Copyright and Reference Information: This material (preprint, accepted manuscript, or other author-distributable version) is provided to ensure timely dissemination of scholarly work. Copyright and all rights therein are retained by the author(s) and/or other copyright holders. All persons copying this work are expected to adhere to the terms and constraints invoked by these copyrights. This work is for personal use only and may not be redistributed without the explicit permission of the copyright holder. The definite version of this work is published as
[·] Michael T. Beck, Hermann De Meer, Stefan Schuster and Martin Kreuzer. Estimating photo-voltaic power supply without smart metering infrastructure. In *Proc. of the 2nd Int'l Workshop (E2DC2013)*, Volume 8343 of *Lecture Notes in Computer Science (LNCS)*, pages 25–39. Springer-Verlag, May 2013. The original publication is available at www.springerlink.com (2013).
See http://www.net.fim.uni-passau.de/papers/Beck2013b for full reference details (BibTeX, XML).

Estimating photo-voltaic power supply without smart metering infrastructure

Michael Till Beck[1,3], Hermann de Meer[1],
Stefan Schuster[2], and Martin Kreuzer[2]

[1] {michael.beck,hermann.demeer}@uni-passau.de,
Computer Networks and Communications, University of Passau, Germany
[2] {stefan.schuster,martin.kreuzer}@uni-passau.de,
Symbolic Computation, University of Passau, Germany
[3] Stadtwerke Passau GmbH, Passau, Germany

Abstract. Due to the lack of appropriate grid communication infrastructure, many energy providers can only measure a very limited subset of their PV plants and therefore have only limited knowledge of the power flow inside their grid. Existing approaches to estimate the total amount of PV energy produced at present time ("nowcasting") require external data such as sun radiation or temperature that are often not available online. Using approximate computational algebra, we construct polynomial models to derive grid-specific formulae estimating the PV power provisioning without the need of additional data. We evaluate our approach based on real data from a German energy provider and demonstrate the accuracy of the derived models. Besides nowcasting, two additional application scenarios, snapshot provisioning and simulation of power flow, are discussed.

1 Introduction

Recently, many European countries integrate huge amounts of renewable based power sources into the grid. They intend to reduce dependencies on fossil sources like coal, oil, and nuclear resources. Some countries like Germany, Italy, and Belgium even aim at completely eliminating these dependencies on nuclear power. In many other European countries political ambitions are similar, since many countries now bolster the integration and utilization of renewables and debate nuclear power phase-outs.

Mainly driven by political objectives, the amount of renewable energy sources that are fed into the power grid heavily increased in Germany lately. The German government advocates the integration of renewable based power generation in favor of reducing dependencies on nuclear power sources. In 2011, the German government announced their objective to shut down all German nuclear power stations by 2022. For this reason, large arrays of wind turbines and solar panels have been installed. In addition to that, the German government assured monetary incentives to citizens that install photo-voltaic panels on their residence's roof. Most of these private power plants are rather small in size and only produce a limited amount of power. However, many citizens decided to participate in the initiative due to monetary incentives granted. So in total, across Germany, photo-voltaic power production rapidly increased (and is still continuously increasing) within the power grid.

Traditionally, the power grid was not intended to cope with the integration of large amounts of highly fluctuating energy sources. Therefore, the main challenge with power plants that are based on renewables like wind or solar radiation is that power production is highly fluctuating and, in general, hard to predict. Fluctuations, however, directly influence and destabilize the frequency in the grid, if the energy provider can not cope with them in terms of adapting its other power plants in-time or (with respect to demand/response mechanisms) negotiating with flexible customers. In this case, the energy provider has to fall back to (negative or positive) balancing reserve power which comes with very high costs. For these reasons, precise prediction of energy that is expected to be available in the (near) future is essential for the energy provider due to economical and ecological reasons.

Photo-voltaic power generation is drastically increasing in Germany. During the year 2010, approximately 7 GWpeak of solar power plants were installed, which, at that time, rapidly increased the total available capacity by 70%. Power generated by these new photo-voltaic power plants was often underestimated by the distribution system operators (DSOs). In Germany, this became most conspicuous in September 2010, where an unexpected imbalance of +7 GW occurred for several hours. The German DSOs were not able to predict this rapidly increasing amount of positive imbalance, as they underestimated the impact of the newly installed panels. Thus, adaption of their power plants could not be performed in time. Since overproduction exceeded all of the available negative balancing reserve power (4300 MW), a huge amount (~2800 MW) had to be paid to other countries. This resulted in high costs and almost in a break down of the grid [1] [2].

The reason for these kinds of underestimations lies in the current infrastructure of the grid. Energy providers just have started to upgrade grid infrastructure. This includes, but is not limited to additional power lines and communication channels. In almost every part of Germany (and also in many other European countries), smart metering devices have not yet been deployed due to excessive additional expenses and efforts. This also means that most of the photo-voltaic panels can not be measured directly. In fact, approximately 75% of a total of 900.000 photo-voltaic power plants can not be measured directly due to technical limitations [2].

