eTransportation 18 (2023) 100283

Contents lists available at ScienceDirect

eTransportation

journal homepage: www.journals.elsevier.com/etransportation

Data selection framework for battery state of health related parameter estimation under system uncertainties

Jackson Fogelquist, Xinfan Lin *

Department of Mechanical and Aerospace Engineering, University of California, Davis, CA 95616, USA

ARTICLE INFO

Keywords:
Parameter estimation
Lithium-ion battery
State of health
Uncertainty
Data selection
Sensitivity

ABSTRACT

Data selection is a practical technique for improving parameter estimation accuracy through the strategic selection of information-rich data for use in the estimation algorithm. Traditional selection criteria have been either heuristic or sensitivity-based, without consideration of uncertainties in measurement, model, or parameter. In this paper, we propose an uncertainty-aware data selection framework that selects data segments based on the potential of the ingrained data structures to mitigate the influence of system uncertainties on the estimation result. The framework comprises two components: the data quality rating and data selection algorithm. The data quality rating is a metric for evaluating the uncertainty-propagating data structures of a data segment, and the data selection algorithm automatically integrates the data selection into the estimation procedure. Furthermore, a novel adaptive approximation of model/measurement uncertainty is derived and implemented in the data quality rating formula to enhance performance in the presence of time-varying sensor bias/noise and unmodeled system dynamics. The framework is validated through an advanced battery management system application, where two lithium-ion battery health-related electrochemical parameters are separately estimated under random drive-cycle input data to emulate battery state of health monitoring for an electric vehicle. We show that the drive-cycle data, which are frequently used for battery state of health estimation as the only available data during battery operation, may not provide accurate estimation results due to the existence of large portions of low-quality data (low sensitivity and high uncertainty). By extracting the high-quality data segments, the data selection framework reduced experimental estimation errors by one order of magnitude when compared with the conventional approach of estimating without data selection.

1. Introduction

Data-based parameter estimation is the practice of using measured input–output data to determine the parameters of a system model. It is vital for the reliable modeling and control of dynamic systems because the quality of a model (and any model-based functionality that may rely upon it) is dependent upon the accuracy of its parameters. This is especially important for advanced battery management systems (BMSs) as they monitor state of charge (SOC) and state of health (SOH) with an increasing reliance on complex physics-based electrochemical battery models that have dozens of parameters [1–3].

Of the three components of a parameter estimation problem – model, data, and estimation algorithm – data is a growing topic of interest, as researchers seek to understand how to quantify and optimize the quality of data to maximize estimation accuracy. This is motivated by the fact that data is the fundamental input to the estimation problem; specifically, a poor data set will limit the achievable estimation accuracy regardless of the complexity of the model or algorithm. The Fisher information is often regarded as the standard metric for data quality, as its inverse yields the Cramér–Rao bound, i.e., the lower bound of achievable estimation error (co)variance for an unbiased estimator [4,5]. Accordingly, the Fisher information is routinely implemented as the criterion for data optimization and optimal experiment design [6–8]. This is exemplified in the field of battery modeling and control, where works have centered on analytical [9,10], experimental [11], and computational [12–14] data design using the Fisher information, as to optimize current input excitations for improved parameter estimation accuracy.

The majority of existing data optimization research is applicable to offline parameter estimation, where input excitations are *designed* and administered in a laboratory setting. However, if no control authority exists over the data, as in the case of online estimation where data are passively generated by system operation under random load, the practical question arises: can high-quality data be strategically *selected* from a data stream to improve parameter estimation accuracy? Some early works took a temporal approach and empirically established data

* Corresponding author.
E-mail addresses: jbfogelquist@ucdavis.edu (J. Fogelquist), lxflin@ucdavis.edu (X. Lin).

https://doi.org/10.1016/j.etran.2023.100283
Received 2 September 2022; Received in revised form 16 August 2023; Accepted 7 September 2023
Available online 12 September 2023

sampling rates for each target parameter, based on the time-scales of their variations [15–18]. Another approach empirically assigned weights to different data based on their expected reliability [19]. Others used input excitation as the criterion for data selection. Specifically, the framework proposed in [20] empirically truncates low-magnitude and unvarying input data because these structures are often dominated by measurement noise; the frameworks proposed in [21,22] prioritize data diversity by selecting samples of input data that are distributed across the amplitude range of a data set. The aforementioned methodologies are empirical in the sense that they rely on heuristic metrics that are not directly related to the suitability of the data for estimation. As such, selection criteria are often hand-tuned from experience, and may even degrade estimation accuracy if ill-chosen [19]. Recent works have explored sensitivity-based data selection, which seeks to effectively maximize the Fisher information by selecting data that are highly sensitive to the estimated parameter(s) [23,24]. These works argue that insensitive data add little value to the estimation, yet can introduce substantial errors through ingrained parameter uncertainty [25] and model/measurement uncertainty, i.e., due to unmodeled system dynamics [26,27] and measurement noise [28–30].

The state-of-the-art sensitivity-based methodologies overcome the heuristic limitations of the empirical approaches by basing the selection metric on the Fisher information and Cramér–Rao bound. However, although widely implemented for data optimization, the Fisher information and Cramér–Rao bound feature several theoretical limitations that restrict their use as criteria for evaluating data quality. These include the assumption that the estimator is unbiased, neglect of model and parameter uncertainties, and the possibility that the estimator cannot achieve the best-case error (co)variance specified by the Cramér–Rao bound in practice [4,5]. To address these limitations, an equation was derived in [31] to directly quantify the parameter estimation error under uncertainties in model/measurement and parameter for the widely used least-squares estimation objective. It was found that for each type of system uncertainty, there is a specific data structure, represented in terms of parameter sensitivity, that governs the propagation of the uncertainty to the estimation error. The estimation error equation also provides a critical insight—estimation accuracy is dependent on the quality of the data, rather than the quantity. This insight casts new light on the conventional view that follows from the mathematical definition of the Fisher information, i.e., more data provide more information and thus always improve estimation accuracy. We envision that these findings can be leveraged to overcome the restrictions of the state-of-the-art sensitivity-based data selection criteria.

The objective of this paper is to enhance battery SOH-monitoring by establishing a data selection framework for parameter estimation that focuses on the quality of data. Our recent work explored the potential for the uncertainty-propagating data structures identified in [31] to indicate the quality of data, with promising results in simulation [32]. Here, we leverage these data structures and incorporate a novel adaptive approximation of model/measurement uncertainty to select data segments that can achieve the best estimation accuracy. The key contributions of this work include the following aspects. First, it is the first attempt, to the best of our knowledge, at a data selection approach that explicitly addresses system uncertainties in the process of estimation. Second, we propose a mechanism to integrate automatic data selection into the estimation procedure, which is necessary for the data selection framework to be practically implemented in BMS applications. Third, we establish a rating formula as a new criterion for evaluating data quality, which can be efficiently applied to battery voltage data segments of arbitrary length under generic input current. Finally, we derive a new adaptive approximation of time-varying model/measurement uncertainty that can be conveniently incorporated into the data selection criterion, to consider the limitations of BMS sensor resolution and model fidelity. It will be shown through an SOH-monitoring application in simulation and experiment that the framework is capable of significantly improving parameter estimation accuracy, with experimental error reductions of one order of magnitude when compared with the conventional approach of estimating without data selection.

The remainder of the paper is organized as follows. Section 2 reviews the lithium-ion (Li-ion) battery dynamics and model associated with our application of estimating health-related electrochemical parameters. Section 3 details the proposed data selection framework through the introduction of the data quality rating concept and the design of the data selection algorithm. Section 4 proposes a preliminary data quality rating formula and demonstrates the functionality of the framework in simulation and experiment. Section 5 extends the capability of the data quality rating through an ingrained approximation of the time-varying model/measurement uncertainty, with experimental validation. Lastly, Section 6 presents concluding remarks and potential applications.

2. Li-ion battery dynamics and modeling

The SOH of a Li-ion battery is a critical quantity that characterizes the remaining cyclable capacity and/or power capability, which directly affect the range and performance of a battery-powered device, e.g., an electric vehicle [3,33]. The SOH can be efficiently monitored with an electrochemical model by routinely estimating health-related parameters, i.e., physical parameters that are intrinsically linked to the SOH. Therefore, we will develop and validate the data selection framework under the objective of accurately estimating health-related electrochemical parameters using random drive-cycle data, to emulate the cycling conditions in an electric vehicle. This section provides a brief overview of the parameters, battery dynamics, and parameter sensitivities that play critical roles in our SOH-monitoring data selection application.

The two electrochemical parameters that will be targeted for estimation are the solid-phase cathode lithium diffusion coefficient $D_{s,p}$ and the cathode active material volume fraction $\varepsilon_{s,p}$. Physically, $D_{s,p}$ characterizes the rate at which lithium ions can diffuse through the cathode electrode particle material, while $\varepsilon_{s,p}$ represents the proportion of the total cathode volume capable of storing lithium ions. Both parameters play a vital role in battery performance and serve as key indicators of battery degradation and SOH [34–36]. Specifically, several works have employed electrochemical impedance spectroscopy to show that $D_{s,p}$ decreases over time with SOH, in accordance with the increasing cell impedance [37,38]. In the same way, $\varepsilon_{s,p}$ decreases over time with SOH because it is directly proportional to the decreasing cell capacity through the relation $Q_p = FA_p\delta_p c_{s,p}^{max}\varepsilon_{s,p}$, where Q_p is the cathode capacity, F is the Faraday constant, $c_{s,p}^{max}$ is the ionic concentration limit of the cathode material, and A_p and δ_p are the cathode area and thickness, respectively. These trends in $D_{s,p}$ and $\varepsilon_{s,p}$ are attributed to degradation mechanisms such as reaction-induced mechanical stress [39,40], transition metal dissolution [41,42], and solid-electrolyte interphase (SEI) layer growth [43,44]. While other electrochemical parameters may also be related to SOH, the $D_{s,p}$ and $\varepsilon_{s,p}$ correlations are well established in the literature and thus commonly studied in SOH estimation applications [44–46]. In addition, $D_{s,p}$ is a weakly sensitive parameter while $\varepsilon_{s,p}$ is strongly sensitive, which makes $D_{s,p}$ conventionally challenging to estimate, especially under uncertainty in $\varepsilon_{s,p}$ [47,48].

