Impact of Model and Forecast Uncertainties on the Performance of the Model Predictive Control of a PV-Battery-Heat Pump-Heat Storage System

Ronny Gelleschus(✉) and Thilo Bocklisch

Technische Universität Dresden, Chair of Energy Storage Systems, Dresden, Germany
ronny.gelleschus@tu-dresden.de

Abstract. Recent research has shown that model predictive control (MPC) is a practical tool for the realization of an intelligent single- or multi-use energy management for both single and hybrid energy storage systems. Based on a system model and forecasts of external influences, such a controller will find the supposedly optimum decision to take in the immediate future. However, this decision will only be optimal for the given forecast and model. The inevitable model and forecast uncertainties may lead to decisions that are mathematically infeasible. Usually, underlying control loops ensure system stability and safety. However, uncertainties can be detrimental to the performance of the MPC, especially in multi-use applications, which have been shown to be preferable in practice due to a more economical usage of the storage devices.

For this study, the authors carried out various analyses on the impact of both model and forecast uncertainties on the performance of the MPC in the case of a PV-Battery-Heat Pump-Heat Storage system in a single-family house providing self-consumption optimization and grid relief. Concerning the impact of model uncertainties, the use case was simulated repeatedly, varying both structure (linear and quadratic) and parameters of the optimization model. The impact of forecast uncertainties was investigated by simulating with real and ideal forecasts and identifying "typical" forecast errors that led to deviations in the system's behaviour using statistical methods. The results show that the influence of forecast uncertainties is usually higher than that of model uncertainties, but large model uncertainties may drastically alter the MPC's usage of a hybrid energy storage system. The identification of the most influential uncertainties forms the basis for developing a more robust MPC-based energy management technique.

Keywords: energy management · model predictive control · PV battery energy storage system · heat pump · *modeling*

1 Introduction

Decarbonisation of all energy sectors depends on the wide-spread use of renewable energies. For many countries, wind and solar power are the most promising technologies. With the supply of both fluctuating on various time scales and resources being scarce,

P. Schossig et al. (Eds.): IRES 2022, AHE 16, pp. 162–192, 2023.
https://doi.org/10.2991/978-94-6463-156-2_13

the expansion of renewable energy production has to be accompanied by an expansion of energy storage capacities and their intelligent use.

One important sector of energy consumption is private households, making up about 28% of Germany's end use energy demand [1]. This paper is concerned with a single-family house in Germany with a rooftop solar photovoltaics (PV) plant, a lithium-ion battery and a heat pump and heat storage for domestic hot water (DHW) supply, forming a hybrid energy storage system (HESS) that enables the partial replacement of fossil fuels in electricity and heat consumption with locally produced renewable energy. The application of HESS allows to couple various energy sectors in a decentralized manner and adds degrees of freedom that can be used to optimize the use of energy resources for costs, efficiency, component lifetime and other objectives [2, 3].

For many different stationary energy storage systems (ESS) it has furthermore been shown that including multiple tasks into their energy management strategies maximizes the benefit that can be drawn from a given ESS in its lifetime. This is usually referred to as multi-use energy management. In [4], a large-scale battery ESS serving a commercial consumer was investigated. The study included ageing effects and shows that combining spot market trading, frequency containment reserve and load peak shaving into one multi-use energy management system (EMS) increases profitability while only slightly increasing battery ageing.

For single-family houses, it is evident that the primary goal is to ensure a highly secure and economic operation. If only PV and a battery system are involved, and the electricity price is constant over time, maximizing the use of locally produced PV power minimizes costs [5]. For systems with an air-source heat pump, a further dimension to optimizing costs is introduced by volatile heat source temperatures. Optimizing the use of the heat pump can decrease energy costs significantly [6, 7]. A further goal has been imposed by German legislation, demanding that PV systems never feed more than 60% (using certain subsidies: 50%) of their nominal power into the grid. This reduces the maximum power flowing into the distribution grid, usually with a high coincidence of several PV plants in the same grid. Several publications addressed this restriction by suggesting to incorporate some kind of feed-in peak shaving into the usual PV self-consumption optimization operation [8–10]. In [11], this was expanded to a HESS with a heat pump and heat storage device while also adding load peak shaving with the same motivation of reducing stress on the distribution grid. For an industrial consumer, load peak shaving can reduce costs due to capacity pricing of their grid connection [12]. For household consumers, however, feed-in and load peak shaving do not increase the profitability of a given energy storage device under German legislation. Nevertheless, there is an evident benefit for the distribution grid operator (deferral or avoidance of grid expansion). Furthermore, batteries are operated at lower powers and spending more time at medium states of charge (SOC), potentially increasing their lifetime. However, if legislation allows it, there are also possibilities to directly increase profitability by combining several use cases. The authors of [13] investigated adding the provision of frequency containment reserve to the operation of a residential PV-battery-heat pump-heat storage system. They found a 15% reduction of the negative annuity of the system. There is also a vast body of literature on the optimal use of ESS for cost optimization in scenarios with time-of-use and fully flexible electricity prices [14–16].

Multi-use energy management is a complex optimal control problem subject to uncertainty. Such control problems can be solved in various ways [17, 18]. A very popular method in recent research is model predictive control (MPC). In MPC, the optimization problem is modelled and solved explicitly for a finite horizon (e.g. one day) into the future. The first part of the solution (e.g. quarter of an hour) is applied. After an adaptation interval passed, the horizon is moved forward and the optimization is carried out again using any newly available information. Although some of these studies do not explicitly name their strategy accordingly, references [4, 6, 7, 10–12] use the MPC scheme and demonstrate its ability to handle complex control problems.

However, MPC in an energy management application relies on both a model of the controlled system and on forecasts of uncertain variables. Both models and forecasts can never be absolutely accurate, which means that decisions taken on the basis of such imperfect information will be imperfect as well. Several of the aforementioned studies compared the performance of their proposed methods when using real forecasts to the performance when using ideal forecasts [10, 11] or even when using different optimization models and solvers [7]. These comparisons show that forecast errors and model errors impact performance of the MPC negatively to an extent that depends on the single- or multi-use application that is investigated and model types, forecast methods and energy time series that are used.

These comparisons, however, usually remain superficial, remaining unhelpful for identifying the most pressing concerns that should be addressed first when designing improved energy management concepts. Therefore, the aim of this paper is to investigate the effects of model and forecast uncertainties on the performance of the investigated system in detail. The results contribute to the development of an EMS that is more robust against the most impactful uncertainties.

Section 2 of this paper presents the system to be controlled, the MPC scheme, the optimization objectives and formulations and the performance criteria used in this study. In Sect. 3 the investigation of the impact of model uncertainties is shown. Section 4 contains the investigation of the impact of forecast uncertainties. Finally, Sect. 5 summarizes the findings, discusses the limitations of the study and gives an outlook on future research.

2 HESS Energy Management

2.1 Controlled System

This paper studies the operation of a battery and a combined heat pump and heat storage device in a single-family house with 4 MWh annual electrical load, 5 MWh local PV production and 2.55 MWh of domestic hot water (DHW) consumption. Figure 1 shows the layout of the system. According to German legislation, no more than 50% of the 5.7 kW nominal PV power can be fed into the grid. Any surplus that remains unused within the system is curtailed. The electricity price is assumed to be 0.30 €/kWh and the feed-in tariff is 0.10 €/kWh.

2.2 EMS Objectives and MPC Scheme

While the system remained the same, due to new insights some minor details of the control concept changed between the analyses carried out for Sects. 3 and 4. In both

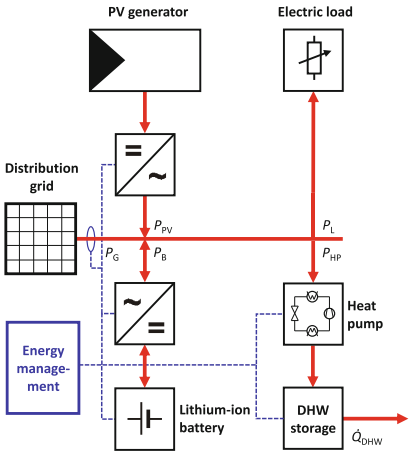

Fig. 1. System overview.

cases, the EMS is a model predictive control as shown in Fig. 2. The formulation for the optimization problem is presented in the following Subsection, with the equations describing the battery being varied as described in Subsects. 3.1 and 4.1.

The EMS for both analyses primarily aims to keep the DHW storage temperature at above 50 °C and to minimize operation cost. The optimizer directly controls the heat pump power.

For the analysis in Sect. 3, the EMS directly sets a grid power and additionally aims to minimize both the maximum power fed into the grid as well as that drawn from the grid. This results in a type of dynamic double peak shaving with load levelling features. All forecast errors and variability within an adaptation interval of the MPC are handled by the battery, as long as it is not completely charged or discharged. The adaptation interval is 30 min, and the prediction horizon is 18 h.

For the analysis in Sect. 4, the EMS only sets a maximum grid power, allowing grid power to vary between zero and that variable maximum, while additionally to the common goals aiming to minimize only the maximum power fed into the grid. This results in a dynamic feed-in peak shaving without load levelling features. Some of the forecast errors and variability within an adaptation interval of the MPC are handled by the grid. The adaptation interval is 15 min and the prediction horizon is 16 h.

