Impact of PV forecasts uncertainty on batteries management in microgrids

Andrea Michiorri, Arthur Bossavy, Georges Kariniotakis, Robin Girard

Impact of PV forecasts uncertainty in batteries management in microgrids

Andrea Michiorri
Arthur Bossavy

Georges Kariniotakis
Robin Girard

MINES ParisTech
PERSEE : Centre for Processes, Renewable Energies and Energy Systems
Sophia Antipolis, France
andrea.michiorri@mines-paristech.fr

Abstract—This paper is motivated by the question of the impact that uncertainty in PV forecasts has in forecast-based battery schedule optimisation in microgrids in presence of network constraints. We examine a specific case where forecast accuracy can be impacted by the lack of enough data history to finetune the forecasting models. This situation can be expected to be frequent with new PV installations. A probabilistic PV production forecast algorithm is used in combination with a battery schedule optimisation algorithm. The size of the learning dataset of the forecast algorithm is modified in order to simulate the application of the system to new plants and the impact on the performance in the management of the battery is analysed.

Index Terms-- Photovoltaic systems, Batteries, Forecasting

I. INTRODUCTION

This paper presents first results of a study on the effect of the size of forecast algorithm datasets on the management of energy storage carried out within the framework of the project NiceGrid. The objective of this project is to develop an active network management system able to control distributed electric storage devices and flexible loads in order to facilitate the penetration of solar power on the distribution network and to increase power quality and security of supply. One of the main challenges found is the necessity to provide individual forecasts for each solar plant on the network in order to identify localised voltage constraints violations. This information is then used to schedule optimized battery charge and discharge sequences.

In general solar power production forecast algorithms use up to date meteorological forecasts and historical production data as input to generate the forecasts. Two problems relative to the size of the learning datasets are found: little datasets may be not suitable for advanced statistical methods, resulting in poor performance. On the other hand large datasets may increase considerably the computational intensity and time of the forecast algorithm. Historical data are necessary in particular when information on the characteristics of the solar plant, such as tilt and orientation are not available making a so-called physical approach difficult to apply. In that case a

statistical approach might me more appropriate. This is a common situation for residential solar plants, for which often only information about the rating of the inverter is available. In these cases it is not possible to estimate the solar plant peak production period, when its output can create constraints violations on the network.

The scheduling of batteries is strongly influenced by errors in input forecasts. In the system studied, batteries are called to increase their consumption in hours of the day when high solar production is forecasted. When high photovoltaic (PV) production levels are not forecast, batteries are potentially used for day/night arbitrage. The battery is therefore called to de-optimise its typical cycle in order to provide a service to the network and the difference between the daily gain of the undisturbed schedule and the schedule optimized for absorbing the excess solar production represents the cost of the flexibility offered. If a high production event is not forecasted, solar panels can be disconnected, with a loss of renewable energy and potentially additional charges for the distribution network operator. On the other hand if forecasted high production events do not occur, the cost of de-optimisation of the battery has been lost.

The study does not aim at identifying the optimal or the minimal training dataset size of solar power production forecasts, but to characterise its influence on battery operation performance. This information can then be used to facilitate planning decisions by reducing uncertainty about the effect of lack of historical information on new plants. This is particularly important considering the current growth of installed solar power, the relative small size of solar plants and the cost of storage systems which requires correct evaluation of the risks associated with their utilization.

II. BACKGROUND

Short term forecasts of PV plants output are based on: i) meteorological forecasts coupled with a physical model, ii) time series analysis or iii) a combination of the two approaches. Research is particularly active on the second approach because of its possibility to use only measurements resulting from the PV system as inputs. Examples of this

approach can be found in [1], [2] and [3]. Two [1], [2] of these three works make use of neural networks whilst the third proposes the use of an autoregressive moving average (ARMA) method coupled with a Kalman filter. An example of the third approach can be seen in [4], where meteorological solar radiation forecasts are used as an alternative to local solar radiation measurements, in combination with a nonlinear autoregressive exogenous model (NARX) neural network.