The battery dynamics are modeled with the widely adopted single particle model with electrolyte dynamics (SPMe) [49,50], which predicts the output terminal voltage (V) from the input current (I). The SPMe is a simplified version of the full-order Doyle–Fuller–Newman (DFN) electrochemical model [51], operating under the assumption that lithium intercalation current density (and thus ionic concentration) is uniform across each electrode. Accordingly, the electrochemical mechanisms in each electrode (e.g., diffusion, intercalation) are represented with a single particle, and both electrode particles are interfaced with the electrolyte diffusion dynamics. Mathematically, the

single-particle assumption decouples the governing partial differential equations (PDEs) of the DFN model, which significantly reduces the computational complexity and makes the SPMe suitable for use in real-time BMS applications [3]. This is facilitated through model-order reduction techniques that enable computationally-efficient and high-fidelity solutions to the decoupled PDEs [52,53].

The output terminal voltage is expressed as

$$V = U_p(c_{se,p}) - U_n(c_{se,n}) + \phi_{e,p}(c_{e,p}) - \phi_{e,n}(c_{e,n}) + \eta_p(c_{se,p}, c_{e,p}) - \eta_n(c_{se,n}, c_{e,n}) - I R_l, \quad (1)$$

which includes the difference between the cathode and anode (denoted by subscripts p and n respectively) open-circuit potentials (OCPs) U, electrolyte potentials ϕ_e, and overpotentials η. The OCPs U represent the equilibrium potential of each electrode as a nonlinear function of the electrode particle surface lithium concentration c_{se}, which is governed by Fick's second law of diffusion. The electrolyte potentials ϕ_e are driven by the ionic concentration gradient across the electrolyte, which is characterized by the dynamic electrolyte lithium concentration at each electrode boundary c_e, according to Fick's second law. The overpotential η drives the intercalation reaction at the electrode particle surface according to the Butler–Volmer equation, in function of c_{se} and c_e. Finally, the voltage drop across the various Ohmic resistances (i.e., of the SEI layer, electrolyte, and current collectors) is incorporated through the lumped resistance term R_l. The reader is referred to [50] for the full details of the model.

The data quality rating formulas, to be introduced in subsequent sections, rely upon the sensitivity of battery voltage to various parameters. To facilitate the computation of sensitivity, we employ the analytical sensitivity expressions derived in [50] for the SPMe, which efficiently capture the sensitivity dynamics through sensitivity transfer functions. For example, the sensitivity of the output voltage V to $\varepsilon_{s,p}$ can be derived by taking the partial derivative of Eq. (1) with respect to $\varepsilon_{s,p}$ as

$$\frac{\partial V}{\partial \varepsilon_{s,p}}(t) = \frac{\partial \eta_p}{\partial \varepsilon_{s,p}} + \left(\frac{\partial \eta_p}{\partial c_{se,p}} + \frac{\partial U_p}{\partial c_{se,p}} \right) \cdot \frac{\partial c_{se,p}}{\partial \varepsilon_{s,p}}(t). \quad (2)$$

The first term reflects the non-dynamic sensitivity of η_p to $\varepsilon_{s,p}$, which can be easily obtained based on the model as a nonlinear function of current, while the second term captures the dynamic sensitivity of η_p and U_p to $\varepsilon_{s,p}$ through the solid-phase diffusion mechanism in the cathode. Regarding the second term, $\frac{\partial \eta_p}{\partial c_{se,p}}$ and $\frac{\partial U_p}{\partial c_{se,p}}$ are the slopes of overpotential and OCP, while $\frac{\partial c_{se,p}}{\partial \varepsilon_{s,p}}(t)$ reflects the dynamic nature of the sensitivity due to the dynamic diffusion process. A sensitivity transfer function has been derived to characterize $\frac{\partial c_{se,p}}{\partial \varepsilon_{s,p}}$,

$$\frac{\partial C_{se,p}}{\partial \varepsilon_{s,p}}(s) = \frac{7R_{s,p}^4 s^2 + 420 D_{s,p} R_{s,p}^2 s + 3465 D_{s,p}^2}{s(R_{s,p}^4 s^2 + 189 D_{s,p} R_{s,p}^2 s + 3465 D_{s,p}^2)} \cdot \frac{I(s)}{F \varepsilon_{s,p}^2 A_p \delta_p}, \quad (3)$$

which allows the dynamic sensitivity to be conveniently computed, e.g., by converting to a linear state-space model. Here, $R_{s,p}$ is the cathode particle radius. Similarly, the $D_{s,p}$ sensitivity expression can be derived in the same way as

$$\frac{\partial V}{\partial D_{s,p}}(t) = \left(\frac{\partial \eta_p}{\partial c_{se,p}} + \frac{\partial U_p}{\partial c_{se,p}} \right) \cdot \frac{\partial c_{se,p}}{\partial D_{s,p}}(t), \quad (4)$$

which relies entirely upon the diffusion dynamics through $c_{se,p}$. The associated sensitivity transfer function is

$$\frac{\partial C_{se,p}}{\partial D_{s,p}}(s) = \frac{43R_{s,p}^4 s^2 + 1980 D_{s,p} R_{s,p}^2 s + 38115 D_{s,p}^2}{(R_{s,p}^4 s^2 + 189 D_{s,p} R_{s,p}^2 s + 3465 D_{s,p}^2)^2} \cdot \frac{21 R_{s,p}^2 I(s)}{F \varepsilon_{s,p} A_p \delta_p}. \quad (5)$$

These sensitivity transfer functions were derived using Laplace transforms and Padé approximations, with the full procedure detailed in [50].

3. Data selection framework

The proposed data selection framework comprises two core elements, namely, the data quality rating formula and the data selection algorithm. The rating formula predicts the extent to which system uncertainties are propagated to the estimation error, based on the sensitivity dynamics of the uncertain parameters. Accordingly, the rating serves as a metric for data segment quality. Two rating formulas are presented in this paper—a simple preliminary rating formula is derived in Section 4 for instructional purposes and to motivate the derivation of the improved adaptive rating formula in Section 5.

The purpose of the whole framework is to maximize parameter estimation accuracy by integrating the data selection (based on the data rating formula) into the estimation procedure. Specifically, the algorithm seeks to select high-quality data segments from a given data set, perform the estimation using the selected data segments, reevaluate the quality of the selected segments based on the estimation results, and return an accurate final estimate based on the updated quality rating. This methodology is illustrated in Fig. 1 and compared with the conventional univariate estimation approach. An additional benefit of integrating the data selection and estimation procedures is that the estimation only needs to be performed for the selected data segments (rather than for all data segments), which maintains computational tractability for practical implementation in advanced BMSs.

In this work, the data selection framework is proposed for the scenario of univariate estimation, where one health-related parameter θ (i.e., $D_{s,p}$ or $\varepsilon_{s,p}$) is estimated in the presence of uncertainties in model, measurement, and other parameters. This is an important problem for two reasons. First, it is a common scenario in practice, as many applications only require the estimation of a certain parameter, rather than all of them. For example, battery capacity fade can be indicated by monitoring $\varepsilon_{s,p}$ [39,45]. Meanwhile, parameters that are not estimated may not be perfectly known, as some of them may undergo large variations due to changing operating conditions and/or system degradation [54–56]. These parameters need to be assumed with nominal values and become an unavoidable source of uncertainty. An alternative to assuming values for uncertain parameters is to estimate them simultaneously alongside the target parameter. However, estimating more parameters may make the problem ill-posed while increasing the computational complexity—a vital factor for online estimation applications. It will be shown in Section 4.2.2 that jointly estimating even one additional parameter can cause the problem to become ill-posed and lead to significant estimation errors under random drive-cycle data. The second reason regarding the importance of this type of problem is that it can be very challenging to solve, as estimating a weakly sensitive parameter (e.g., $D_{s,p}$) under the shadow of uncertainty in strongly sensitive parameters (e.g., $\varepsilon_{s,p}$) is traditionally extremely difficult [47,48]. The data quality rating seeks to facilitate a solution by evaluating the extent to which an estimation result may be affected by system uncertainties. Nevertheless, the methodology proposed in this paper can be extended to multivariate estimation problems in future work.

The data input to the algorithm is a sequence of input–output measurements, making the framework suitable for both offline and online applications. In the offline case, the data sequence can be retrieved from an existing database or acquired through laboratory measurement. In online applications, the data sequence can be a window of an incoming passive data stream, and the estimation can be performed recursively on the moving window as new data become available.

The integrated data selection and estimation procedures are summarized in Algorithm 1 and each step is subsequently detailed:

1. *Data Segment Quality Evaluation*: The data quality rating is computed *a priori* to evaluate the quality of each data segment that is extracted from the data set. This step of computation is considered *a priori* because it occurs before the estimation and uses the initial guess of the target parameter ($\hat{\theta}^-$) when evaluating the rating formula.

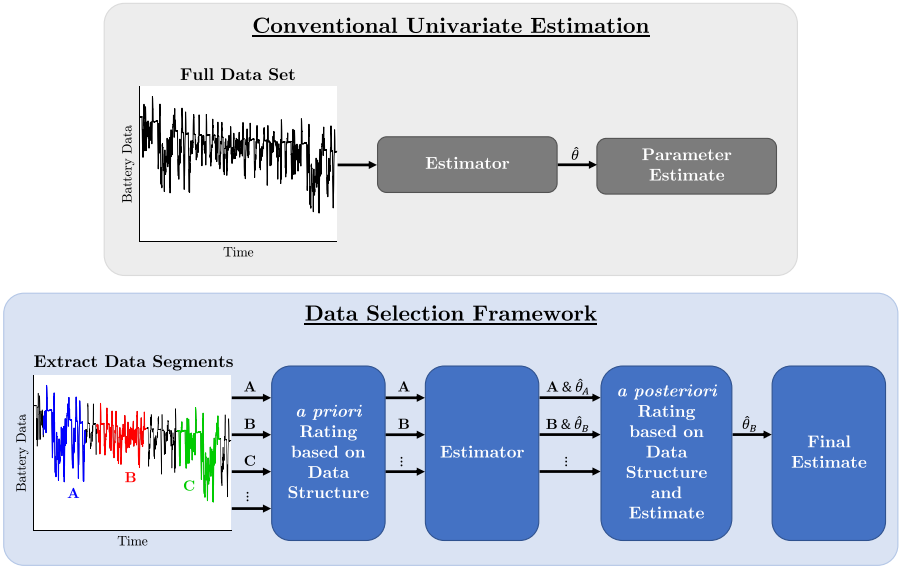


Fig. 1. Univariate estimation under conventional approach and data selection framework.

