© 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Forecasting Issues for Managing a Portfolio of Electric Vehicles under a Smart Grid Paradigm

R.J. Bessa and M.A. Matos, *Member, IEEE*

***Abstract*—In order to participate in the electricity market, electric vehicles (EV) need to be aggregated by a market agent, since the current rules do not allow the participation of small loads. The EV aggregator purchases electrical energy for charging its clients, and can offer reserve services. This activity requires forecasting methods for several variables. This paper presents a global view of the relevant variables for an EV aggregation agent participating in the electricity market and discusses the associated forecasting issues. The load forecast problem for direct and indirect control of the EV charging process is discussed. Variables from the market-side, such as reserve price and direction, are also addressed. Existing approaches are reviewed, discussed and tested according to different metrics.**

***Index Terms*—Electric vehicle, forecasting, electricity market, demand dispatch, demand response, reserve.**

I. Introduction

THE integration of flexible loads in the electricity market, in particular for supplying ancillary services, is gaining importance inside a smart grid paradigm with advanced metering and bidirectional communication [1][2]. The electric vehicle (EV) is a highly flexible load that when aggregated by a market agent can supply [3]. A discussion about the participation of EV in future electricity markets designed for accommodating this type of load is described in [4].

The EV aggregator acts as an intermediate entity between drivers, transmission system operator (TSO), distribution system operator (DSO) and the electricity market. This aggregation agent, depending on its business model, will control its clients' consumption rates with two possible modes: direct control by sending control signals to the charging process; indirect control by sending price signals to induce a certain reaction from the load. The first control mode is called *demand dispatch* [5], while the second is called *demand response* (DR).

In both cases, the aggregator needs to forecast the effective consumption from the EV under contract. Specialized forecasting algorithms are needed for each control mode.

The forecasted EV consumption is an input of optimization problems that derive the optimal purchasing bid for the electricity market. For an EV aggregator with direct control over the charging process, Wu *et al.* [6] described an algorithm for purchasing electrical energy at the lowest possible cost in the day-ahead market. For an aggregator with indirect control, Wu *et al.* [7] discussed pricing schemes for promoting the participation of EV in frequency regulation services.

The minimization of the purchasing cost also requires forecasts for the electrical energy price of day-ahead, intraday and real-time sessions. There are several publications about this topic [8][9].

Furthermore, for selling reserve services, in addition to the EV consumption and electrical energy price, forecasts of the direction of mobilized reserve (i.e., upward or downward) and the price of available and used reserve capacity, are very valuable for several optimization problems [10]-[12]. To our knowledge, only Jónsson [13] proposed models for forecasting the regulation direction in the Danish regulation power market. The goal was to use this information for computing the optimal selling bids of a wind farm [14]. In the present paper, similar variables are addressed, but related to the participation of EV in secondary and tertiary reserve markets.

In some DR programs, the bids for load reduction are only activated if the market price is above a threshold point [15]. Thus, in this DR markets it is useful to have a forecast for the market price divided by classes (i.e., price thresholds) or for price spikes, instead of a typical numerical forecast.

The possibility of forecasting these variables is discussed in this paper by presenting examples from real data and reviewing work from the state-of-the-art. The main contribution from this paper is a global view of the relevant variables for demand optimization in the electricity market, as well as pointing research directions for further work. Note that the EV is given as an example, but the ideas discussed in this paper can be generalized to other type of loads.

The paper is organized as follows: section II discusses the variables related with the demand-side; section III discusses the variables related with the electricity market; section IV presents the conclusions.

II. Variables on the Demand-Side

The EV aggregator will participate in the reserves market under the same rules as generating units. Thus, for defining bidding strategies, forecasts for the EV consumption are necessary. This section discusses forecasting methods for the *demand dispatch* and *demand response* modes.

The authors R.J. Bessa and M.A. Matos are with INESC TEC - INESC Technology and Science (formerly INESC Porto) and FEUP - Faculty of Engineering, University of Porto, Portugal (emails: rbessa@inescporto.pt, mam@fe.up.pt). The work of R. J. Bessa was supported by Fundação para a Ciência e Tecnologia (FCT) Ph.D. Scholarship SFRH/BD/33738/2009.

The vehicle-to-grid (V2G) mode is not considered in this paper. Instead, the reserve is supplied by considering a preferred operating point from which the charging rate is increased (downward reserve) or decreased (upward reserve) [5].

A. Controllable EV Consumption – Demand Dispatch

The objective of an aggregator with direct control over the charging rates, and participating in the electrical energy market, is the minimization of wholesale cost. The problem is solved with an optimization algorithm that places the charging in the hours with lowest prices [6][10].

However, it is necessary a forecast of how much electrical energy will be needed in the next hours or days. Forecasting the load is a usual task in problems related with power system operation and electricity markets. However, the problem is different for an aggregator with direct control. This means that the classical methods for forecasting the load in each time interval cannot be strictly followed because the aggregator is forecasting a variable that controls partially at the same time.

The alternative approach is to forecast the two variables illustrated in Fig. 1: the charging requirement and availability period. The EV availability is the time-period where the EV is plugged-in for charging. Note that the EV might be parked but not available for charging. The charging requirement of each EV is the total electrical energy needed to get from the initial (i.e., when the EV arrives for charging) state-of-charge (SOC) to the target SOC defined by the EV driver for the next trip, and including the losses from the charger. A charging requirement value is always associated to an availability period.

This forecasting problem resembles the intermittent demand forecasting for service parts inventories or retail store sales, addressed by the pioneer work of Croston [16] and further explored by other authors [17]. As in the EV case, the demand appears randomly and with periods of zero demand.

The forecasting algorithm for the EV problem consists in two steps: first, forecast the availability period of each EV, and then, the charging requirement for the forecasted period. The forecasting method was inspired by [18], and a detailed description is provided in [19].

