Bottom Up Approach for the Prediction of the Deployment of Distributed Energy Resources

Dr. Oliver POHL
Schleswig-Holstein Netz AG
Germany

Dr. Torsten SOWA
Schleswig-Holstein Netz AG
Germany

Maximilian ROSE
Schleswig-Holstein Netz AG
Germany

Dr. Tobias Markus PLETZER
Schleswig-Holstein Netz AG
Germany

Dr. Xiaohu TAO
Schleswig-Holstein Netz AG
Germany

ABSTRACT

This paper presents a novel approach for the prediction of the deployment of distributed energy resources (DERs), e.g. photovoltaic (PV), in the distribution grid. A Markov chain model is defined to predict the spatio-temporal evolution of DERs. The predictor is trained with extensive data sets of PV deployment as well as geographic grid data. The prediction performance is validated on real data and found to be superior to standard Monte Carlo prediction. In a case study it is shown how DER prediction can be used for grid planning and quantification of self-sufficiency of regional energy communities.

INTRODUCTION

The increasing electrification of heat and mobility sectors together with rapidly spreading renewable energies enable new, multimodal operation concepts. Prominent examples are local energy communities [1, 2, 3]. Such communities are self-sufficient using heat pumps, storage and distributed energy resources (DERs). Knowledge of the evolution of these DERs enables the planning and design of those communities. In particular, a high spatio-temporal resolution, neighboring effects, and existing distribution grids need to be taken into account.

In this study, a bottom-up approach based on predictive analytical algorithms is proposed to analyze the spatio-temporal growth of DERs using the example of residential PV. It is shown that this approach performs significantly better than standard Monte Carlo based methods. The latter are commonly used to disaggregate DERs on distribution grids by making use of independent, identical distributions [4, 5].

Based on a sample prediction, a power flow calculation is performed and implications for the grid planning are discussed. Finally, the potential for self-sufficiency of an energy community is investigated.

BOTTOM UP PREDICTION METHOD

In this section, the temporal evolution of PV activation $a_i(t) \in \{0,1\}$ for all households is defined by means of a Markov chain. Here, “1” stands for at least one existing PV in household i, “0” for none. Defining a as the vector of all house activations, the Markov chain is fully defined via its transition probability $p[a(t+1)|a(t)]$. However, solving for these probabilities is numerically intractable. The state space of the Markov chain is $\{0,1\}^{|a|}$, where the number of households is $|a| \approx 500.000$. Hence, for the matrix p 2^{10^6}entries would have to be calculated. In the following, the problem is reduced. The state space is confined to relevant input features and machine learning techniques are applied.

In the context of diffusion of innovation major impact factors for the spread of novel technologies have been studied [6, 7]. Relevant features are:

- (i) Neighbor behavior
- (ii) Type (farms vs. private households [8])
- (iii) Regional aspects (existence of innovators etc.).

This empirical knowledge is used by considering adequate interactions within the model.

To abstract from a specific geographic structure, households are mapped as nodes on a graph. Based on the geographic distance of the households, the k-nearest neighbors of each node are defined [9]. Then, the feature vector x_i contains the neighbor activation $a_j(t)$ up to the 7th neighbor including its own state $a_i(t)$ (0^{th} neighbor). Second, the type as well as the quantity of the grid connections per household enter the feature vector $x_i(t)$. Third, a regional technology factor $k_{reg}(t) = |N_i|^{-1} \sum_{i \in N_i} a_i(t)$ is defined, where N_i comprises all households of the town of household i.

The output is defined to be the difference of PV activation of two subsequent time points:

$$C_i(t+1) \coloneqq a_i(t+1) - a_i(t)$$

With this definition one obtains:

$$\begin{aligned} p\big(a(t+1)\big|a(t)\big) &\approx \textstyle\prod_{i=1}^{N} p\big(a_i(t+1)\big|a(t)\big) \\ &\approx \textstyle\prod_{i=1}^{N} p\big(a_i(t) + C_i(t+1)\big|x_i(t)\big) \\ &= \textstyle\prod_{i=1}^{N} p\big(C_i(t+1)\big|x_i(t)\big), \end{aligned}$$

where the first approximation accounts for independent updating of PV between households, while the second reflects information reduction when using x_i instead of a. Finally, the resulting conditional probabilities can be approximated with the aid of machine learning techniques.

