Journal of Power Sources 279 (2015) 694–706

Contents lists available at ScienceDirect

Journal of Power Sources

journal homepage: www.elsevier.com/locate/jpowsour

Review

From a novel classification of the battery state of charge estimators toward a conception of an ideal one

Jana Kalawoun [a, b, *], Krystyna Biletska [a], Frédéric Suard [a], Maxime Montaru [b]

[a] CEA, LIST, Laboratoire d'Analyse de Données et Intelligence des Systèmes, 91191 Gif-sur-Yvette, France
[b] CEA, LITEN, Laboratoire du Stockage ElectroChimique, 73375 Le Bourget-du-Lac, France

HIGHLIGHTS

- Ideal model provides reliable *SoC* for any battery type and cycling condition, online.
- None of the existing estimation methods offer an ideal *SoC* model.
- Novel classification facilitates the identification of to-be improved aspects.
- Methods using closed loop processing are promising candidates for ideal *SoC* model.
- Machine learning online techniques adapt the model's parameters when a drift occurs.

ARTICLE INFO

Article history:
Received 25 August 2014
Received in revised form
16 December 2014
Accepted 6 January 2015
Available online 14 January 2015

Keywords:
State of charge
Estimation
Modeling
Machine learning
Online estimation

ABSTRACT

An efficient estimation of the State of Charge (*SoC*) of an electrical battery in a real-time context is essential for the development of an intelligent management of the battery energy. The main performance limitations of a *SoC* estimator originate in limited Battery Management System hardware resources as well as in the battery behavior cross-dependence on the battery chemistry and its cycling conditions. This paper presents a review of methods and models used for *SoC* estimation and discusses their concept, adaptability and performances in real-time applications. It introduces a novel classification of *SoC* estimation methods to facilitate the identification of aspects to be improved to create an ideal *SoC* model. An ideal model is defined as the model that provides a reliable *SoC* for any battery type and cycling condition, online. The benefits of the machine learning methods in providing an online adaptive *SoC* estimator are thoroughly detailed. Remaining challenges are specified, through which the characteristics of an ideal model can emerge.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

At present time and in the foreseeable future, electrical batteries will continue to be used in real-time applications such as cell phones and laptops, hybrid and electrical vehicles, as well as in non real-time applications like energy storage systems.

The battery state of charge (*SoC*) is essential to calculate the autonomy and the available energy of the battery. An accurate *SoC* is fundamental to obtain an efficient control strategy to manage energy, as well as to guarantee a safe utilization of the battery by preventing under or over-charge that may lead to permanent damage. Energy management also plays a significant role in extending and optimizing the lifetime of the battery.

The battery being a complex electrochemical system, neither its remaining capacity nor its *SoC* can be directly measured. In addition, battery behavior depends on its utilization conditions like current profile, ambient temperature and state of health. Therefore one needs to develop a *SoC* estimation method, reliable and adaptable for real-time applications.

Two difficulties constrain the performances of a real-time *SoC* estimator. The first comes from the limited storage capacity and calculation resources of the Battery Management System (BMS). The second comes from the fact that the battery behavior depends on its technology and the cycling conditions.

Hence, we point out the need for an efficient model able to

* Corresponding author. CEA, LIST, Laboratoire d'Analyse de Données et Intelligence des Systèmes, 91191 Gif-sur-Yvette, France.
E-mail addresses: jana.kalawoun@cea.fr (J. Kalawoun), krystyna.biletska@cea.fr (K. Biletska), frederic.suard@cea.fr (F. Suard), maxime.montaru@cea.fr (M. Montaru).

http://dx.doi.org/10.1016/j.jpowsour.2015.01.038
0378-7753/© 2015 Elsevier B.V. All rights reserved.

estimate the *SoC* of any battery, regardless of its technology, under any cycling conditions in real-time contexts and applications. Such a model will be referred to hereafter as "ideal *SoC* estimator".

By taking a closer look at the existing methods, it is clear that none possess the characteristics of this ideal *SoC* estimator. In order to obtain it, a suitable approach must be identified among the large number of existing ones. Thus this identification can be achieved through a comprehensive classification of existing methods.

The *SoC* estimation methods can be classified with respect to different criteria. The first one is the nature of the input variables, either measured or estimated. The second one is the type of the *SoC* estimation model, which is a relationship between the input variables and *SoC*: physical, electrochemical or statistical regression model. The third criterion deals with the temporal dimension: static methods like those based on *SoC*−*OCV* lookup tables and methods able to provide a real-time *SoC* estimate. Also the methods can be classified according to the battery technology: Li-ion, Ni-MH, Lead-acid and so on. Finally, the classification can be made based on the mathematical tools used by the estimation method: Kalman filter, artificial neural network, fuzzy logic, etc. However it is important to distinguish between the tools applied to the *SoC* estimation and those used to estimate the input variables like *OCV* and electrical impedance. Indeed, the more the classification criteria are relevant, the more easily we can identify the methods that can be improved in order to provide an ideal *SoC* estimator and flesh out new ways of developing it.

Several reviews of the existing *SoC* estimation methods are available in the literature. The authors of [1–4] give an overview of the methods without classifying them. The drawbacks and advantages of each method are presented by the authors, but this is not sufficient to provide an exhaustive and well structured vision on the path to be followed to develop an ideal *SoC* estimator. Pop et al. [5] give a chronological review of the estimation methods before classifying them under three categories: direct measurement methods, book-keeping systems that involve basic and modified Ah-counting, and adaptive systems which are supposed to be self-designed and to adjust automatically following the battery agingaging and online changes in battery and user's behavior. Kalman filter, artificial neural network and fuzzy logic approaches were allocated to this category, but the authors acknowledge that these methods have some important limits and cannot be considered as adaptive to all cycling conditions.

Chang [6] gives a similar classification while adding to it a fourth category of hybrid methods, each corresponding to a combination of the first three categories.

Hence the classification of Pop [5] and Chang [6] doesn't make a distinction between the nature of *SoC* models and input variables, focusing the attention on the temporal and technological criteria.

Subsequently, the above classifications of the *SoC* estimation methods does not strictly abide by all earlier mentioned criteria, thus rendering difficult the careful examination of the aspects to be improved.

In this paper we introduce a novel classification of the *SoC* estimation methods based on their concept, their adaptability and their performances in real-time applications.

This novel classification shows the importance of machine learning methods in providing an ideal *SoC* estimator. This estimator is capable of providing precise *SoC* values in real-time configurations, and automatically adapts to the evolution of the battery behavior, all of this while being fully independent of the battery technology.

The rest of the paper is organized as follows. Section 2 recalls the definition of the battery state of charge and addresses the limitations of the classical definition. Section 3 introduces a novel classification of the existing *SoC* estimation methods. Section 4 gives an analysis of the most important aspects of these existing methods and study their ability to become a generalized *SoC* estimation method. Before concluding, a discussion of the characteristics of an ideal *SoC* estimator and the benefits of a machine learning approach in providing this ideal *SoC* estimator are conducted in Section 5. The conclusion sums up the findings of this paper and the challenges that remain to be addressed.

2. Battery state of charge

2.1. Definition of the state of charge

The state of charge of a battery is defined as the ratio between the available capacity and the reference capacity, which is the maximum capacity that can be withdrawn from the fully charged battery under reference conditions. The reference conditions are generally a constant current rate and a specific ambient temperature. A battery being a chemical energy storage system, there is no sensor that directly measures. These reference and available capacities must be calculated.

2.2. Challenges in estimating the battery capacity

One way to compute the battery capacity is the "discharge test". It consists of discharging the battery under reference conditions to reach the end of discharge criterion, i.e. the cutoff voltage.

However the discharge test cannot be applied in real-time application, as well as in off-line application as it leads to a loss of energy. The state of charge can be calculated based on the Ah-counting equation:

$$SoC_t = SoC_{t_0} + \frac{\int_{t_0}^{t} I_\tau \, \mathrm{d}\tau}{C_{ref}} \tag{1}$$

where SoC_{t_0} is the initial *SoC*, *I* the algebraic current measurement: positive for a charge current and negative for a discharge current and C_{ref} the reference capacity. A numerical implementation requires a temporal discretization, and then the *SoC* is calculated using the following formula:

$$SoC_t = SoC_{t-\Delta t} + \frac{I_t \times \Delta t}{C_{ref}}, \tag{2}$$

where Δt is the sampling interval, which can be constant or variable. It is clear that the precision of this method depends on the accuracy of the current sensor as well as on the sampling interval.

Nevertheless, the reference capacity is not constant during the battery charge/discharge; it depends on the state of health and the cycling conditions like the current profile and the ambient temperature. In a real-time context, the cycling conditions are uncontrolled as they depend on the user's behavior, weather conditions, road conditions, etc. Accordingly, in some situations, the state of charge can be lower than 0 or higher than 100. The establishment of a deterministic function to provide a reliable value of the reference capacity is a challenging problem.

3. Novel classification of the SoC estimation methods

From a global point of view every estimation method is characterized by its input variables, the *SoC* estimation model and the type of the *SoC* estimation processing, see Fig. 1.