Therefore, in the age of renewable energy sources, energy providers have to monitor grid stability, e.g., by estimating the amount of power provided by the photo-voltaic plants that can not be measured directly. Being able to obtain more accurate information, energy providers can react more precisely to discrepancies, which leads to an increase in overall, trans-regional grid stability.

In this paper, we introduce a novel nowcasting methodology to estimate the total available photo-voltaic power by mathematically analyzing correlations of power provisioning characteristics. In contrast to others, our approach does not depend on external information like solar radiation or additional information on the type and alignment of panels. Therefore, our approach is expected to be more accurate, especially in regional contexts. Furthermore, we are confident that accuracy of existing forecasting algorithms can be significantly improved by also taking into account the hidden interdependencies of PV plants. This is especially interesting for demand/response mechanisms that are subject of current research: Instead of just tailoring energy provisioning to the demand, demand/response mechanisms are currently also integrated into grid infrastructures: In case of power shortages/surplus, energy demand of flexible customers can be reduced by notifying them to adapt power consumption accordingly. Up to now, demand/response mechanisms have been deployed for major customers like big factories only, since they show the most potential of power adaption capabilities.

However, current research also focuses on integrating mid-size consumers like data centres into the grid (or even small customers like private homes)[3] [4]. Since data centres are expected to be able to adapt power demand much more fine-grained than factories, more accurate information of power surplus/shortage is becoming more and more important. Using nowcasting techniques, a *snapshot* of the current state of the grid can be derived, which is a valuable input for these adaption mechanisms.

We run our evaluation on data provided by an energy provider located in Bavaria, Germany. First results show that the approach seems to be quite promising, even for small-scale grids and small geographical distances between power plants.

2 Background

The integration of renewable power sources into traditional power grids comes with several difficulties and challenges, since they notably differ from traditional, fossil based power sources. During the day, power quality needs to be maintained in the grid continuously for the grid to remain stable. High power quality means that voltage and frequency do not (or only within very tight boundaries) vary from specific values. Disruptions and disturbances caused by unforeseen effects have to be avoided by all means to ensure that power provisioning is working properly and to avoid grid failure [4]. Traditionally, this has to be ensured by the responsible energy provider by switching off unneeded power plants in time in case there is additional power fed into the grid by renewable power sources. Similarly, the energy provider has to react timely if these additional power sources

[4] The All4Green project, which is funded by the European Union, aims at integrating data centres into the smart grid. It introduces new service level agreements for data centres (and also for its end users) to define the degree of flexibility in terms of quality of service. For further information, please refer to http://www.all4green-project.eu

disappear. A major challenge in this respect is that adapting these ”’traditional”‘ power sources has to be done quickly, i.e., as fast as the renewables appear or disappear. Therefore, energy providers have to carefully plan ahead hours of operation of their power plants, since provisioned power has to match demands.

For this reason, energy providers have to take weather forecasts into account, which introduce uncertainties in terms of predictability. Predictability, however, is of significant importance for the energy provider for operating economically and ecologically. Power that is provided by various kinds of power plants always relies on constraints that come with these differing power plants and the power sources they use. For example, due to technical reasons, several models of diesel generators can only be used for a limited number of times a year; nuclear power plants can not provide power immediately after they are switched on. Thus, in general, there are four types of power plants that distinguish in reactiveness, ecologic footprint, and economic costs:

1. Base Load Power Plants:
   Power plants delivering high amounts of inflexible energy. They are inflexible in terms of reactiveness, but power provisioning in general is relatively cheap. These power plants can not be used to cover Medium Load, nor can they compensate quickly fluctuating power sources.
   Examples: nuclear power plants, coal power plants

2. Medium Load Power Plants:
   These power plants can react quicker than Base Load Power Plants to changing demands. I.e., they can cover the lack of energy that was forecasted to be provided by a photo-voltaic plant if the changed weather conditions were predicted hours or several minutes before. However, Medium Load Power Plants can not be used to cover on-demand events.
   Examples: combined gas and steam energy generators

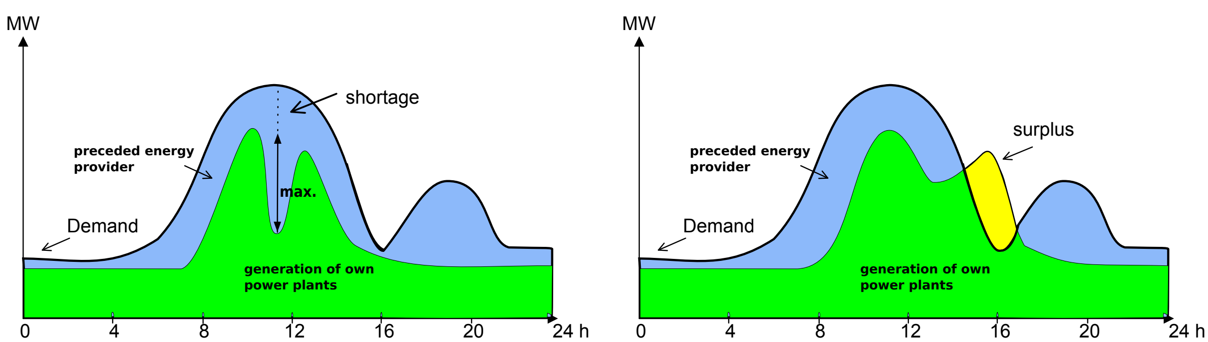
3. Peak Load Power Plants:
   Peak Load Power Plants react quickly and almost in-time to unforeseen changes. However, the usage of these plants is costly and in general also involves high CO2 emissions.
   Examples: diesel generators and gas turbines