2. *Data Selection*: Data segments are selected for estimation based on the *a priori* data quality rating.
    - *Offline Estimation*: Data segments are selected from the data set to span a range of good *a priori* rating values.
    - *Online Estimation*: A data segment is selected from a moving window of incoming random data if the *a priori* rating satisfies a given threshold.
3. *Estimation*: The target parameter is estimated from each selected data segment.
4. *Data Quality Update*: The accuracy of the data quality rating is improved by recomputing it *a posteriori* for the selected data segments. Specifically, this step of the computation uses the estimated value of the target parameter ($\hat{\theta}^+$) to reevaluate the rating formula. It will be demonstrated that this step can significantly reduce the chance of mis-selection.
5. *Return Estimation Result*: The final estimation result is returned based on the *a posteriori* data quality rating.
    - *Offline Estimation*: The estimation results from the data segments with the best *a posteriori* ratings are averaged and returned.
    - *Online Estimation*: The estimation result from a data segment is returned if the *a posteriori* rating satisfies a given threshold.

Algorithm 1 Data Selection Framework

Input: Measured input–output data set/stream

1. Evaluate quality of each data segment with *a priori* data quality rating
2. Select high-quality data segments:
   Offline: Select from specified data set
   Online: Select from window of data stream
3. Perform estimation with each selected data segment
4. Reevaluate quality of selected data segments with *a posteriori* data quality rating using estimation results
5. Return final estimation result:
   Offline: Return mean of estimates from highest-quality data segments
   Online: Return estimates from data segments with acceptable quality

Output: Final estimation result $\hat{\theta}^+$

4. Preliminary data quality rating formula: Derivation, verification, & validation

The preliminary data quality rating formula is subsequently derived for implementation in the data selection framework. Simulation verification and experimental validation follow, which highlight key benefits of the data selection process while motivating the derivation of the improved adaptive rating formula in Section 5.

4.1. Derivation

The purpose of this subsection is to develop the preliminary data quality rating formula—the metric for evaluating the quality of candidate data segments for selection. We begin with the discrete-time least-squares estimation objective in Eq. (6),

$$\min_{\hat{\theta}^+} J = \sum_{k=1}^{N} \left(y_k^m - y_k(\hat{\theta}^+, \hat{\boldsymbol{\phi}}, \boldsymbol{u}_k)\right)^2, \tag{6}$$

which is one of the most widely used objectives for parameter estimation. The purpose of this function is to determine the estimate of one target parameter $\hat{\theta}^+$ that minimizes the sum of squared errors between the measured system output y_k^m and that predicted by a certain model $y_k(\hat{\theta}^+, \hat{\boldsymbol{\phi}}, \boldsymbol{u}_k)$ over a time sequence indexed by k. The modeled output y_k is driven by the input excitation sequence $\boldsymbol{u}_k = [u_1, \ldots, u_k]^T$ (with any state dynamics contained implicitly), and parameterized by θ and a set of other parameters that are not being estimated $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_m]^T$. Since the exact parameter values in $\boldsymbol{\phi}$ are not necessarily known, $\hat{\boldsymbol{\phi}}$ is used to denote the assumed values in the estimation problem, which may contain uncertainty. The output y_k is treated as a scalar in this work, as is the case for most battery applications, but it may be readily extended to the multidimensional case. The measured system output y_k^m is represented as

$$y_k^m = y_k(\theta, \boldsymbol{\phi}, \boldsymbol{u}_k) + \delta y_k, \tag{7}$$

where $y_k(\theta, \boldsymbol{\phi}, \boldsymbol{u}_k)$ is the modeled system output under the true target parameter θ and true non-target parameter set $\boldsymbol{\phi}$, and δy_k is the varying uncertainty between the modeled and measured system outputs (e.g., due to unmodeled system dynamics and/or sensor noise).

It is noted that without uncertainties (and structural unidentifiability), the solution of the least squares problem should yield the exact

value of the target parameter, as it would minimize the sum of squared error to zero. It is the inevitable uncertainties in practice that deviate the estimation result from the true value. In our prior work [31], we derived an equation to quantify the estimation error induced by different types of system uncertainties for the least-squares objective in Eq. (6). The results revealed the data structures that form the basis of the data quality rating. The derivation was performed by approximating the modeled system output $y_k(\theta, \boldsymbol{\phi}, \boldsymbol{u}_k)$ in Eq. (7) under the unknown true parameter set $(\theta, \boldsymbol{\phi})$ with a first-order Taylor series expansion about the estimated/uncertain parameter set $(\hat{\theta}^+, \hat{\boldsymbol{\phi}})$, i.e.,

$$y_k(\theta, \boldsymbol{\phi}, \boldsymbol{u}_k) \approx y_k(\hat{\theta}^+, \hat{\boldsymbol{\phi}}, \boldsymbol{u}_k) + \frac{\partial y_k}{\partial \hat{\theta}}(\hat{\theta}^+, \hat{\boldsymbol{\phi}}, \boldsymbol{u}_k)(\theta - \hat{\theta}^+) + \frac{\partial y_k}{\partial \hat{\boldsymbol{\phi}}}(\hat{\theta}^+, \hat{\boldsymbol{\phi}}, \boldsymbol{u}_k)(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}). \tag{8}$$

The first-order optimality condition $\left(\frac{\partial J}{\partial \hat{\theta}^+} = 0\right)$ was then applied to Eq. (6) to complete the derivation. The reader is referred to [31] for the full details.

The final form of the derived error equation is reproduced in Eq. (9), where $\Delta\theta = \theta - \hat{\theta}^+$ denotes the estimation error as the difference between the true value of the target parameter and the estimated value. The right-hand side of the equation shows the errors induced by each type of uncertainty, associated with summations of certain sensitivity terms, where $\frac{\partial y_k}{\partial \hat{\theta}}(\hat{\theta}^+)$ and $\frac{\partial y_k}{\partial \hat{\phi}_i}(\hat{\theta}^+)$ denote the sensitivity of the modeled system output y_k to $\hat{\theta}$ and $\hat{\phi}_i$, respectively, under the estimate of the target parameter $\hat{\theta}^+$. The sensitivities can be calculated with the derived expressions for specific parameters, e.g., Eqs. (2)–(5). Note that each sensitivity term is also dependent on the non-target parameter set and input excitation, e.g., $\frac{\partial y_k}{\partial \hat{\phi}_i}(\hat{\theta}^+, \hat{\boldsymbol{\phi}}, \boldsymbol{u}_k)$, but the notation in Eq. (9) excludes the $\hat{\boldsymbol{\phi}}$ and $\boldsymbol{u}_k$ terms for brevity.

$$\Delta\theta = -\frac{\left(\sum_{k=1}^{N} \frac{\partial y_k}{\partial \hat{\theta}}(\hat{\theta}^+)\delta y_k\right) + \sum_{i=1}^{m}\left[\left(\sum_{k=1}^{N} \frac{\partial y_k}{\partial \hat{\theta}}(\hat{\theta}^+)\frac{\partial y_k}{\partial \hat{\phi}_i}(\hat{\theta}^+)\right)\Delta\phi_i\right]}{\sum_{k=1}^{N}\left(\frac{\partial y_k}{\partial \hat{\theta}}(\hat{\theta}^+)\right)^2} \tag{9}$$

The most important insights from Eq. (9) are the set of data structures that govern the propagation of system uncertainties to the estimation error. The numerator reveals that each system uncertainty, namely the varying model/measurement uncertainty δy_k and constant parameter uncertainties $\Delta\phi_i = \phi_i - \hat{\phi}_i$, is multiplied by a summation of sensitivity terms. Thus, a data structure with a minimal summation of the respective sensitivities will minimize the effect of the associated uncertainty on the estimation result. For example, consider the first numerator term, where a data segment with a structure that minimizes $\sum_{k=1}^{N} \frac{\partial y_k}{\partial \hat{\theta}}\delta y_k$ will significantly reduce the estimation error induced by δy_k. In the same way, data structures that minimize $\sum_{k=1}^{N} \frac{\partial y_k}{\partial \hat{\theta}}\frac{\partial y_k}{\partial \hat{\phi}_i}$ (i.e., with a high degree of orthogonality between the sensitivities of the target and non-target parameters), will attenuate the influence of $\Delta\phi_i$ on the estimation result. Finally, data structures that yield a large denominator term, i.e., $\sum_{k=1}^{N}\left(\frac{\partial y_k}{\partial \hat{\theta}}\right)^2$, can also be effective at reducing the overall estimation error. Note that the denominator term is the Fisher information, simplified under i.i.d. Gaussian noises [57,58], which reflects the data information content about the target parameter.