Regarding battery schedule optimisation, dynamic programming or mixed integer dynamic programming are largely the most common approaches as seen in [5], [6]. An interesting aspect of battery schedule optimisation is the necessity of the algorithm to be robust and able to incorporate stochastic inputs as shown in [7], [8].

III. METHODOLOGY

In this study, the methodology reported below is adopted.

1. The forecast algorithm is trained with a dataset D

2. Forecasts are calculated for N test days

3. For each test day, an optimal schedule for the battery is calculated taking into account i) the forecast calculated at point 2 and ii) the objective function

4. The performance of the forecast-based schedule optimisation algorithm is evaluated over the test period

5. The points from 1 to 4 are repeated with different sizes of the training dataset D

A description of the details of the algorithms for PV power forecast and for the battery optimisation used in this study are provided below.

A. PV plant power output forecast algorithm

The most complete information we can get from the future PV production is given through its probability distribution. Such a distribution can be characterized through a probability density function, a cumulative distribution function or equivalently through a set of quantiles. If $F_{t+h|t}$ denotes the cumulative distribution function and $q_{t+h|t}^{\alpha}$ the α-quantile of the production p_{t+h}, forecast for the horizon h, conditionally to the information set Φ_t available at instant t, we got:

$$F_{t+h|t}(x) = \mathbb{P}(p_{t+h} \leq x | \Phi_t) \qquad (1)$$

$$q_{t+h|t}^{\alpha} = F_{t+h|t}^{-1}(\alpha) \qquad (2)$$

Quantiles may represent a good alternative to forecasts of the mean production level, as considered in [9], since it has been demonstrated they are optimal predictors for a large class of decision loss functions [10]. In [11], the authors used a recursive algorithm, initially proposed in [12], to make time-adaptive forecasts of the PV production distribution. In their approach however, quantiles are estimated independently from one another, which might lead to undesirable phenomena, e.g. quantiles crossing [13].

In this paper, we propose to forecast the PV production distribution using the Quantile Regression Forest (QRF) algorithm [14]. The QRF algorithm estimates the whole cumulative distribution function of the production, then allowing to derive the desired quantiles all at once. Such estimation is made conditionally to the projected surface global solar irradiation at the level of the solar plant. We considered solar irradiation numerical weather predictions from the European Centre for Medium-Range Weather Forecasts (ECWMF). They are available at a 3-hour time resolution, but for the purposes of the study they have been linearly interpolated to an hourly resolution and cover forecast horizons from 18 h to 30 h ahead (i.e. from 6 am to 6 pm the next day).

For each forecast horizon, the forecasting model is repeatedly trained using a sliding window. This approach permits to update the model's parameters and provides time-adaptive forecasts. For the purpose of the study, we considered different training periods with size ranging from 7 days to 4 months, with one week increment.

B. Battery schedule optimisation algorithm

The objective of the battery schedule optimisation algorithm is to minimise the cost associated to the exploitation of the battery. For a complete schedule of the battery PLAN = [P_1, ..., P_T], with P_i being the power charged or discharged by the battery at time i, the overall cost can be calculated as the sum of three main components, as proposed in [15] and reported in Equation (3).

$$C(PLAN) = C_{source} + C_{loss} + C_{lol} \qquad (3)$$

where, C_{Source} represents the cost of importing/exporting power from/to the main grid, C_{loss} represents the cost associated to the losses in the inverter and the efficiency in the charge/discharge of the battery and C_{lol} represents the cost of the loss of life of the battery.

