

Bayesian inference in non-Markovian state-space models with applications to battery fractional order systems

Pierre E. Jacob, S.M. Mahdi Alavi, Adam Mahdi, Stephen J. Payne and David A. Howey

Abstract—Battery impedance spectroscopy models are given by fractional order (FO) differential equations. In the discrete-time domain, they give rise to state-space models where the latent process is not Markovian. Parameter estimation for these models is therefore challenging, especially for non-commensurate FO models. In this paper, we propose a Bayesian approach to identify the parameters of generic FO systems. The computational challenge is tackled with particle Markov Chain Monte Carlo methods, with an implementation specifically designed for the non-Markovian setting. Two examples are provided. In a first example, the approach is applied to identify a battery commensurate FO model with a single constant phase element (CPE) by using real data. We compare the proposed approach to an instrumental variable method. Then we consider a non-commensurate FO model with more than one CPE and synthetic datasets, investigating how the proposed method enables the study of various effects on parameter identification, such as the data length, the magnitude of the input signal, the choice of prior, and the measurement noise.

Index Terms—Parameter estimation, System identification, Bayesian Inference, Fractional order systems, Batteries.

I. INTRODUCTION

Fractional Order (FO) models are important in the study of electrochemical and biological systems [1].

P.E. Jacob is with the Department of Statistics, Harvard University, Science Center 7th floor, 1 Oxford Street, Cambridge, MA 02138-2901, USA. Email: pjacob@fas.harvard.edu

S.M.M. Alavi was with the Energy and Power Group, Department of Engineering Science, University of Oxford. He is now with the Brain Stimulation Engineering Laboratory, Department of Psychiatry & Behavioral Sciences, Duke University, Durham, NC 27710, USA. Email: mahdi.alavi@duke.edu

A. Mahdi and S.J. Payne are with the Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Old Road Campus Research Building, Oxford, OX3 7DQ, United Kingdom. Emails: {adam.mahdi, stephen.payne}@eng.ox.ac.uk

D.A. Howey is with the Energy and Power Group, Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, United Kingdom. Email: david.howey@eng.ox.ac.uk

The source code is available online at <https://github.com/pierrejacob/BatteryMCMC>

Certain features in FO models make their identification challenging. Before reviewing specific issues, we recall that a continuous-time state-space model of a FO system is given by [2]:

$$\begin{aligned} \frac{d^\alpha x(t)}{dt^\alpha} &= \bar{A}(\beta) x(t) + \bar{B}(\beta) u(t), \\ y(t) &= M(\beta) x(t) + D(\beta) u(t), \end{aligned} \quad (1)$$

where $x \in \mathbb{R}^n$ is the state vector; $u \in \mathbb{R}$ and $y \in \mathbb{R}$ are input and output signals, respectively; $\bar{A}(\beta) \in \mathbb{R}^{n \times n}$, $\bar{B}(\beta) \in \mathbb{R}^{n \times 1}$, $M(\beta) \in \mathbb{R}^{1 \times n}$ and $D(\beta) \in \mathbb{R}$ are system matrices which depend on the parameter vector $\beta \in \mathbb{R}^q$ to be identified. Moreover,

$$\frac{d^\alpha x(t)}{dt^\alpha} = \left[\frac{d^{\alpha_1} x_1(t)}{dt^{\alpha_1}}, \dots, \frac{d^{\alpha_n} x_n(t)}{dt^{\alpha_n}} \right] \quad (2)$$

is the vector of FO derivatives, with unknown FOs denoted by $\alpha_\ell \in (0, 1)$, $\ell = 1, \dots, n$.

By defining the parameter vector as

$$\theta = [\alpha_1 \quad \dots \quad \alpha_n \quad \beta], \quad (3)$$

the corresponding model in the discrete-time domain is given by [2]:

$$\begin{aligned} x_{k+1} &= \sum_{j=0}^k A_j(\theta) x_{k-j} + B(\theta) u_k, \\ y_k &= M(\theta) x_k + D(\theta) u_k, \end{aligned} \quad (4)$$

with

$$\begin{aligned} A_0(\theta) &= \text{diag}\{\alpha_1, \dots, \alpha_n\} + \\ &\quad \text{diag}\{T_s^{\alpha_1}, \dots, T_s^{\alpha_n}\} \bar{A}(\beta), \\ A_j(\theta) &= (-1)^j \text{diag} \left\{ \binom{\alpha_1}{j+1}, \dots, \binom{\alpha_n}{j+1} \right\}, \\ &\quad \text{for } 1 \leq j \\ B(\theta) &= \text{diag}\{T_s^{\alpha_1}, \dots, T_s^{\alpha_n}\} \bar{B}(\beta), \end{aligned} \quad (5)$$

where T_s is the sample time, $k \in \mathbb{N}$ is the time index, $\text{diag}\{\cdot\}$ denotes the diagonal matrix and $\binom{\alpha_\ell}{j}$ is the binomial coefficient given by

$$\binom{\alpha_\ell}{j} = \frac{\Gamma(\alpha_\ell + 1)}{\Gamma(j+1)\Gamma(\alpha_\ell + 1 - j)}, \quad (6)$$

where, $\Gamma(\cdot)$ denotes the gamma function

$$\Gamma(\alpha_\ell) = \int_0^\infty z^{\alpha_\ell-1} e^{-z} dz, \text{ for } \alpha_\ell \in \mathbb{C} \text{ with } \Re(\alpha_\ell) > 0,$$

where \Re denotes the real part of a complex number.

Definition 1: an FO system is said to be commensurate if for all $\ell \in \{1, \dots, n\}$, there exists $\rho \in \mathbb{N}$, such that $\alpha_\ell = \rho\alpha$, where $\alpha \in \mathbb{R}$; otherwise it is said to be non-commensurate [1]. \square

Some issues associated with parameter estimation of FO systems are now more evident: 1) the state vector x_{k+1} depends on all the past states x_0 up to x_k , therefore the FO system is non-Markovian¹, and 2) the model is nonlinear, with respect to the parameters. Therefore, FO systems constitute a non-standard class of state-space models, for which there is no generic parameter estimation method.

Several least-squares estimation methods have been proposed in [3]–[5] for the identification of continuous-time FO transfer functions of the form

$$F(s) = \frac{Y(s)}{U(s)} = \frac{b_m s^{\beta_m} + \dots + b_1 s^{\beta_1} + b_0 s^{\beta_0}}{a_n s^{\alpha_n} + \dots + a_1 s^{\alpha_1} + 1}, \quad (7)$$

where U and Y are Laplace transforms of the input and output (observation) signals, respectively. These methods have been modified in [6], [7], and developed into the Crone toolbox [8], [9]. The Crone toolbox is mainly based on the instrumental variable filter (ivsvf) method and the simplified refined instrumental variable for continuous-time fractional (srivcf) method, which are both based on instrumental variable concepts [10]–[12]. In these methods, the FO model is simplified by approximating all fractional differentiation operators by higher-order rational transfer functions [6]. The coefficients of the transfer function are identified by using the coefficient map that exists between the original and approximated models. Manual search [6], gradient descent [7], or interior-point [13] optimizations are combined with the ivsvf and srivcf functions for the estimation of the fractional orders. In the non-commensurate case, the approximate model is of high order so that the coefficient map between the original and approximated model may be intractable. This issue highlights the need for novel tools to directly identify general FO models.