The availability period is a binary time series that can be modelled and forecasted with a generalized linear model (GLM), where the response variable follows a binomial distribution [20]. For producing multi-step ahead, a binomial GLM is fitted for each look-ahead time step as follows:

$$p(y_t = 1 \mid y_{t-l}) = 1 \Bigg/ \left(1 + \exp\left(- \begin{pmatrix} \phi_0 + \phi_1 \cdot y_{t-1} + \\ \phi_2 \cdot y_{t-2} + \phi_3 \cdot y_{t-3} \\ + \phi_4 \cdot y_{t-48} + \phi_5 \cdot y_{t-336} \end{pmatrix} \right) \right) \quad (1)$$

$$p(y_{t+1} = 1 \mid y_{t-l}) = 1 \Bigg/ \left(1 + \exp\left(- \begin{pmatrix} \phi_0 + \phi_1 \cdot y_{t-1} + \\ \phi_2 \cdot y_{t-2} + \phi_3 \cdot y_{t-3} \\ + \phi_4 \cdot y_{t-47} + \phi_5 \cdot y_{t-335} \end{pmatrix} \right) \right) \quad (2)$$

$$\ldots$$

note that the lags related with the daily and weekly seasonal periods change with the look-ahead time steps, while the first three lags are fixed for each model. The output of the GLM is the posterior probability *p(y=1|.)* that is a function of lagged variables of the response variable *y*.

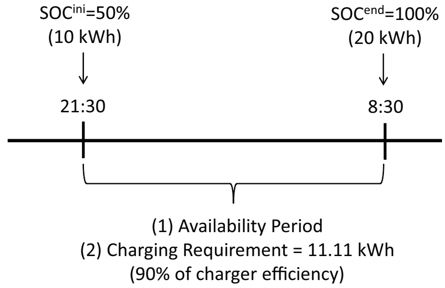

Fig. 1. Availability period and charging requirement of an EV.

For a time horizon with *k* time steps, it is necessary to fit *k* GLM models to each EV. This task, using non-linear models such as neural networks (NN) [22] or support vector machines (SVM) [23], demands a high computational effort. Thus, the use of simple and linear models is recommended, and if needed, generalized additive models (using splines) can be used for capturing non-linear relations between variables [24].

After forecasting the availability period, the corresponding charging requirement is estimated with non-parametric bootstrapping [20]. The bootstrap samples are conditioned to the number of hours the EV is plugged-in. For example, for the first time interval of the availability period, the bootstrapping technique only resamples from the historical consumption from the same time interval. The same process is followed for the subsequent hours. The sum of the bootstrap samples over the complete availability period gives the charging requirement forecast.

The output of the forecasting algorithm is illustrated in Fig. 2 where the forecast (in grey) and the realized value (in black) of the availability period for one EV in a 100 half-hours time horizon period are depicted. The EV time series are synthetic and generated with the method described in [25].

These forecasts show three different situations: in the first period, the forecasted departure instant is earlier; in the second period the forecasted arrival and departure instant do not match with the realized ones; in the final period the forecasted departure time is later than the realized one.

Based on the forecasted periods, the bootstrapping approach estimated a charging requirement of 11.52 kWh for the period between intervals 1 and 14 (the realized value was 17.03 kWh for a period 1-17), for the period between intervals 45 and 67 the estimated charging requirement was 11.36 kWh (the realized value was 11.15 kWh for a period 41-62), and for the period between 73 and 90 intervals the estimated charging requirement was 10.43 kWh (the realized value was 9.11 kWh for the period 73-86).

1) Case-Study Results

The forecasting methodology was tested in synthetic time series of 844 EV. These drivers only charge the EV at the end of the day and in slow charging points. Fig. 3 depicts the accuracy (for a 100 half-hours time horizon) of the availability forecast for each EV. The accuracy is computed as follows:

$$Accuracy = \sqrt{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}} \cdot 100 \qquad (3)$$

where TP is the number of correct plugged-in predictions (true positives), FN is the number of wrong zero predictions (false negative) and FP is the number of wrong plugged-in predictions (false positive).

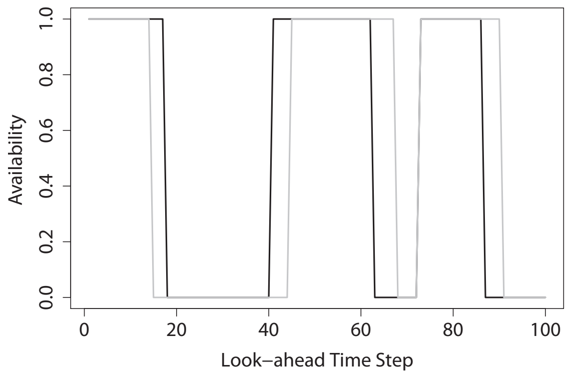

Fig. 2. Forecast and realized values for the EV availability (grey line is the forecast and the black line is the realized value).

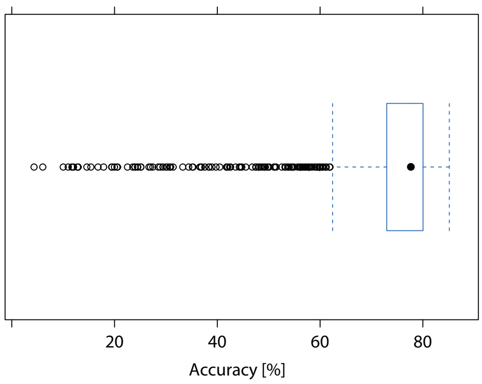

Fig. 3. Boxplot of availability forecast accuracy of a fleet with 844 EV. The boxplot have five statistics: lowest datum [within 1.5 Inter-quantile range (IQR)] of the lower quartile, lower quartile, median, upper quartile, and the highest datum (within 1.5 IQR) of the upper quartile. The outliers are also identified on the boxplot.