2.3 Optimization Problem Formulation

The optimization part of an MPC consists of a description of the physical system and its socio-economic environment, as well as an objective function. While some minor details changed, the optimization problem formulation of both analyses presented in this paper is similar to the one in [11]. The aims of the EMS are expressed using a weighted sum F,

$$\min F = C + \alpha P_{\mathrm{G,B,max}} + \beta P_{\mathrm{G,FI,max}} + \gamma F_{\mathrm{thermal}}, \tag{1}$$

where C is the operation cost that would result from the predicted grid import und feed-in, $P_{\mathrm{G,B,max}}$ and $P_{\mathrm{G,FI,max}}$ are auxiliary variables representing the dynamic peak shaving limits, F_{thermal} is a penalty term for the heat storage temperature violating a lower limit of 50 °C and α through γ are weighting factors. For the analysis in Sect. 3, $\beta = -\alpha = 0.1$ and $\gamma = 10$. For the analysis in Sect. 4, $\alpha = 0$, whereas the rest is unchanged. The penalty term is calculated as

$$F_{\mathrm{thermal}} = \sum_{k=1}^{K} U_{\mathrm{penalty}}(k) \cdot \Delta t \tag{2}$$

with

$$U(k) + U_{\mathrm{penalty}}(k) \geq U_{\mathrm{min}} \forall k, \tag{3}$$

where $U(k)$ is the thermal energy content of the DHW storage at time step k, U_{min} is the energy content corresponding to the lower temperature limit (50 °C) and U_{penalty} is an auxiliary variable. This construct is also known as a soft constraint.

In order to represent the physical system behaviour, several energy balances are used, which translate to equality constraints. The battery energy balance is varied in Sect. 3, which is why it is described there. The DHW storage energy balance is described by

$$\frac{U(k) - U(k-1)}{\Delta t} = \epsilon_{\mathrm{COP}} P_{\mathrm{HP}}(k) - \dot{Q}_{\mathrm{dhw,f}}(k) - R_{\mathrm{th,sd}} \forall k, \tag{4}$$

where ϵ_{COP} is the average heat pump COP, P_{HP} is the electrical heat pump power, $\dot{Q}_{\mathrm{dhw,f}}$ is the forecasted energy flow due to DHW being drawn and $R_{\mathrm{th,sd}}$ is the self-discharge rate of the DHW storage. The forecasts are described in Sect. 4, because they are still assumed to be perfect in Sect. 3. The electrical power balance is described by

$$P_{\mathrm{PV,f}}(k) - P_{\mathrm{L,f}}(k) = P_{\mathrm{B,c}}(k) - P_{\mathrm{B,d}}(k) + P_{\mathrm{HP}}(k)$$

$$+ P_{\mathrm{PV,ct}}(k) + P_{\mathrm{G,FI}}(k) - P_{\mathrm{G,B}}(k) \forall k, \tag{5}$$

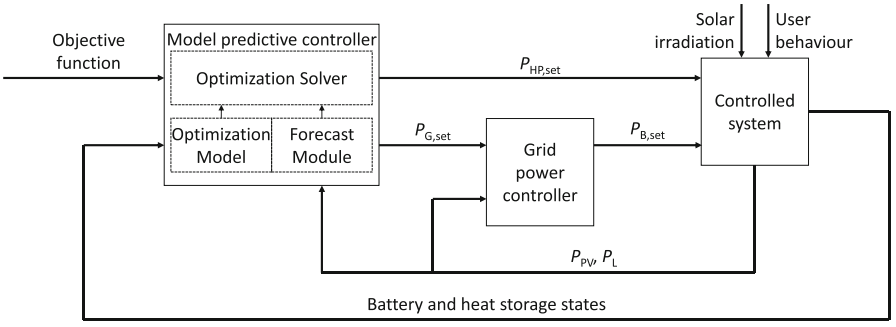

Fig. 2. Structure of the control system with the MPC as the outer control loop and the grid power controller as the inner loop.

where $P_{\mathrm{PV,f}}$ and $P_{\mathrm{L,f}}$ are the forecasted PV and load powers, $P_{\mathrm{B,c}}$ and $P_{\mathrm{B,d}}$ are the battery charging and discharging powers, $P_{\mathrm{PV,ct}}$ is PV curtailment and $P_{\mathrm{G,FI}}$ and $P_{\mathrm{G,B}}$ are powers fed into and imported from the grid.

Several inequality constraints are needed in order to correctly represent the behaviour and interplay of the previously introduced variables. Inequalities (6) and (7) ensure that the dynamic peak shaving limits are respected.

$$P_{\mathrm{G,B}}(k) \leq P_{\mathrm{G,B,max}} \forall k \tag{6}$$

$$P_{\mathrm{G,FI}}(k) \leq P_{\mathrm{G,FI,max}} \forall k \tag{7}$$

Inequalities (8) through (10) ensure that the battery cannot be charged and discharged at the same time.

$$P_{\mathrm{B,C}}(k) \leq s_{\mathrm{B,C}}(k) \cdot P_{\mathrm{B,c,max}} \forall k, \tag{8}$$

$$P_{\mathrm{B,D}}(k) \leq s_{\mathrm{B,D}}(k) \cdot P_{\mathrm{B,d,max}} \forall k, \tag{9}$$

$$s_{\mathrm{B,C}}(k) + s_{\mathrm{B,D}}(k) \leq 1 \forall k, \tag{10}$$

where $s_{\mathrm{B,c}}$ and $s_{\mathrm{B,d}}$ are binary variables indicating that charging or discharging takes place in the respective time step and $P_{\mathrm{B,c,max}}$ and $P_{\mathrm{B,d,max}}$ are the maximum charging and discharging power, respectively.

Finally, inequalities (11) and (12) ensure that the heat pump has to either be operated between a minimum and a maximum power or be switched off:

$$P_{\mathrm{HP}}(k) \geq s_{\mathrm{HP}}(k) \cdot P_{\mathrm{HP,min}} \forall k, \tag{11}$$

$$P_{\mathrm{HP}}(k) \leq s_{\mathrm{HP}}(k) \cdot P_{\mathrm{HP,max}} \forall k. \tag{12}$$

where s_{HP} is another binary variable indicating the on/off state of the heat pump, $P_{\mathrm{HP,min}}$ is the minimum heat pump power when switched on and $P_{\mathrm{HP,max}}$ is the maximum heat pump power.

For the analysis in Sect. 3, the EMS applied the grid power of the first time step of the optimization result as the set point for the inner control loop. For the analysis in Sect. 4, the feed-in limit was set to the grid power of the first time step of the optimization if it was positive or to zero otherwise. If there is a deficit of PV production, the inner control loop will always try to discharge the battery before using the grid, regardless of the optimization result.

2.4 Performance Criteria

For the described system, many criteria can be used to assess the performance of the EMS. For this paper, only five criteria, which are closely related to the common goals of both EMS presented in Subsect. 2.2, are examined:

- operation cost: $k_{\mathrm{oc}} = \sum_i P_{\mathrm{G,b}}(\tau_i) p_{\mathrm{G,b}} - P_{\mathrm{G,fi}}(\tau_i) p_{\mathrm{G,fi}}$

- self-sufficiency: $k_{\mathrm{ss}} = 1 - \frac{\sum_i P_{\mathrm{G,b}}(\tau_i)}{\sum_i P_{\mathrm{L}}(\tau_i) + P_{\mathrm{HP}}(\tau_i)}$
- self-consumption: $k_{\mathrm{sc}} = 1 - \frac{\sum_i P_{\mathrm{G,fi}}(\tau_i)}{\sum_i P_{\mathrm{PV,max}}(\tau_i) - P_{\mathrm{PV,ct}}(\tau_i)}$
- PV curtailment: $k_{\mathrm{pvc}} = \frac{\sum_i P_{\mathrm{PV,ct}}(\tau_i)}{\sum_i P_{\mathrm{PV,max}}(\tau_i)}$
- fraction of DHW drawn at less than 50 °C:
- $k_{\mathrm{dhw}} = \frac{\sum_i \dot{Q}_{\mathrm{dhw}}(\tau_i) I_T(\tau_i)}{\sum_i \dot{Q}_{\mathrm{dhw}}(\tau_i)}$ with $I_T(\tau_i) = \begin{cases} 1, \textit{if } T_{\mathrm{HS}}(\tau_i) < 50\,°C \\ 0, \textit{else} \end{cases}$

3 Impact of Model Uncertainties

3.1 Models and Simulations

In order to assess the impact of model uncertainties, the system was simulated varying both the structure and the parameters of the description of the battery within the optimization model used by the MPC. As a plant model representing the controlled system, a simple quadratic fit for the energy losses of the battery was used:

$$\frac{\Delta E_{\mathrm{B}}}{\Delta t} = P_{\mathrm{B,c}} - P_{\mathrm{B,d}} - P_{\mathrm{B,l}} \tag{13}$$

$$P_{\mathrm{B,l}} = a + b_1 |P_{\mathrm{B}}| + b_2 P_{\mathrm{B}}^2 + c_1 E_{\mathrm{B}} + c_2 E_{\mathrm{B}}^2 + d |P_{\mathrm{B}}| E_{\mathrm{B}} \tag{14}$$

where E_{B} is the battery energy content, $P_{\mathrm{B,l}}$ are battery losses, a through d are regression parameters, $P_{\mathrm{B}} = P_{\mathrm{B,c}} - P_{\mathrm{B,d}}$ and $|\cdot|$ denotes the absolute value, respectively.

This description can also be used in a mixed-integer quadratically constrained program (MIQCP) resulting in a situation where the optimization model and the plant model of the battery are identical. However, it is much more common to use (mixed-integer) linear programs (MILP) within the MPC framework to control such a system. It is possible to use piecewise approximations in order to accurately represent the nonlinearities within the system, but it is also common to use a simpler linear representation of the battery:

$$\frac{\Delta E_{\mathrm{B}}}{\Delta t} = \eta_{\mathrm{c}} P_{\mathrm{B,c}} - \frac{P_{\mathrm{B,d}}}{\eta_{\mathrm{d}}} - P_{\mathrm{B,sd}}, \tag{15}$$

where η_{c} and η_{d} denote the charging and discharging efficiencies of the battery (approximated as constants) and $P_{\mathrm{B,sd}}$ is the battery's self-discharge rate.

Comparing the performance of the MPC with this simple MILP representation and the "perfect" (i.e. plant and optimization model are identical) MIQCP representation, both with ideal forecasts (i.e. forecast for prediction time step = mean of the actual powers that will occur in the simulation time steps corresponding to the prediction time step), allows to assess the impact of the uncertainty of the model structure within the optimization model.