In [14], an identification method based on swarm optimization has been proposed to identify a battery non-commensurate FO model.

¹In Markovian systems, x_{k+1} can be written as functions of x_k and inputs.

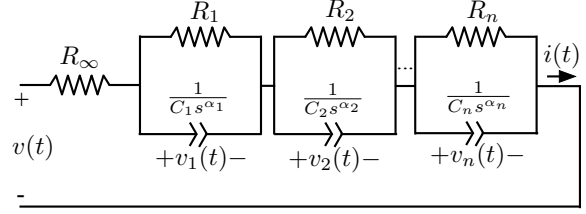


Fig. 1. The general battery electrochemical impedance spectroscopy model.

In this paper, a novel method based on Bayesian inference is presented. Bayesian inference provides a comprehensive framework for statistical analysis, including the investigation of parameter identifiability. Its main challenge is computational and often involves Markov chain Monte Carlo methods (MCMC) to approximate the posterior distribution. Advances in MCMC methods have made Bayesian inference possible for generic state space models [15]. We will develop this approach for non-Markovian models.

Recently, there has been a significant interest in the design of model-based battery systems to improve the efficiency and reliability of electric vehicles and renewable energies [16]. Among the employed models, Electrochemical Impedance Spectroscopy (EIS) models with FO dynamics have received much attention. They are more accurate than conventional lumped models and are computationally less expensive than electrochemical models defined by highly coupled partial differential equations. A comprehensive survey of battery models has been given in [16]. A general schematic of the battery impedance FO models is shown in Figure 1, where i and v denote the battery current and voltage, respectively. R_∞ represents the battery ohmic resistance at high frequencies. Each parallel pair is employed to model diffusion processes (e.g. charge transfer resistance and double layer capacitance) more accurately compared to the lumped models, [17] and [13]. The number of parallel pairs depends on the required accuracy. The terms $C_\ell s^{\alpha_\ell}$, $\ell = 1, \dots, n$, are called constant phase elements (CPEs). In low frequency ranges, the impedance frequency response may show constant phase behaviour such that the associated parallel resistor can be considered as an open circuit. This is referred to as a Warburg term in the literature [17]. Reference [17] provides more information on battery EIS and associated FO models.

By defining $u = i$ and $y = v$, and the voltage across

the CPEs as the state variables,

$$x \triangleq \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}, \quad (8)$$

it is easy to show that A_j , B , M and D in the state-space model (4) are given by:

$$\begin{aligned} A_j(\theta) &= \text{diag} \{a_{1,j}(\theta), \dots, a_{n,j}(\theta)\} \\ B(\theta) &= \begin{bmatrix} b_1(\theta) & \cdots & b_n(\theta) \end{bmatrix} \\ M(\theta) &= \begin{bmatrix} m_1 & \cdots & m_n \end{bmatrix} \\ D(\theta) &= d(\theta), \end{aligned} \quad (9)$$

with

$$\begin{aligned} a_{\ell,0}(\theta) &= \alpha_\ell - \frac{T_s^{\alpha_\ell}}{R_\ell C_\ell}, \quad a_{\ell,j}(\theta) = (-1)^j \binom{\alpha_\ell}{j+1} \\ b_\ell(\theta) &= \frac{T_s^{\alpha_\ell}}{C_\ell}, \quad m_\ell = 1, \quad d(\theta) = R_\infty \end{aligned} \quad (10)$$

for $\ell = 1, \dots, n$ and $j = 1, 2, \dots, T$,

where T is the data length.

In order to estimate the parameter vector θ from v and i , models are typically fitted to frequency domain impedance spectra that are obtained through EIS, [18]–[20]. It may be preferable outside of the laboratory to estimate parameters of FO models directly from time domain data, rather than convert such data into the frequency domain for fitting. The latter would involve a window function, which may introduce bias [13].

In this paper, we explore the use of Bayesian inference to identify battery FO models. First, we compare the proposed approach to an instrumental variable method, on a battery model with a single CPE and real data. Then we consider a model with more than one CPE and synthetic datasets, investigating how the proposed method enables the study of various effects on parameter identification, such as the data length, the magnitude of the input signal, the choice of prior, and the measurement noise. We also note a lack of identifiability for some parameters of the model using the methods and data herein. The source code is available online at <https://github.com/pierrejacob/BatteryMCMC>.

The main contributions of the paper are:

- introduction of a Bayesian framework to investigate the identification of general FO systems,
- reduction of the computational cost of particle filters in non-Markovian settings, using a tree representation of the particles.

II. BAYESIAN INFERENCE

A. Inference in state-space models

Given measurements $y_{0:T} = y_0, \dots, y_T$, we consider the task of estimating the parameter θ in the state-space model

$$\begin{aligned} x_0 &\sim \mu(\cdot \mid \theta), \\ \forall k \in \{1, \dots, T\} \quad x_k &\sim f_k(\cdot \mid x_{0:k-1}, \theta), \\ \forall k \in \{0, \dots, T\} \quad y_k &\sim g_k(\cdot \mid x_k, \theta), \end{aligned} \quad (11)$$

where $T \in \mathbb{N}$ denotes the total number of observations. The transition f_k might depends on all the past states $x_{0:k-1}$, and thus $\{x_0, \dots, x_k\}$ is not necessarily a Markov chain. Both f_k and g_k could be nonlinear functions. Therefore, the FO model (4) is in the state-space model framework (11), upon specifying a state noise and an observation noise. For simplicity, we will consider additive Gaussian noise and the model is thus given by:

$$\begin{aligned} x_{k+1} &= \sum_{j=0}^k A_j(\theta) x_{k-j} + B(\theta) u_k + \sigma_x \varepsilon_k, \\ y_k &= M(\theta) x_k + D(\theta) u_k + \sigma_y \eta_k, \end{aligned} \quad (12)$$

where (ε_k) and (η_k) are sequences of independent standard Gaussian variables, and σ_x, σ_y are positive values.