The accuracy on average is around 78%. However, some forecasts, that have in common a low number of hours with the EV plugged-in, present a low accuracy. For example, an EV driver with a forecast of accuracy equal to 4.31% is only plugged-in during 24.60% of a one-year period. For example, an EV with 80% of accuracy is plugged-in during 52.5% of the time. It is well known from the literature that unbalanced datasets are more difficult to predict. Thus, future work consists in developing methods for unbalanced binary time series.

To evaluate the quality of availability forecast of aggregated EV, the modified mean absolute percentage error (mMAPE) is used:

$$mMAPE = \frac{\sum_{j=1}^{N}\left(\left|y_j - \hat{y}_j\right|\right)}{\sum_{j=1}^{N}\left(y_j\right)} \qquad (4)$$

Fig. 4 depicts the mMAPE as a function of the aggregation size. For each aggregation size, the mMAPE is computed for different combinations of EV, and the results are depicted by a boxplot. As expected, the error decreases when the aggregation size increases. For 844 EV, the mMAPE is 13.64%.

For evaluating the charging requirement estimated with the bootstrapping technique, a perfect forecast for the EV availability is used. This removes the influence of the availability forecast errors. Fig. 5 depicts the boxplots for the mMAPE as a function of the aggregation size. For each aggregation size, the mMAPE was computed for different combinations of EV. The mMAPE decreases when the aggregation size increases, and its value is 8.26% for an aggregation size of 844 EV.

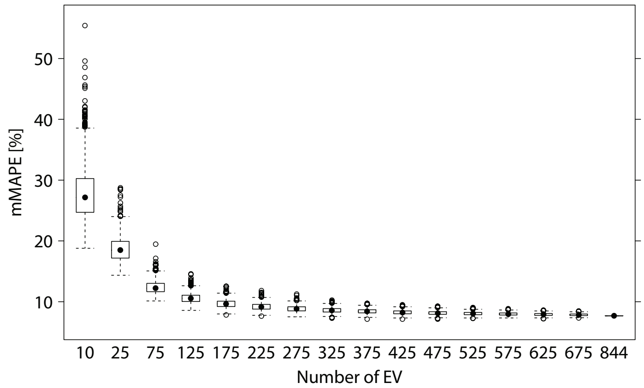

Fig. 4. mMAPE of the EV availability forecast with GLM for different aggregation sizes.

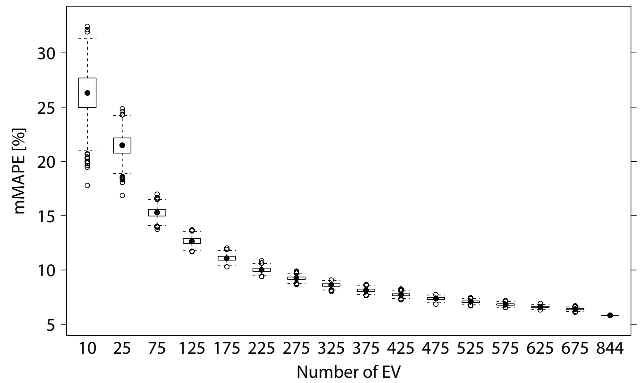

Fig. 5. mMAPE of the EV charging requirements forecast with bootstrap (and perfect availability forecast) for different aggregation sizes.

B. *Price Responsive EV Consumption – Demand Response*

In the DR control mode, the aggregator sends price signals for indirect control. The price signal can be the real-time price, or the forecasted day-ahead electricity market price, or any tariff value defined by the aggregator for getting a certain behavior from the EV. With the adequate price, the aggregator can induce the following behavior: load shift from high to low price periods; consumption reduction that can be offered as upward reserve; consumption increase that can be offered as downward reserve.

The idea of using price signals for controlling small loads and supplying services in the Danish regulation power market is being studied in the Flexpower project [26]. A one-way price signal that changes every five minutes is transmitted to the loads. For example, when upward reserve is needed a price signal with high price for consuming is transmitted, and for downward reserve the price is low. This framework was identified in [4] as a promising opportunity for EV.

In this problem, the aggregator needs to forecast, based on historical data, the impact of the price signal on the EV consumption. In other words, the aggregator needs to estimate the load as a function of the price signal to be defined. This task is complex because changes in one hour affect all the other hours. David [27] proposed the concept of cross-time price elasticity that might be used for assessing how the EV load will redistribute over a specific time-period in response to a price signal. Kirschen *et al.* [28] also explores similar concepts: self-elasticity and cross-elasticity. The self-elasticity relates demand change with the price in that interval, and cross-elasticity relates the demand change in one interval with respect to the prices in the other intervals.

The analytical calculation of these elasticity coefficients might be mathematically impracticable. Therefore, the approach should consist in using machine-learning algorithms for learning these relations from historical data. Khotanzad *et al.* [29] described an algorithm based on fuzzy logic for extracting rules relating the load behavior with the price signal.

Only with time-varying prices, or voluntary participation, it might be difficult to offer reserve with acceptable reliability, mainly because the client might not adjust its consumption based on the price signal. An alternative approach is to have the aggregator negotiating directly with the EV (that has installed software for negotiation) the provision of this service. The participation is mandatory, when there is agreement between aggregator and client. This resembles the direct control, however in this framework some EV will only participate if the price offered during the negotiation is attractive, while in the direct control is the aggregator who decides who participates.