In order to assess the impact of model parameters within the optimization model, the MIQCP structure was chosen and the six parameters of Eq. (14) are varied using three scaling parameters v_1 through v_3:

$$P_{\mathrm{B,l}} = v_1 a + v_2 b_1 |P_{\mathrm{B}}| + v_2 b_2 P_{\mathrm{B}}^2 + v_3 c_1 \mathrm{SOE}$$

$$+v_3c_2\mathrm{SOE}^2+v_2v_3d|P_\mathrm{B}|\mathrm{SOE} \tag{16}$$

These scaling parameters where set to the levels 0, 0.5, 0.9, 1.0, 1.1, 1.5, and 2.0 allowing to control the level of simulated uncertainty of the model parameters.

3.2 Results

3.2.1 Model Structure

Most performance criteria, presented in Table 1, showed very little sensitivity to changing the optimization model structure (k_oc, k_ss and k_sc) or are at a very low level (k_pvc, due to the use of ideal forecasts). The exception was the fraction of DHW drawn at less than 50 °C, changing from about 6.7% for the quadratic model to about 14.6% for the linear model.

Further analysis demonstrates that the MILP and MIQCP use the two storage devices differently. Figure 3 shows that the MIQCP charges the battery and the heat storage at the same time much more often than the MILP, while reducing the occurrences of charging the heat storage using energy stored in the battery.

Figure 4 and Fig. 5 also implicate that the MILP prefers to charge the battery first, while the heat storage temperature remains low, whereas the MIQCP charges the heat storage first before charging the battery. This obviously reduces the occurrences of DHW drawings at low heat storage temperatures.

Table 1. Performance criteria when varying optimization model structure.

	MILP	**MIQCP**
operation cost, k_oc	475.12 €	472.33 €
self-sufficiency, k_ss	53.86%	53.99%
self-consumption, k_sc	61.83%	61.69%
PV curtailment, k_pvc	$1.23 \cdot 10^{-4}$	$1.20 \cdot 10^{-4}$
fraction of DHW drawn at less than 50 °C, k_dhw	14.57%	6.71%

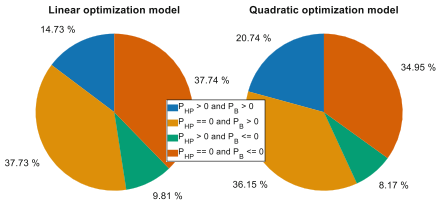

Fig. 3. Joint operation of heat pump and battery over one year. The percentages are time steps, where each combination of positive or negative power with zero or positive heat pump power occurred divided by the total number of time steps.

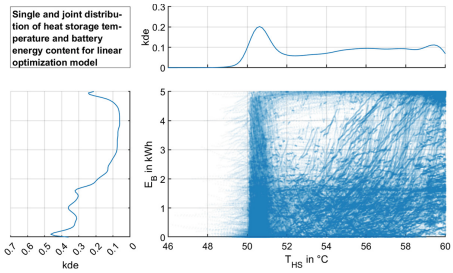

Fig. 4. Jointplot (combination of two kernel density estimate [kde] plots and a scatterplot) visualizing the single and joint distributions of the two system states for the linear optimization model.

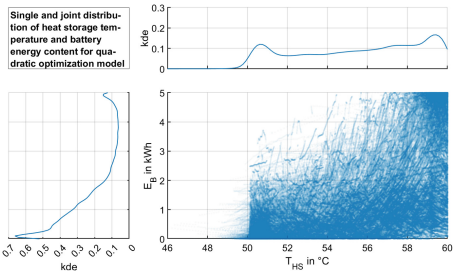

Fig. 5. Jointplot visualizing the single and joint distributions of the two system states for the quadratic optimization model.

3.2.2 Model Parameters

The overall sensitivity of the performance criteria towards the scaling parameters varied between the criteria. Figure 6 shows four of the five criteria normalized to their value at $v_1 = v_2 = v_3 = 1.0$ for "small" (all $v_i \in [0.9, 1.1]$), "moderate" (at least one v_i outside $[0.9, 1.1]$ but still $v_i \in [0.5, 1.5]$) and "large" (at least one v_i outside "moderate" range) errors. PV curtailment is not shown because the base level is so small that the relative variability would be high although the absolute values do not matter at all in practice. It is clear that, the larger the errors get, the larger the spread of the criteria becomes. Furthermore, the fraction of DHW drawn at less than 50 °C again has the largest variability.

Figure 7 shows the fraction of DHW drawn at less than 50 °C when varying both v_2 (under- or overestimating the effects of battery power on losses) on one axis and v_1 together with v_3 (primarily associated with self-discharge) on the other. The white area represents simulations that could not be carried out because the matrix of quadratic and collinear elements in the battery loss model was not positive semi-definite, which is a prerequisite for finding a solution to an MIQCP. Underestimating the influence of battery power on the losses (low v_2) and overestimating self-discharge (high v_1 and v_3) decreases this fraction. This is due to the model suggesting to the MPC that it is more attractive to store surplus PV energy in the heat storage, which leads to the heat storage being at a higher temperature more often than in the case of the "correct model". Underestimating

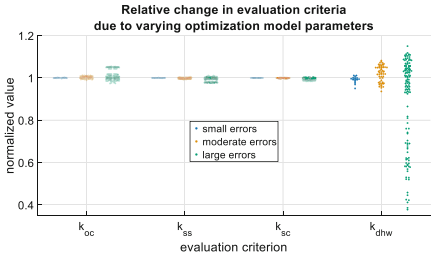

Fig. 6. Bee swarm plot of relative change of evaluation criteria, highlighting their variability, in presence of small, moderate and large model parameter errors.

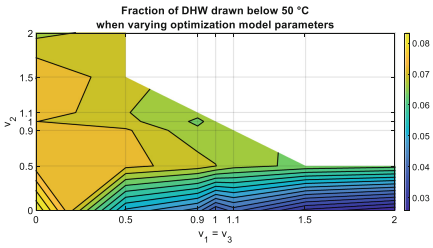

Fig. 7. Fraction of DHW drawn at less than 50 °C when varying the optimization model parameters.

self-discharge, however, increases the fraction due to the MPC preferring to store surplus energy in the battery because it believes it to have much smaller losses than the heat storage.

In contrast, self-sufficiency, shown in Fig. 8, decreases when the influence of the battery power on losses is underestimated. This behaviour seems to be almost unchanged by an under- or overestimation of battery self-discharge. Note that, confirming the interpretation of Fig. 6, the total difference between the highest and lowest self-sufficiencies obtained from these simulations was rather small in comparison to the change of the domestic hot water drawn at less than 50 °C.

It can be argued that the decrease of domestic hot water drawn at less than 50 °C is worth the trade-off in self-sufficiency. This is similar to multi-objective optimization, where many solutions can be optimal forming a pareto-front and choosing "the best"

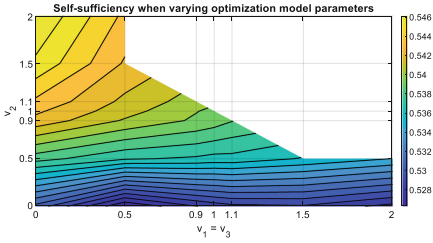

Fig. 8. Self-sufficiency when varying the optimization model parameters.

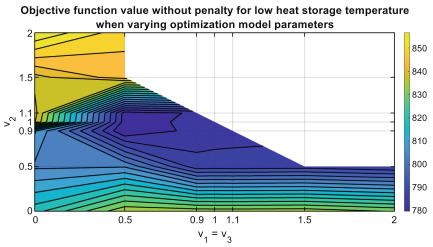

Fig. 9. Objective function without soft constraint term and scaled to one year when varying optimization model parameters.

solution is subjective or needs further well defined criteria. In this case, it makes sense to refer back to the objective function of the optimization solved by the MPC in Eq. (1). This objective function was scaled up from the prediction horizon to one year. The term penalizing DHW drawn at less than 50 °C, was neglected due to it dominating the result although it is only a soft constraint, not an actual objective.

Figure 9 shows that the scaled objective function has a plateau around the "correct" model, with the optimum being close to it. The objective function increases in all directions from there, especially for large model parameter errors. This suggests that small errors in the optimization model parameters, as they are to be expected in a real application, are uncritical, whereas large errors, i.e. inappropriate models, will result in a drastically different behavior of the system, which will cause some performance criteria to deteriorate considerably.

4 Impact of Forecast Uncertainties

4.1 Methods

In order to assess the impact of forecast uncertainties, the MILP version of the MPC in Subsect. 3.1 with the modifications to the objective function, prediction interval and horizon and the grid power control discussed in Subsect. 2.2 was simulated both with real and ideal forecasts. The real forecasts used in this study are based on [19], with minor modifications.

4.1.1 Forecasts

The PV forecasts use the equation

$$P_{\mathrm{PV,f}}(t + k\,\Delta t) = k_{\mathrm{scale}}(t) \cdot P_{\mathrm{PV,env}}(t + k\,\Delta t), \tag{17}$$

where t is the current time, $\Delta t = 0.25\mathrm{h}$ the time step width, $k \in \mathbb{N} \wedge 1 \leq k \leq 64$ the time index, k_{scale} a scaling parameter roughly representing the current cloud cover and $P_{\mathrm{PV,env}}$ an empirical envelope of the PV power under clear sky conditions. These two parameters are calculated as follows:

$$k_{\mathrm{scale}}(t) = \frac{1}{12} \cdot \sum_{k=1}^{12} \frac{P_{\mathrm{PV}}(t - k\,\Delta t)}{P_{\mathrm{PV,env}}(t - k\,\Delta t)} \tag{18}$$

$$P_{\mathrm{PV,env}}(t + k\,\Delta t) = \max_{j} P_{\mathrm{PV}}(t - j \cdot 24h + k\,\Delta t), \tag{19}$$

where $j \in \mathbb{N} \wedge 1 \leq j \leq 10$, i.e. the mean of the measured PV power divided by the corresponding envelope of the previous three hours (different from [19], where the energies are summed up first, and divided afterwards, making this version more dynamic) and the maximum measured PV power of the corresponding time of day during the last ten days. For each time step during the night, the fraction within the mean in Eq. (18) would be undefined. Therefore, it is replaced by the last k_{scale} value of the previous evening, capped to values between 0.5 and 1, with the reasoning that both higher and lower values than that are more likely to be numerical artefacts than a real weather situation that still persists the following morning.