The likelihood function, defined by $\theta \mapsto p(y_{0:T} \mid \theta)$, where $p(y_{0:T} \mid \theta)$ is the density of the observations evaluated with the dataset $y_{0:T}$ for the parameter θ , can be written

$$\begin{aligned} p(y_{0:T} \mid \theta) &= \int \mu(x_0 \mid \theta) g_0(y_0 \mid x_0, \theta) \times \\ &\quad \prod_{k=1}^T f_k(x_k \mid x_{0:k-1}, \theta) g_k(y_k \mid x_k, \theta) dx_{0:T}. \end{aligned} \quad (13)$$

Maximizing the likelihood function $\theta \mapsto p(y_{0:T} \mid \theta)$, we obtain the classic maximum likelihood estimator. Confidence intervals can be constructed via asymptotic arguments, relying on regularity conditions. The bootstrap method to construct confidence intervals is not readily applicable to state-space model settings. We will see in numerical experiments that the likelihood function associated with FO battery models shows signs of partial non-identifiability, which call for alternative approaches.

Bayesian inference starts from a prior distribution on the parameters θ , and uses the likelihood to obtain a posterior distribution [21], using Bayes formula. In the numerical experiments, we will be particularly interested in the changes between the prior and the

posterior, which inform us about how much has been learned from the data, and thus give a practical notion of identifiability. Indeed, contrarily to point estimation methods, including maximum likelihood and least squares estimators, Bayesian inference delivers a collection of parameter values that represent the landscape of the posterior distribution. This framework enables the investigation of identifiability, by looking at the potential concentration of the posterior on some subset of the parameter space, or the lack thereof.

Bayesian inference for general state-space models, that is, generic choices of μ , f_k and g_k , the integral in (13) can be envisioned with recent advances in Monte Carlo methods that combine MCMC algorithms and particle filters [15]. For our purposes, particle filters [22] are algorithms that provide unbiased estimators of the likelihood in (13), for any $\theta \in \Theta$, up to a multiplicative constant. The user sets a number of particles $N \in \mathbb{N}$. Larger values of N yield lower variance but the computational cost increases linearly with N . The particles are first sampled from $\mu(\cdot | \theta)$, and denoted by $x_0^{1:N} = (x_0^1, \dots, x_0^N)$. Then, each particle is weighted: for each $\ell \in \{1, \dots, N\}$, $w_0^\ell = g_0(y_0 | x_0^\ell, \theta)$. Next, the resampling consists in selecting some particles and discarding others, according to their weights. For instance, one draws for each $\ell \in \{1, \dots, N\}$, an ancestor variable $a_0^\ell \in \{1, \dots, N\}$ according to a categorical distribution with parameters $(w_0^1 / \sum_{j=1}^N w_0^j, \dots, w_0^N / \sum_{j=1}^N w_0^j)$. Once the ancestors $a_0^{1:N}$ are drawn, the new particles are sampled and weighted according to $x_1^\ell \sim q_1(\cdot | x_0^{a_0^\ell}, \theta)$ and

$$w_1^\ell = \frac{f_1(x_1^\ell | x_0^{a_0^\ell}, \theta) g_1(y_1 | x_1^\ell, \theta)}{q_1(x_1^\ell | x_0^{a_0^\ell}, \theta)}.$$

Introduce the paths $\bar{x}_{0:1}^\ell = (x_0^{a_0^\ell}, x_1^\ell)$, for each $\ell \in \{1, \dots, N\}$. Then the weighted paths $(w_1^\ell, \bar{x}_{0:1}^\ell)_{\ell=1}^N$ are approximately distributed according to the path distribution $p(x_0, x_1 | y_0, y_1, \theta)$. The algorithm then proceeds in a similar fashion for the subsequent steps, until all the data have been assimilated. Importantly, the quantity

$$\hat{p}(y_{0:T} | \theta) = \prod_{k=0}^T \left(\frac{1}{N} \sum_{\ell=1}^N w_k^\ell \right) \quad (14)$$

is an unbiased estimator of the likelihood $p(y_{0:T} | \theta)$ [15]. A rich literature is devoted to the theoretical study of the algorithm and the properties of this type of estimator [23], [24]. In the Markovian setting, these

results indicate that, as the number of observations T goes to infinity, the relative variance of the likelihood estimator $\hat{p}(y_{0:T} | \theta)$ can be bounded independently of T if one chooses N proportionally to T . The choice of the proposal distribution $q_k(\cdot | x_{0:k-1}, \theta)$ impacts the variance of the likelihood estimator, and we give details on our choices in Appendix V-A.

B. Particle filters for non-Markovian models

Given the non-Markovianity of the latent process, sampling each particle x_k^ℓ at time k requires computing a function ψ_k of a trajectory, $(x_0^\ell, \dots, x_{k-1}^\ell)$. In the models considered in the article, such as (12), this function takes the form of a weighted sum: $\psi_k(\bar{x}_{0:k}^\ell) = \sum_{t=0}^k \alpha_{k,t} \bar{x}_{k-t}^\ell$, where $(\alpha_{k,t})_{t \geq 0}$, for each k , are coefficients that can be computed given the parameter θ . Naively computing this weighted sum for each particle would yield a cost of order k , at time k ; and thus an overall computational cost of order $N \times T^2$, where T is the number of observations to assimilate and N the number of particles. Furthermore, the naive memory cost of storing all the trajectories would be of order $N \times T$.

We can reduce both the memory cost and the computational cost, by representing the trajectories as branches of a tree [25]. At each step of the particle filter, new leaves are added to the tree, corresponding to the new generation of particles. The ancestors give the list of new branches. Some existing branches of the tree can be cut, if the corresponding particles have been discarded in the resampling step. It was shown in [25] that the number of nodes in the tree is of order $k + CN \log N$ at time k , in expectation. The constant C does not depend on N and T . This theoretical expected cost applies directly to the non-Markovian setting (the assumption in [25] only involves the measurement distribution). Therefore, the memory cost can be reduced from $N \times T$ to $T + CN \log N$. Furthermore, computing the N sums at time k only requires browsing the whole tree once, from the root to the leaves, for a cost of order $k + CN \log N$ at step k . Performing this operation for each k in $\{1, \dots, T\}$ yields an overall computational cost of order $T^2 + CTN \log N$, instead of $N \times T^2$ with the naive implementation. This is to be compared with a computational cost of NT and a memory cost of N for Markovian models. Since N is typically comparable to T , and $\log N$ is a small value, the tree structure gives an order of magnitude improvement, both computationally and in terms of memory, over the standard implementation.

Therefore, the particle filter machinery is directly implementable in the non-Markovian setting [26]. The unbiasedness of the likelihood estimator of (14) holds, as can be easily proven by induction over the time steps, in a slight extension of the Markovian setting [23]. An important open question is how the variance of the likelihood estimator behaves with the number of observations in non-Markovian settings. Our numerical experiments indicate that the algorithm is applicable in realistic scenarios.

Unbiased likelihood estimators $\hat{p}(y_{0:T} | \theta)$ can be used to obtain samples from the posterior distribution, using particle MCMC algorithms [15]. We will use the Particle Marginal Metropolis–Hastings (PMMH) variant. Each iteration of PMMH involves a run of particle filter adapted to the non-Markovian setting. Algorithmic tuning details are given in Appendix V-B.