Based on these results and insights, a preliminary data rating formula ($Q_\theta^\pm$) is proposed to evaluate the quality of an arbitrary data segment according to its potential of propagating system uncertainties to the estimation result,

$$Q_\theta^\pm = \frac{\alpha_\Delta\left|\sum_{k=1}^{N} \frac{\partial \bar{y}_k}{\partial \hat{\theta}}(\hat{\theta}^\pm)\right| + \sum_{i=1}^{m}\left(\alpha_{\phi_i}\left|\sum_{k=1}^{N} \frac{\partial \bar{y}_k}{\partial \hat{\theta}}(\hat{\theta}^\pm)\frac{\partial \bar{y}_k}{\partial \hat{\phi}_i}(\hat{\theta}^\pm)\right|\right)}{\sum_{k=1}^{N}\left(\frac{\partial \bar{y}_k}{\partial \hat{\theta}}(\hat{\theta}^\pm)\right)^2}. \tag{10}$$

Smaller rating values indicate reduced uncertainty propagation and thus higher-quality data. The superscript $^\pm$ indicates that the rating formula can be evaluated both *a priori* and *a posteriori*, under the initial guess $\hat{\theta}^-$ and estimate $\hat{\theta}^+$, respectively, which is needed in the first and fourth steps of the aforementioned data selection algorithm in Section 3. The design of the rating formula is based on Eq. (9) under the following considerations.

- The sensitivity terms are normalized via Eq. (11),

$$\frac{\partial \bar{y}_k}{\partial \hat{\phi}_i} = \hat{\phi}_i \frac{\partial y_k}{\partial \hat{\phi}_i}, \tag{11}$$

  to account for large variations in magnitude among parameters. For example, Li-ion electrochemical battery models typically have diffusion parameters that are on the order of $10^{-10} - 10^{-15}$ m^2/s, while volume fraction parameters are of order 10^{-1}. Normalizing the sensitivity terms makes the rating nondimensional so that it reflects the impact of each uncertainty in proportion to the value of the target parameter.
- The amount of uncertainties, i.e., δy_k and $\Delta\phi_i$ in Eq. (9), are unknown in practice. Thus, the rating formula leverages the uncertainty-propagating data structures in Eq. (9) to indicate the extent to which the unknown system uncertainties may influence the estimation result. For instance, in each $\alpha_{\phi_i}\left|\sum_{k=1}^{N} \frac{\partial \bar{y}_k}{\partial \hat{\theta}}\frac{\partial \bar{y}_k}{\partial \hat{\phi}_i}\right|$ term, which is associated with the parameter uncertainty, $\left|\sum_{k=1}^{N} \frac{\partial \bar{y}_k}{\partial \hat{\theta}}\frac{\partial \bar{y}_k}{\partial \hat{\phi}_i}\right|$ accounts for the potential of the data to propagate the uncertainty in ϕ_i according to the observed data structures. Meanwhile, α_{ϕ_i} is a weight reflecting the magnitude of the uncertainty, which can be either estimated based on any rough knowledge of the parameter uncertainty magnitude, or tuned to specify a relative ratio between different uncertainties. Similarly, the first term in the numerator, $\alpha_\Delta\left|\sum_{k=1}^{N} \frac{\partial \bar{y}_k}{\partial \hat{\theta}}\right|$, accounts for the impact of the model/measurement uncertainty on the estimation result. This term is formulated by approximating the time-varying δy_k as its average value over the data segment, represented as bias weight α_Δ. This simplification is performed because of the unknown nature of the time-varying uncertainty, which is usually difficult to even have a moderate knowledge of. Selecting the value of α_Δ is a major challenge, as the measurement and model uncertainties (especially the latter), are dependent on the operating conditions (e.g., the input) and hence vary among data segments. Therefore, beyond the difficulty of hand-tuning (or guessing) the bias weight α_Δ, using a universal α_Δ for all data will intrinsically cause inaccuracy in the data rating, which will be illustrated in Section 4.2.2. This is addressed in Section 5 through the derivation of an improved rating formula that incorporates an adaptive approximation of the varying model/measurement uncertainty.
- The absolute value is applied to each numerator term in Eq. (10) so that a small rating can only be achieved if the data segment can mitigate the influence of all uncertainties. This is necessary because each uncertainty may have an unknown and even changing sign, which can cause different uncertainty terms to either add up or (partially) cancel out. Without applying the absolute value, the rating formula may substantially overestimate the quality of the data segment if the numerator terms erroneously cancel out. To put it another way, we consider the worst-case scenario, where the errors caused by different types of uncertainties add up.

It is important to note that the rating formula in Eq. (10) requires the sensitivity terms to be computed efficiently, which is often challenging through the conventional methods of manual perturbation [59] or solving sensitivity differential equations [13,24]. Therefore, the analytical sensitivity expressions introduced in Section 2 for the electrochemical parameters will be used to facilitate the sensitivity computation so that the data rating formula may be evaluated with minimal computational expense. Finally, the rating formula is not limited to the ordinary gradient-based least-squares algorithm, but applicable to general estimation methods with similar objectives of minimizing the squared error or variance, e.g., Kalman filter and moving horizon observer, among others.

4.2. Verification & validation

The data selection framework with the preliminary data rating formula is verified through simulation and validated through experiment in two Li-ion battery electrochemical parameter estimation problems. The first problem targets the solid-phase cathode lithium diffusion coefficient $D_{s,p}$ for estimation under uncertainty in the cathode active material volume fraction $\varepsilon_{s,p}$. The second problem is the converse, where $\varepsilon_{s,p}$ is estimated under uncertainty in $D_{s,p}$. As detailed in Section 2, both physical parameters are intrinsic to the modeled battery dynamics and serve as key indicators of battery SOH.

For simulation verification of the data selection framework, the battery output voltage data set is generated using the SPMe under various input current profiles and a true parameter set denoted as $(\theta, \boldsymbol{\phi})$. For experimental validation, the battery voltage data are attained through physical measurement of an LGM50T INR21700 Li–Nickel–Manganese–Cobalt (NMC) cell subjected to the same input current profiles. In this work, the true parameter set is adopted from [60], which implemented a variety of electrochemical measurement techniques to experimentally parameterize an LGM50 INR21700 cell. Several parameter values were adjusted according to [48] to incorporate the subtle differences between the LGM50 and LGM50T cells. This parameter set thus serves as a benchmark for the estimation results determined through both simulation and experiment. Alternative methods of obtaining a benchmark parameter set include acquiring it from the cell manufacturer (if possible) or system identification from experimental input–output data, as in [61–64]. In both simulation verification and experimental validation, the SPMe serves as the model used for estimation, which provides the modeled output voltage under the input current and estimated/uncertain parameter set $(\hat{\theta}^{+}, \hat{\boldsymbol{\phi}})$. Results are generated with an initial SOC of 50% under four drive-cycle input current profiles, namely the Federal Urban Driving Schedule (FUDS), Urban Dynamometer Driving Schedule (UDDS), US06 Highway Driving Schedule (US06), and Dynamic Stress Test (DST). These current profiles were selected to emulate the estimation of electric vehicle battery SOH from online operation data, as each profile represents a typical battery operation scenario in an electric vehicle application. Drive-cycle profiles typically provide limited information about the health-related parameters due to the characteristic rapid current fluctuations and associated shallow discharges [65]. Each profile has a duration of 1800 s with a time step of 0.3 s, yielding 6000 samples.

In this section, the preliminary data rating formula only considers uncertainty in one parameter $\Delta\phi$, while omitting the model/measurement uncertainty δy_k (with $\alpha_{\Delta} = 0$) due to the aforementioned variability and difficulty of tuning α_{Δ}. Accordingly, in simulation verification, the voltage data are generated under only parameter uncertainty with no model uncertainty to verify the effectiveness of data rating and selection to accommodate the former. Then in the subsequent experimental validation, data selection will be subject to both parameter and model/measurement uncertainty to reveal that the latter can substantially reduce the effectiveness of the rating, motivating the need to improve the rating formula with an adaptive approximation of model/measurement uncertainty (detailed in Section 5). For this scenario of estimation under uncertainty in one non-target parameter, the rating formula contains one numerator term and hence the weight α_{ϕ} can be arbitrary (i.e., α_{ϕ} uniformly scales the rating for all data segments and thus plays no role in discerning data quality).

The data selection framework is evaluated under two scenarios of uncertainty for each target parameter, according to the uncertain parameter sets summarized in Table 1. Each parameter set contains uncertainty in both $D_{s,p}$ and $\varepsilon_{s,p}$, represented as deviations from the true values of 4.0×10^{-15} m^2/s and 0.5616, respectively. The remaining parameter values are consistent with the true parameter set. For example, in Uncertain Parameter Set II of Table 1, the initial guess of target parameter $D_{s,p}$ is 20% smaller than the true value, i.e., $\hat{D}^{-}_{s,p} = 0.8 D_{s,p}$, which will be used in the *a priori* data rating formula. This is represented in the parameter uncertainty notation of Eq. (9) as $\Delta D_{s,p} = D_{s,p} - \hat{D}^{-}_{s,p} = D_{s,p} - 0.8 D_{s,p} = 0.2 D_{s,p}$, or $\Delta D_{s,p} = 20\%$ of the true value. The uncertainty in $\varepsilon_{s,p}$ is −10%, where the assumed value $\hat{\varepsilon}_{s,p}$ is $0.9\varepsilon_{s,p}$, or $\Delta\varepsilon_{s,p} = 10\%$ of the true value. Note that, physically, a 10%–20% uncertainty in $\varepsilon_{s,p}$ is substantial because $\varepsilon_{s,p}$ is directly related to cell capacity [39,45], which typically degrades only 20% throughout the cell operating life in an electric vehicle application [3]. The variation in $D_{s,p}$ throughout the cell operating life is expected to be similar.

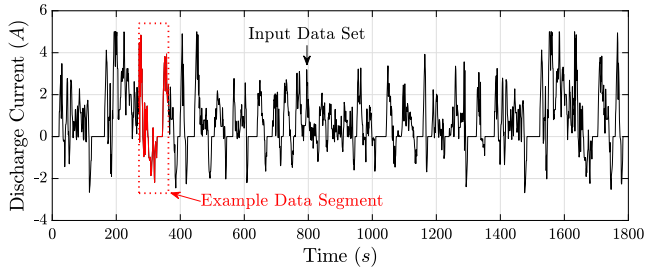
Table 1
Summary of uncertain parameter sets.

