# Multiple-imputation-particle-filtering scheme for Uncertainty Characterization in Battery State-of-Charge Estimation Problems with Missing Measurement Data

David E. Acuña[1], Marcos E. Orchard[1], Jorge F. Silva[1], and Aramis Pérez[1]

[1] *Universidad de Chile, Department of Electrical Engineering. Av. Tupper 2007, Santiago, Chile*
*davacuna@ing.uchile.cl*
*morchard@ing.uchile.cl*
*josilva@ing.uchile.cl*
*aramis.perez@ing.uchile.cl*

## ABSTRACT

The design of particle-filtering-based algorithms for estimation often has to deal with the problem of missing observations. This requires the implementation of an appropriate methodology for real-time uncertainty characterization, within the estimation process, incorporating knowledge from other available sources of information. This article presents preliminary results of a multiple imputation strategy used to improve the performance of a particle-filtering-based state-of-charge (SOC) estimator for lithium-ion (Li-Ion) battery cells. The proposed uncertainty characterization scheme is tested and validated in a case study where the state-space model requires both voltage and discharge current measurements to estimate the SOC. A sudden disconnection of the battery's voltage sensor is assumed to cause significant loss of data. The results show that the multiple-imputation particle filter enables reasonable uncertainty characterization for the state estimate as long as the voltage sensor disconnection continues. Furthermore, when the voltage measurements are once more available, the level of uncertainty adjusts to levels that are comparable to the case where data was not lost.

## 1. INTRODUCTION

During the last century there has been an increase in production and development of electronics that has changed the way of living. Due to the increasing scarcity of oil, an imminent migration to alternative kinds of energy becomes relevant. The automotive industry has been putting research efforts into the development of energy storage devices (ESDs) for the production of hybrid electric vehicles (HEV) or fully electric vehicles (EV). As a result, ESDs have been playing a crucial role regarding autonomy of systems. This last fact has impulsed research on Li-Ion battery cells due to advantages over other types of ESDs, being its larger charge density by unit of mass or volume one of the most important features. From the automotive industry, the concept of *"Battery Management Systems"* (BMS) (Pattipati, Sankavaram, & Pattipati, 2011) rises naturally looking for systems capable of providing protection and optimal operating conditions for batteries while accounting for life predictions through monitoring acquired data and interfacing external modules. Regarding this, the *"State-of-Charge"* (SOC) (Pattipati et al., 2011) -quantifing the remaining available energy stored-, the *"State-of-Health"* (SOH) (Pattipati et al., 2011) -describing the degree of degradation-, and the *"Remaining Useful Life"* (RUL) (Orchard & Vachtsevanos, 2009) generate important information about the actual battery cells for optimal management. Unfortunately, due to incapability to measure them directly in an online framework, BMS systems must incorporate real-time estimation and prediction routines to carry out their objectives.

These routines heavily depend on real-time measurements for their implementation, and when measuring data from any device it is possible to miss information due to, for example, transmission problems within sensor networks. Completing the acquired data set is not just as simple as filling in the missing information with averaged values. In this regard, many strategies may be adopted to solve the problem of sequential state estimation with incomplete data sets. Among them, single imputation methods fail due to the lack of uncertainty characterization. In (Rubin, 1987) the idea of multiple imputations was proposed. This method considers different values for each missing datum and combines their induced probability distributions into a single solution for parameter estimation. This led to the multiple imputation particle filter (Housfater, Zhang, & Zhou, 2006), where particle-

filtering methods (Andrieu, Doucet, & Punskaya, 2001) were used, taking into account uncertainty of missing data through a multiple imputation strategy.

This work presents an improvement of the particle-filtering-based Bayesian approach adopted by (Orchard, Cerda, Olivares, & Silva, 2012) for real-time uncertainty characterization in SOC estimation for Li-Ion batteries, based on a multiple-imputation strategy. The validation case for this proposed Multiple Imputation Particle Filter algorithm considers a situation where 1000 sequential voltage measurements are assumed to be lost, emulating the disconnection of the associated sensor during the execution of a specific discharge cycle. Obtained results show that the uncertainty associated to the state estimate due to lost data is bounded. Furthermore, those uncertainty bounds are smaller than those obtained when simply discarding incomplete measurements and applying n-step prediction to generate the prior state density function.

The article is structured as follows. In Section 2, a theoretical background is presented reviewing the underlying concepts of particle filters and the multiple imputation strategy. In Section 3, a new multiple-imputation-based particle filter is applied in Li-Ion battery cells for SOC estimation when voltage and discharge current are measured. Sudden disconnections of the battery's voltage sensor are simulated and uncertainty characterization is analyzed. Finally, conclusions and future work are presented in Section 4.

## 2. THEORETICAL BACKGROUND

Real world systems are commonly dynamic, nonlinear, and may involve a high dimensionality relationship between variables. In this regard, state-space models offer a good treatment for these systems; for example, when monitoring critical system components which physical phenomenology may be modeled directly under the state-space form. Moreover, uncertainty due to noisy measurements associated with sensors constrains or other sources of disturbances such as the lack of knowledge about the actual system dynamics, can be incorporated into the state-space form with ease. This allows to adopt a Bayesian approach, where the main objective is to estimate the underlying probability distribution in order to perform statistical inferences. Since the analytical solutions may be founded under certain conditions, the real problem to be addressed is that of evaluating complex integrals where numerical methods tend to breakdown, even more when high dimensional systems are involved. An alternative to address this problem is the use of particle filters, which is presented in the following section. Later, an introduction to multiple imputation for dealing with missing data and the way multiple imputation particle filter is presented.