4. Plants relying on uncontrollable power sources:
   Power plants that are not controllable in terms of power provisioning times. These power plants depend on renewable energy sources like sun or wind. Since these sources are highly fluctuating and their local availability and disappearance are hard to predict, huge amounts of additional power might rapidly disappear in the grid, so the energy provider has to adapt power provisioning of its other plants in-time. Weather forecasting techniques are an essential tool for reasonable, economically arrangement.

Thus, grid stability is a complex task, and power demands and power supply have to be planned carefully. For this reason, the European grid is not just a collection of individual, national (or even local) grids, but is highly interconnected. This allows to transfer cheap energy (that is currently available in the south of Europe, for example) to the north, if it is not needed. In general, this also helps to stabilize the grid, since partners can help compensating local instabilities. I.e., a local energy provider is connected to its preceded energy provider. The preceded provider might also provide a major contribution to the energy supplied locally, usually within stipulated bounds.

As a consequence of integrating fluctuating renewable sources like solar power plants and wind turbines, energy providers have to cope with two main challenges: Power Shortage and Power Surplus.

1. Power Shortage
   Power shortage is often caused by special mainstream events (e.g., football matches), changing weather conditions (sudden drop of renewable energy due to upcoming, sky-covering clouds or reduced wind), or due to seasonal conditions (e.g., heating in winter, usage of air-conditioning during summer months). To cover power shortages, economically and ecologically expensive peak load power generation has to be activated, or additional power needs to be bought from third-party suppliers (that might even reside in a foreign country). Otherwise, the grid will become unstable and break down.
   Basically, there are two reasons for a power shortage: Failures or forecast deviations. In case of a power plant's breakdown, power supply suddenly drops within the grid and other plants have to compensate the lack of power generation. The same happens if there are forecast deviations, i.e., if forecasted reneable based energy generation differs significantly from their real power supply. Figure 1a depicts a shortage at 12pm. The maximum amount of power supplied from the preceded energy provider exceeds the stipulated bounds and the lack of power can not be compensated by own power plants.



(a) power shortage

(b) power surplus

Fig. 1: Deviations of power supply and power demand

2. Power Surplus
   If more power is available than demanded, the energy provider has to deal with power surplus. This can be caused by sudden increases of uncontrollable power plants (e.g., wind and no clouds) or the fact that power demand is decreasing (e.g., on sunny Sundays or holidays, where industrial production is paused and people go outside). In this case, power plants have to be switched off (either completely or partly). Sometimes, unneeded power resources even have to be sold to third party suppliers/foreign countries, mostly with negative prices. Otherwise, a power surplus will lead to grid instability.
   Reasons for a power surplus are similar to the ones described for power shortages: One reason is that there are failures, i.e., on the side of a (big) customer. In this case, power demand suddenly decreases and power plants have to be adapted accordingly. Another reason is that there might be deviations from forecasts so that power plants could not be arranged accordingly in time. Figure 1b depicts a surplus of energy: At around 4pm, there is more energy generated by, e.g., solar power plants than needed.

3 Related Work

Up to now, research has investigated in two closely related, but still different areas in this field: *Forecasting* and *Nowcasting* techniques. Forecasting aims at predicting the available amount of power provisioning that will be available in the future. Based on weather forecasts, energy providers can estimate how much energy will probably be available within the next couple of hours (or even days). Then, based on this information, they can derive how much additional energy they have to buy from third party suppliers or on the stock market – or, in case of a power surplus, how much energy they have to sell.

Therefore, prediction methods are used for estimating the demanded amount of power. In this case, forecasting mostly focuses on inter-day predictions, i.e., on expected weather conditions for the next day, or even on intra-day to predict weather changes within the next few hours. However, from the energy provider's perspective, short-term forecasting techniques are getting more and more attractive. This is mainly due to the rapid integration of photo-voltaic panels in several parts of Europe. Weather forecasting is highly complex and, therefore, prediction of available energy is very error prone in this case. Especially, forecasting of direct sunlight beam is much more inaccurately to perform than global irradiance. Several approaches have been proposed to predict future weather conditions. Also, based on weather forecasting, several methods have been introduced to estimate the actual amount of power that will be provided by solar panels or wind turbines. [5] evaluates and compares several of them.