The load forecast combines an exponential smoothing of the current load power with daily and weekly persistence:

$$P_{\mathrm{L,f}}(t + k\,\Delta t) = k_{\mathrm{exp}}(k) P_{\mathrm{L}}(t) + \left(1 - k_{\mathrm{exp}}(k)\right) \frac{P_{\mathrm{L}}(t + k\,\Delta t - 24\mathrm{h}) + P_{\mathrm{L}}(t + k\,\Delta t - 7\mathrm{d})}{2} \tag{20}$$

The exponential smoothing factor is calculated as

$$k_{\mathrm{exp}}(k) = \exp\left(\frac{1 - k}{10}\right). \tag{21}$$

DHW consumption of a single household is much more erratic than electricity consumption. An exponential smoothing of the current DHW flow rate or heat flow would not represent the patterns well. As a simple approach to get at least more or less correct forecasts for the necessary thermal energy, a persistence forecast of the measured heat flows of one day and one week ago is used:

$$\dot{Q}_{\mathrm{dhw,f}}(t + k\,\Delta t) = \frac{\dot{Q}_{\mathrm{dhw}}(t + k\,\Delta t - 24\mathrm{h}) + \dot{Q}_{\mathrm{dhw}}(t + k\,\Delta t - 7\mathrm{d})}{2} \tag{22}$$

4.1.2 Forecast Errors

Forecast errors for each of the three physical quantities are a two-dimensional matrix: at each adaptation of the MPC, i.e. every quarter of an hour for one year (8760 h), one average power value for every quarter hour interval over the next 16 h is forecasted, resulting in 35040 × 64 values.

Owing to the structure of all three forecasts, the errors of the 64 forecasted powers of one adaptation step are causally related to each other. For the PV power either the scaling parameter or the envelope or both might be erroneous, which could cause PV powers at many time steps to be over- or underestimated. Because of the use of one scaling parameter for the whole prediction horizon, the forecast usually cannot successfully

predict days with varying weather conditions. Similarly, when there is an exceptionally high peak in the load power, it will not be predicted beforehand (underestimation) and, because of the exponential smoothing, at the time when it occurs, it often causes overestimation of load power for several more hours into the future. Similarly, errors committed at different adaptation steps are often causally related to each other, especially where periodic (daily/weekly) persistence is used.

4.2 Feature Engineering

It is nearly impossible to identify individual forecast error impacts by just looking at individual time series (trajectories of powers and states) resulting from differences in the decisions of an MPC with real and one with ideal forecasts. This can be attributed to two reasons:

- As discussed in the previous Subsection, forecast errors are causally related to each other, happening in a two-dimensional space (adaptation steps and prediction horizon) and occurring simultaneously for three physical variables.
- A difference in heat storage temperature or battery state of charge trajectory does not necessarily mean that the two trajectories do not reach the main goals of the EMS equally well (e.g. charging both storages simultaneously vs. first charging the heat storage and later charging the battery).

It is more promising to look at the performance criteria defined in Subsect. 2.2. These have been shown to differ between simulations with real and ideal forecasts [10, 11]. However, these criteria are cumulative, meaning that usually they are only calculated once at the end of a simulation or in real operation for past time intervals (days, months, years). This does not allow to identify the impact of individual forecast errors, only the impact of there being forecast errors at all. Therefore, for this study, all performance criteria have been calculated for every time step in which the MPC made a new decision from the start of the simulation until that point in time. The resulting time series of the five performance criteria for the operation with real and ideal forecasts were subtracted from one another so that a positive value in the time series means that the MPC with real forecasts had a worse performance than the MPC with ideal forecasts. In the case of the operation cost, the resulting time series of cost differences is not stationary. It was detrended by calculating the change from one time step to the next (compare the integral part in the ARIMA modeling framework). For some criteria negative values are possible for a short time because of short-sighted decisions that improve performance for a few hours but impact the ability of the MPC to reach its goals later on. For the same reason, the time series of criteria differences are not monotonically increasing.

With 3×64 values being forecasted every quarter of an hour, the forecast errors themselves are hard to keep track of. However, it can be expected that an error far into the prediction horizon will not have a large individual impact on the performance of the MPC. It is much more likely that there are certain features of the forecast errors that have an impact. For this analysis, the error in the forecasted power in the next time step, the error in the energy content of a power variable over the prediction horizon and the error of the predicted maximum power of that variable within the prediction horizon are taken into account.

Figure 10 shows the spearman correlations between these forecast error features. There is practically no correlation between the forecast errors of the three different physical variables. The correlations between the forecast errors of the next powers to the corresponding errors of forecasted energy contents and maximum powers are rather small with values ranging from 0.11 to 0.27. The correlations between the error of the maximum forecasted powers and the related energy contents are rather high, ranging from 0.57 to 0.81. This means that a potentially observed correlation between the error of a maximum power with a criterion could be due to the correlation between these two features and the actual causal relationship being between the error of the forecasted energy content and the criterion - or vice versa.

For these features and the time series of criteria differences, classical methods of time series analysis led to inconclusive results. There was no clear influence of individual lags of forecast error features on deterioration of criteria differences with absolute values of spearman rank correlation coefficients ranging from 0 to about 0.3. Figure 11 shows the spearman correlations of lagged copies of all the forecast error features to the change of cost difference as an example. Most correlation coefficients are insignificant or close to insignificant with only very few small and broad peaks for some forecast error features. Even for the significant values of spearman correlation, such small absolute values indicate no real predictive power of the forecast errors. This makes sense because, depending on states of storage devices and load situation, the exact time passing between an impactful forecast error occurring and its effect materializing can vary.

Therefore, one further abstraction step has been carried out. Instead of looking at the numerical correlation between lagged copies of forecast error features and the criteria differences, two new time series were defined, looking just at extreme forecast errors

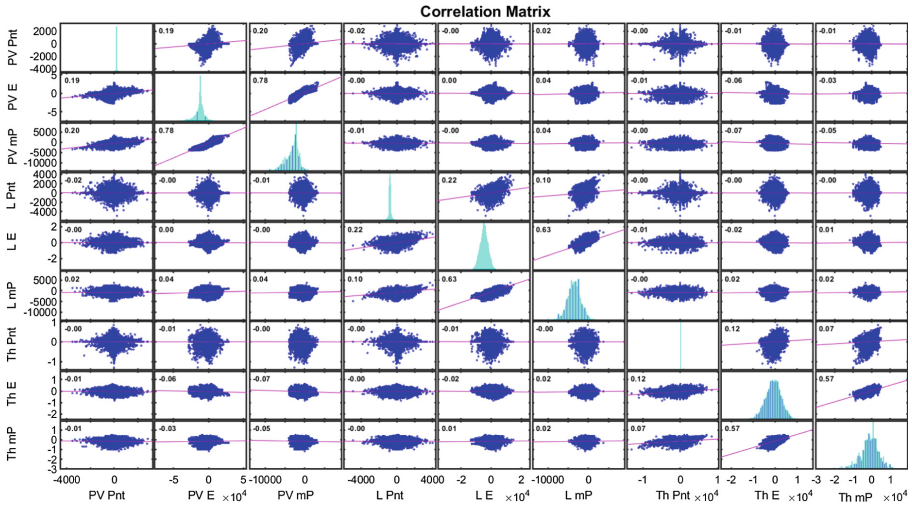

Fig. 10. Plot of the matrix of Spearman rank correlation coefficients between all forecast error features. Due to space constraints, the forecast error features had to be abbreviated. L stands for the electric load, Th for the thermal load, Pnt is power in the next time step, E is energy content and mP is maximum power in the forecast horizon.

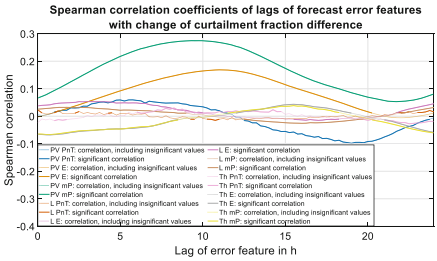

Fig. 11. Spearman correlation coefficients of lagged copies of forecast error features with the change of cost difference.

and extreme deteriorations of the criteria difference. For the purpose of this study, an extreme forecast error is one where one of the previously defined features is greater than the 97.5th percentile or smaller than the 2.5th percentile of that feature. Deteriorations of criteria often happen in situations that span over multiple time steps, often hours, in small steps. Calculating the 95th percentile or the 99th percentile of the change of the criteria difference often leads to detecting only some of these events that don't necessarily account for the majority of the criteria difference. Therefore, in order to identify extreme deteriorations of a performance criterion, the 2 h changes of a criterion difference are divided by its rolling 7-day mean first. An extreme deterioration of a performance criterion then is one where this relative 2 h change is above the 99th percentile of all relative 2 h changes.

This results in a reduction of the high complexity to a series of two binary variables for each pair of forecast error feature and performance criterion. A forecast error feature is now considered to have an impact if, in some horizon x (up to 24 h) before an extreme deterioration of a criterion, it is more likely that there was an occurrence of that error feature than if there was no relationship between forecast error feature and performance criterion at all. Figure 12 and Fig. 13 summarize all the manipulations carried out for the forecast error features and the criteria (leaving out the detrending of the operation cost), respectively.