The input variables can be either directly measured by a sensor, or estimated through a physical, electrochemical or statistical

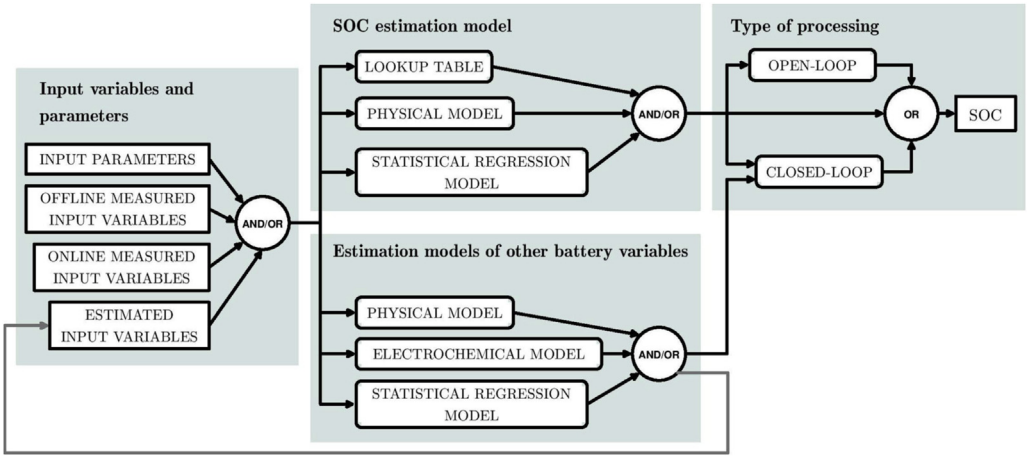


Fig. 1. Novel generalized classification of the *SoC* estimation methods based on the type of the *SoC* model, the nature of the input variables and parameters and the processing type.

regression model. The current is generally measured online. However, the *OCV* can be measured offline as well as be estimated online. Thereby the specification of the type of input variables is the first step when determining if a *SoC* estimation method can be improved to become ideal.

There are three types of *SoC* estimation models: a lookup table, a physical or a statistical regression model. The lookup table represents a direct relation between the *SoC* and a measured physical quantity like the *OCV*, internal impedance and electrolyte density. The physical model is based on the *SoC* definition; we can identify essentially two physical models: the macroscopic model where the *SoC* is calculated using Ah-counting and the microscopic model where the *SoC* is defined as the ratio between available and maximum concentration of active material in the anode. The statistical regression model can be a linear or non-linear relationship between the *SoC* and the input variables and does not necessarily have a physical interpretation. The type of the *SoC* model is the most important classification criterion, it is important not to confuse the *SoC* model with the models used to estimate other battery variables.

Finally, the *SoC* estimation methods can be classified by the type of processing: open- and closed-loop processing. The first one is based only on a *SoC* estimation model. The latter takes into account, in addition to the *SoC* estimation model, an estimation model for other battery variables. We consider that the use of a lookup table for the *SoC* estimation is not related to any category.

In the rest of this paper we present a review of methods organized in five subsections: the improved Ah-counting methods, the methods based on directly measured input variables, the methods based on inputs estimated with physical models, the methods based on inputs estimated with electrochemical models and the *SoC* estimation based on the machine learning techniques. There could as many families and/or sections as there are combinations of types of input variables, *SoC* models, and processing methods, see Fig. 1. We could have comprehensively listed all existing and/or possible families; and allocated a section of the article to each family. However, this leads to a size disparity between the sections; some parts would even be empty. Consequently we have structured the sections, except for Ah-counting, according to the type of the models of input variables as listed above. As a reading aid, at the beginning of each section, we have introduced a table summarizing the novel classification characteristics of the presented *SoC* estimation methods. A list of abbreviations and nomenclatures used in the paper is provided in Table 1.

4. State of charge estimation methods

4.1. Improved Ah-counting

The *SoC* definition given by (2) offers a generic *SoC* estimation method, called Ampere hour counting (Ah-counting), suitable for all battery types and used as a reference method for the evaluation of the performance of any other estimation method. Indeed given a reference capacity this method only requires a continuous measure of the current. This optimizes the volume of information exchanged between the battery and BMS and reduces the allocated memory space as well as the calculation time.

Despite its apparent simplicity, this method holds a number of limitations. Some of them have been successfully overcame, but some remain unsolved largely because the method does not consider the complexity of the electrochemical system. Below we discuss the sources of the imprecisions of model (2) and the improvements that were brought to it until now.

4.1.1. Calculation of SoC_{t_0}

The Ah-counting approach supposes that the initial value of the state of charge SoC_{t_0} is known, but even if the *SoC* is equal to 100% for a fully charged battery, and to 0% for a fully discharged battery, questions remain about the initial state of charge of a partially charged battery. Wang et al. [7] use a Kalman filter to estimate SoC_{t_0}. After a long rest period, SoC_{t_0} can also be estimated using a lookup table establishing a link between the open circuit voltage and *SoC* [8].

4.1.2. Current leakages in rest period

During the rest period, secondary reactions lead to current leakages and therefore reduce the remaining capacity. This phenomenon is known as self-discharge [9–11]. The rate of the self-discharge depends on the duration of the rest period, ambient temperature, state of charge and state of health [12,13]. The Ah-counting method does not take it into account and the *SoC* remains constant at zero current. Wang et al. [7] establish a linear relationship between the rate of the self-discharge and the rest period.

4.1.3. Coulombic efficiency

Undesired reactions, which lead to current loss, can also consume or produce charge at either the positive or negative electrode. The Coulombic efficiency measures the delivered/withdrawn current loss [14]. However its value depends essentially on

ambient temperature, current rate, state of health and even on the *SoC* of the battery [15]. The establishment of a deterministic function to provide the Coulombic efficiency remains a challenging problem. For an accurate estimation of the *SoC*, most algorithms include the Coulombic efficiency η in the formula of Ah-counting:

$$SoC_t = SoC_{t-\Delta t} + \frac{\eta \cdot I_t \cdot \Delta t}{C_{ref}} \quad (3)$$

Usually two values of the Coulombic efficiency are considered: η_c for the charge and η_d for the discharge [7,16,17]. Ng et al. [8] use a variable Coulombic efficiency depending on the *SoC*. Alzieu et al. [18] integrate the influence of the ambient temperature and the current rate on the Coulombic efficiency: $\eta = K_{\Delta T} \cdot K_{\Delta I}$ where $K_{\Delta T}$ and $K_{\Delta I}$ represent the influence of the variation of the temperature ΔT and the current ΔI respectively. Malkhandi [19] uses a fuzzy logic model to estimate the Coulombic efficiency based on current and temperature values.

4.1.4. Error of the current sensor

When integrating the current over time the Ah-counting does not use a feedback loop to offset an eventual estimation error. From a control theory perspective, the system is referred to as "open": the estimation error is accumulated over time and increases the bias of the estimator. This error is mainly a result of the inaccuracy of the current measurement which can be caused by the sensor error and the sampling frequency.

Calibration points are often used to adjust the estimated *SoC*. Hence, these points are available for specific states of the battery: 0% for a fully discharged battery, 100% for a fully charged battery and the *SoC* estimated by an open circuit voltage after a long rest period. However these calibration points are seldom available in a real-time application where the battery is often partially charged and no *OCV* measurement is available.

4.1.5. Variation of the battery capacity

The available capacity as well as the reference capacity change during the charge/discharge of the battery according to the current profile, temperature and state of health [20,21]. Gaddam et al. [22] use a fuzzy logic system to estimate the reference capacity based on the current rate and ambient temperature.

Accordingly the alteration of the available capacity affects not only the remaining capacity at instant t but also the one at $t-1$. This is a fundamental issue in the Ah-counting method considering that it is based on a recursive computation.

In summary, Ah-counting is a simple generic model suitable for all battery technologies and can be used in real-time applications. However this model does not take into account the calculation of the initial state of charge, the precision of the current sensor, the Coulombic efficiency or the variation of the reference capacity. Multiple improvements are developed to overcome these difficulties. Nevertheless they remain limited, especially for online *SoC* estimation, due to the complexity of the battery behavior which depends strongly on uncontrolled cycling conditions.

4.2. SoC estimator based on directly measured input variables

Another concept of *SoC* estimation is based on a direct relation between the *SoC* and a measured physical quantity. This relation can be described using a lookup table or a statistical regression model. Unlike Ah-counting this relation is not generic and depends on the battery technology.

A lookup table is a one-to-one relation obtained in an empirical way. Several laboratory experiments need to be realized in order to construct one table; various tables are required depending on the cycling conditions and the inherent battery properties. Here, an accurate indication of the *SoC* is determined by the precision of the measured physical quantity and by the quality and wealth of the lookup tables.

In the following we describe and analyze the *SoC* estimation methods taking as input variables a measure of open circuit voltage, impedance *Z* or other measurements specific to battery technologies. Table 2 summarizes these methods according to the novel classification, see Fig. 1.

4.2.1. Open circuit voltage measurement

The term voltage refers to the electrical potential difference between the two electrodes of an electrochemical cell. The open circuit voltage (*OCV*) is the measured terminal voltage for zero current [23]. Thus it gives an indication of the available energy and is directly proportional to the *SoC* [24–26]. Kelrich et al. [27] use *OCV* and temperature measurements as input parameters for a lookup table to obtain the *SoC* of the battery.

