to $CLC_{1..N}$. Finally, the result of our interest computes as $\sum_1^N FLC$. This output must be *evaluated* using adequate error measures like the root-mean-square error (RSME) or the symmetric mean absolute percentage error (SMAPE). Hence, the following equation shall be minimized:

$$error(\sum_1^N CLC, \sum_1^N FLC). \tag{1}$$

Depending on the result of the evaluation, either the cluster determination or the energy forecasting step can be repeated, each time applying different methods, models or parameter settings.

For example, reconsidering the scenario presented in Section 2, a simple clustering algorithm for PV-installations can consist in: (1) Aggregate all SOLs with identical geographic position, (2) use them as cluster centroids and (3) assign the nearest SOTs to each centroid. This results in the seven geographical clusters shown in Figure 3, where the number of assigned installations (dark red) and aggregated power capacities (light blue) are presented. Once having computed the aggregated load curve for each cluster, stochastic regression models can be applied in order to predict the expected energy production. Typically, there is a strong correlation between solar energy supply and global radiation (probably as the most relevant external influence), so these time series must not be missing.

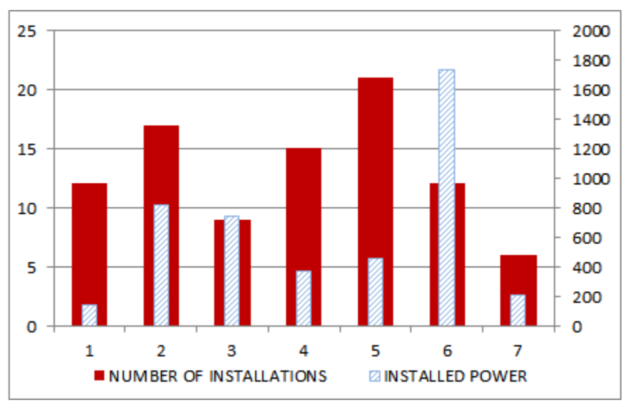


**Figure 3: First-level cluster for PV installations**

# 4. FUTURE WORK CHALLENGES

In order to solve the optimization problem, we have to study the fields of cluster determination and energy forecasting more in detail, as these are considered to be the most promising steps in the defined iterative process chain. In contrast, raw-data processing and load profile generation are typical energy logistics tasks, in practice usually realized by highly-automated energy data management systems [12].

*Cluster Determination.*
Regarding clusters, we focus on three problems: (1) How to choose the optimal clustering methods for a given dataset?, (2) Can hierarchical aggregation levels be defined? and (3)

Final edited form was published in "EDBT/ICDT '13: Joint 2013 EDBT/ICDT Conferences. Genoa 2013", S. 262-266, ISBN 978-1-4503-1599-9
https://doi.org/10.1145/2457317.2457360

Which maintenance strategies are needed to integrate dynamic influences?

*Clustering Methods:* The clustering of time series data is a frequently treated research problem with effective practical applications in various domains. A survey of existing methods is provided by Warren Liao [13], showing also commonly used criteria for clustering performance evaluation and corresponding measures to determine the similarity between two compared time series. In general, existing methods belong to three major categories: (1) clustering using raw-data, (2) feature-vector extraction and (3) model-based clustering. Respective to the energy domain, Misiti et al show how clustering can be used to disaggregate a global energy demand load curve [7] to obtain forecasts of finer granularity. However, to the best of our knowledge, there is no related work available on how to apply such clustering approaches to specific RES output curves. To investigate the possibilities of adopting those standard techniques to our scenario is the first challenge for future work.

*Hierarchy Levels:* Forecast quality will vary depending on the size of each cluster, because the overall impact of single peak values decreases with increasing size of the aggregates. Besides the first-level example shown in Figure 3 resulting in a maximum number of cluster elements, more combinations can be computed following an agglomerative (bottom-up) approach by combining similar centroids on each level, thus forming new aggregates. To determine the best aggregation level for forecast models, hierarchical clustering techniques (e.g. [8]) must be applied, a challenge of special interest in expanded scenarios.

*Maintenance:* Clusters in use need to be maintained, as certain dynamic influences will affect the underlying master data. Adding or removing load profile installations or weather stations to a network might require the redefinition of either single cluster elements or even the whole hierarchical structure. Efficient maintenance strategies must be developed to reduce these calculation efforts while avoiding data integrity violations.

### *Energy Forecast.*

Once suitable RES clusters are found, a statistical model has to be built over each cluster, which computes the corresponding forecast values. Hereby, we face three main challenges: (1) Which forecasting approach should be applied?, (2) How much history length should be used for estimating the model parameters? and (3) How should mixed load curves (e.g. prosumers) be handled?