### 2.1. Particle Filters

Due to the employment of digital computers for signal processing, it is of interest to develop a Bayesian processor where measurements arrive sequentially in time. The recursive estimation of the evolving posterior distribution is the so called *optimal filtering* problem. A mathematical framework is provided below for solving this problem using particle filters.

Let $X = \{X_t, t \in \mathbb{N}\}$ be a first order Markov process denoting a $n_x$-dimensional system state vector with initial distribution $p(x_0)$ and transition probability $p(x_t|x_{t-1})$. Also, let $Y = \{Y_t, t \in \mathbb{N} \setminus \{0\}\}$ denote $n_y$-dimensional conditionally independent noisy observations. The whole system is represented in state-space form as

$$x_t = f(x_{t-1}, w_{t-1}) \quad (1)$$

$$y_t = g(x_t, v_t) \quad (2)$$

where $w_t$ and $v_t$ denote independent random variables whose distributions are not necessarily Gaussian. Since it is difficult to compute the filtering posterior distribution $p(x_t|y_{1:t})$ directly, Bayesian estimators are constructed from *Bayes' rule*.

Under Markovian assumptions, the filtering posterior distribution can be decomposed into

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t) \cdot p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})} \quad (3)$$

In this context, sequential Monte Carlo methods (SMC) offer an alternative to numerical integration techniques that fail due to high computation. SMC methods, also called particle filters, are stochastic computational techniques designed for simulating highly complex systems in an efficient way. In Bayesian estimation, these techniques simulate probability distributions by using a collection of $N$ weighted samples or *particles*, $\{x_t^{(i)}, \mathcal{W}_t^{(i)}\}_{i=1}^N$, that yields to discrete mass probability distributions, as shown in Eq (4).

$$\hat{p}(x_t|y_{1:t}) \approx \sum_{i=1}^{N} \mathcal{W}_t^{(i)} \delta(x_t - x_t^{(i)}) \quad (4)$$

The weighting process is made by applying the *sequential importance resampling* (SIR) algorithm, which is explained in the following subsections.

#### 2.1.1. Sequential Importance Sampling

The concept of importance sampling is used to simulate samples from a proposed distribution in order to estimate a posterior distribution. The key point for a successful sampling is to choose appropriately the importance distribution. Sampling from posterior distributions is a common task in order to get Monte Carlo (MC) estimates. However, it is not feasible most of the time since it becomes computationally intensive.

For example, Eq (5) shows the calculation of expectations.

$$\hat{f}(x_t) = E_{X|Y}\{f(x_t)\} = \int_X f(x_t)p(x_t|y_{1:t})dx_t \quad (5)$$

Drawing $N$ independent identical distributed random samples from $p(x_t|y_{1:t})$, the integral may be approximated by a sum of delta-Dirac functions.

$$\hat{f}(x_t) \approx \frac{1}{N}\sum_{i=1}^{N} f(x_t)\delta(x_t - x_t^{(i)}) \quad (6)$$

$$= \frac{1}{N}\sum_{i=1}^{N} f(x_t^{(i)}) \quad (7)$$

These approximations may not hold when it is not possible to sample directly from $p(x_t|y_{1:t})$, thus the *sequetial importance sampling* (SIS) algorithm avoids these difficulties by drawing samples from an *importance distribution* approximating the targeted posterior distribution by appropriate weighting. The weights are recursively defined as

$$w_t^{(i)} = w_{t-1}^{(i)} \cdot \frac{p(y_t|\tilde{x}_t^{(i)}) \cdot p(\tilde{x}_t^{(i)}|x_{t-1}^{(i)})}{\pi(\tilde{x}_t^{(i)}|\tilde{x}_{0:t-1}^{(i)}, y_{1:t})} \quad (8)$$

where $\{\tilde{x}_t^{(i)}\}_{i=1}^N$ is a set of $N$ random samples drawn from the importance distribution $\pi(\tilde{x}_t^{(i)}|\tilde{x}_{0:t-1}^{(i)}, y_{1:t})$. Also, defining normalized weights

$$\mathcal{W}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}} \quad (9)$$

then the posterior distribution can be approximated by the expression described in Eq (4).

### 2.1.2. Resampling

When the updating process begins, a tendency to increase the variance of particles is seen, setting negligible weights to some of them. These particles become useless as they track low probability paths of the state vector. In order to solve this problem, a *resampling* step is incorporated, which leads to the SIR algorithm.

An analytical expression for measuring how degenerated are the particles is given by the *effective particle sample size* showed in Eq (10).

$$N_{eff}(t) = \frac{N}{1 + Var_{p(\cdot|y_{1:t})}(w(x_t))} \quad (10)$$

As it is not possible to calculate $N_{eff}$, an estimate is given by

$$\hat{N}_{eff}(t) = \frac{1}{\sum_{i=1}^N (\mathcal{W}_t^{(i)})^2} \quad (11)$$

In other words, the resampling step consist of removing small weighted particles while retaining and replicating those of large weights. Thus, whenever $\hat{N}_{eff} \leq N_{thres}$, with $N_{thres}$ a fixed threshold, the depletion of the particles is imminent and resampling must be applied.