However the lookup table utilization/application remains limited due to several reasons. First, *OCV* cannot be directly measured in a real-time application and an accurate value of *OCV* requires the battery to rest an extended period of time [28]. Some techniques can be used to overcome the problem of the rest duration like the interpolation of the *OCV* data during rest time to get a higher resolution of the *OCV* curve, and the extrapolation of the relaxation behavior using a battery modeling [28]. The *OCV*/*SoC* relationship presents a hysteresis behavior according to the charge/discharge history even if the rest time is very long. This behavior is well known but needs specific experimental tests to be deeply characterized [29,30]. Second, this relation is not generic; it depends on the battery technology. For example, there is a slight change in a LiFePO4 battery *OCV* given $SoC \in (20,80)$. A small error on *OCV* induces a large error on the estimated *SoC*. Moreover the *OCV*/*SoC* relationship is not constant: it is influenced by the state of health of the battery [31] and temperature [27].

Hence, to cover all conditions, a lot of lookup tables should be built. This requires a lot of laboratory experiments and a large memory to store all these tables in a real-time application.

Table 1
List of abbreviations and nomenclatures used in the paper.

Abbreviations& nomenclatures	
SoC	State of Charge
OCV	Open Circuit Voltage
BMS	Battery Management System
EIS	Electrical Impedance Spectroscopy
EKF	Extended Kalman Filter
SVR	Support Vector Regression
ANN	Artificial Neural Network
I_t	Battery current at time t
U_t	Battery voltage at time t
T	Ambient temperature
Z	Battery internal impedance
R	Battery internal resistance
C_{ref}	Nominal battery capacity
Ah	Ampere hour
η	Coulombic efficiency
Δt	Sampling interval
x_t	State vector of a Kalman Filter
y_t	Observation vector of a Kalman Filter
u_t	Input vector of a Kalman Filter
z_t	Output vector of a machine learning model
r_t	Input vector of a machine learning model
w	Parameters of a machine learning model
D	Training dataset
$k(\cdot)$	Kernel function

Table 2
Main classification criteria for *SoC* estimator based on directly measured input variables.

Offline measured input variables	OCV, Z, U, electrolyte density
SoC estimation model	Lookup table, statistical model
Other models	–
Type of processing	–

4.2.2. Internal impedance measurements

The electrochemical impedance of a battery characterizes its dynamic behavior, that is, its response to an excitation of a small amplitude [32,33]. We distinguish two categories of methods for impedance measurement: active and passive methods.

The electrical impedance spectroscopy (EIS), the major example of the active methods, involves an excitation of the battery with a small AC signal over a wide frequency range typically from 10 kHz to 10^{-5} Hz. This frequency range depends on the battery chemistry [32,34,35]. This measurement method cannot be used in a real-time application due to hardware complexity and high cost. In addition, the response of the battery is strongly influenced by the temperature [36].

The passive methods search an impulse in the current profile. Once found, this current impulse and the corresponding output voltage are used to estimate the internal impedance. These methods are not reliable since a perfect impulse can hardly be detected in real conditions.

A review of the investigations for the applicability of impedance measurements as a test for the *SoC* of lead-acid and nickel–cadmium batteries is detailed by Huet [37]. These methods attempt to establish a relationship between the real and/or imaginary impedance components and the *SoC*. Thereby the developed models are not generic and remain specific to a particular technology. Huet concluded that the impedance measurement is strongly influenced by the temperature, particularly at low frequencies. This technique seems to be more suitable for lead-acid than nickel–cadmium batteries. Indeed for nickel–cadmium batteries, the ohmic resistance variations with the *SoC* are much lower than variations for lead-acid batteries.

Salkind et al. [38] use a fuzzy logic model to establish a relationship between the *SoC* and the impedance of the battery derived from impedance spectroscopy measurements at different frequencies. Thus the fuzzy logic model limits the effect of the unreliability of the impedance measurement.

Hence the measured values of the impedance depend greatly on the measurement method and they are sensitive to measurement conditions. Furthermore, the type of relationship between this impedance and the *SoC* is not generic for all battery technologies and difficult to establish in real-time contexts.

4.2.3. Other measurements

Some techniques are applicable for particular battery technologies. For example, the lead-acid battery presents a particular behavior called “coup de fouet” which is the initial voltage drop when discharging a fully charged battery [39]. It can be used to calibrate the *SoC* to 100% and provides information regarding the state of health of the battery.

A second specific method for the lead-acid battery is based on a direct relation between the *SoC* and the electrolyte density measurement. However this method is very sensitive to the temperature and the impurities present in the electrolyte [1].

All of the methods mentioned and analyzed above need a direct measurement of a physical quantity inherent to the battery to estimate the state of charge via a lookup table, statistical regression model or fuzzy logic model. In addition, the experimental conditions should be controlled and known. Nevertheless, their measurement is difficult if not impossible to carry out especially in a real-time context. To get around this difficulty, a chosen physical quantity can be estimated by means of a physical, electrochemical or statistical regression model, see Fig. 1.

4.3. SoC obtained from inputs estimated using physical models

The estimation of the input variables is usually based on a physical representation of the battery behavior. It consists of establishing a relationship between an input variable, such as the output voltage or the *OCV*, and the battery model parameters derived from its physical representation.

Two cases are generally considered by the *SoC* estimation methods when using estimated input variables. In the first one an input variable can be estimated by a physical model as well as measured in real-time conditions. The difference between the estimated and measured input variable is then used to improve the *SoC* estimation model via a closed-loop processing. In the second case, an input variable cannot be directly measured but is estimated in real-time by a physical model. The *SoC* is then estimated via a predefined relationship (lookup table or regression model) using the input variable.

In the rest of this Section we present several *SoC* estimation methods where the estimated input variables are either *U* or *OCV*. Table 3 summarizes the variables and models involved in these methods according to the novel classification, see Fig. 1.

4.3.1. Voltage estimation

The output voltage can be measured and estimated at the same time. This is why this input variable is very suitable for a closed-loop processing: the error between the measured and estimated output voltage is used in real-time to adjust the *SoC* obtained by the Ah-counting method. Several battery voltage models are described below, along with examples of closed-loop processing techniques, such as introducing a controller or a Kalman Filter.

Codeca et al. [40] improve the Ah-counting method using a controller. Fig. 2 shows a flow diagram of the *SoC* adjustment using a controller. This is described in the following. The voltage of the battery is estimated using the 2nd order Randle model. The difference between the estimated $\widehat{U}_t$ and the measured U_t voltage is then integrated in the feedback of the controller in order to adjust the Ah-counting value $\widehat{SoC}_t^{AH}$:

$$\widehat{SoC}_t = \widehat{SoC}_t^{AH} + K_p(U_t - \widehat{U}_t), \tag{4}$$

where K_p is the controller parameter.

More efficient than a controller, a Kalman filter takes into account the imprecisions of the voltage measurement and those of the *SoC* modeling. The Kalman filter estimates a dynamical unobserved state x_t given a suitable observed variable y_t:

$$x_{t+1} = A \cdot x_t + B \cdot u_t + w_t \quad \text{(State model)}$$

Table 3
Main classification criteria for *SoC* obtained from inputs estimated using physical models.

Estimated input variables	*U*	*OCV*
SoC estimation model	Ah-counting	Lookup table, regression model
Other models	Physical model of *U*	Physical model of *OCV*
Type of processing	Closed-loop	Closed & open-loop

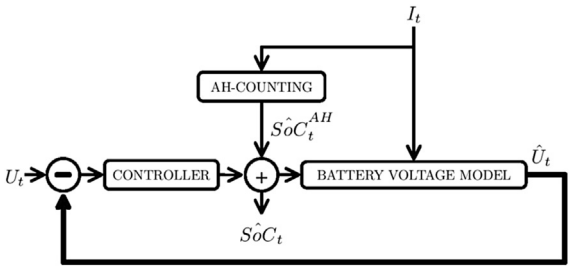


Fig. 2. Concept of *SoC* adjustment through a controller.

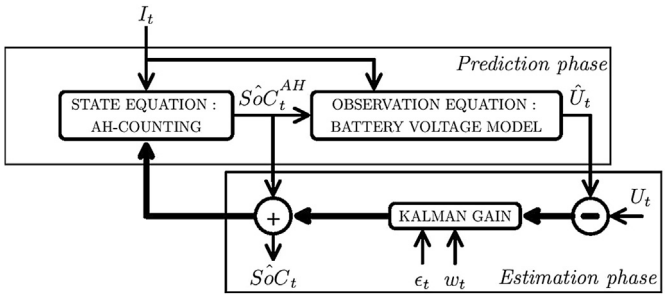


Fig. 3. *SoC* estimator through an extended Kalman filter, where I_t is the input u_t, SoC_t the state variable x_t and U_t the observed variable y_t.

$$y_t = C \cdot x_t + D \cdot u_t + \varepsilon_t \quad \text{(Observation model)}$$

where A and B represent the transition and observation matrix respectively, C and D model the influence of the input u_t on x_t and y_t; and w_t and ε_t are white-noise processes used to model the error of the state and observation equations respectively.

The Kalman filter is an optimal estimator only if the state and observation models are linear and all noises are Gaussian (see Ref. [41] for more information). When one of these models is non-linear one needs to use the Extended Kalman Filter (EKF). This approach is adopted despite its sub-optimal results.