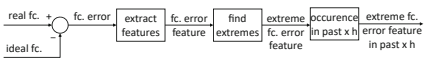

Fig. 12. Abstraction from real and ideal forecasts to binary variable indicating occurrences of "extreme" forecast errors in past 24 h.

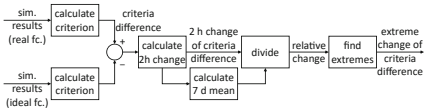

Fig. 13. Abstraction from power and state trajectories to indicator of "extreme" increase in the difference of criteria of simulations with real and ideal forecasts. Note that the signs of the subtraction are swapped when a high value of the criterion is considered advantageous.

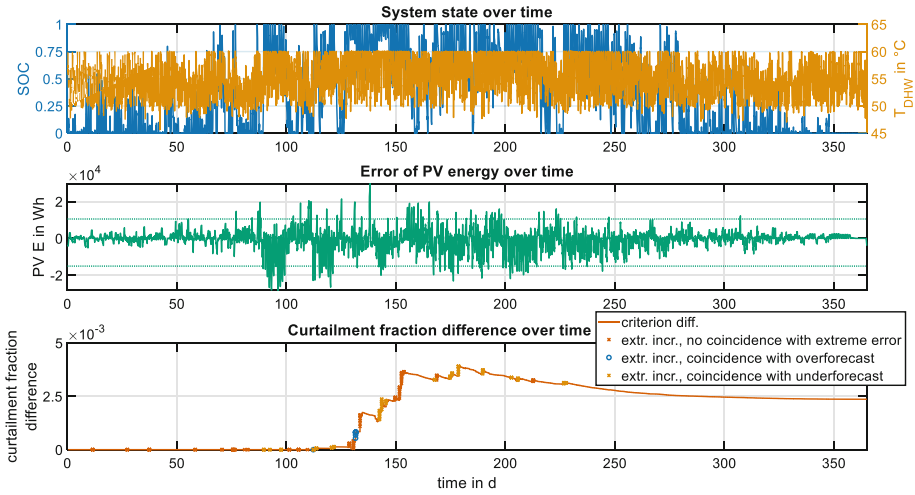

Fig. 14. System states, error of forecasted load energy and difference of curtailment fractions over time.

4.3 Exploratory Data Analysis

In order to answer, which forecast error features impact which performance criteria and how much, graphical methods and accompanying calculations of summary statistics can give a good first impression. Later, statistical tests can answer whether the obtained graphical and numerical results have not just been obtained by chance. In this paper, only a selection of graphics is presented.

Table 2 shows the performance criteria for the adapted EMS without load peak shaving and capacity firming. It can be seen that the costs and fraction of DHW drawn at less than 50 °C are much lower than even in the ideal case of the EMS in Sect. 3. However, the differences between operation cost, self-sufficiency, self-consumption and PV curtailment with real and ideal forecasts are much higher than the differences between linear and quadratic model with ideal forecasts in Sect. 3. Figure 14 shows the system states (SOC and DHW storage temperature) for the simulation with real forecasts, the error of the forecasted PV energy and the curtailment fraction difference for the

Table 2. Performance criteria for the MILP-based MPC for pure feed-in peak shaving with real and ideal forecasts.

	real forecasts	**ideal forecasts**
operation cost, k_{oc}	350.88 €	330.71 €
self-sufficiency, k_{ss}	62.85%	63.90%
self-consumption, k_{sc}	62.53%	62.74%
PV curtailment, k_{pvc}	0.71%	0.48%
fraction of DHW drawn at less than 50 °C, k_{dhw}	3.82%	0.01%

whole simulated period of one year. Extreme increases of the curtailment fraction difference according to the methodology described in Subsect. 4.1 are marked, with different markers for coincidences with extremely over- or underforecasted load energy and for no coincidence with extreme load energy forecast errors. Most increases in the curtailment fraction difference occur in late spring and early summer, between days 130 and 155. At about 142 days, one such considerable increase happens with a clear coincidence with extremely underforecasted PV energy. Two more strong increases at about 133 days and 152 days happen without there being a coincidence with underforecasted PV energy. From days 90 through 100 there is a period with a lot of occurences of extreme underforecasts, but the curtailment fraction difference does not increase considerably.

This illustrates that

a) several factors influence whether a performance criterion will be affected by a forecast error: with the present EMS, curtailment will only occur when the SOC is high and the PV power surplus exceeds 50% of the nominal PV power. The SOC might be high due to other forecast errors (e.g. overforecasted load energy).
b) Furthermore, it is important to keep in mind the long chain of manipulations shown in Figs. 11 and 12. If there is a high PV power surplus and the battery is fully charged, PV power will be curtailed due to the limitations of the battery even when there where ideal forecasts. This explains why the curtailment fraction difference (difference between real and ideal forecasts) does not increase substantially in mid and late summer, although a lot of actual curtailment is expected to happen then. However, this curtailment is due to the physical limitations of the battery, not due to forecast errors, which is why it is of no particular interest for this analysis.

Figure 15 illustrates the single and joint distributions of the battery state of charge and the occurrence of extreme PV energy forecast errors for extreme increases of the curtailment fraction difference (distribution over SOC is shown for non-extreme changes of curtailment fraction difference for comparison). It is clear that the battery is often either fully charged or completely discharged. Most occurrences of extreme deteriorations of the curtailment fraction differences happened when the SOC was close to one, some when it was zero and very few when it was in between.

Comparing this to Fig. 14, the “extreme” increases of curtailment fraction difference at zero SOC correspond to the violet crosses up to day 90. These seem to be numerical artefacts due to there being practically zero curtailment, leading to problems when dividing the 2 h changes by their seven-day rolling mean. The increases of curtailment fraction difference where SOC is between zero and one are probably artefacts of the feature extraction process as well. Using the 2 h change of the criteria, it is likely that these points represent cases where “extreme” curtailment occurred in the preceding 2 h, directly followed by a period of considerable discharge. The remaining points at high or full SOC are expected and are almost evenly split between cases where PV energy was at least once considerably underforecasted in the preceding 24 h and cases where it was either overforecasted or PV energy forecast errors were non-extreme. This observation is commensurate with Fig. 14, indicating that high SOC and PV energy underforecasts are considerable influences on PV curtailment but there being more than that.

Figure 16 shows an extension of the histogram in Fig. 15 for all forecast error features. High orange bars with no other bar being comparably high indicate that the

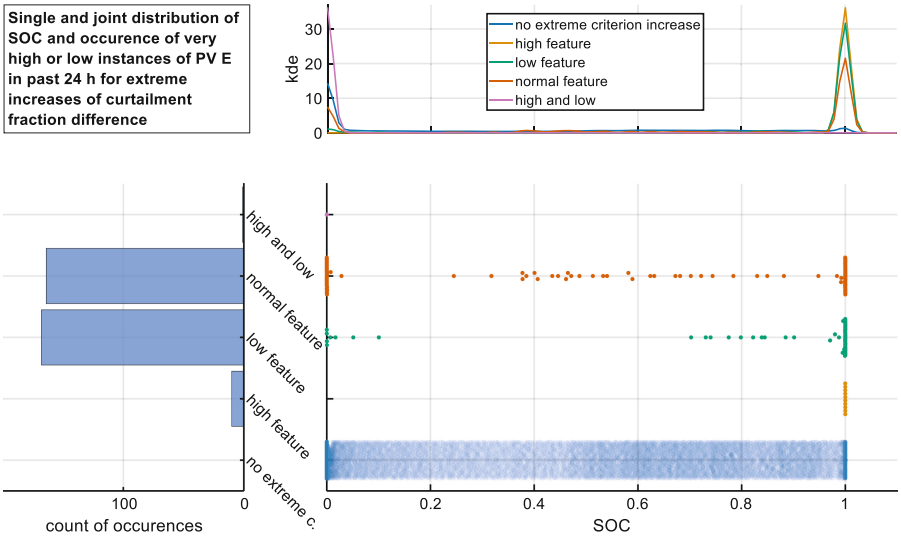

Fig. 15. Jointplot visualizing the single- and bivariate distributions of battery SOC and extreme or non-extreme PV energy forecast errors for extreme and non-extreme changes of the curtailment fraction difference. The histogram on the left has four categories for extreme increases of curtailment fraction difference (PV energy forecast error was non-extreme/normal during the past 24 h, there were extreme overforecasts, extreme underforecasts or both). The fifth category is not in the histogram due to a much higher count of occurrences, but shown in the beeswarm plot in the bottom right for comparison.

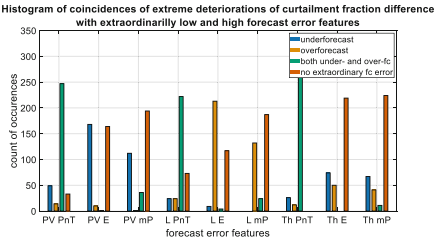

Fig. 16. Histogram of coincidences of extreme deteriorations of curtailment fraction difference with extraordinarily low and high forecast error features.

respective forecast error feature probably has no large impact on the curtailment fraction difference. This seems to be the case for thermal energy and maximum thermal power. PV energy and maximum PV power underforecasts and load energy and maximum load power overforecasts occur quite frequently, pointing to a potential connection. The high green bars for all three powers at the next time step are primarily due to the forecast methods often producing an overforecast shortly after occurrences of underforecasts and vice versa. Therefore, they may not be indicative of real connections, although such connections cannot be ruled out. The corresponding histograms for the other four performance criteria are in the appendix (Figs. 21, 22, 23 and 24).

4.4 Significance Tests Using Bootstrapping

While it makes sense that underforecasted PV energy and overforecasted load energy may lead to avoidable PV curtailment, the observation of there being considerable numbers of these forecast error features coinciding with increased curtailment fraction difference is not sufficient to prove a causal relation. Therefore, a significance test has been carried out for each combination of high or low forecast error feature and performance criteria.