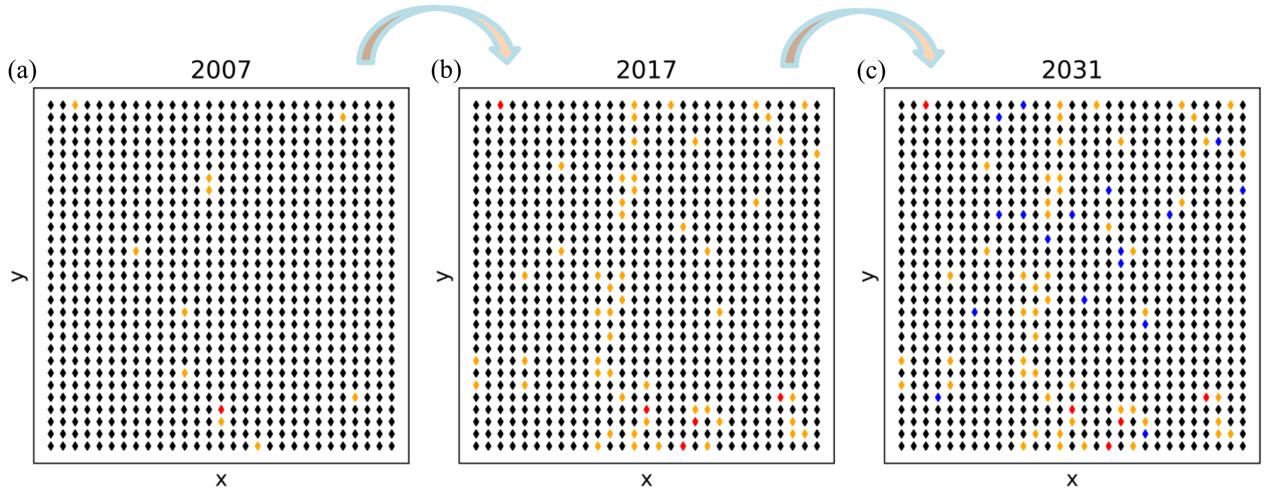

Figure 1: Evolving PV diffusion. PV households marked in red and orange for real data in the years 2007 (a) and 2017 (b), as well as a prediction for the year 2031 (c).

Historic data of PV employment from 2005-2017 is used as training data for a neural network classifier [10]. One detail of the solver is crucial: cross likelihood as loss function is used in order to ensure probabilities, which reproduce realistic relative frequencies. Additionally, a regression neural network [10] is trained in order to predict installed power P_i depending on the features x_i.

By iteratively applying the transition probabilities, sample predictions of future PV diffusion scenarios can be drawn. In the following, we present examples and address the quality of such predictions.

RESULTS

A town with large PV diffusion is selected to depict the data as well as prediction samples. Households with PV (orange) and without a PV system (black) are mapped on a quadratic grid for different years. Households, which added a PV panel at different times, are marked in red. Figures (1a) and (1b) depict the real PV distribution in the years 2005 and 2017. First, one notes a rapid increase in total quantity of PV, which is owed to a massive PV boom within the years 2008 and 2012. Furthermore, spatial correlations are clearly visible, which a posteriori justifies the nearest neighbor approach. Figure (1c) depicts a sample prediction for the year 2031, where predicted PV is marked in blue. Again, PV build-up emerges frequently around existing spots.

Next, the sensitivities of some features are investigated. In Figure 2 the predicted probability of a household to install PV (activation) is plotted for different input features x. The integer values k along the y-axis refer to a PV activation of the k^{th}- nearest neighbor. Thereby, the 0^{th}- nearest neighbor is the household itself. The results indicate that it is most probable to install PV when PV is already existent on the same house. Moreover, the probability gradually decays with increasing k. On the x-axis the regional PV activation factor is varied. A strong positive correlation of k_{reg} and p becomes obvious.

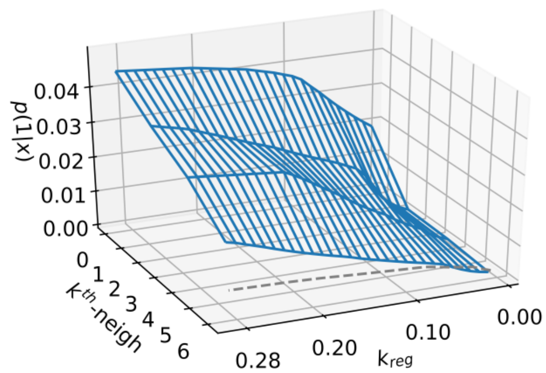

Figure 2: Probability for a PV build-up while varying the regional PV factor k_{reg} as well as PV activation of the k^{th}-nearest neighbor of a given household.

The dotted line below refers to a base curve where no neighbor PV is activated. Although not clearly visible, it also increases notably with increasing k_{reg}.

After getting an impression of the prediction and their dependencies, the quality of prediction is illuminated by means of cross validation tests. First of all, the validation set is defined to be PV diffusion within the years 2012-2014 and the training set is restricted to the years before. Furthermore, communities of 70 up to 130 houses are defined, which resemble typical low voltage grids in Schleswig Holstein. Then, the installed PV power in these communities during the validation time period is recorded. It is compared to the expected PV power deployment of a random predictor based on an independent, identical distribution and to the previously defined data based predictor.

