COMPARING AND IMPROVING RESIDENTIAL DEMAND FORECAST BY DISAGGREGATION OF LOAD AND PV GENERATION

Simon KLOIBHOFER
AIT – Austria
simon.kloibhofer@ait.ac.at

Matthias STIFTER
AIT - Austria
matthias.stifter@ait.ac.at

Fabian LEIMGRUBER
AIT-Austria
fabian.leimgruber@ait.ac.at

Bharath-Varsh RAO
AIT-Austria
Bharath-Varsh.Rao@ait.ac.at

ABSTRACT

In general, measured demand profiles include load and existing photovoltaic generation, which reduces the net demand. Disaggregating load from generation makes it possible to estimate the net load. Furthermore it allows to improve forecast intervals due applying constraints, if both load and PV models are available. This paper introduces a disaggregation method and discuss forecast interval performance improvement.

INTRODUCTION

Motivation

The evolution of residential consumer behavior toward producing energy by e.g. roof-top photovoltaic systems (PV) is changing the typical characteristics of loads. So called prosumers are more weather/irradiation dependent as non-producing consumers, which affects load forecasting for operational applications and planning.
Available data from smart meter systems enable detailed analysis of the impact of the growth of the share of distributed generation. Additionally monitoring at substation level allows improved accuracy of forecasting and load model prediction.

Problem

Photovoltaics for (short term) forecasting is a well-known domain and many approaches exists [1] and have been investigated on various levels of aggregation [2]. Despite the achieved results of forecasting accuracy, a major drawback is the lack of knowledge about the net loading of the grid (load without reduction by PV infeed). Additionally the uncertainty of the forecast is distributed on both the demand and the generation part.

In case of separate measurements of demand and generation profiles, specific forecasting methods can be applied, which take the individual underlying causes of both profiles, behavioural aspects or weather and irradiation, into account. Though most of the time, only measured profiles of the total sum aggregated demand and PV – will be available.

Contributions

Contributions are twofold. Firstly, the authors introduce briefly a methodology to separate or disaggregate PV generation from sum profiles of aggregated demand and PV. It will be assumed for this analysis that demand and PV are separated, either by this method or separate acquisition. Secondly, based on profile separation it is shown, that forecasting of disaggregated demand and PV will decrease forecast interval and error, which is the main result of this paper.

DATA

Data is taken from anonymized smart meter measurements from 40 households with installed PV systems in Köstendorf, Austria, over a period of 12 month in 5 min time intervals. The aggregated household load is typically monitored at e.g., the transformer station or a feeder head. PV systems are metered separately, which makes it possible to determine also the generation profile. Fig. 1 shows box plots of variations of load and PV for each 5 minute interval of all households for one year. It can be clearly seen that the variance during daytime (PV infeed) increases, due to changes in global irradiation, clearness or clouds, temperature and other weather conditions.

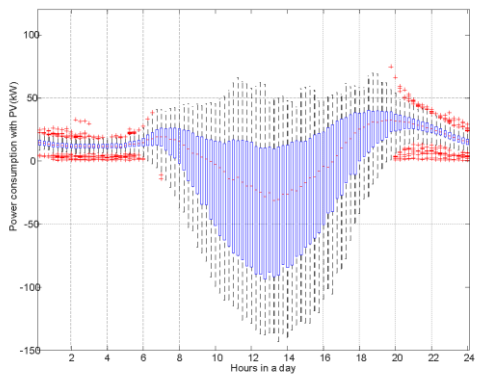

Fig. 1: Characteristics of prosumers illustrated as boxplots (40 households for one year). Variance increases due to PV.

METHODOLOGY

Disaggregation of load and PV from one measurement profile is briefly presented. In a next step, individual forecasts for a 3 hours horizon based on ARIMA for demand, PV and the aggregated sum profile are then generated. Forecast intervals are compared in order to demonstrate the decrease in forecast uncertainty.

