Battery State-of-Charge Estimator Using the MARS Technique

Juan Carlos Álvarez Antón, *Member, IEEE*, Paulino José García Nieto, Francisco Javier de Cos Juez, Fernando Sánchez Lasheras, Cecilio Blanco Viejo, and Nieves Roqueñí Gutiérrez

***Abstract*—State of charge (SOC) is the equivalent of a fuel gauge for a battery pack in an electric vehicle. Determining the state of charge is thus particularly important for electric vehicles (EVs), hybrid EVs, or portable devices. The aim of this innovative study is to estimate the SOC of a high-capacity lithium iron phosphate ($LiFePO_4$) battery cell from an experimental dataset obtained in the University of Oviedo Battery Laboratory using the multivariate adaptive regression splines (MARS) technique. An accurate predictive model able to forecast the SOC in the short term is obtained and it is a first step using the MARS technique to estimate the SOC of batteries. The agreement of the MARS model with the experimental dataset confirmed the goodness of fit for a limited range of SOC (25–90% SOC) and for a simple dynamic data profile [constant-current (CC) constant-voltage charge–CC discharge].**

***Index Terms*—Lithium batteries, modeling, multivariate adaptive regression splines (MARS), nonlinear estimation, state of charge (SOC).**

I. Introduction

ONE of the most important battery parameters is the battery state of charge (SOC). The SOC is defined as the quotient between the difference of the rated capacity and the net amount of charge discharge from a battery since the last full SOC on the one hand, and the rated capacity on the other hand. Due to this definition, the full SOC is reached when the battery current drops below a predefined value at a constant charge voltage and constant temperature [1]. Determining and controlling the SOC is critically important in applications for electric vehicle (EV), hybrid EV, or portable devices. These applications require accurate measurement of battery SOC in order to give users an indication of available runtime. SOC estimation can also be useful to avoid detrimental situations, such as overdischarging and overcharging, which can lead to a reduction in battery life.

Several techniques have been reported for measuring or estimating the SOC of battery cells [1]. The most common methods include current integration techniques, artificial neural networks and fuzzy-logic-based estimations [2], support vector-based estimators [3], Kalman filter-based estimators [4]–[6], and others [7]–[10]. Most of these methods have been widely used and achieve acceptable results in different applications. The most common technique for calculating SOC is the Ampere hour counting or Coulomb counting. In this method, the current entering and leaving the battery must be measured periodically. Hence, SOC of the battery is updated by adding or subtracting the last period's net cumulative charge. The value of the current integral is thus a direct indicator for the SOC. One drawback of current integration techniques is their reduced accuracy. The major problem of this technique is that Coulomb counter is an open-loop SOC estimator [3] that only uses the current variable. Errors in the current detector are accumulated by the estimator. The longer the estimator is operated, the larger the cumulative errors become. For many SOC levels, this estimator carries an average $\pm 15\%$ error. Other methods exhibit greater performance. For instance, the extended Kalman filters reduce the average error for SOC predictions to 5% over a wide range of operation [3], but a high implementation cost is required due to their computational complexity.

This paper presents a novel SOC estimation method based on the multivariate adaptive regression splines (MARS) approach [11]. In this research study, the MARS approach is used to successfully predict the SOC of a high-capacity lithium iron phosphate ($LiFePO_4$) battery cell in an affordable way. The method is also successfully applied to predict the voltage of the same battery cell. Although MARS has been used for classification and regression in highly nonlinear systems, little work has been reported on applying MARS to estimate SOC. In the field of SOC estimation, the MARS technique can be designed to incorporate multiple variables datasets in real time (not only the battery current) with thousand of training data points and reduce them to a simple mathematical expression that can be manipulated by an inexpensive 8-bit microcontroller. If the MARS technique is correctly optimized, it might offer accuracy comparable with more sophisticated techniques at lower computational cost. It should be noted that a simple dynamic data profile [constant-current constant-voltage (CCCV) charge–CC discharge] is used in this study. As a result, for SOC levels between 25% and 90%, an average SOC prediction error less than $\pm$ 1% is obtained using the battery current, voltage, and temperature as independent

Manuscript received July 15, 2012; revised September 6, 2012 and October 27, 2012; accepted November 16, 2012. Date of current version January 18, 2013. This work was supported in part by the Spanish Science and Innovation Ministry, under Grant MICINN-10-IPT-370000-2010-15 and Grant AYA2010-18513. Recommended for publication by Associate Editor Dr. M. Ferdowsi.

J. C. Álvarez Antón and C. B. Viejo are with the Department of Electrical Engineering, University of Oviedo, 33204 Gijón, Spain (e-mail: anton@uniovi.es; cecilio@uniovi.es).

P. J. García Nieto is with the Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007 Oviedo, Spain (e-mail: pjgarcia@uniovi.es).

F. J. de Cos Juez and N. Roqueñí Gutiérrez are with the Department of Mining Exploitation and Prospecting, University of Oviedo, 33004 Oviedo, Spain (e-mail: fjcos@uniovi.es; nievesr@uniovi.es).