III. Variables on the Market-Side

A. Reserve Direction

An important variable for defining the optimal bidding strategy for reserve services is the reserve direction. This variable informs on the probable direction of the reserve and based on this information the aggregator can define a combined strategy for participating in the electrical energy and reserve market. For example, if in a specific hour the probability of downward reserve probability is high, the aggregator can offer a bid with a very low quantity (or zero) in the electrical energy market and then offer the required electrical energy for charging as downward reserve.

The reserve direction consists in two binary time series, one for upward direction and another for downward. Two separate variables are considered because in a specific hour the reserve can be mobilized in both directions.

The objective of this section is to evaluate the forecast's feasibility for the secondary and tertiary reserve directions, using the Portuguese power system as a case-study[1]. The two reserve categories are the same used in Portugal and in the ENTSO-E (European Network of Transmission System Operators for Electricity).

For this task, four different algorithms are tested: GLM with the response variable following a binomial distribution, SVM, NN and naïve Bayes (NB) [30].

The following variables are candidates in the recursive feature selection algorithm from the R package "caret" [31]: lagged variables of the response variable, forecasted electrical energy price, forecasted wind power penetration[2], periodic function for the hour of the day and week day.

The day-ahead price was forecasted with an ARIMA model. The order of the model was selected with the R package "forecast" [32], and the model with minimum AIC (Akaike information criterion) was ARIMA(2,1,1)(7,0,7).

1) Results for Day-ahead Forecasts

First, the four algorithms are tested for day-ahead forecasts. The forecast for the 24 hours of day D+1 is produced at 10 AM of day D. The selected variables for the upward secondary reserve direction were:

- lagged response variables[3]: t-24,t-48, t-72;
- periodic function for the hour of the day.

Fig. 6 depicts a ROC (Receiver Operating Characteristic) curve [33] with the results for the upward secondary reserve obtained with the four algorithms. The diagonal line in the ROC curve corresponds to a method that randomly guesses a class (e.g., the flip of a coin). The x-axis (*specificity*) is the percentage of negative labeled instances (i.e., no upward reserve) that were predicted as negative, and the y-axis (*sensitivity*) is the percentage of positive labeled instances that were predicted as positive. The ROC curve depicts the trade-off between these two metrics. The lower left point (100, 0) corresponds to assume that when $p(y=1|x)>0$ the binary variable gets value 0; the upper right point (0, 100) assumes that when $p(y=1|x)<1$ the binary variable gets value 1.

This ROC curve of Fig. 6 shows that all the forecast models are very close to a random guess predictor, meaning that the predictor only extracts a small amount of information from the data. The results for the downward secondary reserve are not presented here, but the conclusions are similar.

The four algorithms were also tested for the tertiary reserve. The selected variables for the upward tertiary reserve direction were:

- lagged response variables: t-24, t-47,t-48, t-96,t-120, t-168;

[1] The market data can be downloaded from http://www.mercado.ren.pt

[2] The forecasted wind power penetration is the ratio between the forecasted wind power and load for the Iberian Peninsula, which can be downloaded from http://www.esios.ree.es

[3] The lags t-1, t-2 and t-3 are not considered for day-ahead forecasts because its influence in long-term horizons is negligible.

- periodic function for the hour of the day;
- periodic function for the day of the week;
- forecasted wind power penetration;
- forecasted day-ahead electrical energy price.

The curves for the tertiary reserve, depicted in Fig. 7 are much better compared to Fig. 6, since for the same specificity shows higher sensitivity. The results for the downward reserve are analogous.

These curves can be reduced to a single value, the Area Under the ROC Curve (AUC). AUC is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [33].

Table I presents the AUC of the four models, and for the upward secondary and tertiary reserve. This table shows that the best performance is from the NN for the secondary reserve, and from the NB for the tertiary reserve. Nevertheless, the GLM showed an acceptable performance in both reserves. In the secondary reserve, the SVM and NB presents an AUC below 0.5, this indicates a worthless model.

Table II presents the AUC for the downward reserves. In this case, the AUC for secondary reserve are not very close to 0.5, but still below 0.6, and this is considered a poor predictor. For the downward reserve, the GLM and NB present the best performance. The AUC values for the tertiary reserve are around 0.7, which is generally considered a reasonable predictor.

TABLE I
AREA UNDER AN ROC CURVE (AUC) FOR SECONDARY AND TERTIARY UPWARD RESERVE.

	Secondary Reserve	Tertiary Reserve
GLM	51.8	68.2
NN	54.5	61.8
SVM	47.8	57.9
NB	48.4	68.8

TABLE II
AREA UNDER AN ROC CURVE (AUC) FOR SECONDARY AND TERTIARY DOWNWARD RESERVE.

	Secondary Reserve	Tertiary Reserve
GLM	57.3	68.3
NN	57.4	67.8
SVM	57.2	61.1
NB	57.4	69.2

2) *Results for Hour-ahead Forecasts*

Fig. 8 depicts the ROC curve for a one hour-ahead forecast. This time horizon is important for markets where agents can present bids until 45 minutes before the operating hour (e.g., tertiary reserve market in Portugal and the regulation power market in the Nordpool).

As expected, the ROC performance is better than in Fig. 6 and 7. However, the improvement for the secondary reserve was not substantial. The AUC was 0.59 for the secondary reserve and 0.82 for the tertiary reserve.

The results presented in this section indicate that it is possible to forecast the tertiary reserve direction, and it is very difficult to extract information for the time series of secondary reserve direction. This result has a physical explanation: the secondary reserve, in general, handles random and uncorrelated variations in the load-generation balance, while the tertiary reserve (at least in Portugal) covers inter and intra hourly variations, less random and with higher magnitude (e.g., forecast errors, unplanned outages). For example, in [34] it is shown that tertiary reserve deployment is affected by the increasing wind power penetration, while secondary reserve is unaffected.