Load and PV Disaggregation

Recent work on PV and load disaggregation use highly nonlinear machine learning based methods and focus on per household level [3].
Our proposed method for disaggregation considers statistical information of maximum and minimum infeed and compares them with actual measurements. It exploits properties of the stationarity and low dynamic of daily load averages, based on consumer behaviours, and the highly volatile irradiation based generation, which reduces net load. First results are promising and the method is currently further developed for publication.
Results of the disaggregation of PV characteristics from aggregated measured profiles of 40 houses with PV is shown in Fig. 3. The resulting PV profile is between the black lines superimposed of the errors of the method.

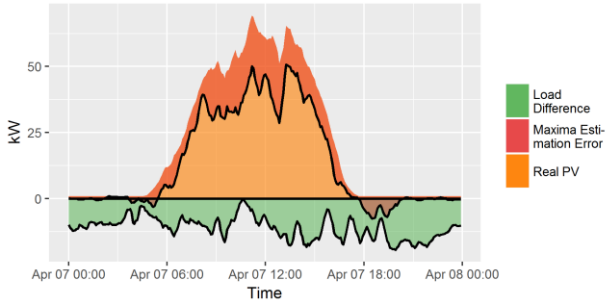

Fig. 2: Load and PV disaggregation for one day, exploiting statistical behaviour of minimum and maximum infeed. Results of PV disaggregation is shown in yellow.

Validation of the method by means of correlation with irradiation is shown as monthly averages in Fig. 3.

Depending on the accuracy of existing information on PV generation (e.g. installed rated powers, direct online measurements from inverter systems, historical irradiation and weather data, or proposed disaggregation method) the accuracy of separate forecasts will change and also impact evaluations results.

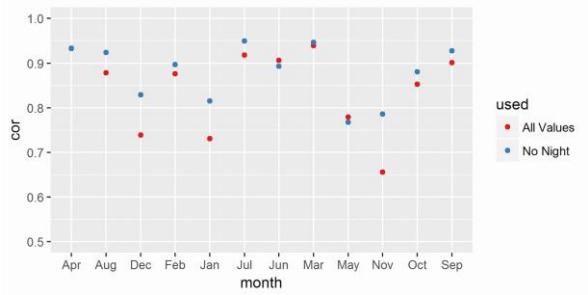

Fig. 3: Correlation of disaggregation method with irradiation data, monthly averaged. Night times reduced for comparison.

For purpose of showing the benefit of forecasting load and PV disaggregated, correct separation - or direct measurements of these two profiles, as available in our data set, is assumed here for further analysis. This can be seen as the error free upper bound or maximum gain of forecast improvement by disaggregation, investigated within the chosen method in this paper.

Forecast Methods

Forecasting is done in a three step process, equally for forecasting PV, load separately and the total (sum) load. The first step removes the daily patterns, considered as seasonality. Then the remainders are forecasted by an ARIMA model. If applicable, constraints are then imposed on the forecasts, e.g., demand cannot go beyond zero. The forecast is trained every month with a historical data set of three months. This is decided to have still some season information, while having enough data to model all different regimes of change. The load and total forecasts are furthermore divided into one model for weekdays and another one for weekend (Saturday and Sunday).

In a first step the daily patterns apparent in all three forecasts have to be modeled. Seasonal trend decomposition by Loess (STL) is used. With this method data is separated into additive parts, a nearly static daily cycle and a remainder, showing the difference to the daily cycle. Only the remainder is modelled by ARIMA [4].

The advantage to more specialized ways of handling seasonality opposed to other methods like one based on clear sky index for PV is, that different daily patterns can be modelled. This is import when using the same forecast method also for load.

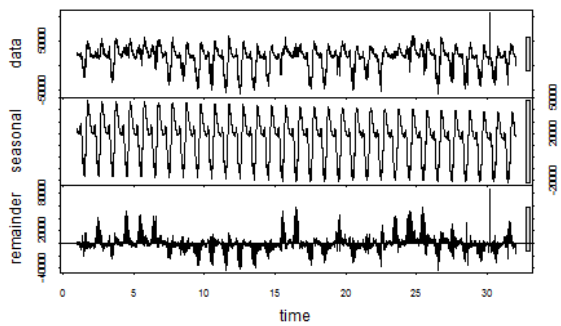

Fig. 4: The aggregated sum measurement (top) and its two additive components obtained from a robust STL decomposition with flexible trend superimposed on the fixed seasonality.