Uncertain Parameter Set	Target Parameter	$\Delta D_{s,p}$	$\Delta\varepsilon_{s,p}$
I	$D_{s,p}$	−20%	−20%
II	$D_{s,p}$	20%	10%
III	$\varepsilon_{s,p}$	20%	10%
IV	$\varepsilon_{s,p}$	−20%	10%

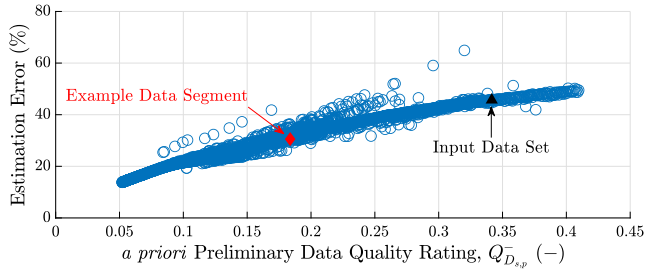
4.2.1. Simulation verification

The simulation verification entails applying the data selection algorithm to solve each aforementioned estimation problem under simulated output data, i.e., data generated via simulation under the benchmark parameter set. First, the effectiveness of the preliminary data quality rating formula will be assessed by examining the correlation between the quality rating and the parameter estimation error for different data segments within a data set. This correlation is presented in Fig. 2 for the estimation of $D_{s,p}$ under the FUDS input current profile. Specifically, Fig. 2(a) shows the FUDS current profile, from which different data segments can be extracted with different combinations of starting point and length. One such selected data segment is boxed in red as an example. The *a priori* rating was computed via Eq. (10) for every data segment that was extracted from the data set, and 3000 segments were examined as demonstration. These 3000 segments span the full range of observed *a priori* rating values as representative samples. The $D_{s,p}$ estimates were computed via Eq. (6) for each selected data segment and the results are plotted in Fig. 2(b), with the example segment marked by the red diamond and the full FUDS cycle by the black triangle. The *a posteriori* rating was then computed for each data segment and the resulting rating-error correlation is shown in Fig. 2(c). Four important insights can be drawn from this plot:

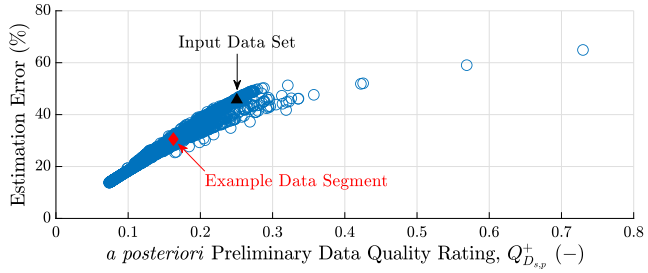
- The black triangles in Figs. 2(b) and 2(c) indicate that the estimation using the full FUDS cycle yielded an estimation error of 46%. The majority of the data segments (blue circles) yielded smaller estimation errors (as low as 14%), which attests to the benefit of data selection. Since the full FUDS cycle is the longest data segment, it is evident that longer segments do not necessarily lead to higher estimation accuracy in the presence of system uncertainties. This can be explained by examining the estimation error equation in Eq. (9), where both the denominator (Fisher information) and the numerator (reflecting the impact of uncertainties) can increase with the number of data points N. For example, consider a data segment with low ($\ll 1$) parameter sensitivity $\left(\frac{\partial y_k}{\partial \hat{\theta}}\right)$ and a high amount of uncertainty (δy_k and/or $\Delta\phi_i$). The growth of the numerator (first-order with respect to the product of $\frac{\partial y_k}{\partial \hat{\theta}}$ and the uncertainties) may outpace that of the denominator (quadratic with respect to $\frac{\partial y_k}{\partial \hat{\theta}}$) as the number of data points N increases, leading to an increasing estimation error.
- There is a good correlation between the estimation error and the data rating, which indicates that the rating can be used as an effective metric for evaluating data quality. The large spread of ratings and estimation errors among segments reveals that data quality can vary significantly throughout a given data set. Thus, it is critical for data segments to be carefully selected to achieve optimal/adequate estimation accuracy.



(a) FUDS current profile and example data segment.



(b) *A priori* rating-error correlation for different data segments, including full data set and example segment from (a).



(c) *A posteriori* rating-error correlation.

Fig. 2. Data rating results using preliminary rating formula for $D_{s,p}$ estimation under FUDS profile in simulation, with $\Delta D_{s,p}$: −20% (in initial guess) and $\Delta\varepsilon_{s,p}$: −20%.

- The minimum achievable estimation error was 14%, indicating that estimation accuracy is limited by the data structures present in the data set. This is a fundamental limitation—the data selection framework seeks to optimize the use of available data, but can only provide estimates as accurate as the available data will allow.
- By comparing Figs. 2(b) with 2(c), it is seen that incorporating knowledge of the estimates through the *a posteriori* rating evaluation renders a more monotonic and cleaner rating-error correlation than the *a priori* correlation. This can be explained by examining the estimation error equation in Eq. (9), which specifies the use of the estimated (instead of guessed) target parameter value. The observed refinement facilitates the selection (or confirmation) of high-quality data segments for determining the final estimation result, as the highest-quality data segments are more likely to be distinguished by the smallest ratings.

Fig. 3 shows two more examples of *a priori* and *a posteriori* rating-error correlations for the estimation of $D_{s,p}$ and $\varepsilon_{s,p}$ under different current profiles. These plots reinforce the insights from Fig. 2; specifically, the large spread of estimation errors indicates that data segment quality can vary substantially throughout a given data set, the strong rating-error correlations attest to the effectiveness of the preliminary rating formula for evaluating data segment quality, and the improved monotonicity of the *a posteriori* rating-error correlations indicate that *a posteriori* rating evaluations can facilitate the selection of high-quality

Table 2
$D_{s,p}$ estimation results using preliminary rating formula for data selection in simulation.

Case			$D_{s,p}$ Estimation Error	
$\Delta D_{s,p}$	$\Delta\varepsilon_{s,p}$	Input Profile	Data Selection Framework (Preliminary Rating)	Conventional Univariate Approach
−20%	−20%	FUDS	−13.8%	−45.7%
		UDDS	0.4%	−45.9%
		US06	−20.5%	−46.1%
		DST	−18.9%	−45.4%
20%	10%	FUDS	7.9%	54.6%
		UDDS	0.4%	55.6%
		US06	12.8%	56.7%
		DST	12.1%	53.4%

Table 3
$\varepsilon_{s,p}$ estimation results using preliminary rating formula for data selection in simulation.

Case			$\varepsilon_{s,p}$ Estimation Error	
$\Delta D_{s,p}$	$\Delta\varepsilon_{s,p}$	Input Profile	Data Selection Framework (Preliminary Rating)	Conventional Univariate Approach
20%	10%	FUDS	0.003%	5.9%
		UDDS	0.003%	6.0%
		US06	0.005%	6.0%
		DST	0.003%	6.0%
−20%	10%	FUDS	−0.005%	−4.1%
		UDDS	−0.005%	−4.2%
		US06	−0.003%	−4.2%
		DST	−0.005%	−4.2%

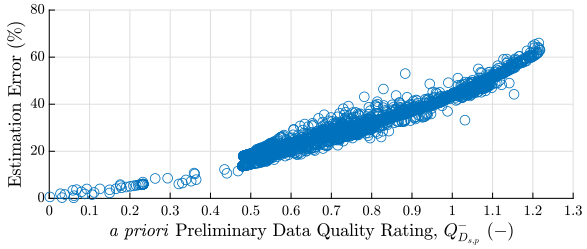
data segments. Notably, Fig. 3(b) indicates that a small portion of data segments achieved excellent estimation accuracy when $D_{s,p}$ was estimated under the UDDS profile, with errors as small as 0.25%. A comparison with the correlation in Fig. 2(c) exemplifies how a different (albeit seemingly similar) input data set, i.e., UDDS vs. FUDS, can significantly improve estimation accuracy by providing data structures that suppress the influence of system uncertainties on the estimation result.

For the estimation of $\varepsilon_{s,p}$, both rating-error correlations in Figs. 3(c) and 3(d) have relatively small estimation errors across all segments. These rating-error correlations are also stronger than those of $D_{s,p}$ (Figs. 2(c) and 3(b)). Both of these characteristics were attributed to the fact that battery voltage is substantially more sensitive to variations in $\varepsilon_{s,p}$ than $D_{s,p}$. This is illustrated in Fig. 4, which indicates that the RMS of the normalized $\varepsilon_{s,p}$ sensitivity (i.e., $\varepsilon_{s,p}\frac{\partial V_k}{\partial \varepsilon_{s,p}}$) is four times larger than that of $D_{s,p}$ under the FUDS current profile and true parameter set. Highly sensitive parameters like $\varepsilon_{s,p}$ are often estimated with relative ease, despite the presence of system uncertainties. This can be explained by examining the estimation error equation in Eq. (9), which reveals that highly sensitive parameters will yield a large denominator term (i.e., sum of squared sensitivity, or Fisher information), which can drive down the estimation error despite moderately-sized uncertainty terms in the numerator.

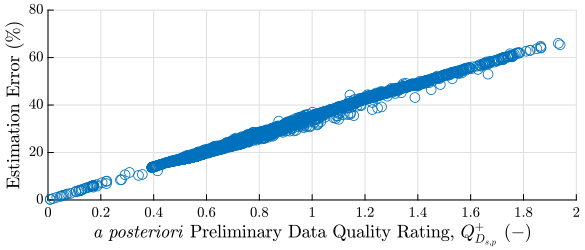
We then compiled the estimation results for the two target parameters under the aforementioned four scenarios of parameter uncertainty and four data sets, yielding a total of 16 cases for thorough evaluation of the data selection performance. For each case, the final estimation result was computed by averaging the estimates from the five data segments with the smallest *a posteriori* ratings. The results are summarized in Table 2 for the estimation of $D_{s,p}$ and Table 3 for the estimation of $\varepsilon_{s,p}$. For comparison, estimation results are included from the conventional univariate approach without data selection, i.e., using each complete data set.