The performance and efficiency of an EKF shall be determined by the reliability of the state and observation models. Fig. 3 shows a flow diagram of the *SoC* estimation using an EKF. This is described in the following. The unknown variable x_t is predicted using the state model given through observation at $t-1$. The observation at t is then predicted using the observation model, based on the predicted value $\widehat{x}_t$. Thus $\widehat{x}_t$ is adjusted based on the gap between the predicted value and the measured value of the observation. The power of the EKF is that this gap is weighed by an adaptive Kalman gain that depends on the correlation between the observed and the unknown variables.

In battery applications, the state and observation models depend on the battery model itself. Chen and Lin [42] review several types of battery models used on electrical vehicles.

Plett et al. [43] develop and analyze several observation models; the *SoC* is considered to be an unknown variable, and the battery voltage an observed variable. To describe the relationship between the *SoC* and the voltage, the authors create a combined model by merging three existing voltage models: Shepherd, Unnewehr universal and Nernst. This combined model estimates the voltage efficiently but suffers from computation problems when the *SoC* is near to 0% and 100%, because of the terms $1/SoC_t$, $\log(SoC_t)$, etc.

Furthermore, the authors consider that these terms can be reduced to a $OCV(SoC)$ function. In this "simple model", the relationship between the battery voltage and its *SoC* is described as follows:

$$U_t = OCV(SoC_t) + R \cdot I_t \tag{5}$$

$OCV(SoC_t)$ plays a central role in the EKF. Plett and al. do not give a theoretical description of this relationship, but a graphical one; the ratio $\partial OCV_t/\partial SoC_t$ is calculated using this graphical representation.

Dai et al. [16] use the 2nd order Randle model to physically represent a Li-ion battery pack:

$$U_t = OCV(SoC_t) + U_t^{RC1} + U_t^{RC2} + R \cdot I_t \tag{6}$$

where U_t, the observed variable, is the pack terminal voltage, U_t^{RC1} and U_t^{RC2} are the voltages across the two RC circuits estimated using the EKF, R is the internal resistance of the battery and $OCV(SoC_t)$ is the relationship between the *OCV* and *SoC*. As in the previously discussed article [43] this relationship is not clearly described despite its importance in EKF.

However, the precision of the voltage estimation depends essentially on the quality of $OCV(SoC_t)$. We have seen earlier (Section 4.2) that this relationship depends on the battery technology. In certain cases the *OCV* has a limited variation on a broad *SoC* range, thus calling into question the performance of the EKF. In addition this relationship varies according to the current rate, the state of health, etc. As a result, this relationship cannot be described using a single equation or a single lookup table.

Also, the output voltage depends on the polarization of the battery. This is highlighted by the hysteresis effect occurring in the battery. In order to describe the voltage hysteresis effect, Plett et al. [43] develop the "zero-state hysteresis model", an improvement of the "simple model" discussed above. The authors add a switching factor which depends on the sign of the current, and the difference between the maximum positive and the minimum negative of the hysteresis.

However, this hysteresis factor flips immediately when the current sign changes. A slow transition may be modeled considering that the battery voltage contains an unknown hysteresis voltage that will also be estimated using the EKF. For a Ni-MH battery, Xuyun and Zechang [44] represent the evolution of the hysteresis factor using an exponential function depending on the current value, a parameter derived from the hysteresis voltage test and the maximum value of the hysteresis voltage. For a Li-ion polymer battery, Plett et al. [43] consider the evolution of the hysteresis voltage as an exponential function of the *SoC*, the rate-of-charge of the *SoC* and the current value.

In a more advanced model called "enhanced self-correcting model", Plett et al. [43] consider the relaxation effect during pulsed current events and rest periods as a low-pass filter on the current. The model forces the estimated voltage to converge to *OCV* after a rest period.

Hence the extended Kalman filter is a strong tool to estimate the *SoC* but a number of issues remains unsolved. First, the EKF is not an optimal estimator because it leads to an inaccurate estimation when the state or observation equations are strongly nonlinear. An optimal model that predicts efficiently the battery voltage without being too complex is needed. Second, the EKF strongest limitation is the need to initialize different filter parameters, like the state vector and the covariance matrix. The EKF may diverge quickly if its parameters are inadequately initialized. To solve this problem Han et al. [45] use an adaptive Kalman filter that estimates automatically the initial covariance matrix.

4.3.2. Open circuit voltage estimation

The open circuit voltage cannot be measured in real-time but may be estimated based on a suitable battery voltage model.

Pan et al. [46] estimate *OCV* based on an EKF whose unknown variable is *OCV* and observed variable is the terminal battery voltage. The authors use a reactive battery circuit model to describe the evolution of *OCV*.

Hirai et al. [47] and Moo et al. [48] calculate OCV_t directly given the current I_t and the terminal voltage U_t based on the Randle equivalent model.

The *OCV* can also be calculated using a state observer [49–51]. A state observer estimates the hidden state $\widehat{x}_t$ of a process, i.e. the unknown *OCV*, based on the following formula:

$$\widehat{x}_{t+1} = f(\widehat{x}_t) + G\cdot(\widehat{y}_t - y_t). \tag{7}$$

In battery applications, $f(.)$ is the transition equation modeling the battery behavior, the observed variable y_t is usually the output voltage, $\widehat{y}_t$ is estimated based on an electrical model and G is the observer gain which can be constant [49,50] or variable [51]. Hu et al. [51] estimate *OCV* using an adaptive Luenburger observer. The observer gain is adaptively adjusted using a stochastic gradient approach so as to reduce the error between estimated and measured battery output voltage, se Fig. 4. The computational cost of a state observer with fixed gain is reduced compared to adaptive gain.

Given the *OCV*, the *SoC* is then estimated using a predefined relationship that is considered as linear [46,48] for Lead-acid batteries and exponential [47] for Li-ion ones.

In order to modulate the hysteresis effect between the *SoC* and *OCV*, Tang et al. [52] use the Preisach operator to decompose this hysteresis in an N elementary hysteresis called hysteron. Two variables are associated with each hysteron i: a SoC_i and a portion w_i. The *SoC* is estimated as:

$$SoC = \sum_{i=1}^{N} w_i \cdot SoC_i. \tag{8}$$

4.3.3. Parameters identification

The identification of the parameters of the state and observation models, specifically the voltage model, can generally be classified into two categories:

1. The time domain analysis, which uses experimental battery voltage and current data, under pulse tests [40,63,47].
2. The frequency domain analysis, for example the impedance measurement of the battery using Nyquist plots [63]. More efficient than time domain analysis methods, the impedance measurement, again for example, suffers from several limitations mentioned in Section 4.2.

In both categories, the estimated parameters do not change according to the cycling conditions, and remain stable during the charge/discharge of the battery. As a result the model is generally valid only in identical experimental conditions.

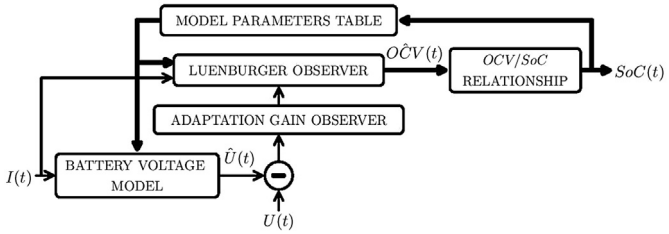


Fig. 4. *SoC* estimator through an adaptive Luenburger observer.

However, these parameters change during the charge/discharge of the battery and depend on the cycling conditions [26,53–55]. To overcome this problem, several researchers have implemented heuristic methods. Hu et al. [51] create a lookup table to identify the model parameters for a specific *SoC*. This method is based on an estimated and therefore inaccurate *SoC*, and on a lookup table whose limitations are presented in Section 4.2. Pang et al. [46] incorporate all model parameters in the vector state of the Kalman filter; so they are estimated at each time step. This method can use a lot of BMS resources despite the fact that these parameters change slowly during the cycling conditions and do not need to be frequently estimated. Hirai et al. [47] identify the model parameters for a specific temperature and *SoC*. This model cannot be efficiently generalized to situations not seen previously. Plett et al. [43] develop a more advanced model by performing joint optimization over the temperature range where every parameter is represented by a continuous polynomial (4th order) of temperature.

The model parameters can also be estimated using machine learning methods. In battery applications, the most used method is the least-squares method [16,43]. More advanced methods take into account the sensor error and can be applied in real-time. Accordingly, the estimated parameters can be updated during the charge/discharge of the battery and adapted according to the cycling conditions. These techniques are designated as online learning and will be discussed in Section 4.5.

All of the presented methods above require a battery modeling in order to provide a terminal battery voltage model. This model is used to improve the Ah-counting method or to estimate *OCV*, and benefits from the strong link between the *OCV* and the *SoC*. To physically establish an efficient battery voltage model, several phenomena have to be modeled like the hysteresis and the relaxation effects. This makes the voltage model too complex and poses computational problems [56]. Moreover the model parameters change during charge/discharge of the battery. The proposed improvements [46,43,51,47] are not enough as they do not take into account the limited hardware resources and the improved model cannot be efficiently generalized to cases not previously seen.