In Figure 3 the aforementioned predictions are plotted for 30 exemplary communities. The data predictor performs better than the random predictor. In particular, it captures some of the very active communities, i.e. potential energy communities.

To quantify this finding the correlation between predictions and real data is determined via the Pearson coefficient. Using sufficient test communities one obtains:

$$\rho(real, data\ based) = 0.41 \pm 0.05$$
$$\rho(real, random) = 0.06 \pm 0.02$$

In conclusion, it has been shown that data exhibits significant correlations of PV build-up. The reasons are collective behavior on neighbor as well as on regional scale. Adequate consideration of these features improves prediction precision significantly; in particular, potential energy communities with large PV deployment can be identified.

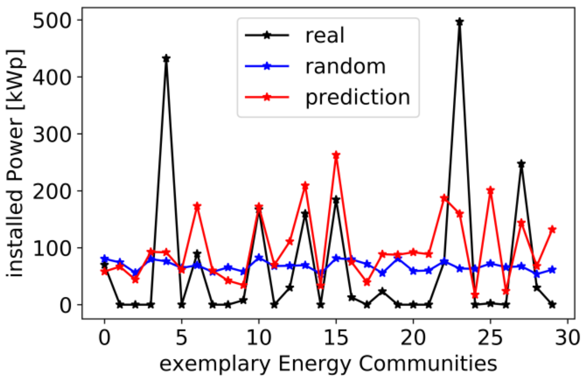

Figure 3: Comparison of validation data and prediction. The installed power (black) within the years 2012 and 2014 is compared to a random (blue) and the data based prediction (red).

CASE STUDY OF A POTENTIAL ENERGY COMMUNITY

In this section, prediction results are used to address two crucial aspects of energy communities: their distribution grid planning as well as potential for autarchy.

Prediction based distribution grid planning

This case study investigates the impact of PV prediction on an exemplary low voltage grid with 100 households. In Figure 4 the distribution grid is depicted. Predicted PV households are marked in blue. In total, the installed power constitutes more than 300 kWp, where some farms contribute notably. Based on this grid model, a power flow is calculated assuming a generation case with minimal load per household and generators producing peak power. Without the predicted PV inflow, the workload of each cable is below 50% of its maximum capacity and node voltages are close to their nominal voltage values. Adding the predicted PV power leads to a different scenario (see Figure 4): There are three cables with critical or even highly critical workload as well as several households with voltage values out of limits. Hence, in order to enable the transformation to a local energy community, the distribution grid provider would have to strengthen this particular low voltage grid.

Note that the collective nature of PV diffusion is crucial here: a spatially correlated PV activation in the bottom right stub cable string leads to voltage and workload violations while other strings operate within their technical limits.

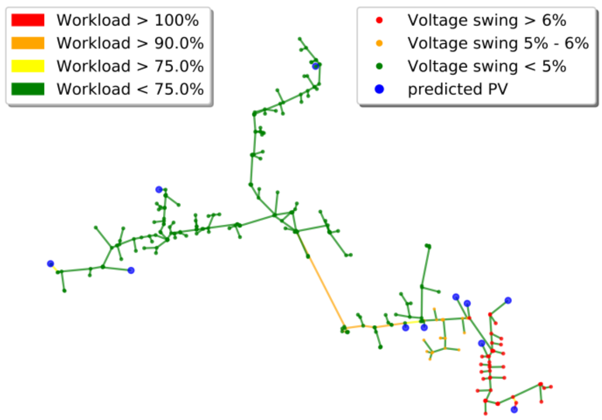

Figure 4: Power flow in a local energy community. Voltages and cable workloads become critical in hot spots, where correlated PV build-up is predicted.

Self Sufficiency of Energy Communities

In this section the potential of self-sufficient energy communities are assessed. Self-sufficient energy communities can be characterized by distributed energy generation and a storage potential. These communities aim for independence of the overlaying grid.

In Figure 5 a load profile measured at the transformer between low and middle voltage grid is depicted (blue) as well as the total solar inflow during a typical summer day in Schleswig Holstein. The solar inflow apparently exceeds load during midday so that the battery storage can optimize self-sufficiency. A mixed-integer linear optimization model is used to analyze the impact of additional storage capacity on the degree of self-sufficiency. Its objective is the optimization of the own consumption considering a given load profile.

As can be seen in the inset of Figure 5 the degree of self-sufficiency without storage is 42%. By integrating a battery, it can be increased by more than 30% up to nearly 75%. Since total energy inflow is limited, the curve eventually saturates at a total battery capacity of 650 kWh.