The results in Tables 2 and 3 were interpreted as follows:

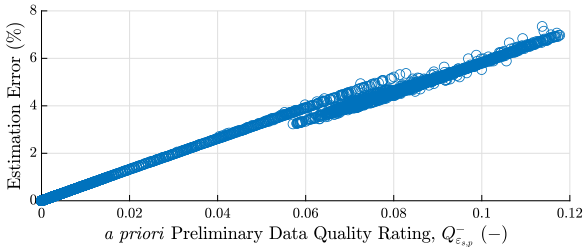
- *Data Selection Performance*: The data selection framework is capable of achieving excellent estimation accuracy by identifying the high-quality data segments present within each data set. This



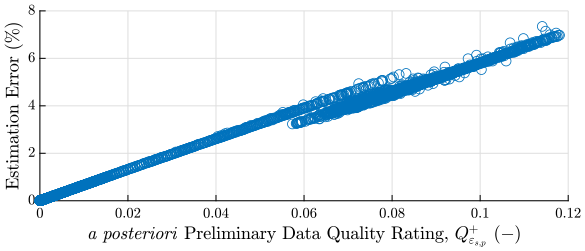
(a) *A priori* rating-error correlation for $D_{s,p}$ estimation under UDDS profile.



(b) *A posteriori* rating-error correlation for $D_{s,p}$ estimation under UDDS profile.



(c) *A priori* rating-error correlation for $\varepsilon_{s,p}$ estimation under FUDS profile.



(d) *A posteriori* rating-error correlation for $\varepsilon_{s,p}$ estimation under FUDS profile.

Fig. 3. Data rating results using preliminary rating formula for $D_{s,p}$ and $\varepsilon_{s,p}$ under different input profiles in simulation, with $\Delta D_{s,p}$: 20% and $\Delta\varepsilon_{s,p}$: 10%.

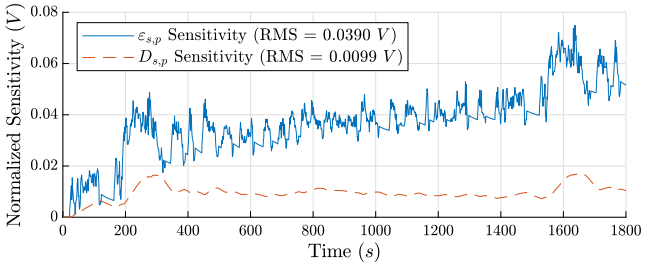


Fig. 4. Evolution of normalized $\epsilon_{s,p}$ and $D_{s,p}$ sensitivities under FUDS profile and true parameter set.

is illustrated by the 0.4% error for the estimation of $D_{s,p}$ under the UDDS profile in Table 2, and the maximum observed error of 0.005% for the estimation of $\varepsilon_{s,p}$ in Table 3. However, the results could only be as accurate as the available data would allow, resulting in relatively large estimation errors for data sets that do not contain uncertainty-suppressing data structures, e.g., the US06 profile under $\Delta D_{s,p}$: −20% and $\Delta\varepsilon_{s,p}$: −20% yielded an error of 20.5% in Table 2.

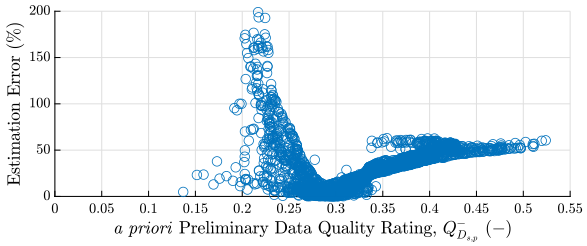
- *Influence of Parameter Sensitivity*: As illustrated in Fig. 4, the voltage output is significantly more sensitive to variations in $\varepsilon_{s,p}$ than $D_{s,p}$. Traditionally, attempting to estimate the weakly sensitive $D_{s,p}$ under the shadow of uncertainty in the strongly sensitive $\varepsilon_{s,p}$ is extremely difficult [47,48]. However, the results in Table 2 show that $D_{s,p}$ can be estimated with excellent accuracy through the selection of data segments that mitigate the influence of uncertainty in $\varepsilon_{s,p}$. Regarding the estimation of $\varepsilon_{s,p}$, the errors in Table 3 are smaller ($\lesssim$ 6%) for every case, even without data selection. This was attributed to the high $\varepsilon_{s,p}$ sensitivity, as strongly sensitive parameters are typically estimated with less difficulty due to their robustness against uncertainties.
- *Comparison with Conventional Approach*: The data selection framework consistently improved estimation accuracy over the conventional approach of estimating without data selection, by as much as two orders of magnitude in the estimation of $D_{s,p}$ and three orders of magnitude in the estimation of $\varepsilon_{s,p}$. Thus, it is not desirable to use arbitrary data sets (e.g., generated online) for parameter estimation without considering the data quality. This also indicates that random data sets, which may yield inaccurate estimation results as a whole, often contain high-quality data segments that can be leveraged to improve estimation accuracy. It is notable that, although the influence of uncertainties may be minor for strongly sensitive parameters like $\varepsilon_{s,p}$, estimation accuracy can still be improved by considering the effects of uncertainties in data selection.

To sum up, the strategy of evaluating the data quality rating *a priori* and *a posteriori* was demonstrated to be effective in simulation under the presence of parameter uncertainty. The *a priori* rating, albeit less precise, is reliable for selecting data segments with the potential to yield accurate estimates, given the available knowledge of the system. The *a posteriori* rating incorporates the knowledge of the estimates to refine the correlation between the estimation error and quality rating, enhancing the reliability of subsequent selections.

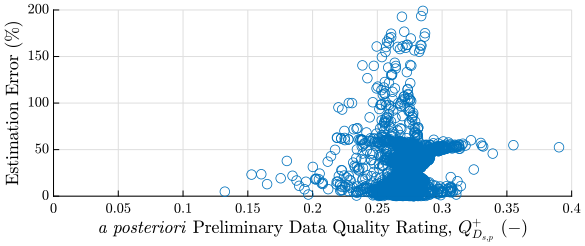
4.2.2. Experimental validation

The experimental validation of the data selection framework with the preliminary data rating formula follows the same procedure as the simulation verification, except that the measured output voltage data set is acquired through physical measurement of an LGM50T INR21700 cell, rather than generated through simulation. Accordingly, model/measurement uncertainty is present due to unmodeled system dynamics and/or sensor noise, which will be shown to adversely impact estimation performance. Voltage data is measured with an Arbin LBT21084 cycler under the same four drive-cycle current profiles, i.e., FUDS, UDDS, US06, and DST. The cell is initialized at 50% SOC for each profile by charging it to the cut-off voltage via the constant-current-constant-voltage protocol, and then discharging it for 30 min at 1C, based on the measured capacity.

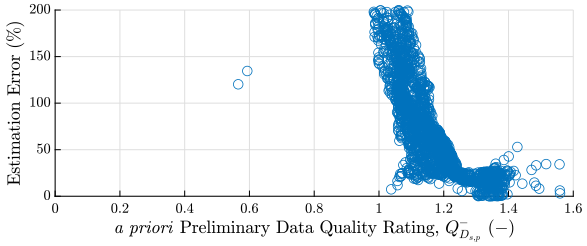
Two examples of *a priori* and *a posteriori* rating-error correlations from experimental data are shown in Fig. 5 for the estimation of $D_{s,p}$. Unlike in simulation, the rating-error relationships are generally much less monotonic and consistent. For example, the *a priori* rating-error correlation in Fig. 5(a) shows that data segments with low-accuracy estimates yielded small (good) ratings while the segments with the most accurate estimates returned mid-range (worse) ratings. Thus, the



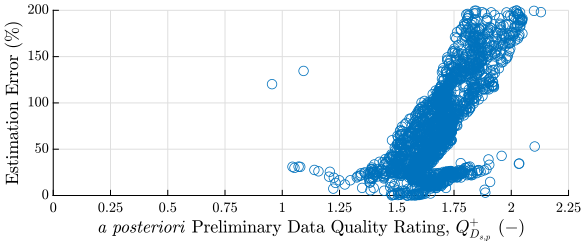
(a) *A priori* rating-error correlation under FUDS profile, $\Delta D_{s,p}$: -20%, and $\Delta\varepsilon_{s,p}$: -20%.



(b) *A posteriori* rating-error correlation under FUDS profile, $\Delta D_{s,p}$: -20%, and $\Delta\varepsilon_{s,p}$: -20%.



(c) *A priori* rating-error correlation under UDDS profile, $\Delta D_{s,p}$: 20%, and $\Delta\varepsilon_{s,p}$: 10%.



(d) *A posteriori* rating-error correlation under UDDS profile, $\Delta D_{s,p}$: 20%, and $\Delta\varepsilon_{s,p}$: 10%.

Fig. 5. Experimental $D_{s,p}$ *a priori* and *a posteriori* rating-error correlations using preliminary rating formula under different input profiles and parameter uncertainties.

a priori rating becomes ineffective for discerning between high- and low-quality segments, as the highest-quality segments can no longer be identified as the segments with the smallest (best) ratings. The *a posteriori* rating-error correlations were slightly more monotonic than the *a priori* correlations, where several high-quality data segments achieved small ratings while most low-quality segments returned large ratings. This separation between high- and low-quality data may be sufficient to return an accurate final estimate by averaging the estimates from the segments with the smallest *a posteriori* ratings, according to the data selection algorithm. However, this is not guaranteed, as Fig. 5(d) indicates that low-quality data segments may remain associated with the smallest ratings.