4.4. SoC obtained from inputs estimated using electrochemical models

During battery cycling, several phenomena occur alongside the main redox reaction, like the diffusion and migration of ions, phase transition and ohmic resistance. Thus the state of charge of the battery is influenced by all these phenomena and cannot be directly measured. An electrochemical model, such as the single particle model [57] and porous electrode model [58], is used to estimate one or several physical quantities based on a micro modeling of the phenomena occurring during battery cycling. This electrochemical model can be used to estimate *OCV* or the concentration repartition of active material in the anode, thus allowing the *SoC* to be calculated using a suitable estimation model. In the following, we present several *SoC* estimation methods where inputs variables are estimated using electrochemical models. Table 4 describes the *SoC* estimation methods of this Section according to the novel

Table 4
Main classification criteria for *SoC* obtained from inputs estimated using electrochemical models.

Estimated input variables	*OCV*, overvolatege, Lithium concentration
SoC estimation model	Physical
Other models	Electrochemical models of *OCV*, overpotential
Type of processing	Closed & open-loop

classification, see Fig. 1.

4.4.1. OCV estimation

Pop et al. [25] estimate two physical quantities, *OCV* and overvoltage, based on an electrochemical model. Thus two *SoC* estimation methods are presented: the first one is based on a *OCV*/*SoC* electrochemical relationship and the second one consists in improving the Ah-counting considering the overpotential. Indeed, due to overpotential, the remaining charge cannot be withdrawn from the battery, because the battery voltage would drop below the cutoff voltage. To incorporate the influence of the ambient temperature, a linear relationship between it and the other model parameters is considered. The main limitation of the two models is that the values of the parameters depend on the determination method and experimental conditions.

4.4.2. Estimation of the lithium concentration

In a Li-ion battery, *SoC* is defined as the ratio between average and maximum concentration of lithium inside the anode. Santhanagopalan et al. [59] model the battery by a single spherical particle which represents each electrode by a single spherical particle. Thus the behavior of each electrode is modeled by a diffusion equation governed by Fick's laws, and then an extended Kalman filter is used to solve these diffusion equations and estimate the concentration repartition of lithium inside the two electrodes.

Di Domenico et al. [60] model the battery by a porous model which considers that each electrode consists of a solid matrix immerged in the electrolyte. Unlike the previous method, it decomposes each electrode into M parts, with a diffusion equation being calculated per part. A Kalman filter with $M-1$ evaluation equations is used to estimate the concentration of lithium inside the anode.

In both methods, a relationship between the output voltage of the battery and the concentration of lithium in the anode provides the observation equation of the Kalman filter.

Nevertheless, the *SoC* estimation based on electrochemical models has several limitations. First, the electrochemical processes depend widely on the battery technology and the fabrication method. Therefore, even for identical technologies and providers, a small disparity can greatly affect the low-level battery parameters. Second, the battery state of health, which affects largely the physical and chemical properties of any battery, is not considered in these models: the ageing of the battery decreases the number of active particles and modifies the conductivity of the electrolyte [11,61,62]. Moreover, electrochemical models require knowledge of low-level battery parameters values which is impossible in some contexts.

4.5. Estimation based on machine learning methods

The majority of *SoC* models presented above have a set of parameters that need to be identified. For instance, the estimation of the parameters of

- the relationship between the *OCV* and *SoC* [46–48] (Section 4.3),
- the state and the observation models of an EKF [63] (Section 4.3),

can be done using machine learning methods. The least squares method, one of the simplest machine learning methods, is already used in the batteries domain in order to estimate the parameters of the battery voltage models [16,43]. However the machine learning domain contains more advanced methods that can efficiently estimate the parameters of a physical model.

In the following, we present the principle of machine learning methods, before we discuss the ones used to estimate the *SoC*.

4.5.1. Principle of machine learning

Consider a system whose behavior is described by:

$$z_t = f(r_t, \mathbf{w}) \tag{9}$$

where z_t is the observed system output, r_t the known input vector and $\mathbf{w}$ the model parameters to be estimated. As shown in Fig. 5, these methods comprise three major phases: the construction of the training dataset phase, the learning phase, and the estimation phase. The first and second phases are often performed offline, while the third phase is achieved online.

In the first phase, input and output variables are considered to be available (directly measured or calculated). A training dataset $\mathscr{D} = \{(r_i, z_i) | i = 1, \ldots, N\}$, where N is the size of the dataset, is collected through experimental tests. In a battery application, z represents the *SoC* and r represents the input vector that can be a set of measured variables like current, voltage and temperature, or estimated variables like *OCV*. A training dataset is generally collected in a laboratory under controlled conditions: ambient temperature, current profile, etc. The *SoC* values in the training dataset are obtained using the Ah-counting method. This is rendered reliable as the current sensor is very efficient and the initial value of *SoC* is known.

During the learning phase, one looks for the model parameters $\mathbf{w}$ in order to have function $f(r,\mathbf{w})$ fitting the output z, given the inputs r. This problem is generally referred to as the regression problem. In other words, the objective is to attain a quality criterion while estimating the parameters based on the training dataset. In the case of the regression problem, this often means minimizing the residual Lp-norm:

$$\min_{\mathbf{w}} \left(\sum_{i=1}^{N} |f(r_i, \mathbf{w}) - z_i|^p \right)^{1/p}. \tag{10}$$

In the most cases one uses the $L2$-norm which is the sum of squares of the residuals. The least squares method is therefore used to estimate $\mathbf{w}$.

In the estimation phase, a new measured input vector r_t is considered, the output $\widehat{z}_t$ is then estimated using the model whose parameters are estimated in the learning phase.

The intent of this technique is not to find an exact model but an approximation of it. It is important to emphasize that these methods can learn the parameters of a non-physical model $f(\cdot)$, especially those of a statistical regression model that establishes a relationship between the *SoC* and the input variables. Therefore such a model is not always interpretable. The main virtue of the machine learning technique is its capacity to be easily extrapolated in different situations as soon as a corresponding training dataset is available.

In the following we present several battery *SoC* estimators based on machine learning methods: the autoregressive moving average, the support vector regression and the artificial neural network. The elements of the *SoC* estimation methods presented in this Section are listed in Table 5 according to the novel classification, see Fig. 1.

4.5.2. Autoregressive moving average (ARMA)

Kozlowski [64] estimates *SoC* using the ARMA model:

$$SoC_t = w_1 \cdot r_t + w_2 \cdot r_{t-1} + w_3 \cdot SoC_{t-1} \tag{11}$$

where r_t is the input vector including the components of the internal impedance, identified online based on pattern recognition of the current and voltage signals. Here the model parameters w_1, w_2

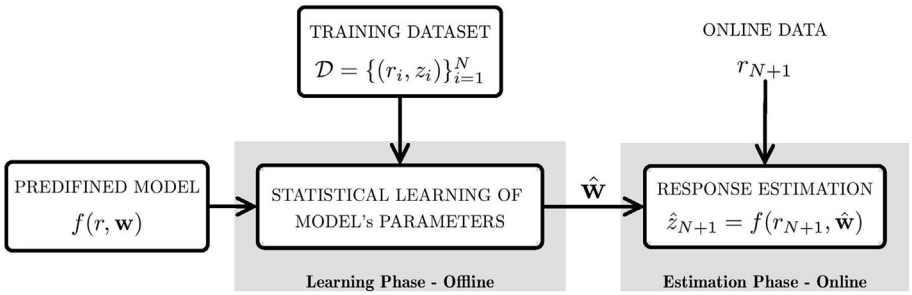


Fig. 5. Estimation of *SoC* based on machine learning methods.

and w_3 are estimated in the learning phase using the least squares method. Multiple training datasets were collected during the charge and discharge phases, under various temperatures and loads. Four types of batteries are considered: primary lithium and alkaline, and nickel–cadmium and lead-acid secondary batteries. Thus there are 300 datasets across the different types and sizes of batteries. The model parameters are then identified separately for each cycling conditions and type of battery. Accordingly, prior to an online estimation of the *SoC*, suitable model amongst the 300 models must be chosen based on the cycling conditions, type and size of the battery.

Unfortunately, the least squares approach is strongly dependent on the training dataset. The learned model may fit fits very well the training dataset but is not generic enough to fit new data that bears slight differences compared to training data. This problem is referred to as "overfitting". To avoid it there are numerous regularization techniques enabling smoother models [76]. This can be usually achieved by:

- adding a penalty term to the cost function (10) like in SVR [65], RVM [75], LASSO [74],
- implementing a cross-validation techniques.

The former technique gives a smoother model, *i.e.* some of the parameters w_i are equal to zero. The latter technique provides an unbiased estimator of the model parameters and is generally coupled with the first technique. Both techniques remain limited when there is a significant difference between the training dataset and the real-time data. For a better understanding of the cross-validation techniques, we describe here *K*-fold technique which is one of the most widely used ones. It consists in splitting the training dataset into *K* subsets. At each iteration, two sets are considered: the learning set, involving all subsets except the *i*th one used to learn the model parameters, and the testing set, also called validation set, including the *i*th subset and used to test the predictive performance of the model. Hence, the best model of the *K* learned models is retained. Fig. 6 shows a detailed flow diagram of the *K*-fold cross-validation. However, the main drawback of this technique is that it has a high computation cost as *K* models need to be learned.