The sourcing cost C_{source} depends on the price Pr of electricity. In this study, the residential day-night tariff applied in the region of Nice is chosen. The cost is calculated both in the charge and discharge phase, where in the second case, it represents a profit for the battery as reported in Equation (4).

$$C_{source} = \Sigma_h (P_h \cdot Pr_h) \qquad (4)$$

The cost C_{loss} associated to the losses in the inverter and in the electrochemical reactions of the battery, is calculated considering an overall charge/discharge efficiency $eff \in [0,1]$, as shown in Equation (5).

$$C_{loss} = \Sigma_h \left(\sqrt{eff} \cdot abs(C_{source,h}) \right) \qquad (5)$$

The cost C_{lol} is relative to the ageing of the battery caused by the charge and discharge cycle. The ageing mechanism is particularly complex, strongly dependent on the Depth of Discharge (DOD) during the discharge, and is described in detail in [16]. In this work, the effect of the temperature is not taken into account and it is considered that the ageing speed increases of a value x $\in [1,2]$, at each doubling of the DOD. It is assumed also that the maximum number $nCycles$ of full cycles (DOD = 100%) that the battery can withstand is known. With these assumptions the cost C_{lol} can be calculated as shown in Equation (6), where B_C [€] is the initial cost of the battery. The sum on negative DOD is done to take into

account that the irreversible reactions contributing to the aging of the battery occur in the discharge phase of the cycle.

$$C_{lol} = \Sigma_h \, (-min(0, DOD_h))^x \cdot B_C \, / \, nCycles \qquad (6)$$

In this study, a gradient-descend optimisation algorithm has been used for minimising the cost function reported in (3). The algorithm, as shown in Equation (7) minimises the sum over the whole period T of the cost for the single horizon h calculated as in Equation (3), subject to the constraints reported in Equation (8) and Equation (9) and returning the optimal battery schedule

$$PLAN_{opt} = argmin(\Sigma_h(C_h(PLAN))) \qquad (7)$$

$$Pd_h \leq P_h \leq Pc_h \qquad (8)$$

$$SOC_{min} \leq SOC_h \leq SOC_{max} \qquad (9)$$

where: P_h [W], is the power exported or absorbed by the battery at horizon h, SOC_h, is the state of charge of the battery at horizon h, Pd_h, Pc_h [W], are respectively the maximum charge and discharge power of the battery that can change in relation to the situation of the network at horizon h, SOC_{min}, SOC_{max}, are respectively the minimum and maximum allowable state of charge of the battery.

C. PV-Battery interaction

Central to the definition of the problem is the relation between the PV plant and the battery. The system considered in this paper is composed by two different entities, where the battery offers a service to the PV producer on a day ahead contract. In this study it is assumed that the export capacity of the PV producer is limited to a threshold *thr* equivalent to the 50% of the peak rating of the PV plant and the exceeding power is either stored in the battery, or not produced. The mechanism is described below:

1) An initial optimal schedule for the battery PLAN(0) is calculated by solving the optimisation problem described in (7-9) without considering network constraints in Equation (8).

2) For each quantile α of the probabilistic forecasts an array of flexibility needed *FLEX = [F₁, ..., Fₜ]* is calculated as described in (10). Negative velues represent energy that must be absorbed by the battery, whilst positive values represent the maximum discharge allowed to the battery.

$$F_h(\alpha) = thr - q_{t+h/t}(\alpha) \qquad (10)$$

3) The corresponding battery schedule PLAN(α) is calculated solving the optimisation problems described in (7-9), imposing the flexibility FLEX(α) as the minimum (in absolute value) discharge power of the battery, as reported in (11), where the maximum discharge capacity of the battery *Pd* is also taken into account. If the total amount of energy of the flexibility demanded is greater than the capacity of the battery, a correction factor is calculated and applied to the flexibility request. The cost associated with the plan C(α) is calculated as the difference between the cost of the schedule PLAN(α) and the schedule PLAN(0), as described in (12)

$$Pd_h = min(F_h(\alpha), Pd) \qquad (11)$$

$$C(\alpha) = C(PLAN(\alpha)) - C(PLAN(0)) \qquad (12)$$

The expected profit EP_{PV} for the PV associated to the flexibility FLEX(h,α) is calculated as described in (13).

$$EP_{PV}(\alpha) = -\Sigma_h(min(0, F_h(\alpha)) \cdot Pr_h) \quad (1-\alpha) \qquad (13)$$

4) The expected profit for the system EP_{SYS} is then calculated as the difference between the expected profit $EP_{PV}(\alpha)$ and the disoptimisation cost of the battery C(α).