The degradation in rating performance is due to the presence of model/measurement uncertainty, which is the only difference between the voltage data sets used for the simulation verification and experimental validation. Since measurement uncertainty (i.e., sensor noise/bias) is negligible in our high-precision testing equipment, we will henceforth refer to the model/measurement uncertainty simply as model uncertainty, as this is the dominant component. However, the discussion still applies to the lumped model/measurement uncertainty. Following from Eq. (7), the model uncertainty is defined as the difference between the measured and modeled system outputs under the true parameter set,

$$\delta y_k = y_k^m - y_k(\theta, \boldsymbol{\phi}, \boldsymbol{u}_k), \tag{12}$$

which is caused by unmodeled or imperfectly modeled system dynamics. As discussed in Section 2, the implemented SPMe battery model is derived from the full-order DFN model through the simplifying assumption that lithium intercalation current density is uniform across each electrode. This simplification has been demonstrated to maintain good accuracy under low current amplitudes, but errors grow as current increases [50]. Regardless, even the DFN model is subject to assumptions (e.g., electrode particles are spherical with uniform radii) and will still yield discrepancies against measured output data [61]. No model can perfectly capture the exact dynamics of a physical system, thus some level of model uncertainty will always be present.

Fig. 6 provides an example visualization of the model uncertainty under the FUDS current profile, in which Fig. 6(a) compares the measured voltage with the modeled voltage under the benchmark parameter set. The model uncertainty was computed according to Eq. (12) as the difference between the measured and modeled voltage responses, and is shown in Fig. 6(b). In this case, the model uncertainty varies in both sign and amplitude with high-frequency fluctuations superposed over a gradual decline. As discussed in Section 4.1, the estimation error equation in Eq. (9) indicates that a data structure will mitigate the influence of model uncertainty on the estimation result if $\sum_{k=1}^{N} \frac{\partial y_k}{\partial \hat{\theta}} \delta y_k$ is small. Since the time-varying δy_k is difficult to predict, the proposed strategy of incorporating it into the rating was to approximate it as a constant (average) value for the entire data set, represented by the bias weight α_Δ in Eq. (10). Disregarding the difficulty of adequately guessing/tuning α_Δ, Fig. 6(b) reveals that approximating δy_k as a constant can introduce considerable error, as the model uncertainty can vary significantly throughout the data set.

The last step of the data selection algorithm was to return the final result by averaging the estimates from the five data segments with the smallest *a posteriori* ratings. The $D_{s,p}$ and $\varepsilon_{s,p}$ estimation results are presented in Tables 4 and 5, respectively. As with the simulation verification, estimation results are included from the conventional approach of univariate estimation without data selection. In addition, results are provided for two joint estimation scenarios (without data selection) to consider the common practice of simultaneously estimating all system uncertainties, including the model uncertainty. The first scenario is the bivariate joint estimation of $D_{s,p}$ and $\varepsilon_{s,p}$, which attempts to effectively eliminate the parameter uncertainty by simultaneously estimating both unknown parameters. The second scenario is the trivariate joint estimation of $D_{s,p}$, $\varepsilon_{s,p}$, and the unknown (and assumed constant) model uncertainty ΔV, which essentially attempts to estimate all system uncertainties,

$$\min_{\hat{D}^+_{s,p}, \hat{\varepsilon}^+_{s,p}, \Delta\hat{V}^+} J = \sum_{k=1}^{N} \left[V_k^m - \left(V_k(\hat{D}^+_{s,p}, \hat{\varepsilon}^+_{s,p}, \hat{\boldsymbol{\phi}}, \boldsymbol{u}_k) + \Delta\hat{V}^+ \right) \right]^2. \tag{13}$$

Tables 4 and 5 provide the following insights:

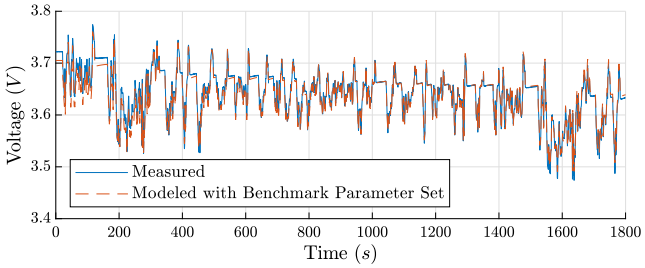
- *Data Selection Performance*: The data selection framework yielded several accurate results (e.g., 0.6% $D_{s,p}$ error in Table 4, 1.3% $\varepsilon_{s,p}$ error in Table 5), though estimation errors were generally higher than those observed in simulation (Tables 2 and 3). This was attributed to the presence of model uncertainty in the experimental data, which degraded the performance of the rating formula and often caused the mis-selection of the highest-quality

Table 4
Summary of experimental $D_{s,p}$ estimation results with preliminary rating formula.

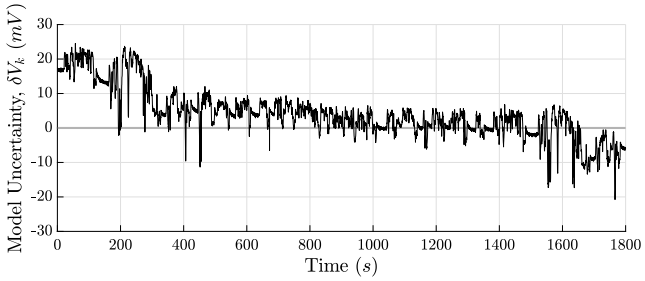
Case			$D_{s,p}$ Estimation Error			
$\Delta D_{s,p}$	$\Delta\varepsilon_{s,p}$	Input Profile	Data Selection Framework (Preliminary Rating)	Univariate Approach ($D_{s,p}$)	Bivariate Approach ($D_{s,p}, \varepsilon_{s,p}$)	Trivariate Approach ($D_{s,p}, \varepsilon_{s,p}, \Delta V$)
−20%	−20%	FUDS	16.8%	−32.0%	26799%	−32.0%
		UDDS	13.7%	−42.3%	26864%	26864%
		US06	−28.4%	−25.9%	26809%	26788%
		DST	−16.5%	−22.9%	26838%	849%
20%	10%	FUDS	52.6%	206%	26799%	312%
		UDDS	−0.6%	117%	26864%	26864%
		US06	−3.7%	337%	26809%	26788%
		DST	21.4%	318%	26838%	849%

Table 5
Summary of experimental $\varepsilon_{s,p}$ estimation results with preliminary rating formula.

Case			$\varepsilon_{s,p}$ Estimation Error			
$\Delta D_{s,p}$	$\Delta\varepsilon_{s,p}$	Input Profile	Data Selection Framework (Preliminary Rating)	Univariate Approach ($\varepsilon_{s,p}$)	Bivariate Approach ($D_{s,p}, \varepsilon_{s,p}$)	Trivariate Approach ($D_{s,p}, \varepsilon_{s,p}, \Delta V$)
20%	10%	FUDS	−1.3%	12.7%	−23.3%	−32.4%
		UDDS	−4.0%	7.5%	−26.7%	−35.7%
		US06	−2.1%	14.5%	−22.0%	−35.8%
		DST	−3.1%	15.7%	−22.3%	−35.4%
−20%	10%	FUDS	−7.7%	1.8%	−23.3%	−32.4%
		UDDS	−9.8%	−3.0%	−26.7%	−35.7%
		US06	−9.2%	3.3%	−22.0%	−35.8%
		DST	−9.7%	4.3%	−22.3%	−35.4%



(a) Comparison of measured and modeled voltage responses under benchmark parameter set.



(b) Model uncertainty computed as difference between measured and modeled voltage responses from (a).

Fig. 6. Visualization of model uncertainty under FUDS current profile.

data segments. Additionally, Fig. 5(d) reveals that the smallest estimation error of 0.6% (for the estimation of $D_{s,p}$ under UDDS) occurred somewhat fortuitously, as the estimates from the five segments with the smallest *a posteriori* ratings are each individually poor, but are centered around the true value such that the average error is small (note that the vertical axis in Fig. 5(d) is the absolute value of the estimation error). In other words, averaging the estimates from a different number of data segments would yield a larger error for this case.

- *Comparison with Conventional Univariate Estimation Approach*: The data selection framework generally delivered smaller estimation errors than the conventional univariate approach without data selection, yielding error improvements in both $D_{s,p}$ and $\varepsilon_{s,p}$ of one order of magnitude. Interestingly, for the estimation of $\varepsilon_{s,p}$, the data selection framework outperformed the conventional approach for every current profile under $\Delta D_{s,p}$: 20% and $\Delta\varepsilon_{s,p}$: 10%, while the conventional approach performed better under $\Delta D_{s,p}$: −20% and $\Delta\varepsilon_{s,p}$: 10%. This inconsistent behavior under different parameter sets was not observed in simulation and is thus attributed to the presence of model uncertainty—the only difference between the two scenarios.
- *Comparison with Bi- and Trivariate Joint Estimation Approaches*: The bivariate joint estimation of $D_{s,p}$ and $\varepsilon_{s,p}$ is generally inaccurate with minimum errors of 26,799% for $D_{s,p}$ and 22% for $\varepsilon_{s,p}$. This is because the estimation problem is ill-posed under the drive-cycle data, and the estimator attempted to reconcile the error between the measured and modeled voltage outputs by varying the values of $\hat{D}_{s,p}$ and $\hat{\varepsilon}_{s,p}$, though the errors were caused by mechanisms beyond parameter uncertainty, i.e., unmodeled system dynamics. Mathematically, we can understand this high sensitivity of joint estimation error to model uncertainty by examining the error equation for the bivariate estimation of two target parameters θ_1 and θ_2, which is presented as Eq. (14) for the first target parameter θ_1. The error equation for the second target parameter θ_2 follows a symmetric form. The equation is derived following the same procedure as the univariate estimation error equation in Eq. (9), and the numerator is abbreviated due to the complicated structure and limited space.

$$\Delta\theta_1 = \frac{\frac{\{\cdots\}}{\sum_{k=1}^{N}\left(\frac{\partial y_k}{\partial\hat{\theta}_2}(\hat{\theta}^+)\right)^2}}{\sum_{k=1}^{N}\left(\frac{\partial y_k}{\partial\hat{\theta}_1}(\hat{\theta}^+)\right)^2 - \frac{\left(\sum_{k=1}^{N}\frac{\partial y_k}{\partial\hat{\theta}_1}(\hat{\theta}^+)\frac{\partial y_k}{\partial\hat{\theta}_2}(\hat{\theta}^+)\right)^2}{\sum_{k=1}^{N}\left(\frac{\partial y_k}{\partial\hat{\theta}_2}(\hat{\theta}^+)\right)^2}} \tag{14}$$