4.5.3. Support vector regression (SVR)

The SVR algorithm estimates the parameters of a non-linear regression model which describes the system behavior as following:

$$z_t = \sum_{i=1}^{N} w_i \cdot \kappa(r_t, r_i), \tag{12}$$

where N is the size of the training dataset, w_i the model parameters estimated by SVR and $\kappa(\cdot,\cdot)$ the kernel function. Thus the regression model is based on a kernel function that measures the similarity between the new sample r_t and the training sample r_i. The SVR algorithm rules out some of the w_i as it adds a penalty constraint on the model parameters. Accordingly, a significant number of w_i are void. The remaining w_i corresponds to specific samples from the training dataset called "support vectors". Only these vectors are then used in Equation (12) to describe the behavior of the modeled system [65].

The SVR model is used to estimate the *SoC* based on a polynomial kernel in Ref. [66] and on an exponential kernel in Ref. [67]. Hansen et al. [66] use the current and output voltage as input vector r_t; Anton et al. [67] add the battery temperature to r_t.

In both studies the training dataset contains only experimental tests under a constant current rate and a single ambient temperature. Accordingly, the estimated model is valid under identical cycling conditions and it cannot be efficiently generalized to cases not seen in the learning phase. In addition, the SVR model is unsuitable for time series processing since it does not take into account the time dimension.

4.5.4. Artificial neural network (ANN)

In machine learning, the ANN is used to describe complex processes by using a non-linear regression model. An ANN is a set of nodes organized in layers which are connected through arrows weighted by w_i. In most cases, a three layers system is used: the input layer includes as many nodes as variables in the input vector which includes the inputs r_t, the hidden layer, containing M nodes, which describes the non-linearity of the process via M intermediary functions $g_i(r_t)$, and the output layer which provides the estimated output z_t via an activation function $K(\cdot)$:

$$z_t = K\left(\sum_{i=1}^{M} w_i g_i(r_t)\right), \tag{13}$$

Based on a training dataset, the connection weights w_i and the parameters of $g_i(\cdot)$ are estimated in the learning phase using such methods as the least squares method or back propagation algorithm [69].

Charkhgard et al. [70] represent the battery voltage via an ANN:

$$U_t = \sum_{i=1}^{M} w_i \exp\left(-\frac{||r_t - t_i||}{\sigma_i^2}\right)^2 + w_0 \tag{14}$$

where t_i and σ_i are the parameters of the intermediary function $g_i(.)$ and r_t=(U_{t-1},I_t,SoC_t). This mathematical representation of the battery voltage is then used as an observation model of an EKF that has the Ah-counting formula as its state model. This model raises computational problems to calculate $\partial U_t/\partial SoC_t$, as it is a sum of M strongly non-linear functions.

Table 5
Main classification criteria for *SoC* estimator based on machine learning methods.

Input variables	Online measured
SoC estimation model	Statistical regression
Other models	–
Type of processing	Closed & open-loop

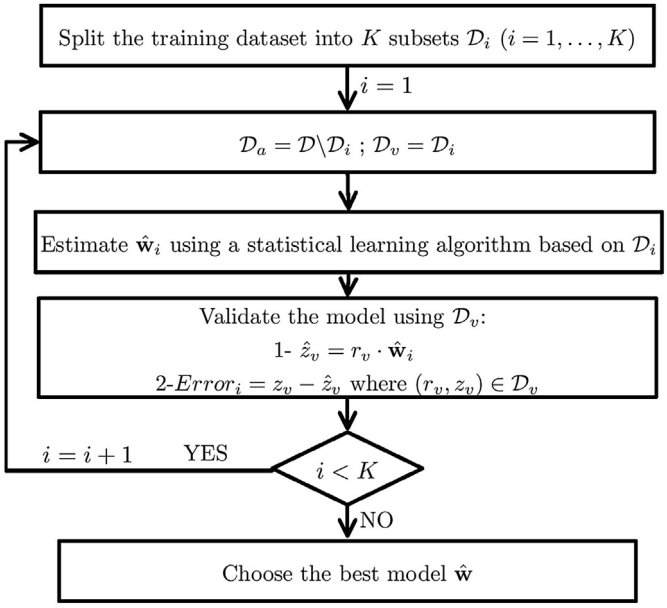


Fig. 6. Principle of *K*-fold cross-validation.

Yanqing et al. [71] develop a more complex model using two successive ANNs. The first one provides a prediction of the battery voltage. The inputs of this ANN are defined based on the combined voltage model [43]. Thus the temperature T_t, the current measurement I_t and SoC_t are used as inputs of the first ANN. The second ANN is used to update the parameters of a PID controller that estimates the *SoC*:

$$SoC_t = K_p \cdot (e_t - e_{t-1}) + K_I \cdot e_t + K_D \cdot (e_t - 2 \cdot e_{t-1} + e_{t-2}) + SoC_{t-1} \quad (15)$$

where K_P, K_I, K_D are respectively the proportional, integral and derivative gain of the controller, and e_t is the gap between the predicted and measured voltage. In this case, the PID controller is called "neuro-controller" as K_p, K_I and K_D are calculated using an artificial neural network. Accordingly the *SoC* model is adapted online when the cycling conditions change and provides an efficient estimation of the *SoC*. Although the model can be generalized more efficiently than the previously presented models, it remains limited when the cycling conditions are very different from the training dataset or when the state of health of the battery decreases. In these cases the model parameters of both ANNs have to be updated. However the use of two successive ANNs increases the complexity of the model and the embedded model can require a lot of the BMS computational resources.

The choice of the input variables directly affects the quality of the ANN. Thus Bo et al. [72] use partial least squares regression (PLSR) to select the input variables. Affanni et al. [73] use the internal impedance, the extracted charge and the open circuit voltage as inputs of the ANN. However these parameters cannot be directly measured; they are estimated using their physical definition as it induces an important error on the input values and affects the performance of the method.

4.5.5. Difficulties in choosing the training dataset

As shown above, in machine learning, the estimated model depends strongly on the training dataset, the learning method and the quality criterion. Thus, in order to obtain a generic model that can give a reliable *SoC* estimation for any battery type under any cycling conditions, as many possible situations must be included in an exhaustive training dataset. Such a dataset requires a large number of experimental tests. We can identify several types of battery behavior according to internal characteristics and external cycling conditions. Indeed, the battery behavior varies following the charge/discharge, *SoC* ranges ([0,20], [20,80], [80,100]), battery technologies, types of current profile (constant, dynamic, etc.), state of health, outdoor temperature and so on. Thus, machine learning methods are able to give a generic *SoC* estimator for all behaviors existing in this dataset. Another approach consists in associating a battery model to each battery behavior like in Ref. [64] or, for example, a model for each voltage rate like in Ref. [68]. The problem of the latter technique is the need for a large storage space that is often limited in real-time contexts. Therefore, our interest is brought below on new and promising machine learning techniques, called "online learning".

4.5.6. Online learning

In a real-time context, the cycling conditions are uncontrolled as they depend on the user's behavior, weather conditions, road conditions, etc. In order to create a generalized model, all possible situations must be included in the training dataset. It is clear that it is difficult if not impossible to carry out as there is an infinite number of cases. Machine learning offers some promising techniques that update the model when necessary, i.e. online. Generally, the model must be updated when the distribution of data changes over time in unforeseen ways. To take into account this change, two strategies are distinguished: adaptive learning and drift detection.

The adaptive learning methods update the model parameters at each time step. These methods have already been used in a battery context by Pang and al. [46] (see Section 4.4.3), but require a lot of BMS resources, even though the parameters change slightly during charge/discharge and do not need to be frequently estimated.

In battery applications, *drift* occurs when the cycling conditions are very different from those of the training dataset or when the state of health of the battery decreases significantly.

Consider an estimated physical quantity which can also be measured in real-time conditions, the battery voltage for example. The drift can be detected when the gap between the estimated and measured values exceeds a predefined maximum threshold.

Another method to detect the drift is to compare the statistical distribution of the observations, the battery voltage for example, in the current time window and also in a predefined reference window. When both distributions are significantly different, a drift is

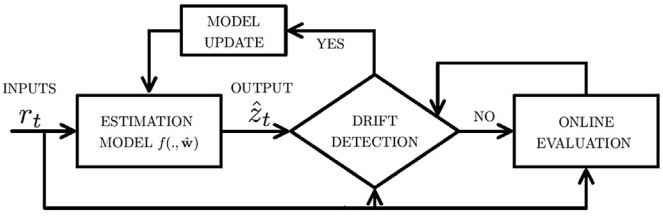


Fig. 7. Online learning procedure.

occurring.

Nonetheless, it is important to distinguish sensor noises from data changes. An efficient method must be robust to noise and adaptive to changes. For this reason, more advanced drift detection is based on the evaluation of the model performance in recent past. The complete online learning procedure is presented in Fig. 7. These methods are yet to be fully developed. The limited hardware resources, *i.e.* computation capacity, as well as the required learning time should be taken into account. In addition, it is necessary to note that no reference *SoC* is available online. At first glance, this may appear as a considerable restraint to this technique, but in practice, additional information like internal impedance, capacity, state of health, *OCV*, etc. is integrated in the online learning procedure and greatly improves the estimated model. Fig. 8 provides an overall and detailed description of creating an ideal *SoC* model.

