Computationally efficient identification of databased models applied to a milk cooling system

Malin Lachmann[a1], Francesca Jung[a] and Christof Büskens[a]

[a]Center for Industrial Mathematics, University of Bremen, Germany

Received on February 04, 2019 / accepted on *****, 2019

Abstract

Along with climate change, the use of renewable energy becomes more important. Farms have a high energy demand as well as space for the installation of renewable energy plants and hence there is a high potential for reducing the use of fossil energy sources by using self-produced renewable energy when available. To best use that energy when it is available, smart energy management systems can reschedule tasks with high energy demand and can charge or discharge storages. For such a system, models describing the behavior for all devices of a farm are required.

We designed a software module using a black box approach to identify general databased models at low computational cost. The software can be used to calculate forecasts for arbitrary generator and storage devices. Using real world data we apply this to model the temperature of a milk cooling system and as a first step we generate additional inputs to improve the model where a next step would be generating these inputs automatically. With these additional inputs, the temperature of that system can be forecasted well meeting all time constraints during the model identification on low cost hardware.

Keywords: databased modeling; least-squares regression; forecasting; renewable energy; computational mathematics & statistics.

1. Introduction

As climate change is progressing, actions need to be taken to prevent irreversible damage for the planet. A shift from conventional to renewable energy sources is a possible action. A downside is that energy from renewable sources is not always available when needed since it often depends on the weather conditions. To tackle that, energy management systems can play an important role by shifting loads that are flexible in their execution time to a time with a surplus of energy. Another option is to install storage

[1]E-mail Corresponding Author: mlachmann@uni-bremen.de

devices and to charge these when there is a surplus of renewable energy and to discharge them when more energy is needed than generated.

The project *SmartFarm* aims to maximize the use of self-produced energy on farms at a small scope. Farms are very eligible for that since they have a high energy demand on the one hand, but also much space to install renewable energy plants on the other hand. Besides, many farms in Germany already have photovoltaic plants installed and sell the energy they do not consume to the network which is not economic for more recent plants due to the grid parity. The project *SmartFarm* consists of two major components: The first, a measurement and control system, measures power data from all generator and certain consumer devices as well as temperatures of thermal storage devices and the states of charge of battery storages. It is also able to implement control signals that arise from the second part, that is a software module to calculate such control signals. It aims to shift loads or charge storages such that the use of self-produced energy is maximal and all constraints, e.g. time constraints or physical laws, are obeyed [1].

To find an operation schedule for a farm, it is essential to know the future energy generation and consumption as well as the behavior of storage devices. The power values, temperature values and states of charge measured need to be modeled in order to calculate forecasts predicting the behavior of the devices for the energy management system.

In the literature, there exist different modeling approaches. They can be classified into physical models on the one hand and databased models on the other hand [2]. Physical models, also called physics-based models, are based on physical correlations. Databased models are calculated based on data measured beforehand and they can find correlations that were not known before. In [3], databased models are further divided into models basing on statistical methods such as regressions and artificial intelligence techniques such as neural networks where identifying models is much faster with an appropriate regression. Both approaches are frequently applied to the forecast of energy generation. In [4], a neural network is applied to estimate the power generated by a wind turbine and in [5] an adaptive neuro-fuzzy inference system is used to forecast power generation of wind plants where both approaches require much computation time to identify models. They are also applied to model state of charges of a battery, for instance in [6]. In [7] and [8], forecasts are made based on regression methods which is much faster.

Therefore we use a databased regression approach to calculate models for the generators and storages. Hence the data measured within the measurement and control system is used and no assumptions on the models need to be made. This has the advantage that models for all kinds of devices can be computed which is very helpful since on farms a variety of devices is installed. Additionally, no expert knowledge is required to transfer the system to other devices. Another benefit is that regression approaches require only small computation times which suits the *SmartFarm* software well since it should be executed on micro computers to ensure low cost for the energy management system.

In [9, 10], the same approach is used to model a photovoltaic plant, a wind plant and a battery storage device with an extension to probabilistic models. In contrast to that, we will model devices with a more complex dynamical behavior and show as a first step how we can improve these by adding additional inputs applying it to the temperature on a milk cooling tank. This requires some expert knowledge and hence, a next step would be to generate these additional inputs automatically.