*Forecasting Approach:* Typical energy forecasting approaches are classified into physical approaches, which rest upon an installations physical and technical character, and stochastic approaches based on time series analysis (e.g. regression, exponential smoothing). Physical approaches do not depend on measured load curves and can quickly be implemented, making them indispensable especially for new installations, where no historical data are available. However, their practical usefulness is limited, as DSOs normally do not have access to all required technical details and as aggregated load curves may combine several RES installations with different technical properties. The advantage of stochastic approaches is the direct availability of historical load profiles from the time series database and the maintenance of data assured by regular updating processes. Various general [6] as well as energy-specific stochastic approaches are available in the literature and selecting the best stochastic approach (or combination of them) for a given cluster exhibits a major challenge. For example, Devaine et al demonstrate in their work how combined specialized experts can be used to obtain an aggregated energy consumption forecast [3]. Additionally, in the case of RES, external information such as wind speed, cloudiness or temperature, have to be efficiently integrated and might be characterized by complex relationships [2]. Most of these values are often highly localized, strongly varying and therefore hard to predict in detail, which will require dynamic short-range adjustments of already computed predictions based on evolving weather situations.

*History Length:* Stochastic approaches require the building of a *forecast model* over the historical load curves, which involves multiple iterations over the time series history to determine the best model parameters. This is an expensive process especially on large data sets i.e. with a long history. Reducing the history length can therefore significantly speed up model creation. However, more importantly, the history length influences the accuracy of the forecasting result. A longer history length might be suitable for learning repeatable patterns, while a shorter history length is more beneficial for strongly fluctuating time series data. The latter requires a continuous adaption of the forecast models and, possibly, of the history length. Previous research in this area has mainly investigated the influence of the historical length for financial data [1]. Ge and Zdonik [11] propose an I/O-conscious skip list data structure for very large time series in order to determine the best history length and number of data points for linear regression models. How such techniques can be applied to more sophisticated models as required in supply forecasting needs to be investigated.

*Forecasting Prosumers:* Finally, small RES installations in private households increasingly self-consume or even store their produced energy. Simultaneously, often only one meter is installed to measure the overall load transmitted to the network, which corresponds to the total produced energy minus the consumed energy. In order to incorporate the output of prosumers in the aggregated production curve, traditional forecasting approaches have to be adapted to handle such special kind of time series data.

As shown in this section, plenty of interesting opportunities for further research directions can be identified, all of them addressing the problem of optimization of RES forecasting in local distribution networks or similar use cases.

## Acknowledgment

The work presented in this paper was funded by the European Regional Development Fund (EFRE) under co-financing by the Free State of Saxony and Robotron Datenbank-Software GmbH. We especially thank Falk Glockner from ENA for supporting our research.

Final edited form was published in "EDBT/ICDT '13: Joint 2013 EDBT/ICDT Conferences. Genoa 2013", S. 262-266, ISBN 978-1-4503-1599-9
https://doi.org/10.1145/2457317.2457360

# 5. REFERENCES

[1] T. G. Andersen, T. Bollerslev, and S. Lange. Forecasting financial market volatility: Sample frequency vis-a-vis forecast horizon. *Journal of Empirical Finance*, pages 457–477, 1999.

[2] L. Dannecker, E. Vasilyeva, M. Boehm, W. Lehner, and G. Hackenbroich. Efficient integration of external information into forecast models from the energy domain. In *Advances in Databases and Information Systems*, pages 139–152. Springer Berlin Heidelberg, 2012.

[3] M. Devaine, P. Gaillard, Y. Goude, and G. Stoltz. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning*, 90:231–260, 2013.

[4] European Commission. Energy roadmap 2050. http://ec.europa.eu/energy/energy2020/roadmap/-index_en.htm, 2011.

[5] W. Glassley, J. Kleissl, C. P. van Dam, H. Shiu, J. Huang, G. Braun, and R. Holland. Current state of the art in solar forecasting. Technical report, California Renewable Energy Collaborative (CREC), 2012.

[6] J. G. D. Gooijera and R. J. Hyndman. 25 years of time series forecasting. *International Journal of Forecasting*, 22:443–473, 2006.

[7] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi. Optimized clusters for disaggregated elecricity load forecasting. *REVSTAT*, 8:105–124, 2010.

[8] P. Pereira Rodrigues. *Hierarchical Clustering of Time Series Data Streams*. PhD thesis, University of Porto, Portugal, 2005.

[9] Renewable Energy Policy Network for the 21st Century. Renewables 2011 - global status report. http://www.ren21.net/REN21Activities/Publications/GlobalStatusReport/GSR2011/tabid/56142/-Default.aspx, 2012.

[10] J. Taylor and A. Espasa. Energy forecasting. *International Journal of Forecasting*, 24(4):561–565, 2008.

[11] G. Tingjian and B. Z. Stanley. A skip-list approach for efficiently processing forecasting queries. *Proceedings of the VLDB Endowment*, 1:984–995, 2008.

[12] R. Ulbricht, U. Fischer, W. Lehner, and H. Donker. Rethinking energy data management - trends and challenges in today's transitioning markets. To appear in Proceedings of BTW 2013, Magdeburg, Germany, 2013.

[13] T. Warren Liao. Clustering of time series data - a survey. *Pattern Recognition*, 38:1857–1874, 2005.