In conclusion, the machine learning methods provide powerful techniques to estimate the parameters of a *SoC* model but they are not well explored. A penalty on the model parameters and a cross validation can provide a simple and smooth model. The main drawback of these methods is the need for an exhaustive training dataset including all possible cycling conditions and types of batteries. Collecting such a dataset is difficult if not impossible to carry out. More advanced techniques, like online learning, can detect variations in battery behavior and adapt the model to new situations. Accordingly, these techniques can provide an ideal model, reliable for all battery types and under all cycling conditions. In addition, it is important to note that having a reliable *SoC* in the training dataset is sometimes difficult; particularly in real-time applications like hybrid electric vehicles.

5. Discussion

The battery being a complex electrochemical system, it is very difficult to provide an exact *SoC* model. Moreover, the electrochemical phenomena of each type of battery are very specific and their description requires the knowledge of the low-level battery parameters which is not possible in some contexts.

Thus in order to create an approximate and generalized *SoC* model, an infinite number of experimental tests must be conducted to estimate its parameters.

Therefore all existing estimation models and methods are not generic and have some limitations. Several researchers develop *SoC* estimators for a specific battery technology, for instance, the method based on electrolyte density for acid batteries and the one based on modeling the migration of materials for Li-ion batteries. In addition, many developed models that do not take into account the influence of the cycling conditions on the battery behavior. For example, the *SoC* model is learned for a fixed temperature and current rate [66] and for a fixed state of health [51]. Several researchers improve the *SoC* model by incorporating the influence of ambient temperature [43] and the variations of the battery characteristics according to the *SoC* [47]. More advanced models can be adapted online [46,71] but they require a lot of BMS resources as the model parameters are updated at each time step. But all of these methods remain limited when the cycling conditions are very different from those of the training data. A summary of all the existing methods and their types, based on the novel classification, is presented in Table 6.

Accordingly, a complete and generic model that provides a reliable *SoC* estimation for any battery technology under any cycling conditions, is needed. The characteristics of the ideal model depend on the goal and the context of the battery usage, and even on the specifications of the system. First, an ideal model must be efficient specifically when the *SoC* is near to 0% and 100% to avoid over-charge/discharge that cause permanent damage and fire risks like, for example, in the case of a Li-ion battery. Also, it must give a linear indication avoiding brutal change in order to give confidence on autonomy prediction. Second, an ideal model must be robust to the sensor noise like the Kalman filter, specifically in commercial products, such as an electric vehicle; an efficient current sensor is very expensive. However in a laboratory, sensors are generally very reliable and there is no need to complicate the model with sensor noise treatment. Third, this model has to estimate the *SoC* in a real-time application. In this case, the battery is often partially charged and reference points that provide a *SoC* calibration are seldom available. In addition, the estimation of the *SoC* of batteries pack, usually used in real-life contexts, introduces more complexity due to the heterogeneity of the elementary battery constituting the pack [77]. The display rate of the *SoC* is also an important factor in real-time applications. In cell phones and laptops, the variation of the *SoC* is slower than the one in electric vehicles. Accordingly, the estimation algorithm must adapt more quickly in applications where the *SoC* changes rapidly. Fourth, a complete model must take into account the limitations in calculation capacity and storage space of the Battery Management System (BMS). Hence, the calculation complexity of this model and its fixed parameters, like the lookup table, is governed. Finally, this model has to incorporate the variation of the battery properties and performances like the state of health, impedance and available capacity, as well as the influence of the cycling conditions on the battery behavior. It must be updated online to keep track of the variation of the battery properties. It must also be able to indicate when it becomes

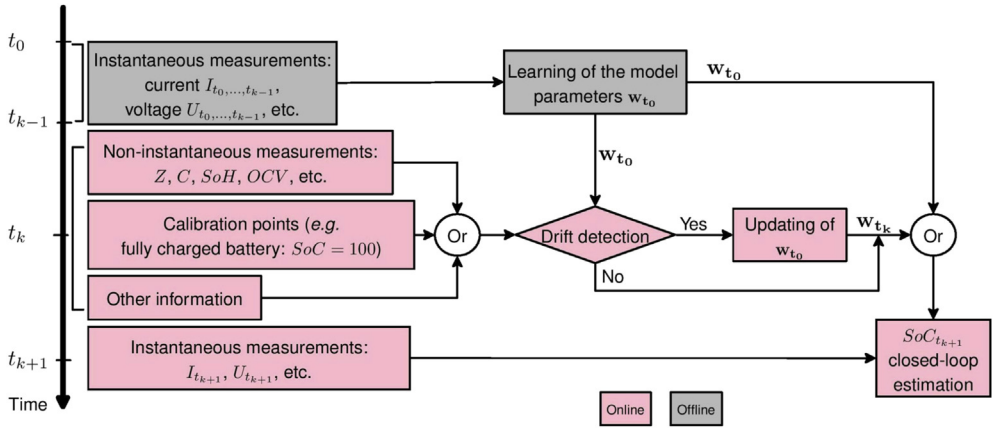


Fig. 8. Complete description of creating an ideal *SoC* model.

Table 6
Summary of the existing models and their types based on the novel classification.

SoC model	Input variables and parameters					Type of processing		References
	Measured online	Measured offline	Estimated using physical model	Estimated using electrochemical model	Estimated using statistical model	Offline	Online	
Lookup table	✓	✓				✓		[27,37,38]
	✓					✓		[1]
	✓		✓				✓	[51]
Physical model	✓		✓				✓	[49,50]
	✓		✓			✓		[8,18,80]
	✓		✓		✓	✓		[19,22]
	✓		✓		✓		✓	[7,43,68,70,78,81]
	✓		✓				✓	[16,17,40,44,45,63]
	✓		✓	✓		✓		[25]
	✓		✓	✓			✓	[59,60]
Statistical model	✓		✓		✓		✓	[46]
	✓		✓		✓	✓		[47,48,52]
	✓				✓	✓		[64,66,67]
	✓				✓		✓	[71–73,79]

inapplicable and when the learning of a new model is necessary.

In light of the above, machine learning methods are able to acquire the behavior of the battery through real data, without the absolute need to have a physical description of the battery. They can be used to identify the parameters of either predefined physical or statistical regression models. The dataset collected during an experimental test like the current, temperature and output voltage, is called training data. Thus the learned model is applicable on identical battery technology with the same experimental conditions of its corresponding training data. Therefore, in order to provide a valid model for all battery technologies under any experimental conditions, exhaustive training data that includes a large variety of situations is necessary. However, the behavior of the battery changes and the cycling conditions are uncontrolled and so, no single estimation model is valid at all times. Hence an online update of this estimation model is crucial. Updating a model online faces two constraints: the first one being limitations on storage and calculation capacities, the second being the absence of a reliable *SoC*. With these problems solved, the machine learning methods can provide an ideal *SoC* estimator.

6. Conclusion

In this paper we have introduced a novel classification of the existing *SoC* estimation methods. The main benefit of this classification is that it enables the identification of the techniques that can be improved to potentially reach an ideal *SoC* estimation model. Such a model can estimate the *SoC* efficiently in a real-time context without being impacted by the battery chemistry or the cycling conditions. We found that methods based on directly measured physical quantities (presented in Section 4.2) cannot be generalized in a real-time context, as these quantities can only be measured offline. Also, methods based on an open-loop processing do not take into account the sensor noises which can greatly decrease the reliability of the estimation, particularly in mainstream commercial products. Moreover, methods based on closed-loop processing, like the Kalman filter and controller, are promising candidates on the way to obtain an ideal *SoC* estimator. The main difficulty of these methods resides in the parameters' identification. Clearly, machine learning methods provide powerful techniques allowing the estimation of these parameters. However, these methods need an exhaustive training dataset to produce a generalized and smooth model. More advanced machine learning techniques can adapt the model online when a drift occurs. Accordingly, the machine learning techniques can provide an ideal *SoC* model that is updated automatically online when the characteristics of the battery or the cycling conditions change.