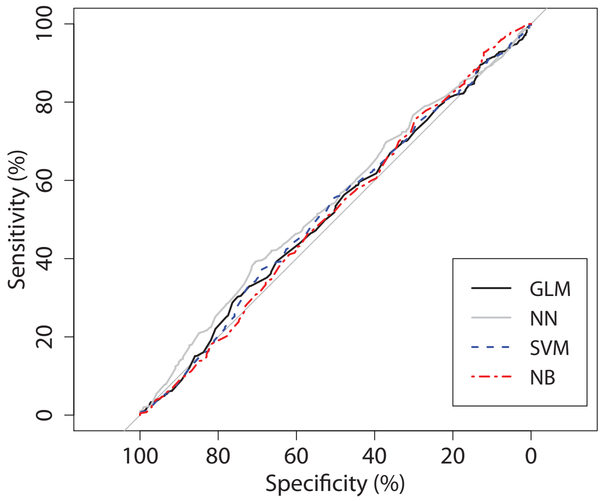

Fig. 6. ROC curve for secondary upward reserve direction day-ahead forecast obtained with GLM (generalized linear model), NN (neural network), SVM (support vector machines), NB (naive Bayes).

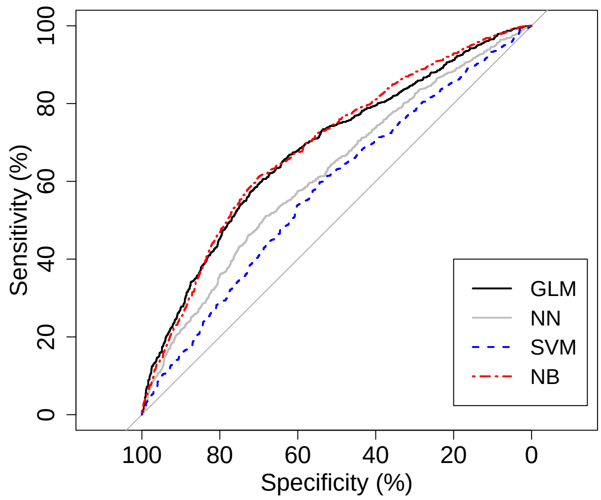

Fig. 7. ROC curve for tertiary upward reserve direction day-ahead forecast obtained with GLM (generalized linear model), NN (neural network), SVM (support vector machines), NB (naive Bayes).

B. *Price for Available Reserve Capacity*

The secondary reserve, in general, is paid by a capacity price (in €/MW) and by a price for used reserve capacity (in €/MWh). The tertiary reserve, in general, is only paid by used reserve capacity.

The price for used reserve capacity is only adressed in the next section. This section tackles the price for available reserve capacity, which is important for defining the hours where the aggregator should offer secondary reserve capacity.

To our knowledge, there are no publications about forecasting methods for this price.

For comparison, Fig. 9 depicts a boxplot for each hour with points of electrical energy price data from 2010 in Portugal. The average pattern resembles the load pattern, with low prices in valley hours, and high prices in peak hours. The boxplot for the reserve capacity price depicted in Fig. 10 shows a completely different pattern with some extreme price values (e.g., 180 €/MW) in valley hours. The high price values occur because there is a high concentration of market agents offering secondary reserve band. These two plots show a distinct behavior from both time series.

A forecasting exercise was conducted with the classical ARIMA models for the reserve capacity price.

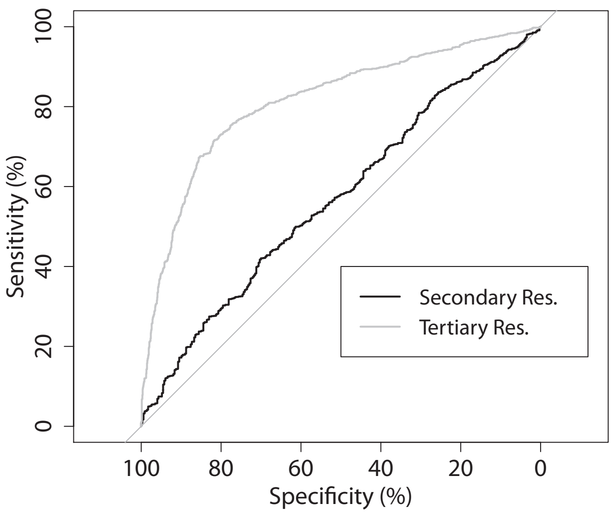

Fig. 8. ROC curve for a one hour-ahead forecast for the upward secondary and tertiary reserve.

1) Case-Study Results

The results in terms of MAE (mean absolute error) and RMSE (root mean square error) for day-ahead forecasts are presented in Table III. The ARIMA model for the reserve capacity price was selected using the R package "forecast" [32]. This secondary reserve is normally scheduled for the next day, so only the day-ahead time horizon is relevant.

The error of the reserve price forecast is lower when compared to the energy price, and its value is acceptable. This shows that these two prices might require different forecasting algorithms, but the forecast accuracy obtained with the ARIMA model for the reserve price is already satisfactory.

C. *Price for Used Reserve Capacity*

For deciding the hours where to make tertiary reserve bids the price for used reserve capacity is very important. For example, the aggregator forecast an electrical energy price of 25 €/MWh for the 4th hour and a price for used downward reserve of 10 €/MWh for the 9th hour. The obvious choice is to consume in the 9th hour as downward tertiary reserve. However, if the realized value was 30 €/MWh, the aggregator was better if it did not offered any reserve on 9th hour and used the 4th for consuming at the electrical energy price. This information for secondary reserve is also important, but in this case there is an associated income for having reserve capacity available in the 9th hour.