In this case, a nonparametric bootstrap test was used. First, the null hypothesis was stated as "there is no (causal) relationship between the occurrence of an extreme value of an error feature in the past x h (with x being varied from 1 to 24 h) and an extreme deterioration of a performance criterion." The test statistic was defined as the number of time steps where there was a recent occurrence of an extreme forecast error feature value (high and low extremes are counted and tested separately) and an extreme increase in the respective criterion. Bootstrap tests simulate the null hypothesis many times and count how often an equally or more extreme (in this case: higher) test statistic was obtained in comparison to the observed test statistic. If the fraction of simulations with more extreme test statistic values is higher than a predefined statistical significance level (here 5%), then there is not enough evidence to refuse the null hypothesis (i.e. it may be that this particular forecast error feature has no influence on the performance criterion). If it is lower than the significance level, there is enough evidence to refuse the null hypothesis (i.e. it is likely that this particular forecast error feature does indeed have an influence on the performance criterion). The statistical simulation is of course dependent on assumptions about the statistical behaviour of the criteria difference and forecast error features. These assumptions need to be informed by some further analysis of the time series to be modelled.

With the null hypothesis being no causal relation between a criterion and a forecast error feature, it is necessary to reasonably "mix up" one of both. However, both are obviously autocorrelated, the error features because of the data-driven forecast approach and the criteria due to their cumulative character. Therefore, results would certainly be incorrect if we just bootstrapped a new time series from the distribution of one of these time series.

Figure 17 shows the absolute and partial autocorrelations of the change of curtailment fraction difference. The partial autocorrelation suggests that it might be possible to model the change of curtailment fraction difference reasonably well by using a simple first order autoregressive model. The same is true for all other performance criteria,

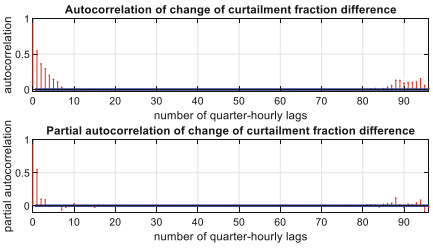

Fig. 17. Autocorrelation and partial autocorrelation of change of curtailment fraction difference.

while the forecast error features seemed to be better modelled by using second and third order autoregressive models. Therefore, for the first iteration of the bootstrap test, the performance criteria were "mixed up" by fitting a first-order autoregressive model and bootstrapping the innovations/errors of that model. In theory, regression implicates the assumption that these model errors are normally distributed. Figure 18 shows that this is almost surely not the case. We still decided to use the autoregressive model and draw the innovations from the empirical distribution of the model errors instead of from a normal distribution, coming closer to the actual behaviour of the performance criteria than a simple bootstrap directly from the observed values.

Figure 19 shows 50 of the 1000 replicated time series of curtailment fraction difference for the bootstrap test in comparison to the actual time series. Obviously, the seasonal behaviour (most of the difference stemming from early summer) is not represented well. The final value of the actual time series, however, is very close to the mean of the distribution of final values of the replicated time series. Considering that any correction of the simulation to represent seasonal behaviour would greatly reduce the degrees of freedom and therefore the probabilistic characteristics of the simulation, we consider this approach sufficient for now.

Table 3 shows the number of coincidences of low and high forecast error features in the past 24 h with extreme deteriorations of the performance criteria, analogous to Fig. 16, with all fields where the previously described bootstrap test rejected the null

Table 3. Number of coincidences of high (overforecast) or low (underforecast) forecast error features with extreme deteriorations of criteria. Statistically significant influences coloured blue. Results for occurences in past 24 h.

	Error of next time step's PV power		Error of PV energy content		Error of maximum PV power		Error of next time step's load power		Error of load energy content		Error of maximum load power		Error of next time step's thermal power		Error of DHW energy content		Error of maximum thermal power	
	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low
k_{ss}	212	153	95	20	141	49	324	325	264	97	129	34	307	341	25	102	45	20
k_{sc}	181	209	66	43	64	44	277	273	170	86	46	48	300	311	103	46	58	58
k_{oc}	172	189	50	50	77	44	302	316	221	87	112	45	326	346	50	106	60	53
k_{pvc}	261	296	11	169	37	148	246	246	217	13	156	24	317	331	50	74	52	78
k_{dhw}	135	148	25	52	47	32	315	318	251	39	107	23	309	343	32	205	11	116

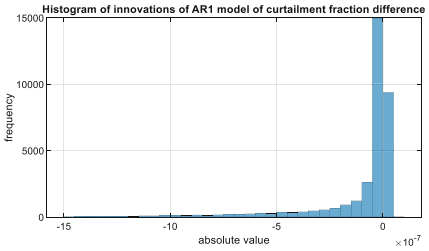

Fig. 18. Histogram of errors of AR1 model for curtailment fraction difference.

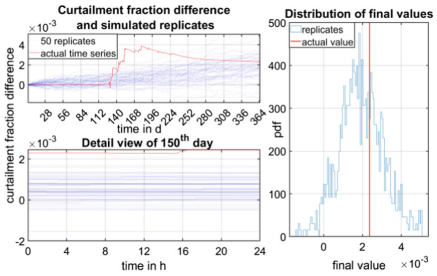

Fig. 19. Actual and replicated curtailment fraction difference.

hypothesis (meaning: "there is not enough evidence to be sure there is no connection") coloured blue.

Interpreting this table has several stumbling blocks:

a) Even small (cor-)relations can be statistically significant, even when engineers wouldn't consider them practically significant.
b) Because the forecast errors of the maximum powers are correlated to the forecast errors of the energy contents, a significant correlation between a maximum power forecast error and the deterioration of a criterion could also point to the actual causality being between the energy content forecast error and the deterioration of the criterion.
c) There might be more complex relations not simulated by the test. Consider the 221 occurences of there being a heavily overforecasted load energy within 24 h before the occurrence of an extreme increase in operation cost. The difference in operation cost can increase when there is not enough energy left in the battery to power all loads or when the battery is fully charged and curtailment occurs. When the load energy is overforecasted, this leads to a reduction of the PV power fed into the grid, so the battery can prepare for the high load by charging as much as possible. If the overforecast occurs when the SOC is low or there is not enough PV energy to fully charge the battery, there will be no unnecessary curtailment and therefore no loss of feed-in remuneration. If, however, the overforecast occurs when the SOC is already high, there will be more curtailment than in the case of ideal forecasts, resulting in a loss of feed-in remuneration.

a) and b) will be addressed in the next Subsection. In order to address c), another test was carried out that preserved the relationship between system states and changes in the performance criteria. In order to achieve this, it was necessary to now "mix up" the time series of forecast error features instead of the performance criteria. We followed the same procedure, building autoregressive models of the forecast error features (now third order, due to the earlier discussion) and simulated their behaviour by randomly drawing innovations from the empirical errors of that model. The test statistic changed from "number of coincidences of extreme deteriorations of criteria with extreme over-/underforecasts" to "number of coincidences of extreme deteriorations of criteria with extreme over-/underforecasts and high/low system state", replacing "system state" with either SOC or DHW storage temperature. Low and high SOC were defined as below 0.2

Table 4. Number of coincidences of high or low forecast error features with extreme deteriorations of criteria. Statistically significant influences of forecast error features by themselves coloured blue and statistically significant influences of forecast errors in certain system states coloured orange. Results for occurences in past 24 h.

	Error of next time step's PV power		Error of PV energy content		Error of maximum PV power		Error of next time step's load power		Error of load energy content		Error of maximum load power		Error of next time step's thermal power		Error of DHW energy content		Error of maximum thermal power	
	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low
k_{ts}	212	153	95	20	141	49	324	325	264	97	129	34	307	341	25	102	45	20
k_{se}	181	209	66	43	64	44	277	273	170	86	46	48	300	311	103	46	58	58
k_{oc}	172	189	50	50	77	44	302	316	221	87	112	45	326	346	50	106	60	53
k_{pvc}	261	296	11	169	37	148	246	246	217	13	156	24	317	331	50	74	52	78
k_{dhw}	135	148	25	52	47	32	315	318	251	39	107	23	309	343	32	205	11	116

and above 0.8, respectively, and low and high temperature as below 51 and above 58 °C, respectively.

This procedure led to many more forecast error features being considered to have statistically significant impact on performance criteria when system states are in these ranges. Table 4 shows the same as Table 3, now with these new statistically significant influences coloured orange. This implicates that, under certain circumstances, almost all forecast errors can have a significant impact on the deterioration of performance criteria. Of course, this further complicates the attribution of deteriorations of criteria to the different forecast error features.

4.5 Impact Estimates Using Logistic Regression

As previously stated, statistical significance does not necessarily imply practical relevance. Therefore, as a last step, logistic regression has been carried out. The predicted variable is the probability that an extreme increase in the performance criterion occurs. The predictor variables are all influences that have been identified as statistically significant (extreme states and forecast error features in the past x h either alone or in conjunction with extreme states), an intercept, the first-order lags of the criterion and the indicator variable for it being extreme, as well as the linear interaction between these last two. In logistic regression, the linear combination of these predictors is the input into a transformation function that translates this weighted sum into a conditional probability. The logistic regression has been carried out for various "horizons" for looking back at extreme forecast error features.

The MATLAB function used for this regression internally carries out a statistical test, testing whether each parameter individually is significantly different from zero. As long as this test reported a p-value above 0.05 for any parameter, the parameter with the highest p-value was removed and the regression carried out again. For each performance criterion, the horizon that yielded the model with the lowest Bayesian Information Criterion (BIC, measuring predictive power in relation to number of explanatory variables) is reported in more detail.

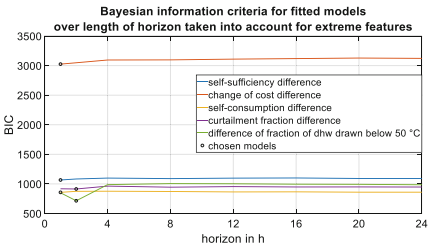

Fig. 20. Bayesian information criteria for fitted models over length of horizon taken into account for extreme features.