In general, since both forecasting approaches (weather forecasting, and, based on this, power forecasting) are highly unreliable, the real amount of power provided by renewable sources can differ significantly from the predicted values. Several Demand/Response protocols like OpenADR guarantee incentives to the customers that are willing to support the energy provider by adapting their power consumption accordingly. These incentives should be based on the current state of the grid, i.e., power consumption that needs to be decreased/increased. However, in many European countries like Germany, grid infrastructure is not ready for the integration of a more complex communication system, i.e., Smart Metering. Therefore, the real amount of currently available photo-voltaic power can not be determined exactly, since measurement data can not be sent to the energy provider. Therefore, Saint-Drenan et al. propose a novel approach to estimate the amount of photo-voltaic power based on sattelite data [2].

In this paper, we propose another nowcasting approach to detect hidden dependencies derived solely from a subset of photo-voltaic power plants. We aim to provide a methodology to derive models that are accurate even for small-scale grids and small geographical distances.

4 Estimating Available Photo-Voltaic Supply in the Grid

We aim at deriving grid-specific formulae to estimate the amount of power fed into the grid by power plants that can not be measured continuously. The approach currently used by German energy providers to estimate this amount assumes a linear relation to the amount of energy which is produced by a small number of directly measured PV plants [6]. However, as discussed in [2], the situation is more complicated: Due to certain characteristics of PV plants (e.g., module orientation) the correlation of produced PV energy between different PV plants is in general rather low. Assuming not linear but more generally *polynomial* relations between PV plants, we therefore suggest in the following an approach based on ideas coming from algebraic geometry and evaluate it against data obtained by a real power grid, located in Bavaria, Germany. After discussing the results we describe possible application scenarios.

4.1 Mathematical Background

Our modeling approach is based on the so called extended ABM algorithm (see [7]), an advancement of a group of algorithms that were developed to obtain polynomial descriptions of physical systems (see [8], [9]) The common assumption hereby is, that a certain set of measured data contains polynomial relations that describe the system under consideration. The goal is to exhibit these relations. The approach is purely data driven since only the data itself and no further assumptions are used to construct the desired models.

To make this idea precise, let $\mathbb{X} = \{p_1, \dots, p_s\} \subset \mathbb{R}^n$ be a finite set of s measured data points, e.g., the power production of n different PV plants, measured at s points in time. The relations in question are polynomials $f \in \mathbb{R}[x_1, \dots, x_n]$ which *vanish ϵ-approximately on* $\mathbb{X}$ for a given $\epsilon \geq 0$, i.e., $(f/||f||)(p) \approx_\epsilon 0$ for each $p \in \mathbb{X}$. Here, $|| \cdot ||$ denotes the Euclidean norm of the coefficient vector of f and $a \approx_\epsilon b$ holds for two real numbers a and b iff $|a - b| \leq \epsilon$. The threshold number ϵ thereby reflects the noise present in the data. Consider for instance the set

$$\mathbb{X} = \{(0, 0), (1, 0.98), (2, 4.01), (3, 8.9), (4, 16.02)\} \subset \mathbb{R}^2 \tag{1}$$

of 5 data points in the plane. Then the polynomial $f = y - x^2 \in \mathbb{R}[x, y]$ vanishes 0.1-approximately on $\mathbb{X}$.

To construct polynomials as desired, Limbeck [7] suggests the Approximate Buchberger-Möller (ABM) algorithm, a new combination of the Buchberger-Möller algorithm for border bases (cf. [9], [10]) and the singular value decomposition, to compute the ϵ-approximate kernel of certain evaluation matrices: Given a set of data points $p_1, \dots, p_s \in \mathbb{R}^n$ and a threshold number $\epsilon \geq 0$, the ABM algorithm constructs a finite set $G = \{f_1, \dots, f_t\} \subset \mathbb{R}[x_1, \dots, x_n]$ of polynomials that vanish ϵ-approximately on $\mathbb{X}$. The Buchberger-Möller algorithm reduces the problem of finding polynomials that evaluate the given points *exactly* to zero to the problem of computing the kernel of linear mappings that come from evaluating just terms at every given point. In the approximate setting the question to compute the approximate kernel of those evaluation matrices is answered by the well organized exploitation of singular value decompositions.

The situation just described is *homogeneous* in the sense that we ask for polynomial equations with right-hand side $\approx_\epsilon 0$. If we allow a non-zero right-hand side, we accordingly ask a more general question, which we can regard as the *inhomogeneous* case. To this end, consider the *tuple* $\Xi = (p_1, \dots, p_s)$ of s data points $p_i \in \mathbb{R}^n$ and assume that $Q = (q_1, \dots, q_s) \in \mathbb{R}^s$ is a tuple of further data points. The goal is now to construct polynomials $f \in \mathbb{R}[x_1, \dots, x_n]$ such that $f(p_i) \approx_\epsilon q_i$ for all $i = 1, \dots, s$, or in other words, such that each f evaluates ϵ-approximately to Q on Ξ. An algorithmic solution to this problem is given by the *extended approximate Buchberger-Möller algorithm* (*extended ABM*, [7]). Given data points $p_1, \dots, p_s \in \mathbb{R}^n$ combined in the tuple $\Xi = (p_1, \dots, p_s)$, a threshold number ϵ and a tuple $Q = (q_1, \dots, q_s) \in \mathbb{R}^s$, the algorithm constructs a finite set $G \subset \mathbb{R}[x_1, \dots, x_n]$ of polynomials such that each $f \in G$ evaluates ϵ-approximately to Q on Ξ. To modify our example from above, we consider the tuple $\Xi = (0, 1, 2, 3, 4)$ of the points $0, \dots, 4 \in \mathbb{R}$ together with $Q = (0,\ 0.98,\ 4.01,\ 8.9,\ 16.02) \in \mathbb{R}^5$. Then, for instance, the polynomial $f = x^2 \in \mathbb{R}[x]$ vanishes 0.1-approximately to Q on Ξ.