F. S. Lasheras is with the Department of Construction and Manufacturing Engineering, University of Oviedo, 33204 Gijón, Spain (e-mail: sanchezfernando@uniovi.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPEL.2012.2230026

TABLE I
HIGH-CAPACITY $LiFePO_4$ BATTERY CELL CHARACTERISTICS

Parameter	Value
Manufacturer Model Nominal capacity	CALB SE100AHA 100 Ah
Nominal voltage	3.2 V
Energy density	105wh/kg at 0.1C
Charging/dischar. cut-off voltage	3.6 V / 2.5 V
Recommend charging current Maximum discharge current (t<10s)	0.3 C 30 A 800 A
Cycle life (at 80% DoD, 0.3 C)	2000 times
Operating temp. (charge/discharge)	0ºC to 45ºC / −20ºC to 55 ºC
Shell material	Plastic
Weight	3.2 kg

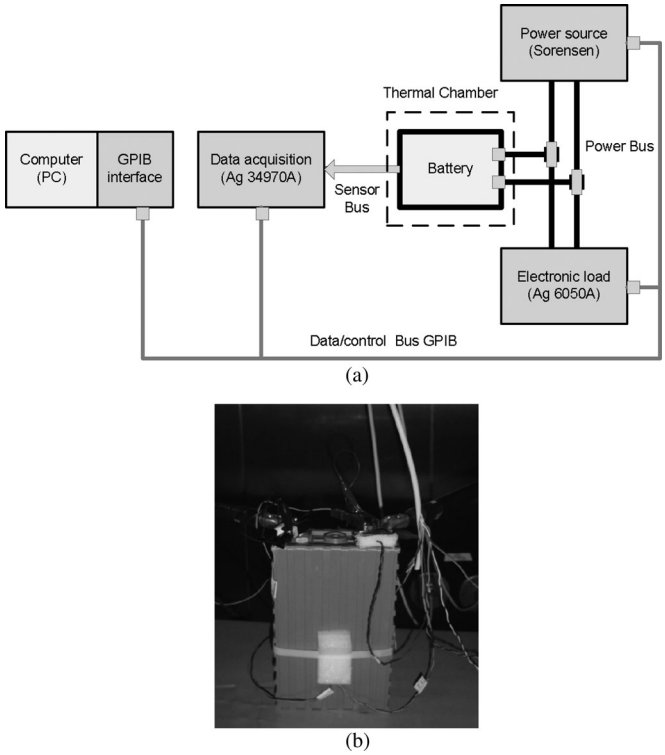

Fig. 1. (a) Test battery workbench block diagram. (b) $LiFePO_4$ battery cell.

variables. The accuracy obtained for SOC levels greater than 90% and less than 25% was 10% and 20%, respectively.

II. MATERIALS AND METHODS

A. Experimental Battery Cell and Test Workbench

Lithium-ion is the newest battery chemistry and is now well understood and widely used in numerous applications thanks to its high cell voltage, high energy density, long lifetime, and exceptional cyclability. A $LiFePO_4$ battery cell with a nominal capacity of 100 Ah is used in this study. The main cell characteristics are summarized in Table I.

The battery-testing platform consists of programmable stand-alone instruments and a computer-control unit using LabVIEW software. The test workbench block diagram is shown in Fig. 1(a). Stand-alone instruments are modular/independent

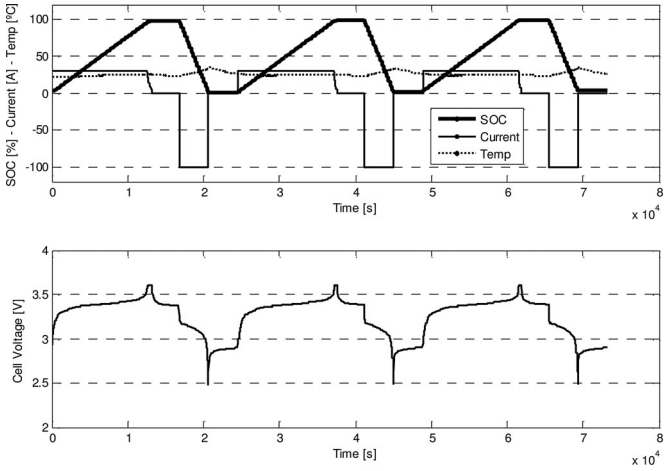

Fig. 2. Training data: cell current, voltage, temperature, and SOC versus time during three charging–discharging cycles. Discharge current is set at C rate (100 A).

instruments that support a wide variety of functions in the power and measurement field. In this case, a programmable power source (Sorensen DHP series, 60–220 M1-M9D, 0–60 V, 0–220 A) is used for cell charging. An electronic load (Agilent 6050 A with three 60504 B modules) is used for discharging purpose. Battery voltage, temperature, and other variables used for charging sensing termination [12] are monitored using a modular data acquisition system (Agilent 34970 A) with a total accuracy of 260 μV for dc measures. Battery cell temperature is measured using a resistance temperature detector attached to the center of the module [see Fig. 1(b)]. Internal power source and electronic load meters are used to provide current measurements with an accuracy of 0.12% $\pm$ 130 mA. All data measurements are taken at a sample rate of 5 s. Battery performance largely depends on temperature, so all tests commence at the initial temperature of 23 °C $\pm$ 2 °C in an electronically controlled environmental chamber. The workbench is controlled by a custom software application written in the National Instrument LabVIEW development tool. A description of the workbench and operating details can be found in [13].