At first, in Section 2, the least squares regression used for modeling in our approach is briefly introduced. Section 3 deals with the application of the method to model the temperature of a milk tank using real world data. Section 4 summarizes the findings.

2. Least squares regression for databased modeling

In this section we sketch how models are identified using a least squares regression as proposed by [11].

Given measured output data $y_i \in \mathbb{R}, i \in \{1, \dots, n\}$ at n different points in time and given corresponding input data $x_i \in \mathbb{R}^m, i \in \{1, \dots, n\}$, at the same points in time, we want to find a model $f : \mathbb{R}^m \to \mathbb{R}$ that fits this data best. As a result of Taylor's Theorem, it is possible to approximate a function f around a point $x_0 \in \mathbb{R}^m$ with a polynomial given that f is regular enough. The coefficients of this polynomial are the parameters we want to identify in the course of this paper using the input and output data.

To describe a polynomial function of degree d in more than one dimension, i.e. when $m > 1$, a notation with multi-indices is required. Let $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}^m$ be a multi-index whose degree $|\alpha|$ is the sum of its components, that is $|\alpha| = \sum_{j=1}^m \alpha_j$. Now, for a vector $x_i = (x_{i,1}, \dots, x_{i,m}) \in \mathbb{R}^m$, where $x_{i,j}$ is the j-th component of the vector x_i, we define $x_i^\alpha := \prod_{j=1}^m (x_{i,j})^{\alpha_j}$. Let $p \in \mathbb{R}^\ell$ be a parameter vector of size $\ell = \binom{m+d}{d}$. A polynomial function f of degree d can now be written as

$$f(x_i, p) = \sum_{\substack{\alpha \in \mathbb{N}^m \\ |\alpha| \leq d}} p_\alpha x_i^\alpha.$$

The parameter vector p, i.e. the vector of coefficients of the polynomial f, is now chosen to best fit the data measured. We assume a normally distributed error of the data. Hence we determine the parameters $p = (p_1, \ldots, p_\ell)$ such that the mean-square deviation between the model $F(X,p) = (f(x_1,p), \ldots, f(x_n,p))^T$ for $X = (x_1, \ldots, x_n)) \in \mathbb{R}^{m\times n}$ with $x_i \in \mathbb{R}^m$ and the measured output data $y = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$ is minimal. In other words, p is the optimal solution of $\min_{p\in\mathbb{R}^\ell} \|F(X,p) - y\|_2$. This optimization problem can be solved very efficiently using the fact that the function F is linear in p. This allows to rewrite the function $F(X,p)$ as the product of a matrix $A(X) \in \mathbb{R}^{n\times\ell}$ and the vector p. This gives following optimization problem:

$$\min_{p\in\mathbb{R}^\ell} \|F(X,p) - y\|_2 = \min_{p\in\mathbb{R}^\ell} \|A(X) \cdot p - y\|_2.$$

To solve this, we use a QR-decomposition method to decompose the matrix $A(X)$, which can be calculated directly from the input data, into an orthogonal matrix Q and an upper triangular matrix R. Then the minimization problem can be rewritten and solved by back substitution at low computational cost. For details on that or on how to decompose the matrix $A(X)$ efficiently, refer to [12]. All in all, the least squares regression allows to identify the coefficients of a polynomial model at low computational cost.

3. Experimental results

In this section we apply the least squares regression to real world data, in particular we model the temperature of a milk tank, a thermal storage device.

3.1 Test setting

The data on hand was measured at a milk tank on a farm in Lower Saxony, Germany between February 18, 2018 and April 10, 2018 and then interpolated to minutely data with a moving average filter to reduce the noise in the measurements. It consists of the temperature inside the milk tank and the measurements of active power consumed by the milk tank. Both are available in a minutely resolution, but for modeling we decide to use a resolution of 30 minutes. Within one minute the temperature often does not change enough to be visible within the data. Hence the actual behavior, that means a slow change of temperature, cannot be learned by a model in that case but in data with a resolution of 30 minutes, the changes become clearly visible.

The 52 days of data are divided into data used for training the model, i.e. the data used to identify the coefficients, and testing data used to evaluate a model's quality. The length of the training horizon is 26 days. On both the training horizon and the test horizon, the rooted mean square deviation (RMSE) is determined and then normalized to the biggest absolute value measured during training and testing horizon (nRMSE).