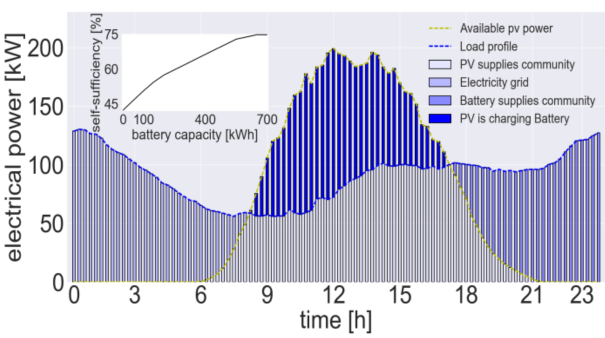

Figure 5: Self Sufficiency of potential energy community. Load and PV power inflow is plotted on a typical summer day in Germany. Inset: Self-sufficiency increases when making use of battery storage.

CONCLUSION AND OUTLOOK

In conclusion, a novel predictor of technology deployment has been presented. The proposed model was applied to residential PV power systems. It captures key aspects of technology diffusion, in particular collective interactions and, accordingly, correlated distributions. These properties have been shown to be crucial in subsequent case studies. In particular, grid planning becomes challenging when PV hot spots emerge. Furthermore, a large degree of self-sufficiency can be reached where these hot spots occur.

The model itself could be extended in various directions. To strengthen knowledge of regional aspects, socioeconomic data sets can be integrated. Furthermore, house specific data such as roof top area and direction would increase the accuracy of prediction. From a theoretical point of view, machine learning might be refined. In this paper a two-step prediction is applied: a classifier followed by a neural regressor predict probabilities of installed PV power values. This way, one profits from the natural probabilistic nature of the neural classifier. However, one might directly predict distributions of installed power, which requires a posteriori Bayesian inference within neural networks [11]. Also, the modelling of temporal evolution of PV diffusion could be elaborated in future models. The training of a Markov chain might be replaced by, for example Gaussian processes [12], in order to learn simultaneously from the entire time series.

Finally, it is noted that the prediction, as presented, emerges “bottom up”. It starts from the last known state iterating up to the time point of interest. However, it can also be applied “top down” for the spatio-temporal disaggregation of DERs within future scenarios. Such scenarios are often generated for political decision making. In this case, one applies transition probabilities until a certain target quantity is reached (i.e. 20% PV) and extracts spatial distributions as before.

ACKNOWLEDGEMENTS

The authors acknowledge funding by the Kopernikus project ENSURE with project number 03SFK1D0. We thank Dr. Joachim Kabs, the boarding member of Schleswig-Holstein Netz AG, for his great support.

REFERENCES

[1] T. Falke, S. Krengel, A.-K. Meinerzhaben, and A. Schnettler, 2016 "Multi-objective optimization and simulation model for the design of distributed energy systems", *Applied energy,* vol. 184, 1508 - 1516.

[2] G. Walker and P. Devine-Wright, 2008, "Community renewable energy: What should it mean?", *Energy policy*, vol. 36.2, 497-500.

[3] G. Doci, E. Vasileiadoua, and A. Petersena, 2015, "Exploring the transition potential of renewable energy communities", *Futures,* vol. 66, 85-95.

[4] S. Simoes, M. Zeyringer, D. Mayr, T. Huld, W. Nijs, and J. Schmidt, 2017, "Impact of different levels of geographical disaggregation of wind and PV electricity generation in large energy system models: A case study for Austria", *Renewable energy,* vol. 105, 183-198.

[5] A. Navarro, L. F. Ochoa, and D. Randles, 2013, "Monte-Carlo-based assessment of PV impacts on real UK low voltage networks", *IEEE,* 1-5.

[6] S. Linder, 2013 "*Räumliche Diffusion von Photovoltaik-Anlagen*", PhD-Thesis, Würzburg, Germany.

[7] E. M. Rogers, 2010 "*Diffusion of innovations*." Simon and Schuster, New York, USA

[8] A. Gonzalez Quintairos, J. Bühler, B. Kleinschmit, and M. Resch, 2015, "Analysis of Potential Distribution and Size of Photovoltaic Systems on Rural Rooftops", *GI_Forum,* 220-224.

[9] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, and A. Procaccini, 2008, "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study*", PNAS,* vol. 105, 1232 - 1237.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, 2011, "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research,*12. Oct., 2825-2830.

[11] D. C. Knill and A. Pouget, 2004, "The Bayesian brain: the role of uncertainty in neural coding and computation", *TRENDS in Neurosciences*, vol. 27.12, 712-719.

[12] C. E. Rasmussen, 2004, "*Gaussian processes in machine learning*", Springer, Heidelberg, Germany.