B. Experimental Dataset

The most reliable test for determining the SOC of a battery is a discharge test under controlled conditions. This test includes a consecutive recharge. Three consecutive cycles of charging and discharging are thus performed using a CALB $LiFePO_4$ battery cell with a nominal capacity of 100 Ah. Cell variables (voltage, current, and temperature) are measured to identify the model parameters. Experimental data are obtained in two steps using the workbench described previously. First, the battery is charged at 0.3C constant current (CC) until 3.6 V. The voltage is then kept constant while the charging process continues in constant voltage mode until the charge current has dropped to 0.05C. Second, after a rest period, the battery is discharged at CC until reaching the cutoff voltage of 2.5 V. This process is repeated three times at different discharging rates, as shown in Figs. 2–4. The main goal of this research work is to obtain a

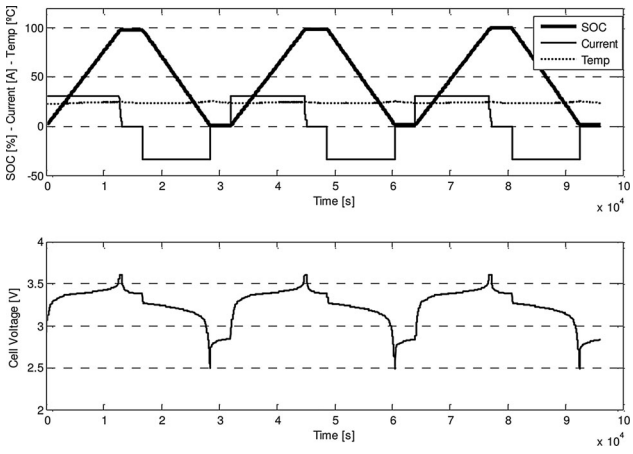

Fig. 3. Training data: cell current, voltage, temperature, and SOC versus time during three charging–discharging cycles. Discharge is performed at $C/3$ rate (33 A).

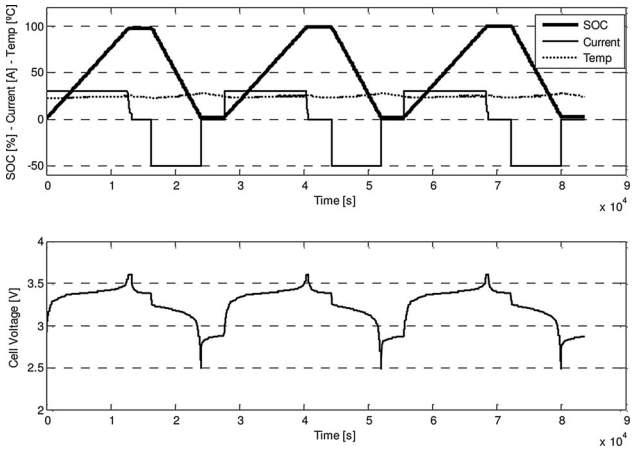

Fig. 4. Validation data: cell current, voltage, temperature, and SOC versus time during three charging–discharging cycles. Discharge is performed at $C/2$ rate (50 A).

model to estimate the SOC using the MARS technique. Figs. 2 and 3 show the SOC, current, voltage, and temperature ranges of the training data for the MARS model. Fig. 4 shows the dataset used to validate the MARS model.

The main difference between the data collection reported in Figs. 2–4 is the load that is used for battery discharge. The current used for battery discharge is fixed at C (-100 A) in Fig. 2, at C/3 (-33 A) in Fig. 3, and at C/2 (-50 A) in Fig. 4. The model training data cover operating from 100% SOC to 0% SOC and back up to 100% SOC. All data points in this SOC range are used for training. The training current goes from 30 to -100 A and 3 to -33 A (see Figs. 2 and 3, respectively). Validation data cover the same SOC range, but current goes from 30 to -50 A. Several plateaus with 0-A current and steady-state SOC are also included as training and validation data.

C. MARS Method

MARS is a form of regression analysis introduced by Friedman in [11]. It is a nonparametric regression technique and can be seen as an extension of linear models that automatically models nonlinearities and interactions between variables. The MARS model of a dependent variable $\vec{y}$ with M basis functions (terms) can be written as ([11], [14], and [15])

$$\hat{\vec{y}} = \hat{f}_M\left(\vec{x}\right) = c_0 + \sum_{m=1}^{M} c_m B_m\left(\vec{x}\right) \tag{1}$$

where $\hat{\vec{y}}$ is the dependent variable predicted by the MARS model, c_0 is a constant, $B_m\left(\vec{x}\right)$ is the mth basis function, and c_m is the coefficient of the mth basis functions. Linear basis functions can take the form of a constant, 1, a *hinge* function, or a product of two or more *hinge* functions in order to form nonlinear functions. *Hinge* functions take the form of $\max\left(0, x-c\right)$ or $\max\left(0, c-x\right)$, where c is a constant called *knot* and $\max\left(a, b\right)$ is a function that returns the largest element. The MARS approach takes the form of an expansion in product spline basis functions, where the number of basis functions as well as the parameters associated with each one (product degree and knot locations) are automatically determined by the data [11]. The MARS technique always converges to the same set of basis functions for the same initial dataset. Therefore, the MARS technique automatically models nonlinearities and interactions as a weighted sum of basis functions [11], [15]. In order to determine which basis functions should be included in the model, MARS employs the generalized cross-validation (GCV) method [15]–[17]. The model-building algorithm builds a model in two phases: forward selection and backward deletion. In the forward phase, the algorithm starts with a model consisting of just the intercept term (a constant) and interactively adds reflected terms of basis functions giving the largest reduction of training error. MARS then repeatedly adds basis functions in pairs to the model. At each step, it finds the pair of basis functions that gives the maximum reduction in sum-of-squares residual error. At the end of the forward phase, we have a large model, which typically over fits the data. Then, a backward deletion phase is engaged. In the backward phase, the algorithm prunes the model. It removes terms one by one, deleting the least effective term at each step until it finds the best submodel. At the end of the backward phase, from those best models of each size, one model of lowest GCV value is selected and considered as the final one. To fix ideas, the backward pass uses GCV to compare the performance of model subsets in order to choose the best subset: lower values of GCV are better. The GCV is a form of regularization: it trades off goodness-of-fit against model complexity. The raw residual sum-of-squares (RSS) on the training data is inadequate for comparing models, because the RSS always increases as MARS terms are dropped. In other words, if the RSS were used to compare models, the backward pass would always choose the largest model, but the largest model typically does not have the best generalization performance. GCV is the mean-squared residual error divided by a penalty that depends on the model complexity. The GCV for a model is calculated as follows [11], [14], [15], and [18]:

$$\mathrm{GCV}\left(M\right) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}_M\left(\vec{x}_i\right)\right)^2}{\left(1 - C\left(M\right)/n\right)^2} = \frac{\mathrm{RSS}/n}{\left(1 - C\left(M\right)/n\right)^2} \tag{2}$$

TABLE II
CELL INPUT VARIABLES USED FOR SOC PREDICTION

Input variables	Variable name
Cell voltage (V)	Voltage
Cell current (A)	Current
Cell temperature (°C)	Temp

TABLE III
LIST OF BASIS FUNCTIONS OF THE SOC MARS MODEL AND THEIR COEFFICIENTS

B_i	Definition	c_i
B_1	1	24.8765
B_2	max(0,Voltage−3.363)	168.2222
B_3	max(0,3.363− Voltage)	−97.9710
B_4	max(0,Current−0)	−1.1051
B_5	max(0,0− Current)	0.5118
B_6	max(0,Voltage −3.439)	−768.5929
B_7	max(0,0− Current)*max(0,Voltage −3.109)	5.5051
B_8	max (0,0− Current)*max(0,3.109− Voltage)	1.5432
B_9	max (0,Voltage −3.185)	332.6715
B_{10}	max (0,Current −0)*max(0,Voltage −3.353)	13. 9872
B_{11}	max (0,Current −0)*max(0,3.353− Voltage)	2.0514
B_{12}	max (0,0− Current)*max(0,Temp−27.843)	−0.1253
B_{13}	max (0,0− Current)*max(0,27.843−Temp)	0.0030
B_{14}	max (0,Current +33)*max(0,3.363−Voltage)	1.6340
B_{15}	max (0,Current −0)*max(0,Temp−25.735)	−1.1097
B_{16}	max (0,Current −0)*max(0,25.735−Temp)	−0.1075
B_{17}	max (0,Current −17.477)*max(0,Voltage −3.185)	−6.2946
B_{18}	max (0,17.477− Current)*max(0,Voltage −3.185)	2.2661

where the numerator is the mean-squared error of the evaluated model in the training data, n is the number of data cases in the training data, and $C(M)$ is the effective number of parameters. $C(M)$ is a complexity penalty that increases with the number of basis functions in the model and which is defined as [11], [14]–[16]

$$C\,(M) = (M+1) + dM \tag{3}$$

where M is the number of basis functions in (1), and the parameter d is a penalty for each basis function included in the model. It can be also regarded as a smoothing parameter. Therefore, the GCV formula penalizes not only the number of basis functions, but also the number of knots by means of the penalty parameter [18]. In this sense, the GCV formula adjusts (i.e., increases) the raw training RSS to take into account the flexibility of the model [11], [14]. In relation with the GCV penalty, the d parameter, theory [11] suggests values ranging from about 2 to 4. In our study, the parameter d equals 2, since this value provides the best result. The maximum interaction level of the basis functions is restricted to 2. This means that the predictive model can take the form of a product of two basis functions, as shown in Table III (see basis functions B_7, B_8 and B_{10} to B_{18}). These multiplicative terms are used to form simple nonlinear functions. These functions keep the computational complexity within reasonable bounds.

TABLE IV
EVALUATION OF THE IMPORTANCE OF THE VARIABLES THAT MAKE UP THE SOC MODEL

Variable	Nsubsets	*GCV*	*RSS*
Cell voltage	17	100	100
Cell current	16	80.80	80.79
Cell temperature	11	16.87	16.87

III. ANALYSIS OF RESULTS AND DISCUSSION

The cell input variables considered in this study are shown in Table II. The total number of predicting variables used to build the MARS model is 3.

A second-order MARS model is used in this study; the basis functions of the model hence consist of linear and second-order types and the maximum number of terms is not limited (no pruning). The model is trained using the dataset shown in Figs. 2 and 3. The results of the MARS model computed using all these data observations are reported in Table III, which shows a list of the 18 main basis functions of the MARS models and their coefficients. Considering (1), $c_0 = 0$, $c_1 = 24.8765$, $B_1 = 1$, $c_2 = 168.2222$, $B_2 = \max(0,\text{Voltage}-3.363)$, and so on (see Table III). The MARS model can be expressed in an explicit mathematical form as

$$\begin{aligned}\text{SOC}_i = {} & 24.8765 + 168.2221 * \max(0, \text{Voltage}_i - 3.363) + \cdots + \\ & + 2.2661 * \max(0, 17.477 - \text{Current}_i) \\ & * \max(0, \text{Voltage}_i - 3.185).\end{aligned} \tag{4}$$