Fig. 11 depicts the boxplots for the price of used tertiary upward reserve capacity. Compared to Fig. 9 and 10, this time series presents a higher number of outliers and variability. Moreover, even in valley hours there are a large set of hours with high upward reserve prices.

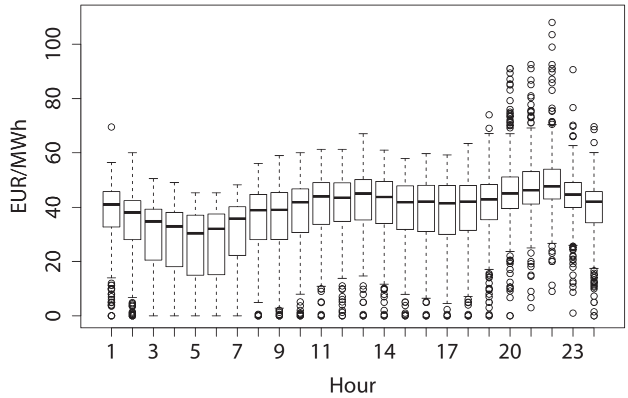

Fig. 9. Boxplot for the day-ahead electrical energy price in Portugal.

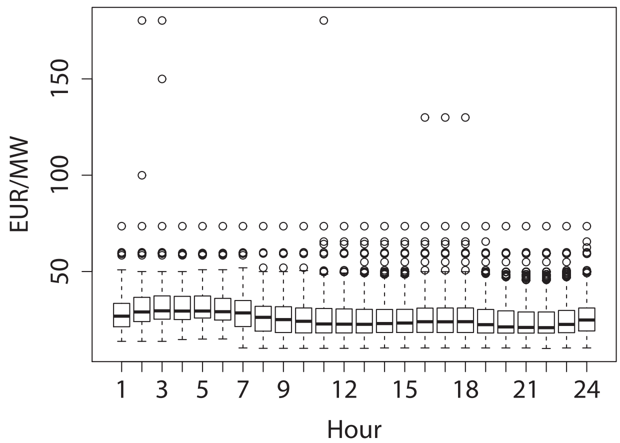

Fig. 10. Boxplot for the reserve capacity price in Portugal.

TABLE III
MAE AND RMSE ERROR OF THE RESERVE CAPACITY AND ENERGY PRICES.

	MAE	RMSE
Energy Price ARIMA(2,1,1)(7,0,7)	6.45	9.47
Reserve Capacity Price ARIMA(7,1,1)(7,0,6)	5.45	7.34

In contrast to the energy and reserve capacity price time series, this variable is an irregular time series because the price only exists when the reserve is used. The literature of algorithms for irregular time series is scarce, in particular for seasonal time series. To our knowledge, the only work about irregular seasonal time series is from Hanzák [35]. The author describes a modified Holt-Winters algorithm for dealing with irregular time series that uses a different representation for the seasonal component.

Fig. 5 depicts an illustrative day with one step-ahead forecasts and realized price of the upward tertiary reserve in

Portugal. Note that in some time intervals the tertiary upward reserve was not used, so there is no price for those intervals.

1) Case-Study Results

Table IV and V present the mean absolute error (MAE) and root mean square error (RMSE) for one-step and multi-step ahead forecasts of the upward and downward tertiary reserve prices in Portugal. The forecast error is significantly high, in particular for the multi-step ahead forecasts.

These results show that irregular time series with a high variability are difficult to forecast and new forecasting algorithms are needed for this type of series. Furthermore, it is conceivable that these prices are influenced by other variables, such as the load, wind power generation and electrical energy price. Thus, future work consists in developing multivariate models for irregular time series.

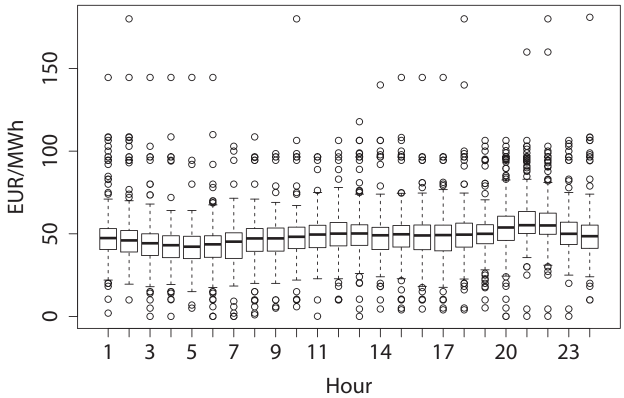

Fig. 11. Boxplot for the price of used reserve capacity in Portugal.

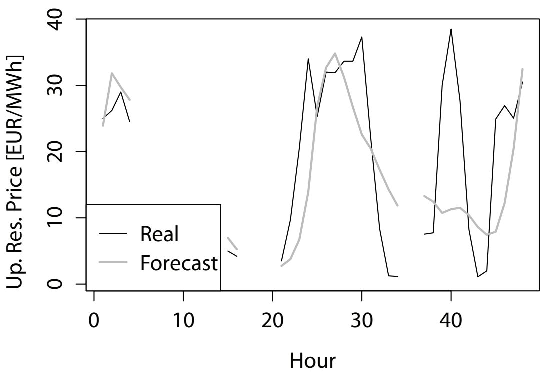

Fig. 12. One step-ahead forecast for the upward tertiary reserve price in Portugal.

TABLE IV
MAE AND RMSE ERROR OF THE TERTIARY UPWARD RESERVE PRICE.

	MAE [€/MWh]	RMSE
One step-ahead	8.21	13.32
Forecast for day D+1 made at 10th hour of day D (38 time steps)	12.98	17.76

TABLE V
MAE AND RMSE OF THE TERTIARY DOWNWARD RESERVE PRICE.