In addition to those remarks on the data, little preprocessing has been conducted before the model identification by removing some obvious errors in the measurements. Up to eight temperature or power values per day, from a total of 1440 values, have not been plausible since their orders of magnitude were much higher than for all other values. These outliers are replaced by the values measured one minute before them since data does not often change from one minute to another.

3.2 Modeling dynamic behavior by iterative forecast computation

The temperature inside the milk tank follows a very regular pattern and can also be modeled by using its periodicity. However, the temperature forecast then would not react to the control given by the energy management system when it aims to shift the power consumed by the milk tank to a different time. Hence such a model for the temperature of the milk tank cannot be used to reschedule power consumption within the framework of the energy management system. Therefore, we use the dynamical behavior of the temperature at the milk tank, i.e. the fact that the temperature at a certain time depends on the temperature one step before that time and the active power as well as potential additional input values.

Then, the temperature value one step before is usually not available when computing a forecast since forecasts for up to 24 hours are required. Hence, after identifying a model also based on the temperature one step before we iteratively compute the forecast for each 30 minutes at a time to obtain a 24-hour forecast. The first value within the forecast horizon, is computed using the latest measured value of the temperature and for all later values in the forecast, the temperature value forecasted one step before is used.

3.3 Numerical results

At first, we model the temperature of the milk tank in a resolution of 30 minutes based on data that is available in the actual application without using expert knowledge. The first input x_1 is the active power consumed by the device between the last time step and the one to be forecasted. It is forecasted using another databased method within the project *SmartFarm* and thus would be available when computing the forecast in an energy management system. However, this is not available yet and thus the model is trained and tested on the active power measurements. The second input x_2 is the temperature one time step before, i.e. 30 minutes before the value to be forecasted. An additional influence to the behavior of the milk tank temperature arises from the fact that within an energy management system the cooling process is controlled, meaning that it is switched on and off at different times than usual. To find an optimal schedule, potential control values are given as an input to calculating forecasts. Hence these are also available and used as a third input x_3 to the model, where $x_3 = 1$ if the cooling is set to be active and $x_3 = 0$ if the cooling is turned off.

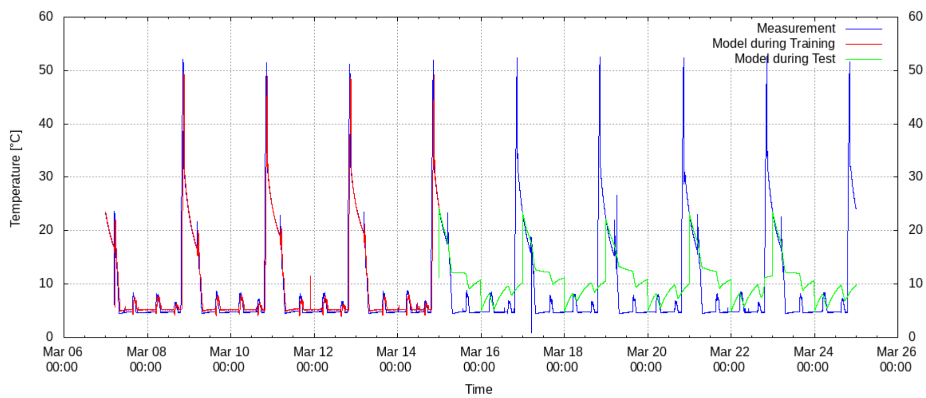

Figure 1 - Model of the temperature inside the milk tank based on input data x_1, x_2, and x_3 during an excerpt of the training horizon (red) and the testing horizon (green) plotted against the actual measurement (blue).

Coefficients of a polynomial model of degree one are identified using the inputs x_1, x_2 and x_3 taking only a few seconds. In the testing horizon, forecasts for 24 hours are calculated at midnight of each day. Computing forecasts at other times than midnight has only little influence on the results and is hence not regarded here. In Figure 1, the model during training (red) and testing (green) is compared with the actual measurements where the forecast is calculated iteratively in the testing horizon. The error value (nRMSE) during the training is good being 4.48 %, but during the testing horizon it is much higher being 15.0 %. This arises since a small error for a value at the beginning of the forecast horizon influences all later values and hence the error grows bigger due to the iterative forecast computation. In general, the error values are okay, since in [13], the state of charge of lithium-ion batteries, a simpler storage device, is estimated with an error of less than 5 %. Nevertheless, the model behavior does not match the actual behavior of the temperature, for instance peaks in the forecast occur at different times than peaks in the measurement. Hence, the model could not