Table IV presents an evaluation of the importance of the variables that make up the SOC model according to the criteria: number of model subsets in which each variable is included (Nsubsets), GCV, and RSS measured on the training data. According to the results shown in Table IV, the cell voltage and cell current are found to be the two most important variables in determining the SOC using this nonparametric technique. The slight preponderance of voltage over current derives from the static charging/discharging current used during the test cycles. In such a situation, the cell voltage explains the nonlinear relation between the input variables and the SOC estimated by the model better than the current. The cell temperature is the least significant variable in this study. A graphical representation of the terms that constitute the model can be seen in Fig. 5. Fig. 5 shows the graphical representation of the terms that constitute the SOC MARS model. After identifying the most relevant predictor variables (see Table IV), the next step is to get an idea of the dependence of the MARS model on each of them. Fig. 5 is a plot of the model's response when varying one predictor, the current in Fig. 5(a) and the voltage in Fig. 5(b), respectively (while holding the other variables predictors constant at their median values). Specifically, Fig. 5(a) plots the predicted SOC as current varies, with the other variables fixed at their median values. Analogously, Fig. 5(b) shows the predicted SOC as voltage changes, with the other variables fixed at their median values. Fig. 5(c) and (d) shows the model response when two variables (current and voltage; and temperature and current, respectively) are changing while holding others at their medians.

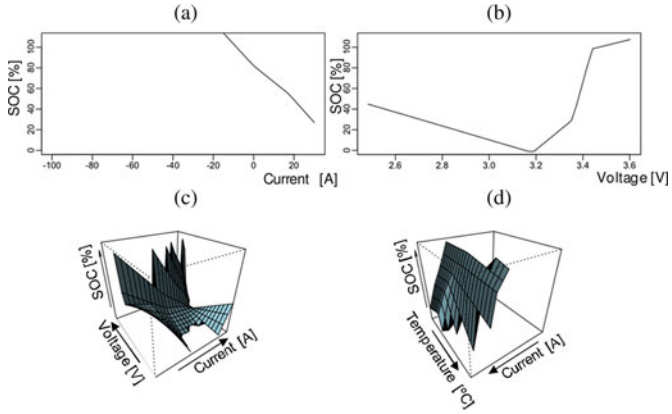

Fig. 5. Graphical representation of the terms that constitute the SOC MARS model. (a) First-order term of the cell *current* variable. (b) First-order term of the cell *voltage* variable. (c) Second-order term of the cell *current* and cell *voltage* variables. (d) Second-order term of the cell *current* and cell *temperature* variables.

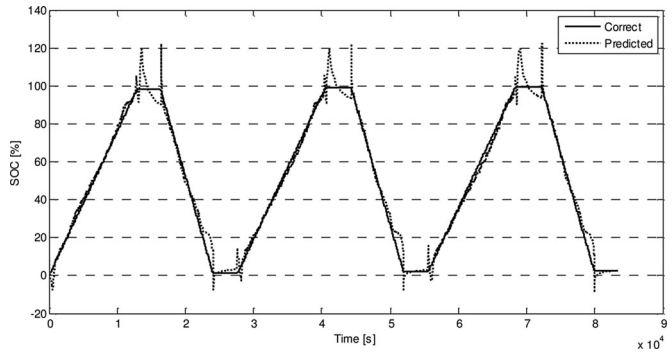

Fig. 6. SOC predicted by MARS versus correct SOC.

If the shape of the dependence on either variable is unaffected by the value of the other variable, this suggests that there is no interaction between these two variables.

An exhaustive cross-validation algorithm is used to guarantee the predictive ability of the MARS model. In this study, tenfold cross validation is used. Cross validation is a way to predict the fit of a model to a hypothetical validation set when an explicit validation set is not available. One round of cross validation involves partition, the dataset is randomly divided into l disjoint subsets of equal size, and each subset is used once as a validation set, while the other $l-1$ subsets are put together to form a training set. In the simplest case, the average accuracy of the l validation sets is used as an estimator for the accuracy of the method.

However, it is important to estimate how well a model performs on *new* data. Thus, in order to obtain a robust testing result, an explicit validation dataset is used in this study. The validation data are obtained from running a simple SOC test (see Fig. 4). The data in Fig. 4 are not used to train the model. The comparison between predicted SOC values using the validation dataset and observed SOC data is shown in Fig. 6. The estimation of "correct" SOC is accurately calculated in Fig. 6 using the internal current meters provided by the power supply (charge sensor) and the electronic load (discharge sensor). The results are significantly good for SOC predictions from 25% to 90%. The MARS approach is a highly accurate SOC estimator that maintains better than $\pm$ 1% accuracy over this range of operation. Additionally, there is no drift over time in SOC estimation by the MARS model. The MARS model estimates SOC with the largest error when the current is zero at full charge. Small peaks can be seen in the SOC prediction curve at the points that match the input current transient points. Note that the input current is constant except at the step-change points, where current jumps from one steady state value to another. From the mathematical point of view, these peaks correspond to points of discontinuity of the current function and reflect the peculiar manner in which the *hinge* functions behave at a jump discontinuity: a Gibbs-like phenomenon [19]. However, these peaks could be corrected using outliers cancellation. The peaks (outliers) appear with an SOC value greater than 100% and then can be easily identified and cancelled. These outliers represent a moderate problem in relation with the global model accuracy. Thus, for SOC levels greater than 90%, the model estimation carries a 10% error using outliers cancellation. For SOC levels lower than 25%, the model carries a 20% error.