**Algorithm 1** SIR Particle Filter

1. Importance Sampling
   **for** $i = 1, \ldots, N$ **do**
   • Sample $\tilde{x}_t^{(i)} \sim \pi(x_t|x_{0:t-1}^{(i)}, y_{1:t})$ and set $\tilde{x}_{0:t}^{(i)} \triangleq (x_{0:t}^{(i)}, \tilde{x}_t^{(i)})$
   • Compute the importance weights $w_t^{(i)} = w_{t-1}^{(i)} \cdot \frac{p(y_t|\tilde{x}_t^{(i)}) \cdot p(\tilde{x}_t^{(i)}|x_{t-1}^{(i)})}{\pi(\tilde{x}_t^{(i)}|\tilde{x}_{0:t-1}^{(i)}, y_{1:t})}$
   • Normalize $\mathcal{W}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}$
   **end for**

2. Resampling
   **if** $\hat{N}_{eff} \geq N_{thres}$ **then**
   **for** $i = 1, \ldots, N$ **do**
   • $x_{0:t}^{(i)} = \tilde{x}_{0:t}^{(i)}$
   **end for**
   **else**
   **for** $i = 1, \ldots, N$ **do**
   • Sample an index $j(i)$ distributed according to the discrete distribution satisfying $P(j(i) = l) = \mathcal{W}_t^{(i)}$ for $l = 1, \ldots, N$
   • $x_{0:t}^{(i)} = \tilde{x}_{0:t}^{j(i)}$ and $w_t^{(i)} = \frac{1}{N}$
   **end for**
   **end if**

In general, the SIR particle filter is divided into two steps. Firstly, a *prediction* is done using the state transition model to generate the prior distribution $p(x_k|x_{k-1})$. Then an *update* step is done to modify the particle weights through the likelihood $p(y_k|x_k)$. If the resulting particles are degenerated, a $resampling$ step is added, as it was shown previously.

## 2.2. Multiple imputations

Missing data is a problem that may be treated mainly from two perspectives. On the one hand, *single imputation* techniques fill the incomplete data set imputing single values at each missing datum. The advantage of this perspective is that it allows standard complete data methods to be used. However, these techniques fail due to the lack of uncertainty characterization of both, the sampling variability and the uncertainty associated with the imputation model. On the other hand, the idea of *multiple imputations* retains the advantages of single imputation techniques and also accounts for the uncertainty of the missing mechanism. Multiple imputations (Rubin, 1987) consist of creating multiple complete data sets imputing $m$ values for each missing datum so that sampling variability around the actual values is incorporated for performing valid inferences. Nevertheless, multiple imputations

has disadvantages like the need of drawing more imputations and larger memory space for storing and processing multiple-imputed data sets.

An important issue is the task of choosing the right number of imputations (Graham, Olchowski, & Gilreath, 2007). Obviously, the computational cost is higher as the number of imputations increases. In this regard, (Rubin, 1987, p. 114) shows that an approximation of efficiency for an estimate is given by

$$(1 + \frac{\gamma}{m})^{-1/2} \tag{12}$$

in units of standard errors, where $m$ is the number of imputations and $\gamma$ is the fraction of missing information in the estimation. Consequently, excellent results may be obtained using only few imputations ($m = 3, 4, 5$).

### 2.3. Multiple Imputation Particle Filter

Originally introduced by (Housfater et al., 2006), the Multiple Imputation Particle Filter extends the PF algorithm by incorporating a multiple imputation (MI) procedure for cases where measurement data is not available, so that the algorithm can include the corresponding uncertainty into the estimation process. The main statistical assumption in this approach is that the missing mechanism is *Missing at Random (MAR)*, thus, it does not depend on the missing measures given the observed ones.

For readability, a change in notation is necessary. As it was stated in (Housfater et al., 2006), lets denote now the measurements as a partitioned vector $U_t = (Z_t, Y_t)$, where $Z_t$ corresponds to the missing part and $Y_t$ is from now on the observed part. Then, the MI PF algorithm performs the same as the SIR PF except that there are missing measures. In this case, a MI strategy is adopted.

An imputation model expressed as a probability distribution $\phi$ is required for drawing $m$ samples -imputations-, that is

$$z_t^j \sim \phi(z_t|y_{1:t}) \tag{13}$$

where $j = \{1, \ldots, m\}$ denotes the imputation index. Similarly to importance sampling, each imputation is associated with a weight $p_t^j$ holding the condition $\sum_{j=1}^{m} p_t^j = 1$. Acording to (Liu, Kong, & Wong, 1994), the filtering posterior distribution may be expressed as

$$p(x_t|y_{1:t}) = \int p(x_t|u_{1:t-1}, y_t)p(z_t|y_{1:t})dz_t. \tag{14}$$

By performing a Monte Carlo approximation yields

$$p(x_t|y_{1:t}) \simeq \sum_{j=1}^{m} p_t^j p(x_t|u_{1:t-1}, u_t^j), \tag{15}$$

where $u_t^j = (z_t^j, y_t)$ are complete data sets formed from imputed values. Additionally, by particle filtering each of these data sets yields

$$p(x_t|u_{1:t-1}, u_t^j) \approx \sum_{i=1}^{N} w_t^{(i,j)} \delta(x_t - x_t^{(i,j)}), \tag{16}$$

where the indexes $i$ and $j$ indicate the particle and the imputation, respectively. Thus, an approximation of the desired posterior distribution is

$$p(x_t|y_{1:t}) \approx \sum_{j=1}^{m} \sum_{i=1}^{N} p_t^j w_t^{(i,j)} \delta(x_t - x_t^{(i,j)}). \tag{17}$$