An important feature of all algorithms we previously mentioned, and the extended ABM in particular, is that they compute in general more than just one model polynomial. Secondly, since the extended ABM proceeds degree by degree, the constructed polynomials are of lowest degree among all ϵ-approximately vanishing polynomials. As a third feature we note, that the constructed polynomials are numerically stable with respect to perturbations in the input data set $\mathbb{X}$.

4.2 Modeling

For evaluating this mathematical approach in the context of photo-voltaic power supply, we build our investigation on real data obtained from a power grid operated by a German energy provider, Stadtwerke Passau GmbH (SWP). SWP is a local energy provider in Bavaria, Germany. More than 50% of energy generated by SWP's power plants is based on solar resources. This is why SWP seems to be a good choice for evaluating our approach. The size of the grid is 7 square kilometers large. This grid connects around 800 small scale photo-voltaic plants, which have a capacity of about 23 MWp. Most of these photo-voltaic power plants can not be measured directly: only 8 power plants are directly accessible by SWP.

Based on measurement results, we observed huge impacts of local weather effects on the amount of provided solar energy. However, no direct linear correlations between individual photo-voltaic power plants can be defined, due to the nature of photo-voltaic power generation. Since characteristics of power plants differ significantly, their relations are not expressible by linear models. These characteristics, e.g., efficiency, age, size, etc., have a high impact on the power provisioning, but there might also be other factors (e.g., local conditions) that are not obvious. Nonetheless, we assume that there might be some "hidden" non-linear correlations between power plants' generation patterns. This assumption might be realistic, since weather conditions for these plants are expected to be similar and sun irradiance does not differ significantly between their geographical locations. But instead of introducing a complex model that covers all of these parameters, we propose a methodology to derive such correlations based on mathematical analysis of power provisioning data. The analysis is purely based on power provisioning patterns obtained from photo-voltaic power plants of SWP and does not depend on any other, additional data. This means that no assumptions on weather conditions, irradiation, geographical positions and so on are being made for the evaluation. For our analysis, we use the extended ABM algorithm which detects dependencies that are approximately polynomial. If there are any correlations between power provisioning data obtained from power plants, these are detected by the algorithm.

Using the extended ABM algorithm, we aim to answer the following questions:

1. Is it, in principle, possible to model some of the photo-voltaic plants in terms of the others?

2. How many different PVs do we need to express the others?

3. How good are the approximations, especially with respect to the number of omitted PVs?

4. Can we observe local effects? Can we explain why some models are in particular good, by looking at the closeness of some PVs?

For this purpose, we observed local weather conditions and traced power supply patterns of all eight directly measured photo-voltaic power plants of SWP for three consecutive days: The first two days are characterized by a highly fluctuating solar irradiation, the third day had a continuously clear sky except early