The coefficient of determination R^2 was chosen in this research study to estimate the goodness of fit of the MARS model [15], [20]. This ratio indicates the proportion of total variation in the dependent variable explained by the model. A dataset takes values t_i, each of which has an associated modeled value y_i. The former are called the observed values and the latter are often referred to as the predicted values. Variability in the dataset is measured through different sums of squares:

1) $\mathrm{SS}_{\mathrm{tot}} = \sum_{i=1}^{n} (t_i - \bar{t})^2$: the total sum of squares, proportional to the sample variance;
2) $\mathrm{SS}_{\mathrm{reg}} = \sum_{i=1}^{n} (y_i - \bar{t})^2$: the regression sum of squares, also called the explained sum of squares;
3) $\mathrm{SS}_{\mathrm{err}} = \sum_{i=1}^{n} (t_i - y_i)^2$: the RSS.

In these sums, $\bar{t}$ is the mean of the n observed data

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i. \tag{5}$$

Bearing in mind the aforementioned sums, the general definition of the coefficient of determination is

$$R^2 \equiv 1 - \frac{\mathrm{SS}_{\mathrm{err}}}{\mathrm{SS}_{\mathrm{tot}}}. \tag{6}$$

A coefficient of determination value of 1.0 indicates that the regression curve fits the data perfectly. In this study, the fitted MARS model has a coefficient of determination equal to 0.98. The results confirm the accuracy of SOC prediction using MARS. Another advantage is that the MARS model can make predictions quickly due to the fact that the prediction function simply has to evaluate the MARS model formula shown in (1). This result means a minimal computation load. Moreover, data preprocessing is not needed to get the MARS model to converge. Additionally, the MARS technique can tackle noisy data, and it has a method for making the tradeoff between goodness of fit to the data and amount of structure in the solution. In

TABLE V
CELL INPUT VARIABLES USED FOR VOLTAGE PREDICTION

Input variables	Variable name
Cell SOC (%)	SOC
Cell current (A)	Current
Cell temperature (ºC)	Temp

TABLE VI
LIST OF BASIS FUNCTIONS OF THE VOLTAGE MARS MODEL AND THEIR COEFFICIENTS

B_i	Definition	c_i
B_1	1	3.2722
B_2	max(0,Current−0)	0.0091
B_3	max (0,0− Current)	−0.0020
B_4	max (0,SOC−17.2627)	$-4.85\text{x}10^{-05}$
B_5	max (0,17.2627−SOC)	−0.0282
B_6	max (0,Current −26.114)* max (0,SOC−17.2627)	−0.00032
B_7	max (0,26.114− Current)* max(0,SOC−17.2627)	$2.74\text{x}10^{-05}$
B_8	max (0,Temp−25.101)* max (0,SOC−17.2627)	0.00012
B_9	max (0,25.101−Temp)* max (0,SOC−17.2627)	$-6.97\text{x}10^{-05}$
B_{10}	max (0,Current −0)* max (0,17.2627−SOC)	0.00052
B_{11}	max (0,0− Current)* max (0,17.2627−SOC)	$8.16\text{x}10^{-05}$
B_{12}	max (0,SOC−98.5364)	0.0116
B_{13}	max (0,Temp−24.562)* max (0,SOC−98.5364)	0.0211
B_{14}	max (0,24.562−Temp)* max (0,SOC−98.5364)	−0.0066
B_{15}	max (0,Current −0)* max (0,Temp−25.746)	−0.0093
B_{16}	max (0,Current −0)* max (0,25.746−Temp)	0.0002
B_{17}	max (0,0− Current)* max (0,Temp−32.766)	−0.0028
B_{18}	max (0,0− Current)* max (0,32.766−Temp)	−0.0002
B_{19}	max (0,Current −0)* max (0,SOC−98.411)	−0.0001
B_{20}	max (0,Current −0)* max (0,98.411−SOC)	$-9.00\text{x}10^{-05}$

particular, MARS use GCV to make this task. In this sense, the backward phase of the MARS technique can be considered as a smoothing process, like a filter. This smoothing process improves numerical stability. This means that GCV reduces the error of the model and also maintains the error bounded even in the presence of noisy data. Furthermore, as the *hinge* functions automatically model portions of the input data, the effect of outliers is contained. As a consequence, the developed model is able to predict the SOC quickly and fairly accurately. All these features make the MARS model suitable for implementation on a low-cost microcontroller. The model can be easily modified to successfully fit data from different batteries, as MARS models are more flexible than linear regression models. Finally, MARS models are also simple to understand and interpret in the sense that it is possible to know the significance of each input variable in the overall model behavior.

To test the ability of the modeling technique, the same MARS method is applied to predict the cell voltage using the same high-capacity $LiFePO_4$ battery cell. Table V shows the model input variables used to predict the cell voltage. The results of the MARS model are reproduced in Table VI, which shows a list of the 20 main basis functions of the MARS models and their coefficients. A graphical representation of the terms that constitute the voltage model can be seen in Fig. 7.

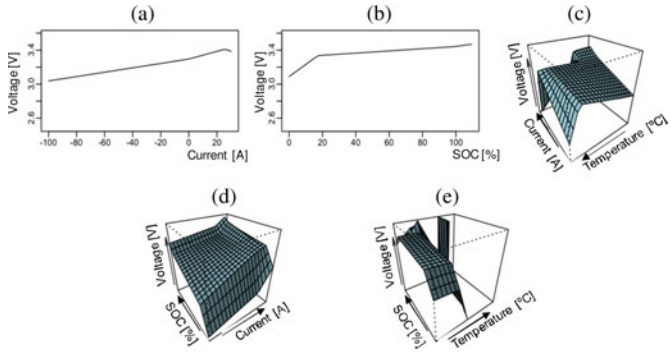

Fig. 7. Graphical representation of the terms that constitute the voltage MARS model. (a) First-order term of the *current* variable. (b) First-order term of the SOC variable. (c) Second-order term of the *current* and *temperature* variables. (d) Second-order term of the *current* and *SOC* variables. (e) Second-order term of the *temperature* and *SOC* variables.