Figure 20 shows the BIC of the final fitted models over the length of the horizon for each criterion and the models that have been chosen for further evaluation. All chosen models apart from the one for change of cost difference reached R^2-values between 0.71 and 0.78. The model for change of cost difference had an R^2-value of 0.11, meaning that the change of cost difference is practically unpredictable by the previously defined predictors. However, costs are closely linked to self-sufficiency, self-consumption and curtailment, where clear influences have been found.

Apart from the intercept, the autoregressive term of the absolute value of the criterion itself and its linear interaction with the indicator value of it being extreme, all terms are binary. This implicates that the values of the regression parameters are the same if two variables have the same impact. In turn, this means that we can measure the relative impact of forecast error features on a criterion by comparing the values of their regression coefficients.

Tables 5, 6, 7, 8 and 9 report the regression coefficients for the five chosen models. For all of the following tables, it should be kept in mind that there is a high correlation between energy forecast errors and maximum power forecast errors. For future research, it would be an interesting question whether excluding either one of them a priori as predictor variables still yields similar results.

Table 5. Estimates of impacts of forecast error features and extreme states in the form of logistic regression coefficients on self-sufficiency difference. A note on units: all but the autoregressive predictors, in this table "last time step's self-sufficiency difference", are dimensionless, meaning that all autoregressive parameters and those of the linear interactions have to be of the dimension $(\text{unitofcriterion})^{-1}$.

Intercept	−9.5310
Low SOC	2.3621
Underestimation of next time step's thermal power	1.8280
High T and underestimation of load energy	1.9488
Last time step's indicator for extreme self-sufficiency difference	7.7539
Last time step's self-sufficiency difference	100.7252

For self-sufficiency difference, shown in Table 5, the 1 h horizon was chosen. The regression ascribes the highest influence to the SOC being low, meaning that there is not enough energy to cover the current loads. Another important influence is seen in underestimating next time step's thermal power. This may lead to not using energy from the PV plant or the battery to increase DHW storage temperature before a consumption event occurs, instead feeding it into the grid, resulting in the need to cover the heating power from the grid. Underestimation of load energy has been included in conjunction with high DHW storage temperature with an impact between the previously mentioned ones. While underestimation of load energy should theoretically be able to cause reductions in self-sufficiency due to similar reasons as underestimations of thermal power, the connection to high DHW storage temperature does not seem so clear. It might be an artefact due to a possible error of the second kind (false negative) excluding load energy underestimations in the first hypothesis test but including it in the second iteration only in connection with extreme system states. It is also interesting to see that no kind of PV forecast error feature has been included. They may be included indirectly by the low SOC which could not only be due to low PV availability but also due to overforecasts dating further back than one hour. This is especially true for the aim of modeling differences in the criteria between real and ideal forecasts (instead of all deteriorations of the criteria themselves), because in the ideal case PV availability would be the same. These two surprising results may point to the variation of the horizon for the past occurences of forecast errors for all features at the same time to be inadequate. In future research, it might be an interesting analysis to vary the horizons for all features independently.

Table 6 shows the regression coefficients for the self-consumption difference, for which also the 1 h horizon was chosen. The highest impact is attributed to underforecasts of PV power in the next time step in combination to high DHW storage temperature, describing situations in which the heat storage is not available for storing surplus PV power that was unexpected. Another influence connected to the supply side, however considered the smallest, is the underestimation of PV energy. On the demand side, underforecasts of load energy or maximum thermal power, the latter connected to high DHW storage temperatures, are considered similarly influential on a medium level.

Table 6. Estimates of impacts of forecast error features and extreme states in the form of logistic regression coefficients on self-consumption difference.

Intercept	−7.1270
Overforecast of PV energy	1.0977
Underforecast of load energy	1.6174
High T and underforecast of next time step's PV power	2.3963
High T and underforecast of maximum thermal power	1.5236
Last time step's indicator for extreme self-consumption difference	9.0397
Last time step's self-consumption difference multiplied with last time step's indicator for extreme self-consumption difference	−167.5227

Table 7. Estimates of impacts of forecast error features and extreme states in the form of logistic regression coefficients on change of cost difference.

Intercept	−7.9707
Low SOC	2.6137
High SOC	2.7450
Low T	1.2584
Overforecast of next time step's PV power	1.3600
Underforecast of next time step's load power	0.6970
Underforecast of next time step's thermal power	1.4452
Underforecast of thermal energy	0.5374
Low SOC and overforecast of PV energy	0.8764
Last time step's indicator for extreme change of cost difference	3.0867
Last time step's change of cost difference	25.8152
Last time step's change of cost difference multiplied with last time step's indicator for extreme change of cost difference	−31.9412

These errors lead the optimizer to feed more power into the grid that in the ideal case are stored to cover later consumption events.

In the case of the change of cost difference, shown in Table 7, there is a rather high number of influences. However, it should be remembered that the R^2-value was very low. The highest influences are the SOC being either high or low. At high SOCs, the feed-in limit may cause PV curtailment, reducing feed-in remuneration. Low SOCs mean that loads cannot be fully covered from the battery. The same is true for the DHW load when DHW storage temperature is low, which is considered to have a medium impact. Supply side overforecasts have low (PV energy) to medium (next time step's PV power in combination with low SOC) individual impacts. Underforecasts of electrical and thermal powers also have low to medium individual impacts, but make up a higher portion than PV overforecasts. All of these are very probably made up of cases where PV energy was fed into the grid when it would have been needed for covering the loads.

Table 8 shows the regression coefficients for the curtailment fraction difference, for which the 2 h horizon was chosen. The highest influence is again ascribed to high SOC. In fact, in the system configuration of this study, only if SOC is high there should be PV curtailment at all, because the nominal power of the battery is high enough for any PV surplus that could occur. Given this necessary condition, it makes sense that underforecasting the PV power in the next time step (the highest of the forecast error feature impacts) or overforecasting the maximum load power (the smallest impact) may lead to decisions where less energy is fed into the grid than what would be possible. Underforecasts of the maximum thermal power (medium impact) may lead to the heat storage not being charged when, in fact, it would make sense to accept higher losses (due to moving energy from one storage to another and self-discharge of the DHW storage).

The second medium influence of low SOC and underestimation of the thermal power in the next time step is another surprise, that may be an artefact.

Finally, Table 9 shows the regression coefficients for DHW drawn at less than 50 °C, for which also the 2 h horizon was chosen. The highest impact by far are underforecasts of the thermal power in the next time step, pointing to situations where the optimizer is not aware that the necessary action to take would be heating the DHW storage. The medium impact by underforecasts of maximum thermal power point to the same situation. As was the case with PV curtailment, the medium impact of low DHW storage temperature is due to it being a necessary condition for the criterion to deteriorate. The rather small impact of low SOC is probably due to the MILP optimizer prioritizing battery storage charging over heat storage charging, as was found in Subsect. 3.2. The medium impact of high SOC in combination with underforecasts of PV energy, however, is another surprising result, that may be due to chance.

Table 8. Estimates of impacts of forecast error features and extreme states in the form of logistic regression coefficients on curtailment fraction difference.

Intercept	−7.4310
High SOC	1.8717
Underforecast of next time step's PV power	1.5695
Overforecast of maximum load power	0.8223
Underforecast of maximum thermal power	1.3928
Low SOC and underforecast of next time step's thermal power	1.3791
Last time step's indicator for extreme curtailment fraction difference	7.6738
Last time step's curtailment fraction difference	−310.4605

Table 9. Estimates of impacts of forecast error features and extreme states in the form of logistic regression coefficients on fraction of DHW drawn at less than 50 °C.

Intercept	−11.0265
Low SOC	0.6462
Low T	2.1187
Underforecast of next time step's thermal power	5.1319
Underforecast of maximum thermal power	1.6589
High SOC and underforecast of PV energy	2.3857
Last time step's indicator for extreme fraction of DHW drawn at less than 50 °C	7.7495

5 Summary and Outlook

5.1 Summarizing the Results

In this study, three different sources of uncertainty in the MPC of a residential PV-battery-heat pump-heat storage system providing DHW, self-consumption optimization and grid power peak shaving have been analysed. These were the structure and parameters of the optimization model and forecast errors. Due to the different natures of these uncertainties, different methods had to be applied.

It was found that, in the case of this HESS, the structure of the optimization model can have a profound impact on the use of the two storage devices. The optimal parameters for a linear model led to the optimizer prioritizing the charging of the battery over the DHW storage. The quadratic model, in that simulation considered identical to the actual controlled system, prioritized charging the DHW storage. While most performance criteria have not been impacted by this difference in priorities, the fraction of DHW drawn at less than 50 °C, representing quality of heat supply, was considerably worse in the case of the linear model.

The impact of small differences of model parameters from the real system, as they will surely occur in real control systems, were found to be small. Large differences of these model parameters, however, had a similar impact to the model structure, causing profoundly different behaviour of the system which, depending on the nature of the differences, caused some performance criteria to improve while deteriorating others. The trade-off represented by the optimization's objective function was shown to be best for low model parameter differences from the real system.

Comparing the numbers in Sect. 4 to those in Sect. 3 shows that the cumulative influence of forecast errors is higher than that of an oversimplified optimization model structure for four out of five performance criteria in this study. For all performance criteria, an effort to identify the most impactful forecast errors was made by a combination of exploratory data analysis, feature engineering, statistical tests and logistic regression. The results show that the forecast error features that have been defined do indeed impact the criteria of interest but don't explain them fully. Furthermore, different forecast errors influence different criteria. While many of the identified influences made sense in the context of the system, some expected influences have not been included by the regression method and some identified influences could not be explained by the authors. These unexpected outcomes warrant further investigation.