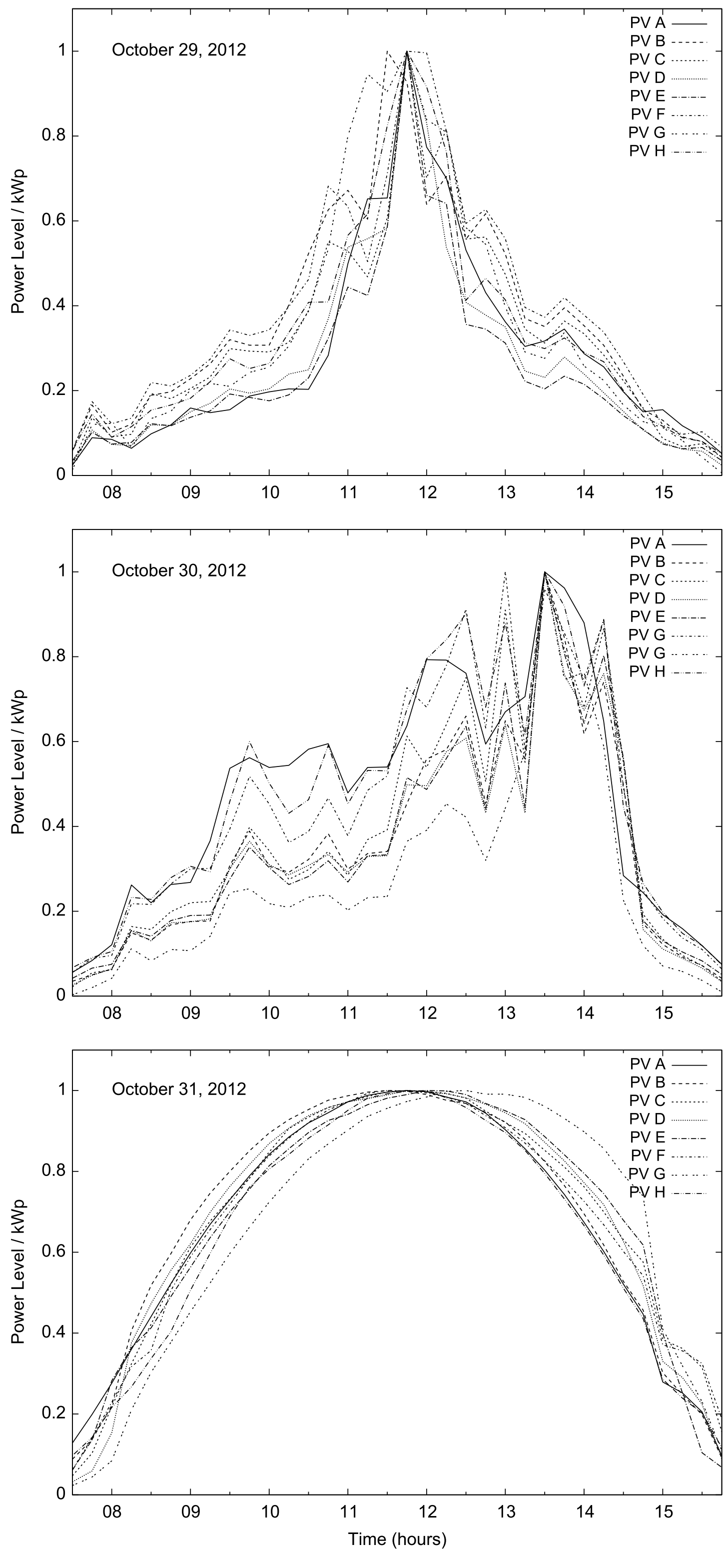


Fig. 2: Normalized measured data of eight PV power plants, taken on three consecutive days from 29 October, 2012 to 31 October, 2012.

morning and evening hours, see Figure 2. To build models using the extended ABM, we proceed as follows.

1. Data selection:
   Divide the data into two sets, a training set and a validation set. In our case, we took the data of one day for training and the other two days for evaluation.

2. Target selection:
   For both the training and validation sets, divide the set of PV plants into two groups, a group of *source data* that will provide the input data $\mathbb{X}$ and a group of *target data* each of which serves as the right-hand side Q in the input of the extended ABM.

3. Preprocessing:
   Typical preprocessing steps consist, e.g., of removing invalid data, filtering and normalization. For the present three days, we removed the values for those times, where the PV plants did not produce any energy and normalized the remaining data series.

4. Model building:
   Use the training data as input for the extended ABM and obtain a set of polynomials G, the set of model candidates.

5. Model selection:
   Using the validation source data from the two selection steps above, evaluate each polynomial and compute its residual error with respect to the validation target. Select the polynomial with the least residual error.

4.3 Results

We applied the method described in the previous section to the data provided by eight directly measured PV systems between 29 October, 2012 and 31 October, 2012. For further reference, we name these eight plants by the letters A to H. The data is measured in intervals of 15 minutes. Since we considere only the time between 7:30 and 15:45, omitting early morning and evening hours where the stations did not produce energy, we receive for each day and each plant 34 data points, measuring the current power production.

Due to the structural similarity of the 8 data series of 31 October, which was a clear day, we did not use that data for training but instead made two different runs: Run 1 using the data from 29 October and Run 2 using the data from 30 October as training set. Table 3 gives an overview of Run 2. We denote by m the number of PV plants that we use as target data, i.e., that we try to model in terms of the remaining $8 - m$ PV plants. There are $s_m := \binom{8}{m}$ possible ways to select those PV plants and in each run we built models for all s_m possible *selections*. Thereby, m ranged from 1 to 7, since we need at least one source data set. For $m = 6$ and $m = 7$ we did not get any reasonable model, the results are therefore omitted.

Table 3: Error statistics for Run 2, standard deviation in parenthesis.

m	**Avg. # models**	**Avg. best error**	**Avg. mean error**	**Best error**
1	56.0 (20.8)	1.7 (3.7)	4.4 (8.0)	0.21
2	47.6 (15.0)	2.8 (6.0)	11.7 (32.3)	0.89
3	36.6 (10.6)	4.5 (10.7)	30.3 (140.4)	1.54
4	25.1 (6.2)	12.5 (59.9)	105.4 (901.0)	2.28
5	15.8 (3.3)	161.8 (1456.6)	1354.7 (13767.4)	4.77

Table 4: Target PV plants for which best residual errors were obtained.