TABLE VII
EVALUATION OF THE IMPORTANCE OF THE VARIABLES THAT MAKE UP THE VOLTAGE MODEL

Variable	Nsubsets	*GCV*	*RSS*
Current	19	100	100
SOC	18	74.654	74.64
Temperature	14	35.88	35.88

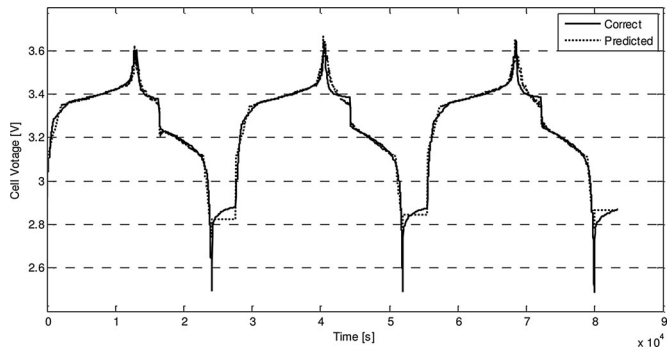

Fig. 8. Predicted and experimental cell voltage using the validation dataset.

Table VII presents the evaluation of the importance of the variables that make up the voltage model according to the criteria: number of model subsets in which each variable is included (Nsubsets), GCV, and RSS measured on the training data. According to the results shown in Table VII, the cell current is found to be the most important variable in determining voltage using this technique. The SOC is the second most influential input variable. The cell temperature is the third most significant variable in this study.

The voltage predicted by the MARS model and the experimental voltage using validation data are shown in Fig. 8. The fitted voltage has a coefficient of determination equal to 0.97. This result also confirms a goodness of fit, including the correct prediction of cell voltage during charging, discharging, and the rest period.

IV. CONCLUSION

The major findings of the study are as follows. First, this paper presents a new SOC estimation method based on the

MARS approach for a high-capacity $LiFePO_4$ battery cell. The proposed SOC estimator extracts model parameters, basis functions, and coefficients from battery charging/discharging testing cycles, using voltage, current, and temperature data as independent variables.

Second, a coefficient of determination as high as 0.98 is obtained when the MARS model is validated with external data not used for training. The results confirm the accurate prediction of SOC using a simple dynamic data profile (CCCV charge–CC discharge). Specifically, an accuracy of 1% was obtained for SOC levels between 25% and 90%. The accuracy obtained for SOC levels greater than 90% and less than 25% was 10% and 20%, respectively.

Third, a quick SOC prediction is also obtained, seeing as the prediction function simply has to evaluate the MARS model formula [see (1)]. Additionally, input data preprocessing is not required. The MARS model may thus be implemented on a low-cost microcontroller-based BMS system to fulfill the function of the SOC prediction.

Fourth, the MARS model is tested in a similar way to predict the cell voltage, obtaining accurate predictions with a coefficient of determination of 0.97.

Fifth, it is possible to establish an order of significance of the input variables involved in the prediction of both the SOC and the cell voltage. The MARS model may thus be more interpretable in the sense that it makes it clear what the effect of each predictor is.

Sixth, this innovative methodology can be applied to other battery types due to the flexible and intrinsic convergence mechanism of the MARS technique.

Finally, future works include supporting of stated MARS model using a complex dynamic data profile.

REFERENCES

[1] P. Sabine, P. Marion, and J. Andreas, "Methods for state-of-charge determination and their applications," *J. Power Sources*, vol. 96, pp. 113–120, 2001.

[2] C. Chan, L. E. Lo, and S. Weixiang, "The available capacity computation model based on artificial neural network for lead-acid batteries in electric vehicles," *J. Power Sources*, vol. 87, pp. 201–204, 2000.

[3] T. Hansen and C. J. Wang, "Support vector based battery state of charge estimator," *J. Power Sources*, vol. 141, pp. 351–358, 2005.

[4] G. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs—II: Modeling and identification," *J. Power Sources*, vol. 134, pp. 262–276, 2004.

[5] H. Dai, X. Wei, and Z. Sun, "Online cell SOC estimation of Li-ion battery packs using a dual time-scale Kalman filtering for EV applications," *Appl. Energy*, vol. 95, pp. 227–237, 2012.

[6] J. Kim, S. Lee, and B. H. Cho, "Complementary cooperation algorithm based on DEKF combined with pattern recognition for SOC/Capacity estimation and SOH Prediction," *IEEE Trans. Power Electron.*, vol. 27, no. 1, pp. 436–451, Jan. 2012.

[7] M. Verbrugge, "Adaptive, multi-parameter based state estimator with optimized time-weighting factors," *J. Appl. Electrochem.*, vol. 37, pp. 605–616, 2007.

[8] Y. Hu and S. Yurkovich, "Battery cell state-of-charge estimation using linear parameter varying system techniques," *J. Power Sources*, vol. 198, pp. 338–350, 2012.