5.2 Limits of the Study

All results have been obtained for just one set of time series and with one HESS configuration. This means that it still remains to be seen if the findings in this study can be generalized for other households (other time series) and for other ESS configurations. With the reason for the impacts of model uncertainties on performance lying in changing priorities between storage devices, it can be expected that the impact of model uncertainties on single ESS is negligible. For other HESS configurations, whether model uncertainties have an impact, may depend on the number and types of energy storage devices involved. It is still expected that forecast uncertainties impact performance similarly for single ESS.

The exact setup of the EMS was different between Sects. 3 and 4, limiting the possibility of comparisons between the impacts of model and forecast uncertainties. In future research both analyses should be carried out for a broader range of multi-use applications, also addressing the topic of generalization capability.

5.3 Conclusions and Outlook

5.3.1 Improving MPC for Residential HESS

The results in Sect. 3 highlight the importance of using optimization models that represent the real system well when controlling a HESS. Using oversimplified models might lead the optimization solver to take decisions that are far from optimal, at least with respect to some goals of the energy management. Small deviations of the model from the real system, as they will inevitably occur, usually should be handled well by the MPC approach with its frequent adaptations to new information on system states.

The impact of errors of the rather simple forecasts used in this study have been shown to be a lot larger than those of model uncertainties. In recent years this has been addressed by a lot of research trying to improve the forecasts themselves. While this approach will surely lead to better results than the simple forecasts used here, there will always remain considerable forecast errors because electric and thermal loads are determined by human behaviour which is very hard to predict.

Therefore, we propose to use probabilistic methods to quantify the uncertainty of forecast uncertainties. This quantitative information could be taken into account in the optimization by various means, including 2-stage stochastic optimization, robust optimization, chance constraints and affine arithmetics. The findings of Subsect. 4.5 should be solidified by further detailed analyses, especially by looking into the surprising results. After that these findings can point to forecast error features that should be the focus in the design of a probabilistic forecast model.

5.3.2 Future Studies on MPC Robustness

The analyses carried out for this study have been rather insightful. After addressing the problems of the investigation of the impacts of forecast uncertainties, they should be carried out for a broader range of multi-use applications and households to test whether and to what degree the findings in this study can be generalized to guide the design of future EMS.

However, they have also been very costly in terms of computing power. Therefore, we propose to find a simpler and less costly method to assess and compare the performance and robustness of different operating strategies.

Authors' Contributions. Conceptualization, definition of the methodology, data curation as well as review and editing of the manuscript have been carried out jointly by R. Gelleschus and T. Bocklisch. Development of the software, validation, formal analysis, investigation, writing and visualization have been carried out by R. Gelleschus. Administration of the resources, supervision, project administration and funding acquisition have been carried out by T. Bocklisch.

Appendix

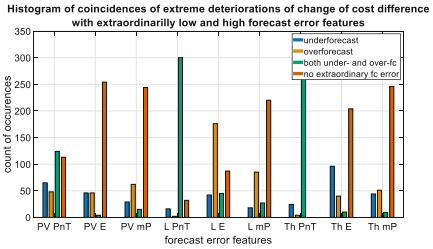

Fig. 21. Histogram of coincidences of extreme deteriorations of change of cost difference with extraordinarily low and high forecast error features.

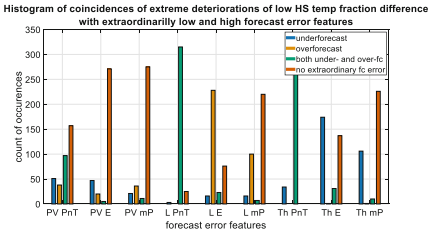

Fig. 22. Histogram of coincidences of extreme deteriorations of low heat storage temperature fraction with extraordinarily low and high forecast error features.

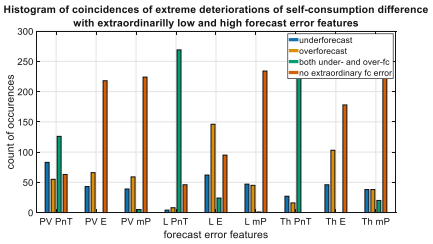

Fig. 23. Histogram of coincidences of extreme deteriorations of self-consumption difference with extraordinarily low and high forecast error features.

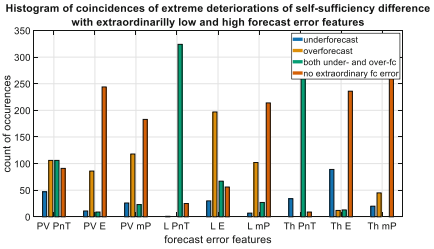

Fig. 24. Histogram of coincidences of extreme deteriorations of self-sufficiency difference with extraordinarily low and high forecast error features.

References

1. AG Energiebilanzen e. V., Energy Flow Chart for the Federal Republic of Germany in 2021. Available online: https://ag-energiebilanzen.de/en/data-and-facts/energy-flow-chart/
2. T. Bocklisch, Hybrid energy storage approach for renewable energy applications, Journal of Energy Storage, Vol. 8, 2016. DOI: https://doi.org/10.1016/j.est.2016.01.004
3. Y. Chen, Z. Xu, J. Wang, P.D. Lund, Y. Han, T. Cheng, Multi-objective optimization of an integrated energy system against energy, supply-demand matching and exergo-environmental cost over the whole life-cycle, Energy Conversion and Management, vol. 254, 2022. DOI: https://doi.org/10.1016/j.enconman.2021.115203
4. S. Englberger, A. Jossen, H. Hesse, Unlocking the Potential of Battery Storage with the Dynamic Stacking of Multiple Applications, Cell Reports Physical Science, 1(10), 2020. DOI: https://doi.org/10.1016/j.xcrp.2020.100238
5. C.N. Truong, Assessment and Optimization of Operating Stationary Battery Storage Systems, Ph.D. dissertation, TU München, 2019. Available online: http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20191121-1463123-1-7
6. S. Kuboth, F. Heberle, T. Weith, M. Welzl, A. König-Haagen, D. Brüggemann, Experimental short-term investigation of model predictive heat pump control in residential buildings, Energy & Buildings, 204, 2019. DOI: https://doi.org/10.1016/j.enbuild.2019.1094 4 4
7. S. Kuboth, F. Heberle, A. König-Haagen, D. Brüggemann, Economic model predictive control of combined thermal and electric residential building energy systems, Applied Energy, 240, 2019. DOI: https://doi.org/10.1016/j.apenergy.2019.01.097
8. K. Büdenbender, M. Braun, A. Schmiegel, D. Magnor, J.C. Marcel, Improving PV-Integration into the Distribution Grid. Contribution of Multifunctional PV-Battery Systems to Stabilised System Operation, in: Proceedings of the European Photovoltiac Solar Energy Conference and Exhibition, pp. 4839–4845, 2010.
9. F. Braam, R. Hollinger, M.L. Engesser, S. Müller, R. Kohrs, C. Wittwer, Peak Shaving with Photovoltaic-Battery Systems, in: Proceedings of the 5th IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), Istanbul, 2014. DOI: https://doi.org/10.1109/ISGTEurope.2014.7028748
10. J. Weniger, J. Bergner, V. Quaschning, Integration of PV power and load forecasts into the operation of residential PV battery systems, in: Proceedings of the 4th Solar Integration Workshop, Berlin, 2014.
11. R. Gelleschus, M. Böttiger, T. Bocklisch, Optimization-Based Control Concept with Feed-in and Demand Peak Shaving for a PV Battery Heat Pump Heat Storage System, Energies, 2019(12), 2019. DOI: https://doi.org/https://doi.org/10.3390/en12112098
12. M. Böttiger, M. Paulitschke, R. Beyer, L. Neumann, T. Bocklisch, Modular hybrid battery storage system for peak-shaving and self-consumption optimization in industrial applications, Energy Procedia, vol. 155, 2018. DOI: https://doi.org/10.1016/j.egypro.2018.11.064
13. G. Angenendt, S. Zurmühlen, D.U. Sauer, Participating in the control reserve market with PV battery energy storage systems and power-to-heat application, in: Proceedings of the 13th International Renewable Energy Storage Conference (IRES 2019), Düsseldorf, 2019. DOI: https://doi.org/10.2991/ires-19.2019.6
14. Z. Wu, H. Tazvinga, X. Xia, Demand Side Management of Photovoltaic-Battery Hybrid System, Applied Energy, 148, 2015. DOI: https://doi.org/10.1016/j.apenergy.2015.03.109
15. C. Chen, J. Wang, Y. Heo, S. Kishore, MPC-Based Appliance Scheduling for Residential Building Energy Management Controller, IEEE Transactions on Smart Grid, 4(3), 2013. https://doi.org/ https://doi.org/10.1109/TSG.2013.2265239

16. D. Azuatalam, K. Paridari, Y. Ma, M. Förstl, A.C. Chapman, G. Verbič, Energy management of small-scale PV-battery systems: A systematic review considering practical implementation, computational requirements, quality of input data and battery degradation, Renewable and Sustainable Energy Reviews, 112, 2019. DOI: https://doi.org/10.1016/j.rser.2019.06.007
17. W.B. Powell, S. Meisel, Tutorial on Stochastic Optimization in Energy—Part I: Modeling and Policies, IEEE Transactions on Power Systems, 31(2), 2016. DOI: https://doi.org/10.1109/TPWRS.2015.2424974
18. W.B. Powell, S. Meisel, Tutorial on Stochastic Optimization in Energy— Part II: An Energy Storage Illustration, IEEE Transactions on Power Systems, 31(2), 2016. https://doi.org/https://doi.org/10.1109/TPWRS.2015.2424980
19. J. Bergner, J. Weniger, T. Tjaden, PVprog-Algorithmus - Algorithmus zur Umsetzung der prognosebasierten Batterieladung für PV-Speichersysteme mit messwertbasierten PV- und Lastprognosen (Version 1.1), 2016. Available online: https://solar.htw-berlin.de/pvprog-algorithmus/