5) The steps 2-4 are then repeated for each quantile of the forecasts and finally the quantile α_{opt} corresponding to the maximum expected profit for the system is identified, along with the optimal schedule PLAN(α_{opt}).

6) The effective profit for the PV TP_{PV} is then calculated as in Equation (13), by considering the effective measured PV production Y instead of the forecasted production $q_{t+h/t}(\alpha_{opt})$.

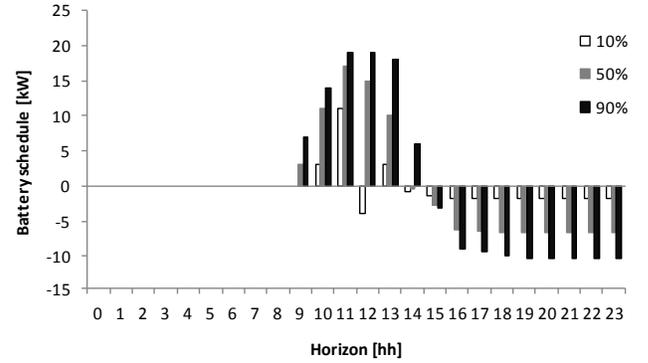A visual representation of the process described is presented in Figure 1 and Figure 2.

Figure 1: Battery schedule for the quantiles 10%, 50% and 90% of the forecasts
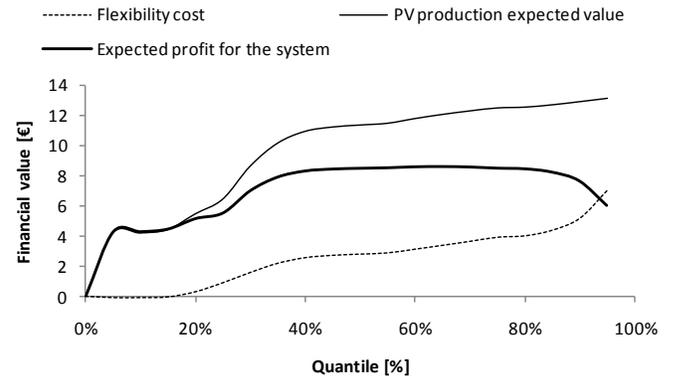

Figure 2: Cost benefit analysis of the flexibility requested at different forecast quantiles.

In Figure 1 the battery schedule for different request of flexibilities relative to the 1st, 5th and 9th decile of the probability distribution of the forecasts is represented. In the three cases the battery is charged in peak hours according to

the expected PV production and the energy is discharged in the later hours. The effect of the constant electricity prices during the day and the night can be seen in the uniform discharge pattern. In Figure 2 the probabilistic cost-benefit analysis performed at each time step is represented: for increasing values of the quantiles, the expected profit of the PV plant EP_{PV} (represented by the light continuous line) increases slowly because of the reduced probability associated to each additional energy level.

On the other side, the cost of the flexibility of the battery C is initially negligible, but increases quickly because of the nonlinearities in the battery ageing mechanism. The resulting profit EP_{SYS} has a characteristic bell shape and the maximum corresponds to the optimal plan maximising the efficacy of the combined PV-battery system.

IV.   RESULTS

A. Datasets

For this work, three main datasets are used: a time series reporting the historical production of a PV plant, the historical Nomerical Weather Predictions of solar radiation for the PV plant location from the ECWMF. The period considered ranges from June 2011 to October 2012. The problem is influenced not only by technical parameters such as the accuracy of the forecasts or the efficiency and lifespan of the battery, but also by financial parameters such as the electricity price and its variation during the day, the eventual incentive for photovoltaic energy and the purchasing cost of the battery. Although a complete sensitivity analysis is out of the scope of this paper, the results of the study have been reported for three different values of the cost of the battery. A summary of the parameters used for the simulations presented in this paper is reported in Table I. The three cases studied involve different values for the initial capital cost of the battery, relative for example to different battery technologies, such as lead-acid or lithium-ions. Other parameters such as the cost per cycle or the ratio between battery cost and electricity price could have been used.