III. BAYESIAN INFERENCE IN BATTERY SYSTEMS

Two examples are provided. In the first, a commensurate FO model of the battery is estimated using real data, and the results are compared with parameter estimates obtained using an instrumental variable (IV) method from the Crone toolbox. In the second example, we add a Warburg term, thus the model includes two CPEs and becomes non-commensurate.

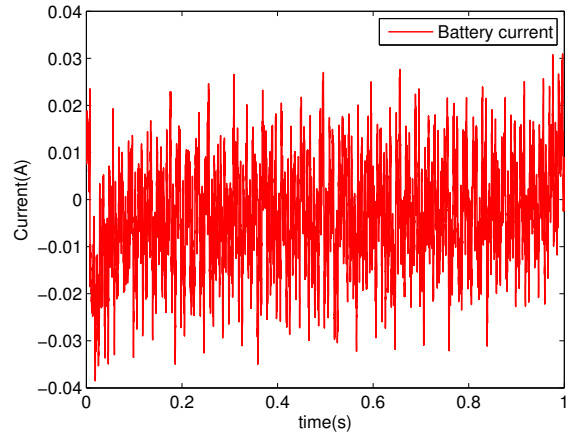
A. Example 1: Identification of a single CPE battery FO model

Consider an FO model with a single CPE without the Warburg term ($n = 1$ in Figure 1). The model is given by (12), where each x_k is univariate and $A_0 = (\alpha_1 - T_s^{\alpha_1}/(R_1 C_1))$, and for $j \geq 1$, $A_j = (-1)^j \binom{\alpha_1}{j+1}$, $B(\theta) = T_s^{\alpha_1}/C_1$, $M(\theta) = 1$, and $D(\theta) = R_\infty$. Our objective is to estimate the parameter vector,

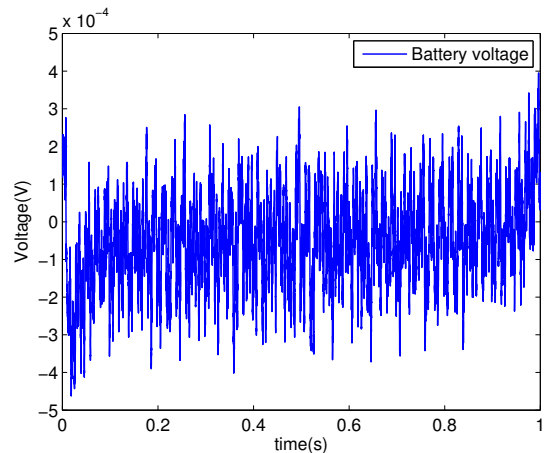
$$\theta = [R_\infty \quad R_1 \quad C_1 \quad \alpha_1],$$

by using the input-output signals u and y through the PMMH approach.

An A123 Systems lithium-ion iron phosphate (LFP) cell was tested, having length 65 mm, diameter 26 mm, nominal capacity 2.3 Ah and nominal voltage 3.3 V. The battery is excited by a multi-sine current u , comprising 80 logarithmically spaced frequencies from 1 Hz to 2 kHz. The output signal y is the battery voltage. The identification data, is recorded for one second, with sampling time $T_s = 1/(20,000)$ s. The data length is then $T = 20,000$. Figure 2 shows the identification data. Reference [27]



(a) Battery current $u(t)$.



(b) Battery voltage $y(t)$.

Fig. 2. The identification data of Example 1.

provides a comprehensive information about the data collection's electronics and signal processing.

In [13], with the same data, we identified the parameter vector by using the IV method (Crone toolbox). Estimation of the parameter vector is given in Table I.

For Bayesian inference, we use a uniform prior on each parameter, with ranges $[1, 50]$ m Ω for R_∞ , $[1, 50]$ m Ω for R_1 , $[1, 5]$ Fcm $^{-2}$ s $^{-\alpha_1}$ for C_1 and $[0.4, 1]$ for α_1 . The intervals are chosen based on the typical EIS curves and the results of the IV method (see [13] and references therein). The initial value of the state vector is set to zero, $x_0 = 0$. We have used $\sigma_x = 10^{-5}$ and $\sigma_y = 2 \times 10^{-5}$ in (12), and we have used the data from times 5,000 to 8,000 instead of the full data to reduce the computational cost. The PMMH