[9] I. Kim, "Nonlinear state of charge estimator for hybrid electric vehicle battery," *IEEE Trans. Power Electron.*, vol. 23, no. 4, pp. 2027–2034, Jul. 2008.

[10] J. Kim, J. Shin, C. Chun, and B. H. Cho, "Stable configuration of a Li-ion series battery pack based on a screening process for improved voltage/SOC balancing," *IEEE Trans. Power Electron.*, vol. 27, no. 1, pp. 411–424, Jan. 2012.

[11] J. H. Friedman, "Multivariate adaptive regression splines (with discussion)," *Ann. Statist.*, vol. 19, pp. 1–141, 1991.

[12] J. C. Viera, M. Gonzalez, B. Y. Liaw, F. J. Ferrero, J. C. Alvarez Anton, J. C. Campo, and C. Blanco, "Characterization of 109 Ah NiMH batteries charging with hydrogen sensing termination," *J. Power Sources*, vol. 171, pp. 1040–1045, 2007.

[13] C. Carballo, M. Gonzalez, J. C. Alvarez Antón, and C. Blanco, "A computerized system for testing batteries in full controlled environment," in *Proc. 17th IEEE Instrum. Meas. Technol. Conf.*, Baltimore, MD, 2000, pp. 395–399.

[14] J. H. Friedman and C. B. Roosen, "An introduction to multivariate adaptive regression splines," *Statist. Methods Med. Res.*, vol. 4, pp. 197–217, 1995.

[15] P. J. García Nieto, F. Sánchez Lasheras, F. J. de Cos Juez, and J. R. Alonso Fernández, "Study of cyanotoxins presence from experimental cyanobacteria concentrations using a new data mining methodology based on multivariate adaptive regression splines in Trasona reservoir (Northern Spain)," *J. Hazard Mater.*, vol. 195, pp. 414–421, 2011.

[16] F. J. de Cos Juez, F. S. Lasheras, P. J. García Nieto, and M. A. Suárez Suárez, "A new data mining methodology applied to the modelling of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women," *Int. J. Comput. Math.*, vol. 86, pp. 1878–1887, 2009.

[17] V. Vapnik, *Statistical Learning Theory*. New York: Wiley-Interscience, 1998.

[18] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*. New York: Springer, 2009.

[19] A. J. Jerri, *Advances in the Gibbs Phenomenon*. New York: Sampling Publishing, 2011.

[20] D. Freedman, R. Pisani, and R. Purves, *Statistics*. New York: Norton, 2007.

Juan Carlos Álvarez Antón (M'08) was born in Veracruz, Mexico, in 1966. He received the M.Sc. degree in computer engineering from the University of Valladolid, Valladolid, Spain, in 1996, and the Ph.D. degree from the University of Oviedo, Oviedo, Spain, in 2007.

He is currently with the Department of Electrical Engineering, University of Oviedo, as an Associate Professor. His research interests include lighting, electronic instrumentation systems, and battery modeling.

Paulino José García Nieto was born in Oviedo, Asturias, Spain, in 1965. He received the Bachelor's degree in mining engineering with specialization in fuels and energy, and the Master's and Ph.D. degrees (overseen by Dr. J. M. F. Díaz) all in mining engineering from the University of Oviedo, Oviedo, Spain, in 1989, 1990, and 1994, respectively.

He was with the Department of Mathematics, University of Oviedo, where since 1996, he has been a Professor. He was involved in research on numerical simulation of the coagulation, condensation, and gravitational settling of atmospheric aerosols as well as the application of the finite element methods in numerous physical and engineering problems with success. He is currently involved in the application of the statistical learning and data mining to several biological, electrical, and physical datasets in order to obtain predictive models.

Francisco Javier de Cos Juez was born in Madrid, Spain, in 1972. He received the B.E. degree in electrical engineering from the University of León, Leon, Spain, in 1995, and the M.Tech. and Ph.D. degrees in electrical engineering from the University of Oviedo, Oviedo, Spain, in 1999 and 2003, respectively.

In 2000, he joined the Department of Mechanical Engineering, University of La Rioja, as a Lecturer. Since December 2001, he has been with the Department of Mining Exploitation and Prospecting, University of Oviedo, where he was an Assistant Professor and became an Associate Professor in 2011. His current research interests include artificial intelligence, mathematical modeling, and data mining.

Fernando Sánchez Lasheras was born in Oviedo, Spain, in 1975. He received the M.Sc. and Ph.D. degrees in industrial engineering from the University of Oviedo, Oviedo, in 2000 and 2008, respectively.

Since 2000, he has been with the aerospace and defense industry. In 2009, he joined the Department of Construction and Manufacturing Engineering, Oviedo University, as a part-time Instructor. His current research interests include artificial intelligence and data mining, with more than 20 technical publications.

Cecilio Blanco Viejo was born in Lada, Spain. He received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Oviedo, Oviedo, Spain, in 1989 and 1996, respectively.

In 1989, he joined the Department of Electrical and Electronic Engineering, University of Oviedo, where he is currently an Associate Professor. His research interests include high-frequency electronic ballast, discharge lamp, and battery modeling.

Nieves Roqueñí Gutiérrez was born in Santander, Spain, in 1964. She received the M.Sc. degree in mining engineering and the Ph.D. degree from the University of Oviedo, Oviedo, Spain in 1989 and 1995, respectively.

In 1991, she joined the Department of Mining Exploitation and Exploration, University of Oviedo, as an Assistant Professor, where she is currently an Associate Professor. Her research interests include environment, mathematical modeling, and data mining.