In the second step an ARIMA model is trained on the remainder of the seasonal decomposition. ARIMA models are working good for forecasting stationary time series, especially for short forecast horizons [4].
Training of the ARIMA model is 30 days and is performed once per month. This is due to train the model for differences in the dynamics of the days, e..g., sunny day or cloudy day. Also the yearly seasonal trend is determined as a superimposed linear trend.
The method `auto.arima` of the GNU R package `forecast` tests different ARIMA model orders iteratively. A Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) for stationarity is used to decide if an integration step is done. Then models up to the order of (p, q) = (4, 4) (can and should still be raised) are fitted with Maximum Likelihood (ML) optimization and the best model is chosen by means of a corrected Akaike's Information Criterion (AIC) [5].

The selected ARIMA model is used in the third step to calculate forecasts every 3 hours for a period of 3 hours. Separation of working days and weekend days for the load and sum profiles are considered as different sets of forecasted results.

Impact on Variance and Forecast Intervals

Variance of the random processes

The variance of the sum of two random variables is given by the sum of the individual variances plus their covariance, expressed by:

$$\mathrm{var}\{P_{sum}\} = \mathrm{var}\{P_{PV}\} + \mathrm{var}\{P_{load}\} + 2 * \mathrm{cov}\{P_{PV}, P_{load}\}$$

(If both random processes are independent and therefore uncorrelated, the variance of the sum is the sum of the variances).

Since the load and the PV profile are negatively correlated (covariance is negative) the variance of the sum P_{sum} decreases. From this we would expect that disaggregation would worsen the forecast intervals, since the variances of the separated are together higher than the sum profile.

Forecast Intervals (Density Forecast)

It should not be confused between confidence interval and prediction interval. The forecast (or prediction) is associated with the random process which is needs to be observed and with a given probability that the forecasted variable will be lying within the (prediction) interval. The confidence interval is according to the classical (or frequentist) statistics non-random, unknown and calculated from data with a probability that a certain interval contains the true value (e.g., 95% of repeated experiments will result in true parameter contained within the interval)

If the forecast error distribution is normal we can use the point forecast and error variance to determine the prediction interval. Conditional forecast error and variance are defined by:

$$e_N(h) = X_{N+h} - \hat{x}_N(h)$$

If unbiased the Prediction Mean Square Error (PMSE) is equal the prediction error variance:

$$E\{e_N(h)^2\} = \mathrm{var}\{e_N(h)\}$$

Forecast intervals are calculated based on the variance of the residuals, and for an e.g., 95% interval for all ARIMA models [4]

$$\hat{x}_{T|T+1} + 1,96\hat{\sigma}$$

Load, PV and sum Forecast Intervals

Next we try to estimate the forecast interval performance from analysis of the STL model remainders of the three different time series. Note that ARIMA forecast intervals are based on assumptions like normal distribution of the remainders and no correlation among the residuals. Showing the histogram and density of remainders for month March. The remainder of the load shows a much better behaviour in terms of variance and improvement of the forecast interval can be achieved.

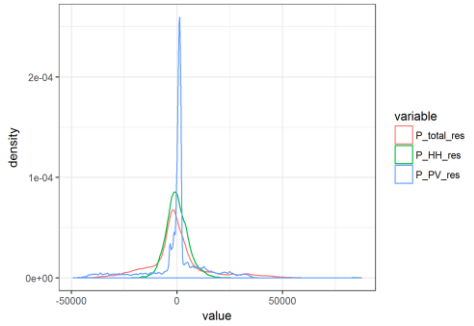

Fig. 5: Distribution of the three STL model remainder on which the ARIMA models are trained. Household (HH) shows a variance. Not much improvement on forecast intervals will be achieved by disaggregation and sum alone.

According to the above prediction error description the variances of the STL model remainders for different months are given in Fig. 6.