Variable	Description
x_1	Active power
x_2	Temperature 30 minutes before current value
x_3	$x_3 = 1$, if milk is actively cooled, else $x_3 = 0$
x_4	$x_4 = 1$, if no milk is in the tank, else $x_4 = 0$
x_5	$x_5 = 1$, if milk from one milking is in the tank, else $x_5 = 0$
x_6	$x_6 = 1$, if milk from two milkings is in the tank, else $x_6 = 0$
x_7	$x_7 = 1$, if milk from three milkings is in the tank, else $x_7 = 0$
x_8	$x_8 = 1$, if milk from four milkings is in the tank, else $x_8 = 0$

Table 1 - Description of all inputs used for modeling.

be used in an energy management system. Also for polynomial degrees of two or three or lower data resolution, the model does not yield a better behavior. A possible reason is that important information is missing, for instance warm milk is filled into the tank twice a day influencing the milk cooling's behavior. However, this information is not available as measured data.

Since the model for the temperature of the milk tank does not describe the actual behavior well and influences to the temperature that are not measured are missing in the model, we now evaluate the influence of additional inputs for the model arising from expert knowledge about the times of the milking processes and the cleaning of the milk tank. The milking processes take place twice each day at 5:45 am and 4:30 pm local time. Every second day, all milk in the tank is picked up at approximately 8:30 pm local time. After that, a cleaning process is started where hot water is conducted through the tank. This will now be used as an input to the model and we will show how using this knowledge improves the model. As a next step the inputs should be generated automatically from the data such that the expert knowledge is no longer required.

We add different input vectors to the data (see Table 1). Five binary input vectors $x_4, \ldots, x_8$ are added. The first, $x_{4,t}$ indicates whether there is milk in the tank ($x_{4,t} = 1$) or not ($x_{4,t} = 0$) at time t. The variables $x_{5,t}$, $x_{6,t}$, $x_{7,t}$ or $x_{8,t}$ are one at time t if one, two, three or four milking processes have taken place since the last milk pickup, respectively and zero otherwise. Evaluations not shown in the course of this paper have shown that combining these five binary variables to one discrete variable does not have a positive influence on the model. Also, adding a variable indicating the cleaning process does not yield better results.

Using all inputs $x_1, \ldots, x_8$ to identify the coefficients of a polynomial model with degree one of the milk tank temperature, we obtain the model depicted in Figure 2. Now, the behavior of the model corresponds to the measurements much better. Only the temperature peaks during the cleaning process cannot be modeled well. The error during training is 4.00 % and during test it is 11.4 % and hence also better than before. All in all, using the additional input improves the model's quality noticeably.

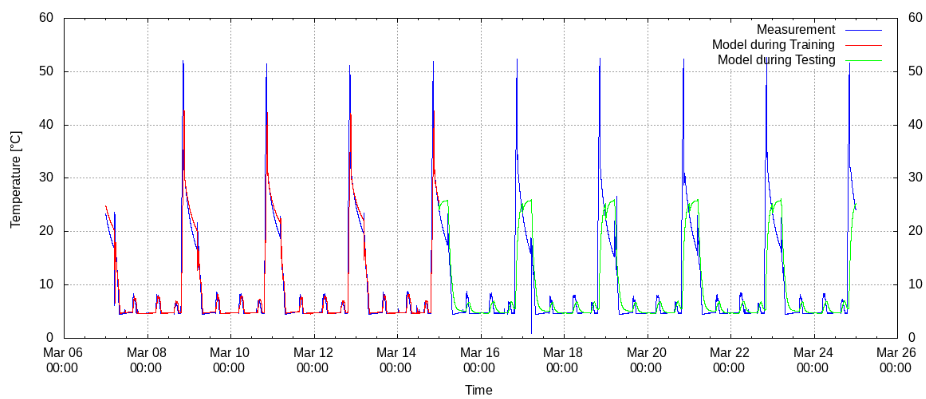

Figure 2 - Model of the temperature inside the milk tank based on input data $x_1, \ldots, x_8$ during an excerpt of the training horizon (red) and the testing horizon (green) plotted against the actual measurement (blue).