## 3. MULTIPLE-IMPUTATION-BASED UNCERTAINTY CHARACTERIZATION FOR SOC ESTIMATION

The SOC is conceived as a quantification of the available energy stored regarding the actual rated capacity, but as a percentage. It conforms an important feature to address for systems' autonomy when they are energized by ESDs, either as main sources or as a backup. As it is not possible to directly measure the SOC, estimation and prognosis algorithms must be addressed for getting valid predictions from usually noisy measurements like current, voltage and temperature, while carrying out a proper management of the system. Actually, knowledge about it is essential for control of autonomous systems where the End-of-Discharge (EoD) time plays a key role.

According to (Orchard et al., 2012), a wide variety of methods have been proposed in the literature for modeling batteries in offline applications; e.g., electrochemical models. Other methods, more suitable for online implementations, are based on open-circuit voltage (OCV) representations. These methods relate directly the SOC and measured voltage but requires large resting periods for batteries, being inefficient for online estimation. The *"Electrochemical Impedance Spectroscopy"* (EIS) method requires costly equipment, being infeasible for practical applications. In this regard, research efforts have focused on developing estimation and prognosis algorithms based on phenomenological relations through fuzzy logic, neural networks and Bayesian frameworks (Orchard et al., 2012), among others. The main problem in all these cases is that these approaches assume complete data sets for state/parameter estimation purposes.

### 3.1. State-Space Model for Lithium-Ion Batteries

One of the main advantages of adopting a particle-filtering approach for estimation under noisy measurement data is that prior knowledge about the systems dynamics can be directly incorporated into the model as well as its associated uncertainties. Also, it is possible to capture critical physical phenomenology directly into the state-space form, relating it to an observation model which enables the convergence to the

true estimates through the likelihood of sequential measurement data.

Proposed by (Pola, 2014), the state-space model for lithium-ion battery cells used is the following.

**State transition model**

$$x_1(t+1) = x_1(t) + w_1(t) \quad (18)$$

$$x_2(t+1) = x_2(t) - \frac{v(t) \cdot i(t) \cdot \Delta t}{E_{crit}} + w_2(t) \quad (19)$$

**Measurement equation**

$$\begin{aligned} v(t) = v_L + (v_0 - v_L) \cdot e^{\gamma \cdot (x_2(t)-1)} + \alpha \cdot v_L \cdot \dots \\ \dots (x_2(t) - 1) + (1 - \alpha) \cdot v_L \cdot (e^{-\beta} - \dots \\ \dots e^{-\beta \cdot \sqrt{x_2(t)}}) - i(t) \cdot x_1(t) + \eta(t) \end{aligned} \quad (20)$$

where $w_1(t) \sim \mathcal{N}(0, \sigma_1)$ and $w_2(t) \sim \mathcal{N}(0, \sigma_2)$ correspond to additive white Gaussian noise and $\eta(t) \sim \mathcal{N}(0, \sigma_{obs})$ is also a normal distributed random variable accounting for measurement uncertainties. The sample time $\Delta t[sec]$ and the current $i(t)[A]$ are considered input variables whereas the battery voltage $v(t)[V]$ is considered the system's output. The state variables $x_1(t)$ and $x_2(t)$ are chosen strategically under physical meaning as the internal resistance and the SOC, respectively. Finally, as the SOC is expressed as a percentage of energy, $E_{crit}$ represents a normalizing constant whose units are $[V A\ sec]$. All other model parameters are assumed to be known constants within each battery discharge cycle. Their values are obtained by following the procedure described in (Pola, 2014), and applying it to data that should be obtained from a complete discharge cycle at constant (nominal) discharge current.

### 3.2. Implementation of a Multiple Imputation Strategy

(Orchard et al., 2012) proposed a detailed procedure for estimation and prognosis for the SOC. However, what happens when sudden disconnections (or data losses) affect sensors' performance? Perhaps, SOC estimates may be eventually biased, affecting deeply the whole estimation stage and providing invalid information, and the system's autonomy would no longer be guaranteed. In this regard, a new approach from the Multiple Imputation Theory is proposed for uncertainty characterization in particle-filtering-based SOC estimators where voltage measurements are missing during extended periods of time (while discharge current measurements are always available). Future work will focus on the case when battery discharge current measurements are lost instead.

The Multiple Imputation Particle Filter uses voltage imputations in a different manner, depending on which stage of the filtering procedure is currently being applied. During the *prediction stage*, and if past voltage measurements are missing, the multiple-imputation algorithm suggests to draw voltage values from a proposal distribution $\phi$. Each one of these imputations will define a different prior distribution for the next time instant, since $x_2(t+1)$ depends on $v(t)$ in Eq (19). However, as the transition equations place particles in different positions of the state-space, Rubin's rule of multiple imputation theory suggests that all those prior transition distributions should be combined into a single distribution by appropriate weighting (Rubin, 1987), yielding an increase in particle population.