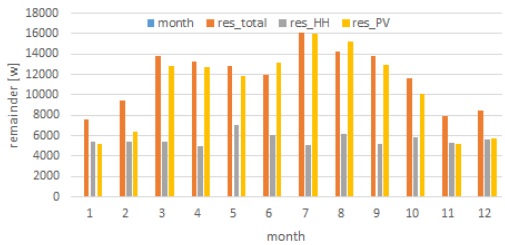

Fig.6: Deviation of the three STL model remainder distribution on which the ARIMA models are trained. Household (HH) forecast interval are improved by disaggregation.

Forecast interval improvement by constraints

Until now the reduced variance of the sum and the distribution of the residuals haven't let us much room for improvement. However by separation of demand and generation we can introduce constraints on them, induced by natural or physical facts. Constraints on the forecast interval applied are:

- No negative demand or negative generation.
- During night-time generation of PV is zero
- Maximum generation is derived from a clear sky model scaled by the maximum power.

This finally leads to potential improvement in the forecast interval, which can now be further analysed.

Comparison metrics

Results for the mean forecasts are compared to real values using Mean Absolute Error (MAE). Forecasts at each 3 hour interval with 3 hours horizon are calculated and compared with real values for the whole data set of one year.

RESULTS

Forecasts are shown for one day horizon with consecutively executed three hours forecast intervals, and the last three days as input data for the model. Evaluation of forecast errors are for one exemplary day.

Fig. 6 shows the forecast of the aggregated measurement which is typically available for monitoring. Fig. 7 shows forecast of load only without PV.

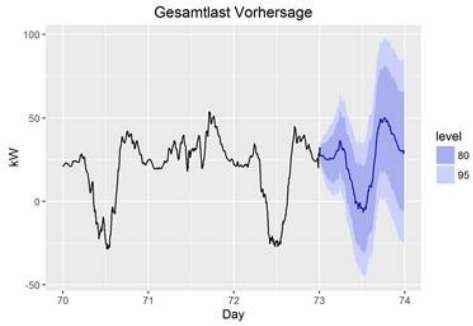

Fig. 6: Forecast of sum profile (load and PV together) with confidence intervals

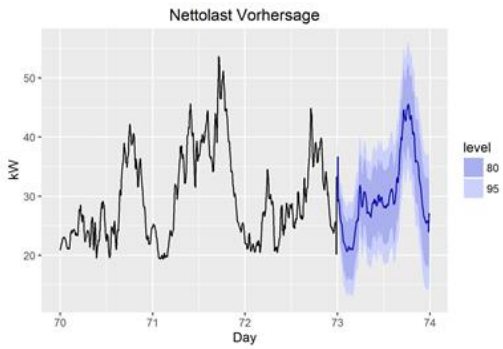

Fig. 7: Forecast of load (without PV) with confidence intervals

Fig. 8 shows forecast of PV generation. Confidence intervals can be bound to non-negative during daytime and zero during night-time (no-infeed).

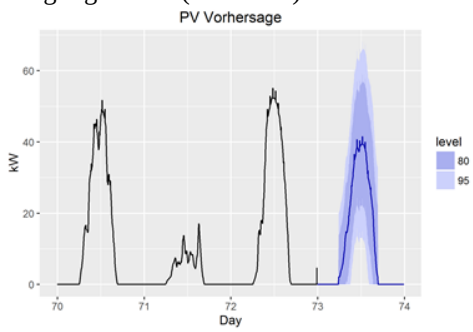

Fig. 8: Forecast of PV (disaggregated from sum profile) with confidence intervals

Forecast Comparison Evaluation

Errors with respect to the 80% the forecast interval of the aggregated total profile and the sum of the disaggregated load and PV are compared for each 3 hour forecast interval for the data set of one year in Fig. 9.

It can be seen that there is a clear advantage in nights where variances of PV can be minimized due to the fact of no infeed. During day-time no short-term improvements are on average over a year are imminent.