References

[1] S. Piller, M. Perrin, A. Jossen, J. Power Sources 96 (2001) 113–120.
[2] J. Zhang, J. Lee, J. Power Sources 196 (2011) 6007–6014.
[3] L. Lu, X. Han, J. Li, J. Hua, M. Ouyang, J. Power Sources 226 (2013) 272–288.
[4] W. Waag, C. Fleischer, D.U. Sauer, J. Power Sources 258 (2014) 321–339.
[5] V. Pop, H. Bergveld, P. Notten, P. Regtien, Meas. Sci. Technol. 16 (2005) R93–R110.
[6] W.Y. Chang, ISRN Appl. Math, ID 953792, 2013.
[7] J. Wang, B. Cao, Q. Chen, F. Wang, Control Eng. Pract. 15 (2007) 1569–1576.
[8] K.S. Ng, C.S. Moo, Y.P. Chen, Y.C. Hsieh, Appl. Energy 86 (2009) 1506–1511.
[9] V. Iliev, D. Pavlov, J. Electrochem. Soc. 129 (1982) 458–464.
[10] B. Wu, R.E. White, J. Electrochem. Soc. 147 (2000) 902–909.
[11] J. Vetter, P. Novak, M. Wagner, C. Veit, K.-C. Moller, J. Besenhard, M. Winter, M. Wohlfahrt-Mehrens, C. Volger, A. Hammouche, J. Power Sources 147 (2005) 269–281.
[12] A.C. Zachlin, Trans. Electrochem. Soc. 82 (1942) 365–373.
[13] M. Broussely, S. Herreyre, P. Biensan, P. Kasztejna, K. Nechev, R.J. Staniewicz, J. Power Sources 97–98 (2001) 13–21.
[14] A.J. Smith, J.C. Burns, S. Trussler, J.R. Dahn, J. Electrochem. Soc. 157 (2010) A196–A202.
[15] T.M. Bond, J.C. Burns, D.A. Stevens, H.M. Dahn, J.R. Dahn, J. Electrochem. Soc. 160 (2013) A521–A527.
[16] H. Dai, Z. Sun, X. Wei, IEEE Int. Conf. Veh. Electron. Safety (2006) 342–347.
[17] Y. Liao, J. Huang, Q. Zeng, Adv. Mater. Res. 152–153 (2011) 428–435.
[18] J. Alzieu, H. Smimite, C. Glaize, J. Power Sources 67 (1997) 157–161.
[19] S. Malkhandi, Eng. Appl. Artif. Intell. 19 (2006) 479–485.
[20] M.F. Chubb, H.R. Harner, Trans. Electrochem. Soc. 68 (1935) 251–259.
[21] D. Linden, Trans. Electrochem. Soc. 92 (1947) 223–228.
[22] V.R. Gaddam, X.Q. Wang, S. Arey, Z.J. Yang, P. Singh, Power Sources Conf. (2000).
[23] P. Kurzweil, Encycl. Electrochem. Power Sources, Elsevier, Amsterdam, 2009, pp. 381–394.
[24] C.C. Christianson, F.R. Bourke, US 3946299A, (1976).
[25] V. Pop, H.J. Bergveld, J.H.G. Op het Veld, P.P.L. Regtien, D. Danilov, P.H.L. Notten, J. Electrochem. Soc. 153 (2006) A2013–A2022.
[26] E.C. Parfitt, E.W. Crofts, R. Buckle, Trans. Electrochem. Soc. 28 (2010) 21–33.
[27] E. Peled, H. Yamin, I. Reshef, D. Kelrich, S. Rozen, US 4725784A, (1988).
[28] M. Petzl, M.A. Danzer, Energy Convers. IEEE Trans. 28 (2013) 675–681.
[29] M. a. Roscher, D.U. Sauer, J. Power Sources 196 (2011) 331–336.
[30] V. Srinivasan, J.W. Weidner, J. Newman, J. Electrochem. Soc. 148 (2011) A969–A980.
[31] V. Pop, H.J. Bergveld, P.P.L. Regtien, J.H.G. Op het Veld, D. Danilov, P.H.L. Notten, J. Electrochem. Soc. 154 (2007) A744–A750.
[32] R. Shalini, N. Munichandraiah, A.K. Shukla, J. Power Sources 87 (2000) 12–20.
[33] E. Willihnganz, Trans. Electrochem. Soc. 79 (1941) 253–258.
[34] S. Buller, M. Thele, R.W. De Doncker, E. Karden, IEEE Trans. Ind. Electron. 41 (2005) 742–747.
[35] J.P. Diard, B. Le Gorrec, C. Montella, P. Landaud, Electrochim. Acta 42 (1997) 3417–3420.
[36] Z. Deng, Z. Zhang, Y. Lai, J. Liu, J. Li, Y. Liu, J. Electrochem. Soc. 160 (2013) A553–A558.
[37] F. Huet, J. Power Sources 70 (1998) 59–69.

[38] J.A. Salkind, C. Fennie, P. Singh, T. Atwater, E.D. Reisner, J. Power Sources 80 (1999) 293–300.
[39] A. Delaille, M. Perrin, F. Huet, L. Hernout, J. Power Sources 158 (2006) 1019–1028.
[40] F. Codeca, M.S. Savaresi, G. Rizzoni, Control Appl. IEEE Int. Conf. (2008) 102–107.
[41] E.R. Kalman, Trans. ASME J. Basic Eng. 82 (1960) 35–45.
[42] Q. Chen, C. Lin, Automob. Technol. 3 (2005) 1–5.
[43] L.G. Plett, J. Power Sources 134 (2004) 262–276.
[44] F. Xuyun, S. Zechang, in: Veh. Power Propuls. Conf. IEEE, 2008, pp. 1–5.
[45] J. Han, D. Kim, M. Sunwoo, J. Power Sources 188 (2009) 606–612.
[46] S. Pang, J. Farrell, J. Du, M. Barth, Am. Control Conf. Proc. 2 (2001) 1644–1649.
[47] T. Hirai, A. Ohnishi, N. Nagaoka, N. Mori, A. Ametani, S. Umeda, Univ. Power Eng. Conf. (2008) 1–5.
[48] C.S. Moo, K.S. Ng, Y.P. Chen, Y.C. Hsieh, Power Convers. Conf. (2007) 758–762.
[49] J. Li, J.K. Barillas, C. Guenther, A.M. Danzer, J. Power Sources 230 (2013) 244–250.
[50] X. Chen, W. Shen, Z. Cao, A. Kapoor, Ind. Electron. Appl. IEEE Conf. (2012) 102–107.
[51] X. Hu, F. Sun, Y. Zou, Energies 3 (2010) 1586–1603.
[52] X. Tang, X. Zhang, B. Koch, D. Frisch, Int. Conf. Progn. Health Manag. (2008) 1–12.
[53] T.K. Dong, A. Kirchev, F. Mattera, J. Kowal, Y. Bultel, J. Electrochem. Soc. 158 (2011) A326–A336.
[54] P.L. Moss, G. Au, E.J. Plichta, J.P. Zheng, J. Electrochem. Soc. 155 (2008) A986–A994.
[55] D. Pavlov, G. Petkova, J. Electrochem. Soc. 149 (2002) A654–A661.
[56] D. Di Domenico, E. Prada, Y. Creff, Control Eng. Pract. 21 (2013) 1851–1859.
[57] M. Guo, G. Sikha, E.R. White, J. Electrochem. Soc. 158 (2011) A122–A132.
[58] J. Newman, W. Tiedeman, AIChE J. 21 (1975) 25–41.
[59] S. Santhanagopalan, E.R. White, J. Power Sources 161 (2006) 1346–1355.
[60] D. Di Domenico, G. Fiengo, A. Stefanopoulou, Control Appl. Proc. IEEE Int. Conf. (2008) 702–707.
[61] A. Barré, B. Deguilhem, S. Grolleau, M. Gerard, F. Suard, D. Riu, J. Power Sources 241 (2013) 680–689.
[62] C. Armenta-Deu, T. Donaire, J. Power Sources 58 (1996) 123–133.
[63] S. Lee, J. Kim, J. Lee, B.H. Cho, J. Power Sources 185 (2008) 1367–1373.
[64] J.D. Kozlowski, Proc. IEEE Aerosp. Conf. 7 (2003) 3257–3270.
[65] N.V. Vapnik, The Nature of Statistical Learning Theory, second ed., Springer-Verlag, New York, 1995.
[66] T. Hansen, C.J. Wang, J. Power Sources 141 (2005) 351–358.
[67] J.C. Alvarez Antona, P.J. Garcia Nietob, F.J. de Cos Juezc, F. Sanchez Lasherasd, M. Gonzalez Vegaa, M.N. Roqueni Gutierrezc, Appl. Math. Model. 37 (2013) 6244–6253.
[68] J. Kim, S. Lee, B.H. Cho, J. Power Sources 196 (2011) 2227–2240.
[69] S. Haykin, Neural Networks: A Comprehensive Foundation, second ed., Prentice Hall PTR, Upper Saddle River, 1998.
[70] M. Charkhgard, M. Farrokhi, IEEE Trans. Ind. Electron. 57 (2010) 4178–4187.
[71] S. Yanqing, Energy Convers. Manage. 51 (2010) 1093–1098.
[72] C. Bo, B. Zhifeng, C. Binggang, Energy Convers. Manage. 49 (2008) 2788–2794.
[73] A. Affanni, A. Bellini, C. Concari, G. Franceschini, E. Lorenzani, C. Tassoni, in: Electr. Mach. Drives Conf. IEEE Int, 2, 2003, pp. 684–688.
[74] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, second ed., Springer-Verlag, New York, 2009.
[75] M.E. Tipping, J. Mach. Learn. Res. 1 (2001) 211–244.
[76] S. Arlot, A. Celisse, Stat. Surv. 4 (2010) 40–79.
[77] Y. Zheng, M. Ouyang, L. Lu, J. Li, X. Han, L. Xu, H. Ma, T.A. Dollmeye, V. Freyermuth, Appl. Energy 111 (2013) 571–580.
[78] G.V. Avvari, B. Balasingam, K.R. Pattipati, Y. Bar-Shalom, J. Power Sources 273 (2015) 185–195.
[79] W. He, N. Williard, C. Chen, M. Pecht, J. Power Sources 62 (2014) 783–791.
[80] K. Kutluay, Y. Cadirci, Y.S. Ozkazanc, I. Cadirci, IEEE Trans. Ind. Electron. 52 (2005) 1315–1327.
[81] H. He, R. Xiong, J. Fan, Energies 4 (2011) 582–598.