Assuming that the prior distribution is known and the actual voltage value is unknown, then voltage imputations may also be considered for the *update stage* of the particle-filtering algorithm. Furthermore, in that case the resulting particles (which represent the posterior distribution) will keep the same location within the state-space. Thus, the number of particles is not increased since Rubin's rule (Rubin, 1987) is applied.

As multiple-imputed data generate an increase of the number of particles during the *prediction stage*, a *reduction stage* has to be incorporated into the algorithm to keep a fixed number of particles throughout time; avoiding a progressive increase of the particle population. This way, the SIR PF will work as it was originally designed, specially after voltage measurements are once more available.

The proposed MI PF implementation treats the problem of missing voltage observations, whereas the discharge current is assumed as an input variable known at each time instant. The imputation model adopted is defined as the probability distribution induced by Eq (20), providing prior knowledge about the voltage variability.

Denoting the multiple-imputed measurement data set as $\tilde{y}^j_{1:t} = \{\tilde{y}_{1:t-1}, y^j_t\}$ where $\tilde{y}_{1:t} = \{\tilde{y}_{1:t-1}, y^1_t, \dots, y^m_t\}$, with $j \in \{1, \dots, m\}$, the MI PF implementation is summarized in Algorithm 2.

## 4. EXPERIMENTAL RESULTS

In this article, the proposed multiple-imputation algorithm is applied to the case of SOC estimation in Li-Ion battery cells. Particularly, this method is intended to improve the way SOC is monitored on a BMS. A complete discharge cycle, containing a total of 2920 samples that were obtained from an experimental setup located at the Advanced Control Systems Laboratory, University of Chile, is analyzed for purposes of algorithm test and validation. To test the algorithm, a loss of 1000 sequential voltage measurements is considered. The estimation results are obtained with the use of 60 particles and 10 imputations. The performance is analyzed considering the average of 30 realizations for three different cases: i) SIR PF with a complete data set, ii) 1000-step prediction procedure along the missing measurements, and iii) MI PF with an incomplete data set.

**Algorithm 2** Multiple Imputation Particle Filter

1. MI Importance Sampling
**if** $y_t$ is available **then**
• SIR PF
**else**
**for** $j' = 1, \ldots, m$ **do**
**for** $i = 1, \ldots, N$ **do**
• Sample $\tilde{x}_t^{(i,j')} \sim \pi(x_t | x_{0:t-1}^{(i)}, \tilde{y}_{t-1}^{j'})$ and set $\tilde{x}_{0:t}^{(i,j')} \triangleq (x_{0:t}^{(i)}, \tilde{x}_t^{(i,j')})$
**end for**
**end for**
• Compute $m$ imputations $y_t^j \sim \phi(\{\tilde{x}_t^{(i,j')}, w_t^{(i,j')}\}, \eta_t)$ and its associated weights $p_t^j$.
• Reduce the particle population from $N \cdot m$ to $N$.
$\tilde{x}_t^{(i,j)} \rightarrow \tilde{x}_t^{(i)}$
**for** $j = 1, \ldots, m$ **do**
**for** $i = 1, \ldots, N$ **do**
• Compute the importance weights
$w_t^{(i,j)} = w_{t-1}^{(i)} \cdot \frac{p(y_t^j | \tilde{x}_t^{(i)}) \cdot p(\tilde{x}_t^{(i)} | x_{t-1}^{(i)})}{\pi(\tilde{x}_t^{(i)} | \tilde{x}_{0:t-1}^{(i)}, \tilde{y}_{1:t}^j)}$
• Apply Rubin's rule
$w_t^{(i)} = \sum_{j=1}^{m} w_t^{(i,j)}$
• Normalize
$\mathcal{W}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^{N} w_t^{(i)}}$
**end for**
**end for**
**end if**

The probability density that was used in this case to draw voltage imputations corresponds to the distribution induced by Eq (20), where particles are obtained from the prior transition PDF shown in Eq. (18)-(19). The imputations are randomly drawn from the aforementioned distribution, and hence their weights are assumed to be equal.

In particular, the problem of reducing the number of particles from $N \cdot m$ to $N$ -where $N$ is the size of the original particle population and $m$ is the number of imputations- could be achieved by resampling. However, this kind of technique fails because of the tendency to retain high probability particles only, discarding the uncertainty characterization provided by the MI strategy. Therefore, a suboptimal solution is proposed. The main focus consists on preserving the probability distribution described by $N \cdot m$ particles using only $N$ of them. Thus, as an attempt to solve this problem, the particles are arranged as a function of the SOC ($\{x_k^{(l)}, \mathcal{W}_k^{(l)}\}_{l=1}^{N \cdot m}$) and clustered into groups of $m$ particles, noting that the SOC corresponds to a state and its dynamic is described in Eq (19). One particle is obtained from each group by a weighted sum and its probability is assumed to be the sum of probabilities of each particle in the group. Therefore, the $N$ new particles are generated as

$$\bar{\mathcal{W}}_k^{(i)} = \sum_{l=m\cdot(i-1)+1}^{m\cdot i} \mathcal{W}_k^{(l)} \quad (21)$$