m	1	2	3	4
Run 1	C	E, F	D, E, F	C, D, E, H
Run 2	C	D, F	B, E, F	B, C, D, F

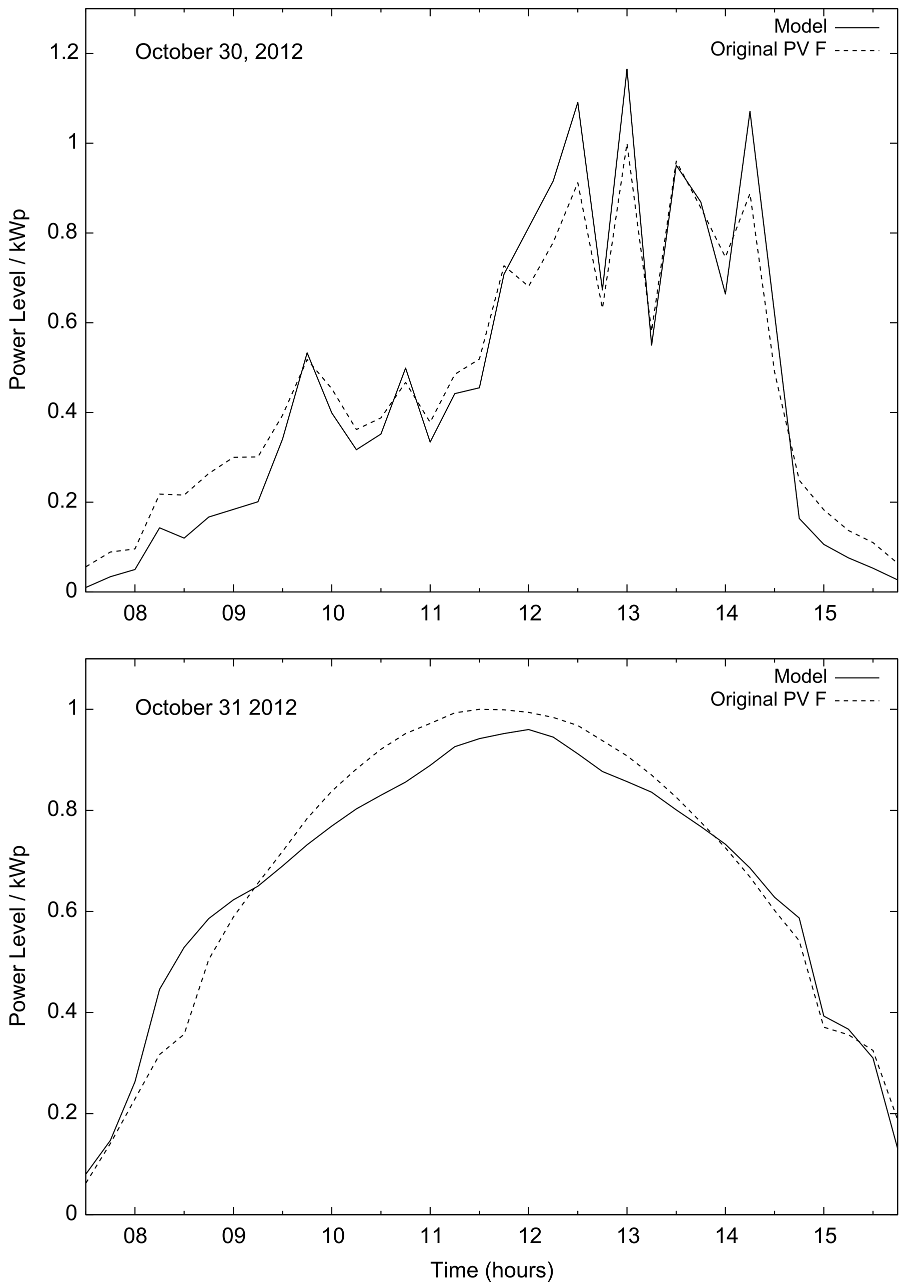
Fixing m and one of the s_m selections, we get three numbers that we use to characterize the models of the given selection: (1) the number of models returned by the AVI, (2) the least (i.e., best) residual error of these models and (3) the mean residual error of these models. Averaging those three values over all s_m selections gives the results of the second, third and fourth column of Table 3. The respective standard deviations are given in parenthesis. The last column shows the overall best error among all selections.

To make the computations comparable, we used a value of $\epsilon = 0.1$ throughout all computations. It is important to note that the value of ϵ must not be confused with the residual errors discussed above. The value of ϵ guarantees bounds within the training data while the residual error is a measure to compare validation data. It is also not surprising that the number of models decreases with increasing m. The reason is that the polynomials are constructed within $\mathbb{R}[x_{i_1}, \ldots, x_{i_{8-m}}]$ (with $i_k \in \{A, \ldots, H\}$) and therefore the number of terms per degree also decreases.

Table 4 shows the selection of target PV stations for the best model of Run 1 and Run 2, respectively. The PV stations B, C, E, F and H are all located within a radius of 1 km. Together with D they lie within a radius of 2 km. Further away are plants A and G, a radius of 3.5 km is required to surround all 8 PV plants. The results meet out expectation that close-by plants can more easily be modeled by each other.