	MAE [€/MWh]	RMSE
One step-ahead	8.55	11.75
Forecast for day D+1 made at 10th hour of day D (38 time steps)	12.63	15.97

D. Price Thresholds and Spikes

In some DR programs instead of forecasting the electricity price magnitude, it might be more useful to forecast a class for the price [15]. For example, forecast the probability of having a price above 100 €/MWh. In these DR programs, with this information the aggregator can plan the charging process for offering a load reduction in situations where the forecasted price is above a certain threshold.

Furthermore, this forecast is also useful to avoid purchasing electricity in high price hours. For example, if the probability of having a high price in a specific hour is high, the aggregator does not submit a bid in that hour. This information is a complement to the numerical forecast for the price.

This consists in a classification problem that could have more than two classes. Zareipour *et al.* [36] addressed this problem and proposed a forecasting framework that combines feature selection algorithms and SVM. Huang *et al.* [37] enhanced this work with the following contributions: a new method for constructing the input variables, and a comparison between three feature selection algorithms and four machine-learning algorithms. The comparison showed that the best feature selection algorithm was the correlation-based. The machine-learning algorithms with best performance were the naïve Bayes and the k-nearest neighbor.

Huang *et al.* [37] applied the price classification forecasts to an industrial load. The load, if the electrical energy price is above a certain threshold (i.e. cost of producing electricity locally), starts a co-generation plant for satisfying the needs and sells the energy surplus. The results showed that the economic losses (with perfect forecast as reference) from using classical point forecasts are higher than a classification approach.

Zhao *et al.* [38] employed a similar approach to forecast price spikes occurrence and value. The proposed method combines a feature selection algorithm with SVM and probability classifier (based on the naïve Bayes algorithm). The results showed that SVM achieves the best performance; the classifier accuracy was 99.3%, against 98.9% of the probability classifier. Mount *et al.* [39] forecast price spikes with a regime-switching model, where the parameters are a function of time-varying variables.

The literature about this topic shows that if the goal is to avoid consuming electricity or placing bids at high price hours the chances of detecting price spikes are higher if a specialized approach (e.g., classification-based) is used, instead of the classical numerical forecast approach.

IV. CONCLUSIONS

This paper shows that in a smart-grid infrastructure and with active participation of EV in the electricity market, a set of load and market variables forecasting algorithms is needed.

The problem of EV load forecasting will gain new attention and new problems such as forecasting the reserve direction or irregular time series deserve more attention in the future. When designing optimization algorithms for an EV aggregator, it is necessary to take into account which variables can be forecasted (and represent additional value), and which variables cannot be forecasted with acceptable quality.

Furthermore, the additional economic value from using these forecasts in decision-making problems is the most important phase of this process.

Tests on existing approaches reveal the need for further improvements in some of these algorithms. For example, specialized algorithms for unbalanced binary time series (i.e., with a low frequency of 1) can improve the EV availability forecast. Furthermore, the forecasting literature is scarce in several problems, such as load forecasting algorithms for price responsive loads and multivariate forecasting algorithms for irregular seasonal time series.

The development of probabilistic forecasting algorithms would also be a valuable contribution for improving the market bidding algorithms.