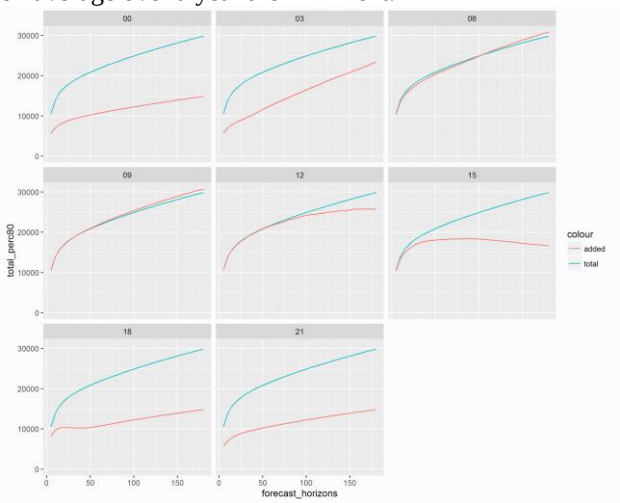

Fig. 9: MAE errors for each 3 hours forecast interval of aggregated and sum of disaggregated forecasts for one year. During daytime PV has the highest forecast variance and almost no improvement is achieved in the very short term.

Fig. 10 shows the MAE of the forecast interval between the aggregated and sum of the disaggregated profiles for the three hour forecast horizon of the one year data set.

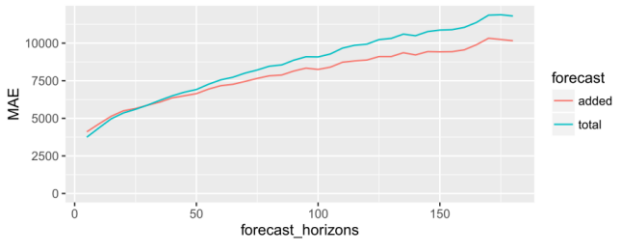

Fig. 10: Summary of MAE errors for the three hour forecast horizon for the full evaluation period of one year.

CONCLUSION

The forecast of demand and PVs are separately computed and the forecast intervals are compared between the total and the sum of the profile. Especially the forecast intervals of the disaggregated approach for the load (household) are meant to be significant smaller as compared to the sum profile.

The improvement in decreasing forecast intervals is twofold. Aggregated household do not have the same dynamic in changes and the STL forecast model have lower variances of the residuals. Due to natural, or physical constraints we can bound PV variances to certain limits (e.g. night time, max. rated power, global irradiation). Without disaggregation, constraints could not be applied.

It was not focus of this paper to improve forecast methods and the chosen forecast model takes only the time series and type of day as input variable.

Further benefits can be expected from separate available load and PV profiles by means of disaggregation. Examples are: knowledge of net load, potential voltage rise due to generation, or reactive power contribution in case of active controls.

OUTLOOK

Other data sets will be investigated which will reflect typical smart meter measurement intervals of 15 minutes. Possible improvement in the domain of PV forecasting through external parameters, like irradiation forecasts will be further explored. The method of load PV disaggregation will be developed and published with data sets from aggregated low voltage and medium voltage networks.

ACKNOWLEDGMENTS

The authors would like to thank Salzburg Netz for providing data of Smart Region Köstendorf.

REFERENCES

[1] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, “Review of photovoltaic power forecasting,” *Sol. Energy*, vol. 136, pp. 78–111, Oktober 2016.

[2] R. Sevlian and R. Rajagopal, “Short Term Electricity Load Forecasting on Varying Levels of Aggregation,” *ArXiv14040058 Stat*, Mar. 2014.

[3] R. Mohan, T. Cheng, and A. Gupta, “Solar Energy Disaggregation using Whole-house Consumption Signals,”, NILM Workshop, 2014.

[4] Hyndman, R.J., Athanasopoulos, G., 2013. Forecasting: principles and practice. OTexts. https://www.otexts.org/fpp

[5] R. J. Hyndman and S. Fan, “Density Forecasting for long-Term Peak Electricity Demand,” IEEE Transactions on Power Systems, May 2010.