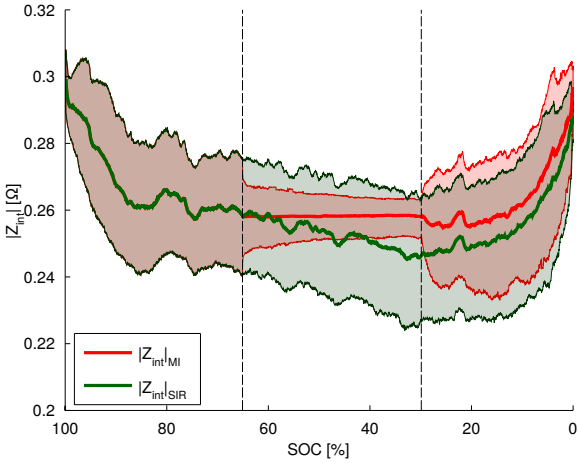
$$\bar{x}_k^{(i)} = \frac{1}{\bar{\mathcal{W}}_k^{(i)}} \cdot \sum_{l=m\cdot(i-1)+1}^{m\cdot i} \mathcal{W}_k^{(l)} x_k^{(l)} \quad (22)$$

$\forall\ i \in \{1, \ldots, N\}$. The biggest assumption adopted for the reduction stage was that the internal impedance remains constant at least when the battery's SOC is over 20%, which in practice makes it almost independent of the SOC. Of course, other factors also affect the value of the internal impedance, for example the battery temperature. In fact, that is the main reason why this parameter has to be estimated from voltage and discharge current measurements. The impact of these factors will not be considered in this particular version of the algorithm, but they will be included as part of future research work.

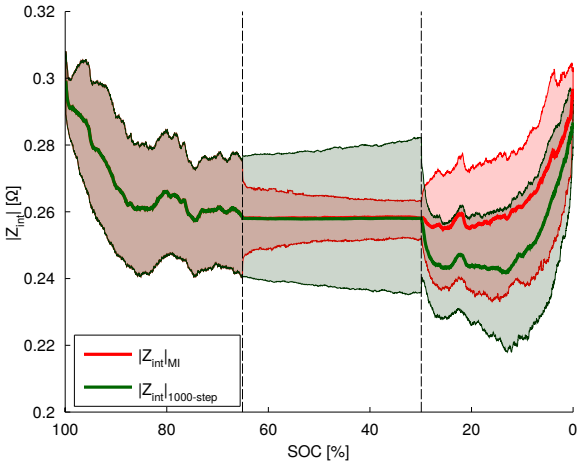
For this case study, the conventional SIR PF is applied in all the cases as long as there are no missing measurements. The focus lays on comparing the MI strategy to a simple n-step ahead prediction algorithm (Orchard et al., 2012) that could be applied when voltage measurements are lost. Also, the MI strategy will be compared to the PF-based estimates that are obtained with no missing data. Both comparisons yield results for internal impedance, SOC and voltage which are exposed in Figures 1, 2 and 3, respectively. For a better analysis, the same conditions are adopted for all the cases up to the time where data starts being lost.

As it is shown in Figure 1a, the assumption of a constant value for the internal impedance becomes invalid along the missing voltage window as the MI PF estimates differ significantly from a complete data estimation, out bounding the confidence intervals depicted by the MI PF estimation. In contrast, Figure 1b shows that the MI PF estimates are very similar to that of the 1000-step prediction but the uncertainties in this case differ among themselves mainly due to the hypothesis in the reduction stage of the MI PF, that leads to a uncertainty dimishment.

Regardless of what it has been mentioned before, the main feature of the proposed MI PF is ensuring a robust and bounded uncertainty characterization for the SOC, which is visualized in Figure 2. Figure 2a shows how the MI PF uncertainty overlaps that of the SIR PF whereas in Figure 2b this last is slightly overlapped by the uncertainty of the 1000-step prediction. It is interesting to note the MI strategy avoids the use of a resampling stage, yielding similar results as a long term prediction. Nevertheless, when voltage measurements are not lost anymore, a bias is added in both cases (MI PF and 1000-step prediction). This problem is generated by the approximately constant estimation for the internal impedance, which