REFERENCES

[1] D.S. Callaway and I.A. Hiskens, "Achieving controllability of electric loads," *Proc. of the IEEE*, vol. 99(1), pp. 184-199, Jan. 2011.
[2] G. Heffner, C. Goldman, B. Kirby, and M. Kintner-Meyer, "Loads providing ancillary services: review of international experience," Tech. Report, Lawrence Berkeley National Laboratory, May 2007.
[3] R.J. Bessa and M.A. Matos, "Economic and technical management of an aggregation agent for electric vehicles: a literature survey," *Eur. Tran. on Elect. Power*, vol. 22(3), pp. 334-350, Apr. 2012.
[4] C. Søndergren, N.C. Bang, C. Hay, M. Togeby, and J. Østergaard, "Electric vehicles in future marker models," Deliverable D2.3, Edison Project, Jun. 2011.
[5] A. Brooks, E. Lu, D. Reicher, C. Spirakis, and B. Weihl, "Demand dispatch," *IEEE Power Energy Mag*., vol. 8(3), pp. 20-29, 2010.
[6] D. Wu, D.C. Aliprantis, and L. Ying, "Load scheduling and dispatch for aggregators of plug-in electric vehicles," *IEEE Trans. on Smart Grid*, vol. 3(1), pp. 368-376, Mar. 2012.
[7] C. Wu, H. Mohsenian-Rad, and J. Huang, "Vehicle-to-aggregator interaction game," *IEEE Trans. on Smart Grid*, vol. 3(1), pp. 434-442, 2012.
[8] A.J. Conejo, M.A. Plazas, R. Espinola, and A.B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Trans. on Power Sys*., vol. 20(2), pp. 1035 -1042, May 2005.
[9] R. Weron and A. Misiorek, "Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models," *Int. J. of For.*, vol. 24(4), pp. 744-763, Oct-Dec. 2008.
[10] R.J. Bessa, M.A. Matos, F.J. Soares, and J.A. Peças Lopes, "Optimized bidding of a EV aggregation agent in the electricity market," *IEEE Trans. on Smart Grid*, vol. 3(1), pp. 443-452, 2012.
[11] Sekyung Han, S. Han, K. Sezaki, "Development of an optimal vehicle-to-grid aggregator for frequency regulation," *IEEE Trans. on Smart Grid*, vol. 1(1), pp. 65-72, Jun. 2010.
[12] E. Sortomme and M.A. El-Sharkawi, "Optimal charging strategies for unidirectional vehicle-to-grid," *IEEE Trans. on Smart Grid*, vol. 2(1), pp. 119-126, 2011.
[13] T. Jónsson, "Forecasting of electricity prices accounting for wind power predictions," MSc Thesis, Technical University of Denmark, 2008.
[14] M. Zugno, P. Pinson and T. Jónsson, "Trading wind energy based on probabilistic forecasts of wind generation and market quantities," *Wind Energy*, in press, 2012.
[15] R. Walawalkar, S. Blumsack, J. Apt, and S. Fernands, "An economic welfare analysis of demand response in the PJM electricity market," *Energy Policy*, vol. 36(10), pp. 3692–3702, 2008.
[16] J.D. Croston, "Forecasting and stock control for intermittent demands," Op. Res. Quart., vol. 23, pp. 289-303, 1972.
[17] R. Fildes, K. Nikolopoulos, S.F. Crone, and A. Syntetos, "Forecasting and operational research: a review," *J. of the Oper. Res. Soc.*, vol. 59, pp. 1150-1172, 2008.
[18] T.R. Willemain, C.N. Smart, and H.F. Schwarz, "A new approach to forecasting intermittent demand for service parts inventories," *Inter. J. of For.*, vol. 20, pp. 375-387, 2004.
[19] R.J. Bessa and M.A. Matos, "Global against divided optimization for the participation of an EV aggregator in the day-ahead electricity market – part I: theory," *Elect. Power Sys. Res.*, under review, Jun. 2012.
[20] S.N. Wood, *Generalized Additive Models: an Introduction with R*, London: Chapman and Hall/CRC, 2006.
[21] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*, New York: Chapman & Hall/CRC, 1994.
[22] B.D. Ripley, Pattern *Recognition and Neural Networks*, Cambridge: Cambridge University Press, 1996.
[23] V.N. Vapnik. *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.
[24] R.J. Hyndman, "Nonparametric additive regression models for binary time series," in *Proc. of the Aust. Meet. of the Econ. Soc., University of Technology*, Sydney, 1999.
[25] F.J. Soares, J.A. Peças Lopes, P.M.R. Almeida, C.L. Moreira, and L. Seca, "A stochastic model to simulate electric vehicles motion and quantify the energy required from the grid," in P*roc. of the 17th PSCC Conf.*, Stockholm, Sweden, Aug. 2011.
[26] C. Bang, F. Fock, and M. Togeby, "Design of a real time market for regulating power," FlexPower WP1 – Report 3, Version 2.0, Ea Energy Analyses, Dec. 2011.
[27] A.K. David, "Load forecasting under spot pricing," *Gen., Trans. and Dist., IEE Proceedings C*, vol.135(5), pp.369-377, 1988.
[28] D.S. Kirschen, G. Strbac, P. Cumperayot, and D. Mendes, "Factoring the elasticity of demand in electricity prices," *IEEE Trans. on Power Sys.*, vol. 15(2), pp. 612-617, 2000.
[29] A. Khotanzad, E. Zhou, and H. Elragal, "A neuro-fuzzy approach to short-term load forecasting in a price-sensitive environment," *IEEE Trans. on Power Sys.*, vol. 17(4), pp. 1273-1282, 2002.
[30] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, pp. 103–137, 1997.
[31] M. Kuhn, "Building predictive models in R using the caret package," J. of Stat. Soft., vol. 28(5), 26 pag., Nov. 2008.
[32] R.J Hyndman, S. Razbash, and D. Schmidt, "forecast: Forecasting functions for time series and linear models," R package version 3.19, 2012.
[33] T. Fawcett, "An introduction to ROC analysis," *Pattern Rec. Letters*, vol. 27, pp. 861-874, 2006.
[34] M. Milligan, P. Donohoo, D. Lew, E. Ela, B. Kirby, H. Holttinen, et al.," Operating reserves and wind power integration: an international comparison," in *Proc. of the 9th Annual Int. Work. on Large-Scale Integ. of Wind Power into Power Sys. and Trans. Net. for Offshore Wind Power Plants Conf.*, Québec, Canada; Oct. 18-19, 2010.
[35] T. Hanzák, "Holt-Winters method with general seasonality," *Kybernetika*, vol. 48(1), pp. 1-15, 2012.
[36] H. Zareipour, A. Janjani, H. Leung, A. Motamedi, and A. Schellenberg, "Classification of future electricity market prices," *IEEE Trans. on Power Sys.*, vol. 26(1),pp. 165-173, Feb. 2011.
[37] D. Huang, H. Zareipour, W.D. Rosehart, and N. Amjady, "Data mining for electricity price classification and the application to demand-side management," *IEEE Trans. on Smart Grid*, vol. 3(2), pp. 808-817, Jun. 2012.
[38] J.H. Zhao, Z.Y. Dong, X. Li, and K.P. Wong, "A framework for electricity price spike analysis with advanced data mining methods," *IEEE Trans. on Power Sys.*, vol. 22(1), pp. 376-385, Feb. 2007.
[39] T.D. Mount, Y. Ning, and X. Cai, "Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters," *Energy Economics*, vol. 28(1), pp. 62-80, Jan. 2006.

BIOGRAPHIES

Ricardo J. Bessa received the Electrical and Computer Engineering (five years degree) and Master degree in Data Analysis and Decision Support Systems from University of Porto, in 2006 and 2008. Currently, he is a researcher at INESC TEC and a PhD student of the Doctoral program in Sustainable Energy Systems (MIT Portugal) at university of Porto.

Manuel A. Matos (El. Eng., Ph.D., Aggregation, M'94) was born in 1955 in Porto, Portugal. He is with the Faculty of Engineering of the University of Porto (FEUP), Portugal, since 1978 (Full Professor since 2000). He is also coordinator of the Power Systems Unit of INESC TEC. His research interests include classical and fuzzy modeling of power systems, reliability, optimization and decision-aid methods.