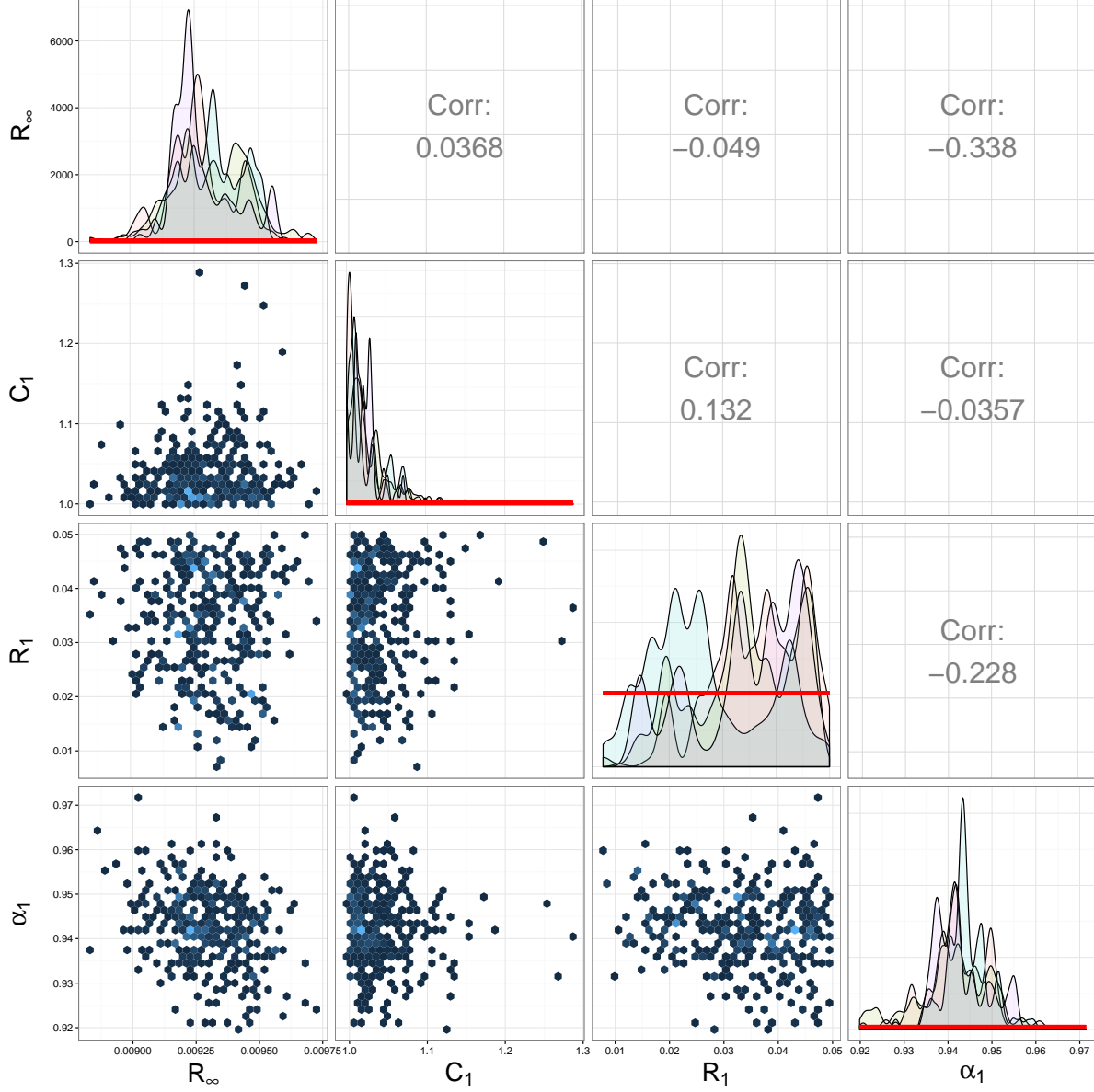


Fig. 3. Results of the PMMH algorithm for Example 1, sampling approximately from the posterior distribution for the model (12). On the diagonal, the posterior density function of each run is overlaid with the prior distribution (red horizontal lines), which is uniform. On the lower triangle, the five runs are pooled together in a histogram with hexagonal binning. On the upper triangle, the correlation coefficients between pairs of parameters are displayed.

TABLE I
ESTIMATION OF PARAMETERS OF THE SINGLE CPE MODEL
USING INSTRUMENTAL VARIABLE (IV), AND 95% CREDIBLE
INTERVALS OBTAINED WITH PMMH, WITH REAL DATA.

Parameter	IV method	PMMH method
R_∞ m Ω	7.40	[9.05, 9.55]
R_1 m Ω	7.50	[13.0, 48.0]
C_1 Fcm $^{-2}$ s $^{-\alpha_1}$	2.04	[1.00, 1.09]
α_1	0.64	[0.93, 0.96]

method was run with $N = 1,024$ particles and 10,000 iterations, with 4 chains in parallel. The tuning of the algorithm is described in details in the appendices. Figure 3 shows the output in a matrix plot. From the diagonal, we see that the posteriors of R_∞ , C_1 and α_1 have concentrated compared to their priors, represented by red curves, while the posterior on R_1 is similar to the prior.

In Table I, we display 95% credible regions computed with quantiles of the posterior samples. The credible regions obtained by PMMH do not contain the IV estimates; this indicates a significant difference. In order to validate the results, we compare both approaches in terms of voltage prediction. To avoid using the data twice, we have run the IV method on the data from time 5,000 to 8,000, and obtained $R_\infty = 6.8$ m Ω , $R_1 = 9.2$ m Ω , $C_1 = 3.7$ Fcm $^{-2}$ s $^{-\alpha_1}$ and $\alpha_1 = 0.54$, which is also outside the 95% regions. For these values, we have generated $y_{1:T}$ according to equation (4) (without adding Gaussian noise), and computed the sum of squared differences between the actual voltage and the generated voltage, from times 8,001 to $T = 20,000$. We have obtained 1.7×10^{-5} . For reference, the IV parameters given in Table I yield 1.5×10^{-5} . On the other hand, lower values are obtained from our posterior samples, with the minimum attained at 6.40×10^{-6} for $R_\infty = 9.3$ m Ω , $R_1 = 35.0$ m Ω , $C_1 = 1.06$ Fcm $^{-2}$ s $^{-\alpha_1}$ and $\alpha_1 = 0.97$. Over the posterior samples, we find a median sum of squares of 7.9×10^{-6} , and a maximum of 9.6×10^{-6} . Note that the minimum value is obtained for an atypical parameter value with respect to the posterior, lying outside the 95% credible regions. Therefore, the Bayesian approach leads to a concentration of the parameters on a set that delivers slightly better predictions compared to parameters obtained with the IV method. On the other hand, the IV method has a computational advantage for large datasets.

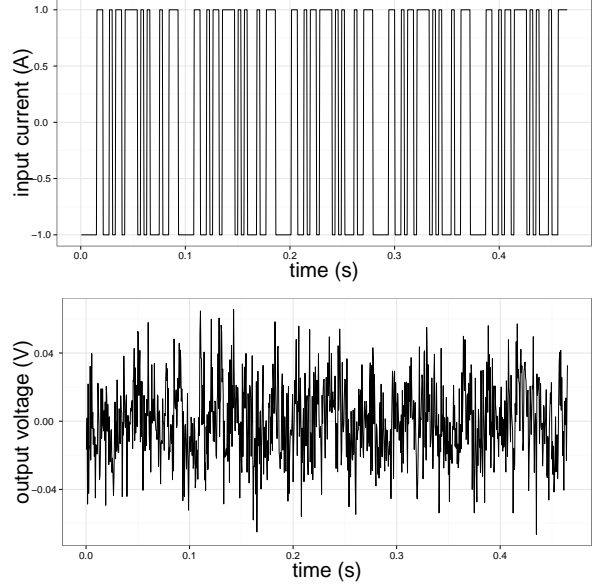


Fig. 4. Identification data for Example 2. Input sequence (left) and observations (right), of length $T = 930$, generated from the model of (4).

B. Example 2: Identification of a battery FO model with two CPEs

We add the Warburg term to Example 1. The model is given by (12) with system matrices (9) and (10) for $n = 2$ and $R_2 = \infty$ (the resistance associated with C_2 is open-circuit to model the Warburg term). The initial value of the state vector is set to zero: $x_0 = [0 \ 0]$. The parameter vector is

$$\theta = [R_\infty \ C_1 \ R_1 \ \alpha_1 \ C_2 \ \alpha_2].$$

We generate synthetic data using the parameter set

$$\theta^* = [0.01 \ 3.0 \ 0.2 \ 0.8 \ 400 \ 0.5]. \quad (15)$$

The data length is set to $T = 930$ samples. The standard deviations are set to $\sigma_x = 0.002$ and $\sigma_y = 0.02$. A pseudo-random binary sequence (PRBS) signal between -1 and +1 was generated for $(u_k)_{k \in \mathbb{N}}$, with sampling time $T_s = 0.5$ ms. The output voltage $(y_k)_{k \in \mathbb{N}}$ is then generated using the model of (4) with the parameter value of (15). Figure 4 shows the input-output data for the base scenario.