The question how good the constructed models are depends on the demands of a concrete application and their particular requirements. We consider the models for $m = 1, \ldots, 4$ as good both with respect to the overall residual error as well as with respect to the coverage of the dynamics of the system. Figure 5 shows that the two quite distinct dynamics of 30 October and 31 October are covered equally well.

All computations were executed on a 2.26 GHz Intel Core 2 Duo laptop using the computer algebra system ApCoCoA [11]. On average, a call of the extended ABM on a 34×4 input matrix took 0.021 seconds.



$$f_F = -1.04x_Ax_B + 0.62x_C^2 + 0.40x_G^2 - 1.2x_Bx_G - 1.38x_Cx_H - 0.81x_Gx_H + 3.31x_H^2$$
$$+ 0.37x_A + 1.43x_B + 0.81x_C - 0.06x_G - 1.52x_H + 0.02 \in \mathbb{R}[x_A, x_B, x_C, x_G, x_H]$$

Fig. 5: Best model for PV plant F in Run 1, $m = 3$ and its evaluation for the validation sets.

4.4 Application Scenarios

We consider two application scenarios for the modeling approach described in the last section: Snapshot provisioning and simulation.

1. *Snapshot provisioning*
   A snapshot of the amount of power available in the grid or smaller sub-areas can help power providers or large consumers such as data centers to adapt feed or consumption parameters, thereby improving load balancing or optimizing economical goals. In this scenario, a small number of PV plants is directly measured and this data is available on-line. The data of PV plants that are not measured on-line is derived by their corresponding model polynomials. To compute these polynomials, an initial calibration phase is necessary: Over a certain period of time, data for every PV station in question is collected. From this data, the model polynomials are constructed.

2. *Simulation of power flow*
   As a straight forward variation of the first scenario, model polynomials once constructed can be used to simulate certain aspects of the grid: Varying the data used as input, one explores the behavior of the grid under different circumstances. Power flow analysis is a substantial tool for grid operators to ensure grid stability. By taking into account these interdependencies, simulation results are expected to become more accurate.

5 Conclusion and Future Work

We introduced a methodology to build models of the production of photo-voltaic power plants. In contrast to other approaches, these models also aim to be suited for small-scale areas in a small scale power grid, in addition to larger-scale scenarios. Furthermore, no external data like sun radiation or justification of the PV panels are required. The models can be used by energy providers to determine available power generated by photo-voltaic power plants (state estimation) that are not directly connected to the communication infrastructure. This *snapshot* of the grid's status can be used for monitoring purposes.

The resulting models met our expectation with respect in two directions:

- According to evaluation results, the approach seems to be quite promising with respect to geographically close-by PV plants.
- The selection of the model plants is sufficiently independent on the choice of training data.

We are confident that there is still great potential to further enhance the approach. Future work is dedicated to enhance the model by considering geographical locations of power plants. Since our evaluation results imply that modeling results for close-by PV plants lead to more precise results, this additional, domain-specific knowledge could lead to accuracy improvements. Furthermore, we think that our methodology leads to improvements in accuracy of power forecast methods. Current models do not analyze hidden interdependencies of PV plants and do not consider them for forecasting available power. This will be subject of future work.

Acknowledgments This work has received support by the EU projects All4Green (FP7 STREP, grant no. 288674) and EINS (FP7 NoE, grant no. 288021).

References

1. K. Thomaschki, *Positionpapier zur verbesserten Prognose und Bilanzierung von Solarstromeinspeisung*, 2011.
2. Y. M. Saint-Drenan, S. Bofinger, B. Ernst, T. Landgraf, and K. Rohrig, "Regional nowcasting of the solar power production with pv-plant measurements and sattelite images," in *Proc. ISES Solar World Congress*, 2011.
3. K. Mares, "Demand response and open automated demand response opportunities for data centers," 2010.
4. B. Kennedy, *Power quality primer*. McGraw-Hill Professional, 2000.
5. H. T. C. Pedro and C. F. M. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, 2012.
6. V. D. E. e.V., "Umsetzung der analytischen lastprofilverfahren – step-by-step," 2000.
7. J. Limbeck, *Computation of Approximate Border Bases and Applications*. PhD thesis, University of Passau, 2013. to appear.
8. D. Held, M. Kreuzer, S. Pokutta, and H. Poulisse, "Approximate computation of zero-dimensional polynomial ideals," *Journal of Symbolic Computation*, vol. 44, no. 11, pp. 1566–1591, 2009.
9. M. Kreuzer and H. Poulisse, "Subideal border bases," *Mathematics of Computation*, vol. 80, no. 274, p. 1135, 2010.
10. H. M. Möller and B. Buchberger, "The construction of multivariate polynomials with preassigned zeros," in *EUROCAM* (J. Calmet, ed.), vol. 144 of *Lecture Notes in Computer Science*, pp. 24–31, Springer, 1982.
11. "ApCoCoA: Applied Computations in Commutative Algebra, see, `http://www.apcocoa.org/`."