Table I: Input parameters used in the simulation

Parameter	Unit	Value
Battery power rating	kW	100
Battery energy rating	kWh	100
Battery cost, Case A	€/kWh	1000
- - - - Case B	€/kWh	500
- - - - Case C	€/kWh	200
Battery efficiency	%	95
Battery max cycles	n	5000
Electricity Price Night (23:00-07:00)	€/kWh	0,0964
Electricity Price Day (07:00-23:00)	€/kWh	0,1391
PV production incentive	€/kWh	0,12

B. Evaluation of the PV production forecasts

The evaluation of probabilistic forecasts generally relies on a reliability / sharpness paradigm [17]. The forecasts

reliability may be define here as the adequacy between the nominal coverage α associated to some forecast quantile $q_{t+h|t}^{\alpha}$, and the true coverage Cov_h^{α} evaluated on a testing data set:

$$Cov_h^{\alpha} = \frac{\#\{p_{t_i+h} \leq q_{t_i+h|t}^{\alpha}\}}{N}; \ i = 1, \dots, N \qquad (14)$$

N being the test set sample size. Reliability is a prerequisite to satisfying forecasts. On the other hand, sharpness refers to the ability of a forecasting system to provide informative forecasts. Different forecasting systems should be compared based on sharpness, *once their respective reliability has been established*. Sharpness is sometimes evaluated through the provision of narrow centered prediction intervals (see [18] for instance).
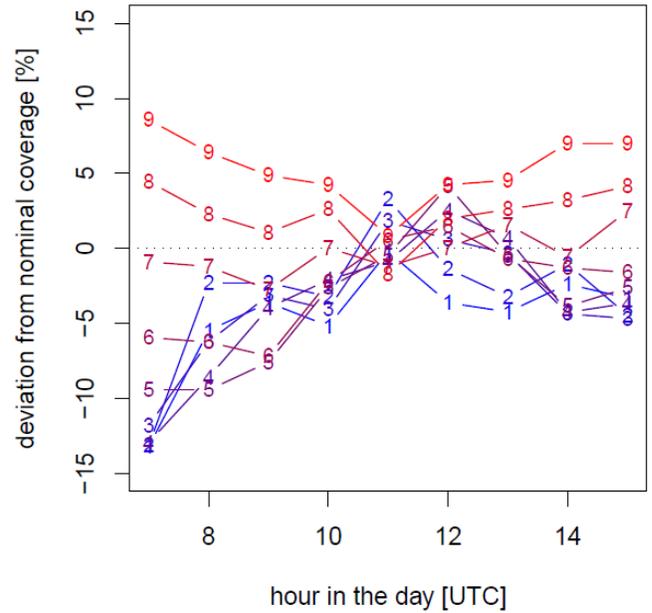

Figure 3: Reliability diagram showing the deviation between nominal and actual coverages associated to the forecast deciles of the PV production distribution, as a function of the hour in the next forecast day. The numbers in the plot indicate the distribution deciles. Colors close to red indicate higher deciles, while those close to blue indicate lower deciles.

In Figure 3 is shown the deviation between the nominal and actual coverages associated to the forecast deciles of the PV production distribution for the next day, considering a maximum training period size of 120 days. The results show that the distribution forecasts have higher reliability for midday hours, while the production is susceptible to reach its maximum value. If one excepts the forecast horizon associated to 11 am, one can notice that the proposed approach tends to underestimate quantiles associated to the upper tail of the distribution (positive difference between α and Cov_h^{α}), while overestimating those associated to the lower tail of the distribution (resp. negative difference). This tendency is amplified when only few data is available to calibrate the models. It results in more narrow centered prediction intervals. One has to be careful here not interpreting this result as a gain in forecasts sharpness, but more as a loss in forecasts reliability. In Figure 4 the mean absolute deviation between nominal and actual coverages associated to the forecast