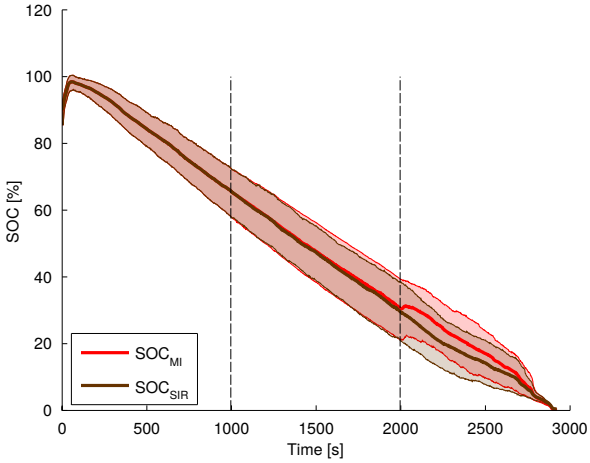



(a) MI and SIR PF

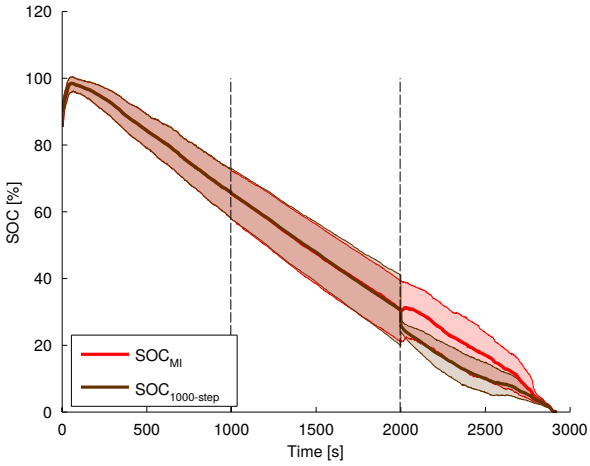



(b) MI PF and Prediction

Figure 1. Internal impedance estimation as a function of the SOC[%] for a disconnection of 1000 sequential voltage measurements denoted in the area between the dashed vertical lines. a) Comparison between the MI PF (red line) and the SIR PF (green line) with 95% confidence intervals. b) Comparison between the MI PF (red line) and the 1000-step prediction algorithm (green line) with 95% confidence intervals.




(a) MI and SIR PF




(b) MI PF and Prediction

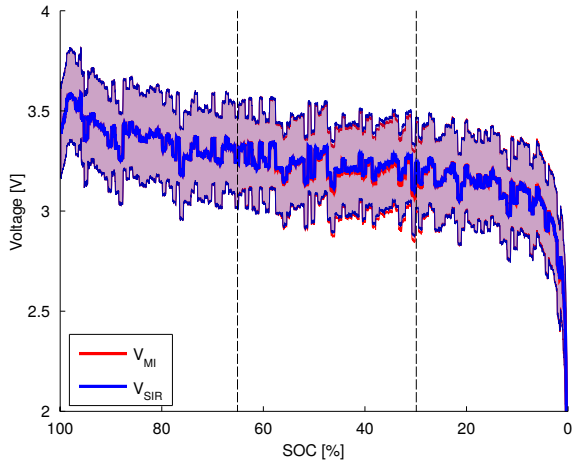
Figure 2. SOC estimation as a function of time[s] for a disconnection of 1000 sequential voltage measurements denoted in the area between the dashed vertical lines. a) Comparison between the MI PF (red line) and the SIR PF (brown line) with 95% confidence intervals. b) Comparison between the MI PF (red line) and the 1000-step prediction algorithm (brown line) with 95% confidence intervals.

introduces a bias affecting the SOC estimation as an attempt to correct the first. Notwithstanding, the uncertainty about the actual value of the internal impedance for the 1000-step prediction affects more intensively its performance when voltage is measured again than that of the MI PF. Consequently, the MI PF approach is the one who experiences better performance.

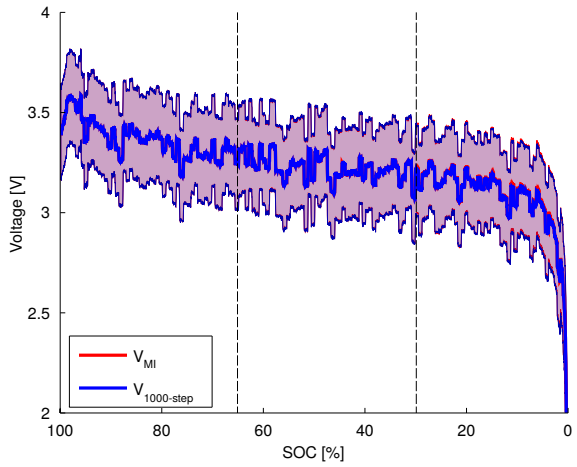
The underlying importance of holding a bounded uncertainty characterization on an estimation stage is that of providing appropiate conditions for a prognosis stage. Simultaneously, this converges into an improved performance of prognostic results due to a bounded uncertainty along the prediction horizon, hence predictions are obtained with a higher degree of certainty.

In the case of voltage estimation, the results are shown in Figure 3. Figure 3a shows that a bias is added to the voltage distribution corresponding to the MI PF. Note that it is obtained from using an imputation model based on the measurement model. The use of a few imputations (10 in this case study) provides a reasonable characterization of the output variability by generating a robust approximation to the true statistics even when data is partially lost. The bias remains negligible considering that the total amount of lost data reaches 1000. However, Figure 3b shows that the 1000-step

prediction shares its behavior, by describing nearly identical curves.




(a) MI and SIR PF




(b) MI PF and Prediction

Figure 3. Voltage estimation as a function of the SOC[%] for a disconnection of 1000 sequential voltage measurements denoted in the area between the dashed vertical lines. a) Comparison between the MI PF (red line) to the SIR PF (blue line) with 95% confidence intervals. b) Comparison between the MI PF (red line) and the 1000-step prediction algorithm (blue line) with 95% confidence intervals.

## 5. CONCLUSION

A new multiple-imputation particle-filtering based scheme for estimation when lost measurements are present is proposed where the Multiple Imputation Theory is the main core for uncertainty characterization. A particular implementation for SOC estimation is presented when voltage measures are sequentially lost along a period of time. Preliminary results show the success of the methodology by incorporating uncertainty by increasing the original number of particles, but then adding a reduction stage. However, a bias is added to the estimation process.

The case study for testing the algorithm includes a missing data window when the SOC is over a 20% of the battery's capacity. This allows the adoption of a simplified way for reducing particles in the algorithm based on the hypothesis that the value of the internal impedance remains constant. The MI strategy is compared to the case without missing data and also to a particle-filtering-based prognosis algorithm for performing a 1000-step prediction. The results show that the uncertainty characterization associated to the estimation stage -once the capacity to acquire data is no longer lost- is more appropiate if the MI PF is used than if the 1000-step prediction is used.