We use a uniform prior on each parameter, with ranges $[5, 100]$ m Ω for R_∞ , $[50, 500]$ m Ω for R_1 , $[1, 5]$ Fcm $^{-2}$ s $^{-\alpha_1}$ for C_1 , $[300, 500]$ Fcm $^{-2}$ s $^{-\alpha_2}$ for C_2 and $[0.4, 1]$ for α_1 and α_2 . The PMMH is tuned with $N = 128$ particles per iteration, and $M = 20,000$ iterations. We first present the results of the base

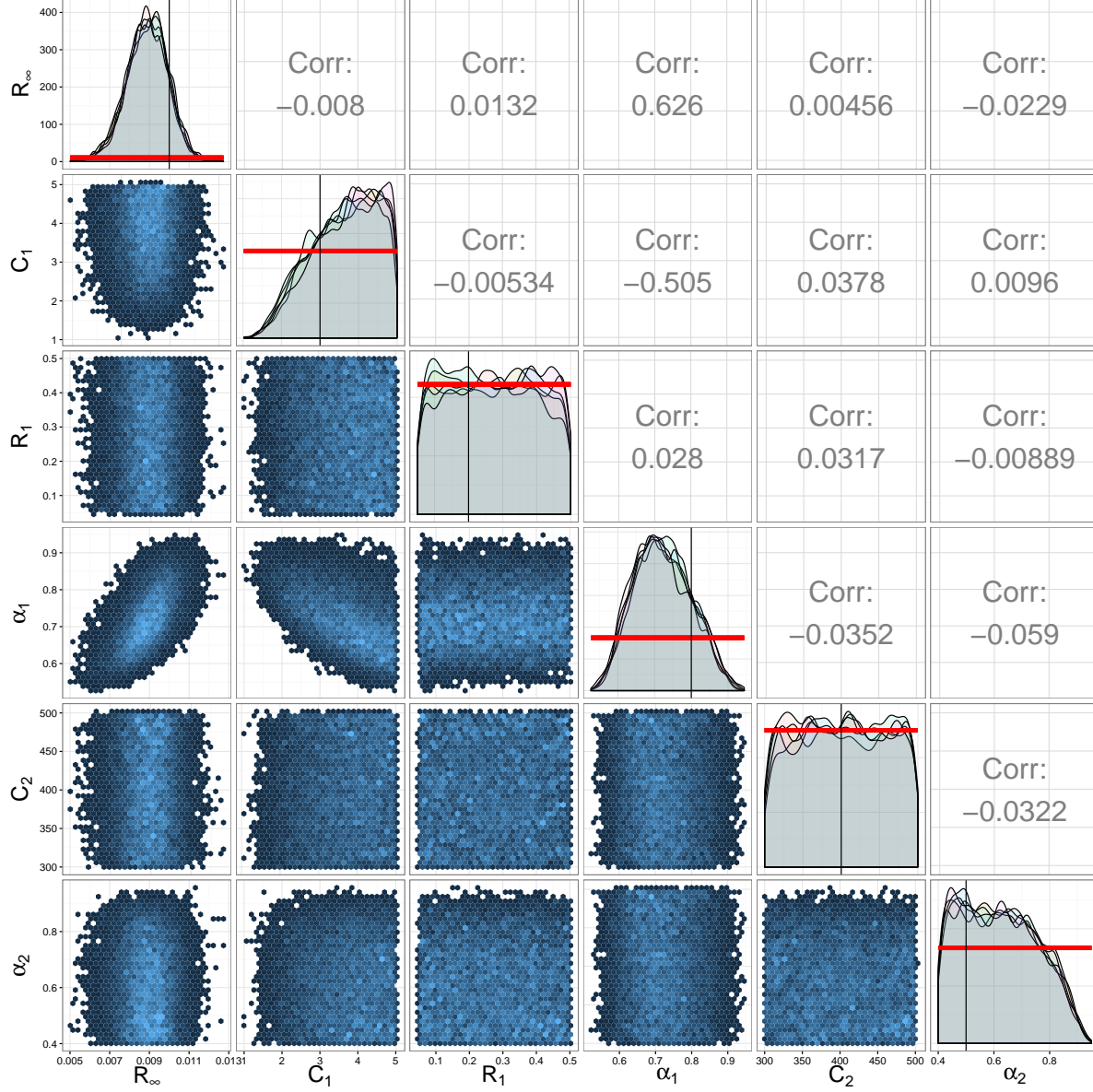


Fig. 5. Results of the PMMH algorithm for the base scenario, sampling approximately from the posterior distribution for the model (12). On the diagonal, the posterior density function of each run is overlaid with the prior distribution (red horizontal lines), which is uniform. On the lower triangle, the five runs are pooled together in a histogram with hexagonal binning. On the upper triangle, the correlation coefficients between pairs of parameters are displayed.

scenario (Section III-C), and then consider various modifications: the number of observations (Section III-D), the magnitude of the input data (Section III-E), the prior distribution (Section III-G) and the state-output noise ratio of the generated data (Section III-F).

C. Posterior samples in the base scenario

The results of the base scenario, computed on five independent runs, are shown in Figure 5. On the diagonal, we see that the five independent runs are consistent, indicating that the PMMH method approximates the posterior distribution in a satisfactory way. Then we see that for some parameters, such as R_∞ , the posterior is significantly different from the prior. This indicates that the data are informative on these parameters; we can expect that these marginal posterior distributions would concentrate around the corresponding values of θ^* in (15), when the number of observations goes to infinity. On the other hand, nothing seems to be learned on some other parameters, such as C_2 , for which the posterior resembles the prior.

Since the prior distribution is uniform, the posterior distribution is proportional to the likelihood, and thus the mode of the posterior would be precisely the maximum likelihood estimate, under the constraints of the prescribed intervals. Since the posterior distribution is clearly flat over some parameters (such as C_2), the mode could be anywhere in the corresponding interval. Thus, a numerical procedure giving only the maximum likelihood estimate would return any value in that range.

The Bayesian approach is closer in spirit to integrated likelihood approaches, where nuisance parameters are averaged over. An advantage of sampling from the posterior distribution is the possibility to investigate correlations between parameters, which are particularly large between R_∞ and α_1 , and between C_1 and α_1 according to Figure 5.

D. Effect of the number of observations

We consider three sets of data of sizes $T = 635$, $T = 930$ and $T = 1890$, generated from three input sequences of these lengths. Instead of showing of all the marginal distributions, we focus on two parameters, R_∞ and C_2 , which are respectively easy and challenging to identify (according to Figure 5). The 95% credible regions are displayed in Table II. We see that adding more data makes the posterior distribution more concentrated for R_∞ , and closer to

TABLE II
EFFECT OF THE NUMBER OF OBSERVATIONS ON 95% CREDIBLE INTERVALS, FOR R_∞ AND C_2 .