quantiles of the PV production distribution, as a function of the training period size is shown. One can see that the highest gain in forecasts reliability is obtained with the first two weeks of data available for training. Further analysis showed that such an evolution in calibration actually depends on the considered distribution quantiles. Indeed, this pattern is particularly observed for quantiles in the upper tail of the distribution, while those in the lower tail did not show much improved reliability with increasing training period size.
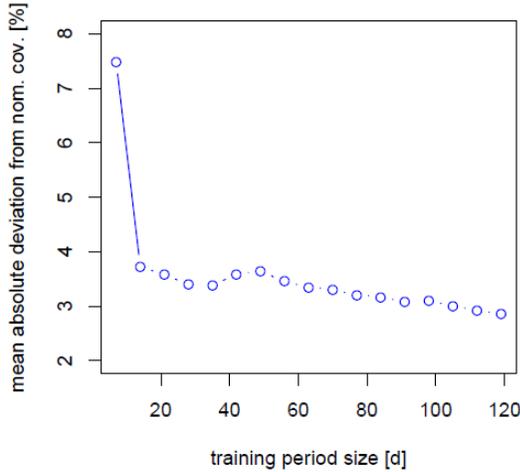

Figure 4: Mean absolute deviation between nominal and actual coverages associated to the forecast quantiles of the PV production distribution, as a function of the training period size.

C. Simulations

The simulations have been carried out applying the methodology described in Section III. In total the performance of the system was tested in 17 cases corresponding to different learning dataset sizes from 7 to 119 days of training. In each case the simulation spanned the 324 days of the training set, considering 19 quantiles for each forecast, equivalent to a percentage step of 5%. The relation between the performance of the learning dataset size, and hence the forecast precision, and the performance of the combined PV-Battery system is shown in Figure 5 and Figure 6 and the main findings are summarised below. From the analysis of Figure 5, for Case B and Case C, the performance in terms of effective revenue roughly doubles after 21 days of training, to become then almost constant, whilst in Case A, this improvement increases more slowly. In all the three cases, this is due mainly to the reduction of the battery management cost, thanks to more precise PV production forecasts. It is possible also to establish a relation between the quality of the forecasts used and the performance of the system. Case B will be used to illustrate this aspect. The initial increase in performance of the forecast algorithm, with a relative reduction of the 49% of the mean absolute deviation between nominal and actual coverages is responsible for a reduction of about the 25% in the operation cost of the battery and an increase of about the 62% of the effective revenue of the system. From the analysis of Figure 6 it is possible to analyse the impact of the initial battery cost on the utilisation of the battery and the quantiles used for the optimisation of the flexibility offered. The behaviour of the system is, as expected, strongly influenced by the financial

parameters of the problem, mainly the difference between night and daytime electricity price and the cost of the battery. The simulation shows that only in Case C, with a cost of the battery of 200€/kWh the system is able to be in profit with the electricity prices used. This is an important observation, since the high cost of the battery and its ageing push the optimisation algorithm to limit considerably its use in Case A and Case B.
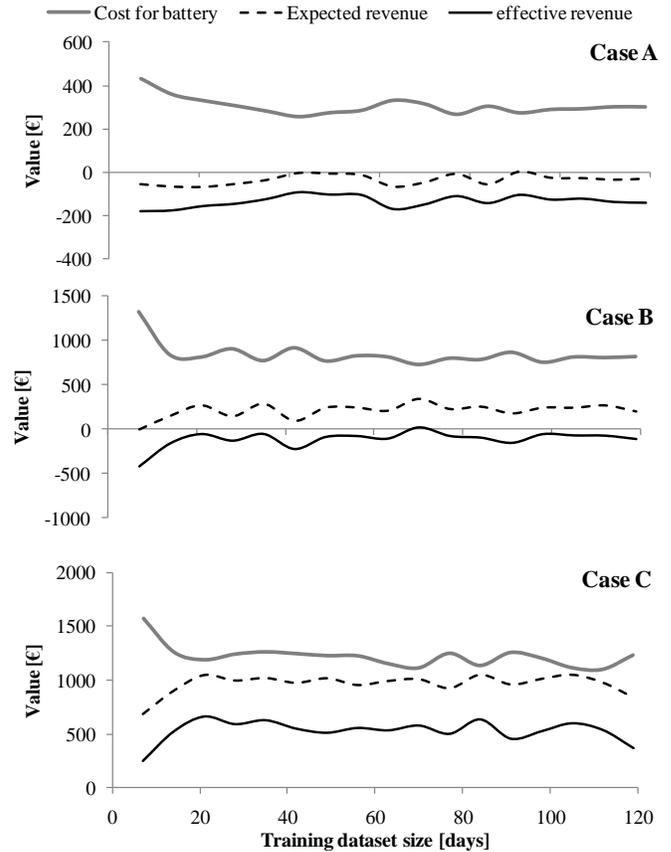