4. Conclusion

In this paper, we introduce a least squares regression to identify models for forecasting the behavior of devices on a farm. These very general polynomial models are obtained at low computational cost. We identify models for a milk tank's temperature based on different real world input data. A model based only on active power, the temperature 30 minutes before and a control given by an energy management system does not fit the data well. If we use additional inputs that are generated from external knowledge, a polynomial model of degree one describes the data well.

Further work will combine the modeling with generating the additional inputs for the models automatically and only from data, for instance by recognizing the different cooling behaviors at different fill levels. Then, the transfer to other devices on farms does not depend on expert knowledge such as milking times and the generality of the approach is no longer affected. Also testing the models for the temperature of the milk tank in a live system where actual controls might affect the behavior is interesting. In that case it could be helpful to consider the adaptation of the models to new data. For an increased robustness, applying the Levenberg-Marquardt method instead of the QR-decomposition could be an interesting approach.

Acknowledgments. This research has been funded by the Federal Ministry for Economic Affairs and Energy of Germany in the project SmartFarm (Project number 0325927). Malin Lachmann acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) within the GRK 2224/1 "Pi^3: Parameter Identification - Analysis, Algorithms, Applications".

References

[1] Heins, W., Büskens, C. Two-Level Forecast-Based Energy and Load Management for Grid-Connected Local Systems Using General Load and Storage Models, 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe): 1-6, IEEE, 2018.
doi:10.1109/EEEIC.2018.8494390

[2] Foley, A. M., Leahy, P. G., Marvuglia, A., McKeogh, E. J. Current methods and advances in forecasting of wind power generation, Renewable Energy 37(1): 1-8, 2012.
doi:10.1016/j.renene.2011.05.033

[3] Pedro, H. T., Inman, R. H., Coimbra, C. F. Mathematical methods for optimized solar forecasting, in: Renewable Energy Forecasting: 111-152, 2017.
doi:10.1016/B978-0-08-100504-0.00004-4

[4] Li, S., Wunsch, D. C., O'Hair, E. A., Giesselmann, M. G. Using neural networks to estimate wind turbine power generation, IEEE Transactions on energy conversion, 16(3): 276-282, 2001.
doi:10.1109/60.937208

[5] Potter, C. W., Negnevitsky, M. Very short-term wind forecasting for Tasmanian power generation, IEEE Transactions on Power Systems, 21(2): 965-972, 2006.
doi:10.1109/TPWRS.2006.873421

[6] Chen, Z., Qiu, S., Masrur, M. A., Murphey, Y. L. Battery state of charge estimation based on a combined model of Extended Kalman Filter and neural networks, in Neural Networks (IJCNN), The 2011 International Joint Conference on (pp. 2156-2163), IEEE, 2011. doi:10.1109/IJCNN.2011.6033495

[7] Tsekouras, G. J., Dialynas, E. N., Hatziargyriou, N. D., Kavatza, S. A non-linear multivariable regression model for midterm energy forecasting of power systems, Electric Power Systems Research, 77(12): 1560-1568, 2007. doi:10.1016/j.epsr.2006.11.003

[8] Tsekouras, G. J., Dialynas, E. N., Hatziargyriou, N. D., Kavatza, S. A non-linear multivariable regression model for midterm energy forecasting of power systems, Electric Power Systems Research, 77(12): 1560-1568, 2007. doi:10.1016/j.epsr.2006.11.003

[9] Jung, F., Büskens, C. Probabilistic Data-Based Models for a Reliable Energy Management, 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe): 1-6, IEEE, 2018. doi:10.1109/EEEIC.2018.8494364

[10] Jung, F. Entwicklung robuster Prognosen fur ein Energiemanagementsystem anhand datenbasierter Modellierungsverfahren unter Berücksichtigung von Unsicherheiten, Dissertationsschrift, Universität Bremen, 2018.

[11] Blume, B. Effiziente Modellgenerierung und optimale Steuerung technischer Systeme am Beispiel von Komponenten von Biogasanlagen, 2012, Mensch & Buch Verlag, Berlin, Germany.

[12] Quarteroni, A., Sacco, R., Saleri, F. Numerical mathematics (Vol. 37), 2010, Springer Science & Business Media.

[13] Eichi, H.R., Chow, M. Adaptive parameter identification and state-of-charge estimation of lithium-ion batteries, Proc. 38th Annual Conference of the IEEE Industrial Electronics Society: 40124017, Citeseer, 2012.