As the MI has been developed for offline applications, there are several aspects to consider for online applications. Some of them include improvements on the imputation model, adaptive estimation for an optimal number of particles and amount of imputations, alternative reduction methods of particle population, better ways for characterizing the internal impedance evolution in time, risk assessment, among others. Furthermore, the development of an optimal particle reduction may enable the connection of asynchronous networks, treatment for missing measurements, and prognosis, to give some examples.

## ACKNOWLEDGMENTS

The authors would like to thank FONDECYT 1140774 and University of Costa Rica (Grant for Doctoral Studies) for their financial support.

## REFERENCES

Andrieu, C., Doucet, A., & Punskaya, E. (2001). *Sequential monte carlo methods in practice* (A. Doucet, N. de Freitas, & N. Gordon, Eds.). Springer-Verlag.

Candy, J. (2009). *Bayesian signal processing: Classical, modern and particle filtering methods*. Wiley.

Crisan, D., & Doucet, A. (2002). A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, *50*(3), 736-746.

Doucet, A., Godsill, S., & Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, *10*(2), 197-208.

Graham, J., Olchowski, A., & Gilreath, T. (2007). How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206-213.

Housfater, A., Zhang, X., & Zhou, Y. (2006). Nonlinear fusion of multiple sensors with missing data. *IEEE International Conference on Acoustics, Speech and Signal*

*Processing*, *4*, 961-964.

Liu, J., Kong, A., & Wong, W. (1994). Sequential imputations and bayesian missing data problems. *Journal of the American Statistical Association*, *89*(425), 278-288.

Orchard, M., Cerda, M., Olivares, B., & Silva, J. (2012). Sequential monte carlo methods for discharge time prognosis in lithium-ion batteries. *International Journal of Prognostics and Health Management*, *3*, 1-12.

Orchard, M., & Vachtsevanos, G. (2009). A particle-filtering approach for on-line fault diagnosis and failure prognosis. *Transactions of the Institute of Measurement and Control*, *31*, 221-246.

Pattipati, B., Sankavaram, C., & Pattipati, K. (2011). System identification and estimation framework for pivotal automotive battery management system characteristics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *41*(6), 869-884.

Pola, D. (2014). *An improved prognosis strategy with temperature-dependent state space models for the analysis of the state-of-health and state-of-charge in lithium-ion batteries*. M.Sc. Tesis. Department of Electrical Engineering, Universidad de Chile.

Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.

## BIOGRAPHIES

**B.Sc. David E. Acuña** was born in Santiago, Chile. He received his B.Sc. degree in Electrical Engineering (2014) from Universidad de Chile. He currently is a Research Assistant at the Lithium Innovation Center (Santiago, Chile). His research interests include signal processing, nonlinear filtering, and prognostics and health management for energy storage devices.

**Dr. Marcos E. Orchard** is Associate Professor with the Department of Electrical Engineering at Universidad de Chile and was part of the Intelligent Control Systems Laboratory at The Georgia Institute of Technology. His current research interest is the design, implementation and testing of real-time frameworks for fault diagnosis and failure prognosis, with applications to battery management systems, mining industry, and finance. His fields of expertise include statistical process monitoring, parametric/non-parametric modeling, and system identification. His research work at the Georgia Institute of Technology was the foundation of novel real-time fault diagnosis and failure prognosis approaches based on particle filtering algorithms. He received his Ph.D. and M.S. degrees from The Georgia Institute of Technology, Atlanta, GA, in 2005 and 2007, respectively. He received his B.S. degree (1999) and a Civil Industrial Engineering degree with Electrical Major (2001) from Catholic University of Chile. Dr. Orchard has published more than 50 papers in his areas of expertise.

**Dr. Jorge F. Silva** is Assistant Professor at the Department of Electrical Engineering, University of Chile, Santiago, Chile. He received the Master of Science (2005) and Ph.D. (2008) in Electrical Engineering from the University of Southern California (USC). He is IEEE member of the Signal Processing and Information Theory Societies and he has participated as a reviewer in various IEEE journals on Signal Processing. Jorge F. Silva is recipient of the Outstanding Thesis Award 2009 for Theoretical Research of the Viterbi School of Engineering, the Viterbi Doctoral Fellowship 2007-2008 and Simon Ramo Scholarship 2007-2008 at USC. His research interests include: non-parametric learning; sparse signal representations, statistical learning; universal source coding; sequential decision and estimation; distributive learning and sensor networks.

**M. Sc. Aramis Pérez** is a Research Assistant at the Lithium Innovation Center (Santiago, Chile) and professor at the School of Electrical Engineering at the University of Costa Rica. He received his B.Sc. degree (2002) and Licentiate degree (2005) in Electrical Engineering from the University of Costa Rica. He received his M. Sc. degree in Business Administration with a General Management Major (2008) and also he is a M. Sc. candidate in Industrial Engineering from the same university. Currently he is a doctorate student at the Department of Electrical Engineering at the University of Chile under Dr. Marcos E. Orchard supervision. His research interests include parametric/non-parametric modeling, system identification, data analysis, machine learning and manufacturing processes.