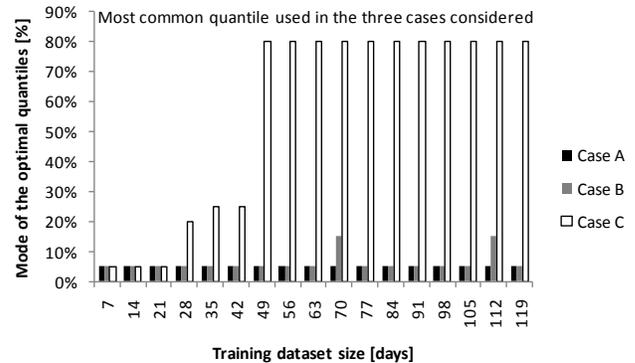Figure 5: Effect of the training dataset size on the performance of the combined PV-Battery system.

Figure 6: Mode of the optimal quantiles selected by the optimisation algorithm for the three cases considered.

The phenomenon is shown more clearly in Figure 7 where the cost benefit analysis described in Section III.C is

represented for Case B. In the chart it is possible to see how the cost of operating the battery becomes greater than the expected value of the PV production that can be saved for quantiles larger than 90%. It is also shown how the profit of the system reaches its maximum value around the quantile of 30%. In Case A in the same condition the two values are respectively 35% and 5%, whilst in Case C (shown in Figure 2), thanks to the lowest cost of the battery the optimum is around 65% and the cost of battery operation is never larger than the expected value of the PV production.
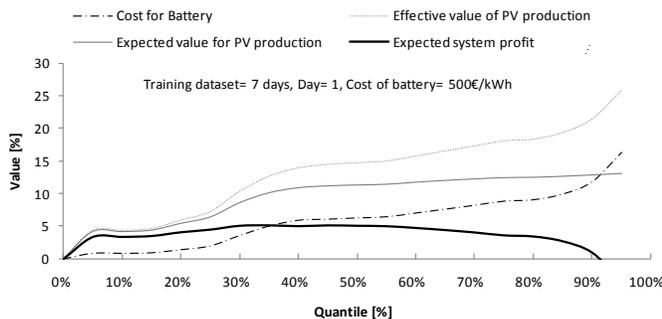

Figure 7: Cost benefit analysis of the flexibility requested at different forecast quantiles for the first day of the simulation using a traning dataset of seven days for Case B.

V. CONCLUSION

The paper has investigated the effect of uncertainty level in PV forecasts on the forecast-based optimal battery scheduling algorithm. Rather than considering artificially different uncertainy levels we consider the realistic case of forecasting models set-up for new PV plants that dispose little data history. This factor that affects accuracy was quantified in the paper and is expected to be frequent in real systems as new PV plants come into operation. The research shows that the performance of the forecasts increase quickly after about 1-2 weeks of training to become roughly constant and this behavior is repeated by the performance of the optimal battery scheduling algorithm, which increases of about the 50% after two weeks of training. This latter conclusion on the data history indicates a time buffer that could be considered for integrating new PV plants into the PV-battery optimization schedules. This is considered a positive result, since it shows that historically trained forecast algorithm can be used with success also after reduced training periods, at least in similar applications.