	$T = 635$	$T = 930$	$T = 1890$
R_∞	[0.008, 0.013]	[0.007, 0.011]	[0.009, 0.011]
C_2	[305, 496]	[305, 496]	[306, 496]

TABLE III
EFFECT OF THE INPUT MAGNITUDE ON 95% CREDIBLE INTERVALS, FOR R_∞ AND C_2 .

	± 1	± 5
R_∞	[0.0067, 0.0108]	[0.0095, 0.0103]
C_2	[305, 496]	[305, 495]

the true value of 0.01, whereas it does not seem to have an effect on the credible regions for C_2 .

E. Effect of the input magnitude

We study the effect of the magnitude of the input. On top of the base scenario, with input data of magnitude 1, we consider an input sequence of magnitude 5 (thus, oscillating between -5 and $+5$). The results are shown in Table III. Increasing the magnitude of the input helps identifying R_∞ , but still does not seem to impact the posterior distribution of C_2 .

F. Effect of the state-output noise ratio

We study the effect of the state to output noise ratio. More precisely, we generate data using $\sigma_y = 0.002$, instead of using $\sigma_y = 0.02$ as in the base scenario. We refer to these state-output noise ratios (i.e. signal to noise ratios) as 1.0 and 0.1 respectively. Again, we find that the effect is very apparent on R_∞ , but that the credible regions for C_2 seem unchanged (Table IV).

G. Effect of the prior distribution

We study the effect of the prior distribution. We consider a Gaussian prior, centered in the middle of

TABLE IV
EFFECT OF THE STATE-OUTPUT NOISE RATIO (SNR) ON 95% CREDIBLE INTERVALS, FOR R_∞ AND C_2 .

	SNR = 0.1	SNR = 1.0
R_∞	[0.0067, 0.0108]	[0.0095, 0.0103]
C_2	[305, 496]	[306, 495]

the range of the previous intervals, and with standard deviation one fourth of the range, so that roughly 95% of the Gaussian mass lands in the range; we truncate the Gaussian distribution outside the range. As expected, we find that the posterior of C_2 is nearly equal to the prior, and thus is very dependent on the choice of prior. On the other hand, the prior has no apparent impact on the posterior distribution of R_∞ , indicating that the posterior is highly driven by the observations (figures not shown but available upon request).

IV. CONCLUSIONS

The ability to identify model parameters depends on the quantity and quality (uncertainties) of the available data. Although there is a rich literature on the experiment design for systems with ordinary differential equations (ODEs), [28]–[31], it remains unclear how to generate informative data for parameter estimation of FO systems. In [32], we applied the persistent excitation concept to the battery Randles circuit model, which is given by ODEs. This method is adopted here. The authors in [33] propose a method for the design of periodic excitation signals to identify the battery models given by ODEs.

In this paper, we have developed a computational approach to Bayesian inference in non-Markovian state-space models, and applied it to the identification of FO systems. The posterior distribution was approximated using particle marginal Metropolis-Hastings, and in particular the use of a tree representation of the trajectories has been proposed to improve the efficiency of numerical computations. We applied the method to battery commensurate and non-commensurate FO models. The results were compared with the instrumental variable identification method. Practical identifiability of model parameters was studied by comparing the prior and posterior distributions. The sensitivity of parameters to the choice of prior distributions, the number of observations, the magnitude of the input signal and the measurement noise were studied.

In summary, the Bayesian framework is a convenient tool to estimate parameters, assess their identifiability and quantify uncertainties in FO models.

V. APPENDICES

A. Locally optimal proposal for the battery model

In the models considered in the article, the following locally optimal proposal can be implemented. The state $x_k \in \mathbb{R}^2$ follows a Gaussian distribution

centered at a function of the past: $x_k \sim \mathcal{N}(\varphi_k, \Sigma_x)$, where $\varphi_k = (\varphi_{k,1}, \varphi_{k,2})^T$ is computed as a deterministic function of the past trajectory $x_{0:k-1}$ and of the parameter, and $\Sigma_x = \text{diag}(\sigma_x^2, \sigma_y^2)$. Furthermore we have $y_k \sim \mathcal{N}(c_k + x_{k,1} + x_{k,2}, \sigma_y^2)$, where c_k is a real value that can be computed given the parameter value. As a result, we can compute $y_k \mid x_{0:k-1} \sim \mathcal{N}(\zeta_k, 2\sigma_x^2 + \sigma_y^2)$, where $\zeta_k = c_k + \varphi_{k,1} + \varphi_{k,2}$ and $x_k \mid x_{0:k-1}, y_k$ has a Gaussian distribution with

$$\begin{aligned} \text{mean} &= \begin{pmatrix} \varphi_{k,1} + \frac{\sigma_x^2}{2\sigma_x^2 + \sigma_y^2} (y_k - \zeta_k) \\ \varphi_{k,2} + \frac{\sigma_x^2}{2\sigma_x^2 + \sigma_y^2} (y_k - \zeta_k) \end{pmatrix}, \\ \text{and variance} &= \begin{pmatrix} \sigma_x^2 - \frac{\sigma_x^4}{2\sigma_x^2 + \sigma_y^2} & -\frac{\sigma_x^4}{2\sigma_x^2 + \sigma_y^2} \\ -\frac{\sigma_x^4}{2\sigma_x^2 + \sigma_y^2} & \sigma_x^2 - \frac{\sigma_x^4}{2\sigma_x^2 + \sigma_y^2} \end{pmatrix}. \end{aligned}$$

B. Particle Marginal Metropolis-Hastings

The Particle Marginal Metropolis-Hastings algorithm of [15] requires a few choices of tuning parameters from the user. First, it requires a particle filter yielding likelihood estimates $\hat{p}(y_{0:T} \mid \theta)$ for all θ . We have described a choice of proposal distribution in Appendix V-A, but one still needs to choose a number of particles N . We adopt the following strategy. For a sample θ from the prior distribution, we run the particle filter n times, given that same θ . This yields a sequence of estimates Z_1, \dots, Z_n of $p(y_{0:T} \mid \theta)$. We then mimic a Metropolis-Hastings scheme with these estimates, and compute the corresponding acceptance rate. It would be 100% if the estimates were all perfect evaluations of $p(y_{0:T} \mid \theta)$. Thus, starting from a small number of particles N , we increase the number N until the average number of acceptances is considered to be enough, for instance 10%. We can perform this procedure on a few parameters θ in parallel, to account for the variability across the parameter space.

Once the particle filter is tuned, we need to choose a proposal distribution to draw the candidate parameters: $\theta^* \sim q_\theta(\cdot \mid \theta)$. A standard choice consists in using a Gaussian random walk: $\theta^* \sim \mathcal{N}(\theta, \Sigma)$, for which we still need to choose the covariance matrix. One approach consists in changing Σ along the run of the algorithm. We take a simpler approach, using a preliminary run. Upon the completion of it, we compute the covariance matrix of the chain, which approximates the posterior covariance, and use it as a choice of Σ .

ACKNOWLEDGEMENTS

Authors would like to acknowledge fundings from the UK's Engineering and Physical Sciences Research Council (EPSRC), S.M.M. Alavi and D.A. Howey under grant EP/K503769/1, A. Mahdi and S.J. Payne under grant EP/K036157/1 and P.E. Jacob under grant EP/K009362/1. We are grateful for the permission from the authors in Ref. [27] to re-use data in Example 1.

REFERENCES

- [1] C. A. Monje, Y. Chen, B. M. Vinagre, D. Xue, and V. Feliu-Batlle, *Fractional-order systems and controls: fundamentals and applications*. Springer Science & Business Media, 2010.
- [2] S. M. M. Alavi, A. Mahdi, P. E. Jacob, S. J. Payne, and D. A. Howey, "Structural identifiability analysis of fractional order models with applications in battery systems," *arXiv preprint arXiv:1511.01402*, 2015.
- [3] L. Le Lay, "Identification fréquentielle et temporelle par modèle non entier," Ph.D. dissertation, Université Bordeaux I, 1998.
- [4] J. Lin, "Modélisation et identification des systèmes d'ordre non entier," Ph.D. dissertation, Poitiers, 2001.
- [5] O. Cois, "Systèmes linéaires non entiers et identification par modèle non entier: application en thermique," Ph.D. dissertation, Bordeaux I, 2002.
- [6] R. Malti, S. Victor, A. Oustaloup, and H. Garnier, "Fractional state variable filter for system identification by fractional model," in *17th IFAC World Congress*, 2008, pp. 1–6.
- [7] S. Victor, R. Malti, H. Garnier, and A. Oustaloup, "Parameter and differentiation order estimation in fractional models," *Automatica*, vol. 49, no. 4, pp. 926–935, 2013.
- [8] A. Oustaloup, P. Melchior, P. Lanusse, O. Cois, and F. Dancla, "The crone toolbox for matlab," in *Computer-Aided Control System Design, 2000. CACSD 2000. IEEE International Symposium on*. IEEE, 2000, pp. 190–195.
- [9] P. Melchior, P. Lanusse, O. Cois, F. Dancla, and A. Oustaloup, "Crone toolbox for matlab: fractional systems toolbox," in *Proceedings of the 41st IEEE CDC*, vol. 2, 2002.
- [10] P. Young and A. Jakeman, "Refined instrumental variable methods of recursive time-series analysis part i. single input, single output systems," *International Journal of Control*, vol. 29, no. 1, pp. 1–30, 1979.
- [11] A. Jakeman and P. Young, "Refined instrumental variable methods of recursive time-series analysis part ii. multivariable systems," *International Journal of Control*, vol. 29, no. 4, pp. 621–644, 1979.
- [12] P. Young and A. Jakeman, "Refined instrumental variable methods of recursive time-series analysis part iii. extensions," *International Journal of Control*, vol. 31, no. 4, pp. 741–764, 1980.
- [13] S. M. M. Alavi, C. R. Birkel, and D. A. Howey, "Time-domain fitting of battery electrochemical impedance models," *Journal of Power Sources*, vol. 288, pp. 345–352, 2015.
- [14] B. Wang, S. E. Li, H. Peng, and Z. Liu, "Fractional-order modeling and parameter identification for lithium-ion batteries," *Journal of Power Sources*, vol. 293, pp. 151–161, 2015.
- [15] C. Andrieu, A. Doucet, and R. Holenstein, "Particle markov chain monte carlo methods," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 3, pp. 269–342, 2010.
- [16] D. A. Howey and S. M. M. Alavi, "Rechargeable battery energy storage system design," *Handbook of Clean Energy Systems*, Ed. J. Yan, vol. 5, pp. 2801–2818, 2015.
- [17] E. Barsoukov and J. R. Macdonald, *Impedance spectroscopy: theory, experiment, and applications*. John Wiley & Sons, 2005.
- [18] U. Tröltzsch, O. Kanoun, and H.-R. Tränkler, "Characterizing aging effects of lithium ion batteries by impedance spectroscopy," *Electrochimica Acta*, vol. 51, no. 8, pp. 1664–1672, 2006.
- [19] J. R. Macdonald, J. Schoonman, and A. Lehen, "Applicability and power of complex nonlinear least squares for the analysis of impedance and admittance data," *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, vol. 131, pp. 77–95, 1982.
- [20] B. A. Boukamp, "A nonlinear least squares fit procedure for analysis of immittance data of electrochemical systems," *Solid State Ionics*, vol. 20, no. 1, pp. 31–44, 1986.
- [21] C. Robert, *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.
- [22] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of Nonlinear Filtering*, vol. 12, pp. 656–704, 2009.
- [23] P. Del Moral, *Feynman-Kac Formulae*. Springer, 2004.
- [24] F. Cérou, P. Del Moral, and A. Guyader, "A nonasymptotic theorem for unnormalized feynman-kac particle models," *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 47, no. 3, pp. 629–649, 2011.
- [25] P. E. Jacob, L. M. Murray, and S. Rubenthaler, "Path storage in the particle filter," *Statistics and Computing*, pp. 1–10, 2013.
- [26] N. Chopin, P. Del Moral, and S. Rubenthaler, "Stability of feynman-kac formulae with path-dependent potentials," *Stochastic Processes and their Applications*, vol. 121, no. 1, pp. 38–60, 2011.
- [27] D. A. Howey, P. D. Mitcheson, V. Yufit, G. J. Offer, and N. P. Brandon, "On-line measurement of battery impedance using motor controller excitation," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 6, pp. 2557 – 2566, 2014.
- [28] L. Ljung, "System identification: theory for the user," *PTR Prentice Hall Information and System Sciences Series*, vol. 198, 1987.
- [29] T. Soderstrom and P. Stoica, *System identification*. Prentice hall London, 1989, vol. 2.
- [30] M. J. Rothenberger, J. Anstrom, S. Brennan, and H. K. Fathy, "Maximizing parameter identifiability of an equivalent-circuit battery model using optimal periodic input shaping," in *ASME 2014 Dynamic Systems and Control Conference*. American Society of Mechanical Engineers, 2014, pp. V001T19A004–V001T19A004.
- [31] M. J. Rothenberger, D. J. Docimo, M. Ghanaatpishe, and H. K. Fathy, "Genetic optimization and experimental validation of a test cycle that maximizes parameter identifiability for a li-ion equivalent-circuit battery model," *Journal of Energy Storage*, vol. 4, pp. 156–166, 2015.
- [32] S. M. M. Alavi, A. Mahdi, S. J. Payne, and D. A. Howey, "Structural identifiability of battery equivalent circuit models," *arXiv preprint arXiv:1505.00153*, 2015.
- [33] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, no. 15, pp. 1923–1929, 2009.