REFERENCES

[1] L. Ciabattoni, G. Ippoliti, S. Longhi, M. Cavalletti, M. Rocchetti, "Solar irradiation forecasting using RBF networks for PV systems with storage", *Industrial Technology (ICIT), 2012 IEEE International Conference on*, pp.699-704

[2] W. Fei, M. Zengqiang, S. Shi, Z. Chengcheng, "A practical model for single-step power prediction of grid-connected PV plant using artificial neural network", *Innovative Smart Grid Technologies Asia (ISGT), 2011 IEEE PES*, pp.1-4,

[3] M. Hassanzadeh, M. Etezadi-Amoli, M. S. Fadali, "Practical approach for sub-hourly and hourly prediction of PV power output", *North American Power Symposium (NAPS)*, 2010, pp.1-5,

[4] T. Cai, S. Duan, C. Chen, "Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement", *Power Electronics for Distributed Generation Systems (PEDG), 2010 2nd IEEE International Symposium on*, pp.773-777

[5] S. J. Kazempour, M. P. Moghaddam, "Economic viability of NaS battery plant in a competitive electricity market", *Clean Electrical Power, 2009 International Conference on*, pp.453-459

[6] Y. Riffonneau, S. Bacha, F. Barruel, "Gestion des flux énergétiques dans un système photovoltaïque avec stockage connecté au réseau   ", (2009)

[7] M. Urbina, L. Zuyi, "A Fuzzy Optimization Approach to PV/Battery Scheduling with Uncertainty in PV Generation", *38th North American Power Symposium,NAPS 2006*, pp.561-566

[8] K-H. Ahlert, C. van Dinther, "Robustness of Scheduling Algorithms for Distributed Storage Systems", *Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI), 2010*, Gottingen, Germany, pp.:2195-2206

[9] Bossavy, A. Michiorri, R. Girard, G. Kariniotakis,; "The impact of available data history on the performance of Photovoltaic generation forecasting", CIRED, 10-13 June 2013, Stockholm

[10] T. Gneiting, "Quantiles as optimal point forecasts", *International Journal of Forecasting*, vol 27(2), pp 197-207, 2011

[11] P. Bacher, H. Madsen, H. Aa. Nielsen, "Online short-term solar power forecasting", *Solar Energy*, vol 83(10), pp 1772-1783, 2009

[12] J.K. Moller, H.Aa. Nielsen, H.Madsen, "Time adaptive quantile regression", *Computational Statistics & Data Analysis*, vol 52(3), pp1292-1303, 2008

[13] R. Koenker, *Quantile Regression,* Cambridge University Press, 2005

[14] N. Meinhausen, "Quantile Regression Forests", *Journal of Machine Learning Research,* vol 7, pp 983-999, 200

[15] L.M. Costa, G. Kariniotakis, "A Stochastic Dynamic Programming Model for Optimal Use of Local Energy Resources in a Market Environment," in *Power Tech, 2007 IEEE Lausanne*, pp.449-454

[16] P. Rong, M. Pedram, "An analytical model for predicting the remaining battery capacity of lithium-ion batteries," Very Large Scale Integra*tion (VLSI) Systems, IEEE Trans*actions on , vol.14, no.5, pp.441-451, May 2006

[17] T. Gneiting, F. Balabdaoui, A.E. Raftery, "Probabilistic forecasts, calibration and sharpness", *Journal of the Royal Statistical Society*, Series B, Statistical Methodology, vol 69, pp 243-268, 2007

[18] P. Pinson, H.Aa. Nielsen, J.K. Moller, H. Madsen, G. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: required properties and evaluation", *Wind Energy*, vol 10(6), pp 497-516, 2007