The form of Eq. (14) mirrors that of Eq. (9), where the denominator is the determinant of the Fisher information matrix [57,58] and each numerator term describes the propagation of one system uncertainty to the estimation result, i.e., δy_k and $\Delta\phi_i$. Thus, the denominator is indicative of estimation robustness to system uncertainties, as it will proportionally reduce (or amplify) the influence of all uncertainty-propagating numerator terms. Comparison of the bivariate estimation error equation in Eq. (14) with the univariate form in Eq. (9) reveals that the denominator of the bivariate form is always smaller by a difference of $\frac{\left(\sum_{k=1}^{N}\frac{\partial y_k}{\partial\theta_1}\frac{\partial y_k}{\partial\theta_2}\right)^2}{\sum_{k=1}^{N}\left(\frac{\partial y_k}{\partial\theta_2}\right)^2}$. Thus, simultaneously estimating a second target parameter will always reduce the denominator of the estimation error equation, causing the estimation result to be less robust against the uncertainties. Similarly, the trivariate joint estimation was also ill-posed and yielded inaccurate results. Fundamentally, this approach of attempting to eliminate all system uncertainties by adding more target parameters for estimation is susceptible to ill-posedness in the absence of adequate information/data. Accordingly, the resulting estimates are often inaccurate or even nonunique (e.g., a denominator of zero in the estimation error equation indicates unidentifiability) [31,61,66].

Evidently, model uncertainty critically influences estimation accuracy for the data selection framework and conventional estimation approaches. The remedy of approximating model uncertainty as a constant bias in the data rating formula or performing multi-variate joint estimation was found to be ineffective due to the substantial variation of model uncertainty among data segments. Thus, we propose a new strategy to fix the issue by adaptively incorporating model uncertainty in the rating formula.

5. Data quality rating adaptive to model/measurement uncertainty: Derivation & validation

The results in Section 4.2 revealed that model uncertainty can significantly degrade estimation accuracy, not only for the data selection framework but for conventional univariate and joint estimation approaches as well. This is a serious concern for EV BMSs, where model fidelity is restricted by the onboard computational resources, and high-precision measurement hardware is typically cost-prohibitive—indeed, model/measurement uncertainty is inevitable in practice. To improve the capability of the data selection framework to differentiate data in terms of the model/measurement uncertainty, an adaptive approximation of the model/measurement uncertainty is derived and embedded into the data quality rating formula. Experimental validation follows. For brevity, we will continue to refer to the model/measurement uncertainty simply as model uncertainty, in reference to the dominant component in our Li-ion battery application.

5.1. Derivation

As discussed in Section 4, the challenge with incorporating varying model uncertainty into the rating formula is that it is varying and depends on the true parameter values, according to Eq. (12). Since the true parameter values are unknown, the model uncertainty can only be approximated. Recent works have developed data-driven approaches for predicting model uncertainty with recurrent neural networks [67], feedforward neural networks [68,69], polynomial regression [70], and Gaussian process regression [70]. However, these techniques cannot be adopted in the data selection framework because they depend on training data that is generated under the unknown true parameter values. For this reason, a new approach is developed for approximating model uncertainty under an uncertain parameter set.

The basic idea of the new approach is to approximate the time-varying model uncertainty as

$$\begin{aligned}\delta y_k &= y_k^m - y_k(\theta,\boldsymbol{\phi},\boldsymbol{u}_k)\\ &\approx y_k^m - \left[y_k(\hat{\theta}^-,\hat{\boldsymbol{\phi}},\boldsymbol{u}_k) + \frac{\partial y_k}{\partial\hat{\theta}}(\hat{\theta}^-,\hat{\boldsymbol{\phi}},\boldsymbol{u}_k)(\theta-\hat{\theta}^-) + \sum_{i=1}^{m}\left(\frac{\partial y_k}{\partial\hat{\phi}_i}(\hat{\theta}^-,\hat{\boldsymbol{\phi}},\boldsymbol{u}_k)\Delta\phi_i\right)\right],\end{aligned} \tag{15}$$

where the modeled output under the true parameter set $y_k(\theta,\boldsymbol{\phi},\boldsymbol{u}_k)$ is approximated by the 1st order Taylor expansion about the *a priori* parameter set $(\hat{\theta}^-,\hat{\boldsymbol{\phi}})$. It is important for the Taylor expansion to be centered about the *a priori* parameter set because the model uncertainty approximation in Eq. (15) will be combined with the estimation error equation in Eq. (9), which was derived through a Taylor expansion of $y_k(\theta,\boldsymbol{\phi},\boldsymbol{u}_k)$ about the *a posteriori* parameter set $(\hat{\theta}^+,\hat{\boldsymbol{\phi}})$. Thus, approximating $y_k(\theta,\boldsymbol{\phi},\boldsymbol{u}_k)$ from the alternative perspective of the *a priori* parameter set will provide the error equation with new information that can be leveraged to characterize the model uncertainty. The adaptive *a posteriori* rating formula is hence derived by combining the model uncertainty approximation in Eq. (15) and the error equation in Eq. (9), and the final form is presented as Eq. (16) after normalizing by the estimate of the target parameter $(\hat{\theta}^+)$ and applying the absolute value.

$$Q_\theta^+ = \frac{\left|\sum_{k=1}^{N}\frac{\partial\bar{y}_k}{\partial\theta}(\hat{\theta}^+)\left[y_k^m - y_k(\hat{\theta}^-) + \frac{\partial\bar{y}_k}{\partial\theta}(\hat{\theta}^-)\left(\frac{\hat{\theta}^-}{\hat{\theta}^+}-1\right) + \sum_{i=1}^{m}\alpha_{\phi_i}\left(\frac{\partial\bar{y}_k}{\partial\hat{\phi}_i}(\hat{\theta}^+) - \frac{\partial\bar{y}_k}{\partial\hat{\phi}_i}(\hat{\theta}^-)\right)\right]\right|}{\left|\sum_{k=1}^{N}\frac{\partial\bar{y}_k}{\partial\theta}(\hat{\theta}^+)\left(\frac{\partial\bar{y}_k}{\partial\theta}(\hat{\theta}^+) - \frac{\partial\bar{y}_k}{\partial\theta}(\hat{\theta}^-)\right)\right|} \tag{16}$$

The adaptive rating formula incorporates the model uncertainty without any need for prior knowledge or tuning. Additionally, the full time-varying model uncertainty is captured without being oversimplified as a constant, as was necessary for the preliminary rating in Eq. (10) and the trivariate joint estimation in Eq. (13). This is enabled by a distinct feature of the adaptive rating, i.e., it involves both the *a priori* and *a posteriori* values of the target parameter. Specifically, besides using the aforementioned sensitivity-based data structures related to uncertainty propagation, the rating also leverages the difference between the *a priori* and *a posteriori* values, featured by $\left(\frac{\hat{\theta}^-}{\hat{\theta}^+}-1\right)$ in the numerator and other terms. This difference essentially indicates the extent to which the initial guess of the target parameter can be changed/improved by the data. In this way, the rating formula not only evaluates the potential of a data segment to amplify/attenuate uncertainty, but also implicitly incorporates the (model) uncertainty itself. However, since the new rating cannot be evaluated *a priori* only, it is recommended that the preliminary rating in Eq. (10) continue to be used for the *a priori* evaluation in the data selection algorithm. Meanwhile, the denominator of the adaptive rating no longer represents the Fisher information of the target parameter, as it is smaller than the preliminary rating denominator by a difference of $\sum_{k=1}^{N}\frac{\partial\bar{y}_k}{\partial\theta}(\hat{\theta}^+)\frac{\partial\bar{y}_k}{\partial\theta}(\hat{\theta}^-)$. Regardless, the denominator can still be large if the target parameter is strongly sensitive, which is indicative of high-quality data through robustness against the uncertainty-propagating numerator terms. Thus, it is still desirable for the target parameter to be strongly sensitive, as it was for the preliminary rating. On the other hand, the $\sum_{k=1}^{N}\frac{\partial\bar{y}_k}{\partial\theta}(\hat{\theta}^+)\frac{\partial\bar{y}_k}{\partial\theta}(\hat{\theta}^-)$ term can substantially reduce the rating denominator, which may cause numerical instability in computation due to the unaccounted errors in the numerator associated with the truncated higher-order terms of the Taylor expansion. For example, if the rating denominator is very small for a given data segment, a minor error in the numerator (due to Taylor approximations) can be amplified to substantially impact the rating value. To mitigate the possibility of mis-rating segments, a threshold will be applied to the rating denominator to enforce a minimum acceptable value.

Table 6
Summary of experimental $D_{s,p}$ estimation results with adaptive rating formula.

Case				$D_{s,p}$ Estimation Error			
$\Delta D_{s,p}$	$\Delta\varepsilon_{s,p}$	$\alpha_{\varepsilon_{s,p}}$	Input Profile	Data Selection Framework (Adaptive Rating)	Univariate Approach ($D_{s,p}$)	Bivariate Approach ($D_{s,p}, \varepsilon_{s,p}$)	Trivariate Approach ($D_{s,p}, \varepsilon_{s,p}, \Delta V$)
−20%	−20%	1/12	FUDS	−4.8%	−32.0%	26799%	−32.0%
			UDDS	−4.7%	−42.3%	26864%	26864%
			US06	−2.6%	−25.9%	26809%	26788%
			DST	−3.5%	−22.9%	26838%	849%
20%	10%	−1/3	FUDS	18.6%	206%	26799%	312%
			UDDS	5.4%	117%	26864%	26864%
			US06	30.8%	337%	26809%	26788%
			DST	37.6%	318%	26838%	849%

Table 7
Summary of experimental $\varepsilon_{s,p}$ estimation results with adaptive rating formula.

Case				$\varepsilon_{s,p}$ Estimation Error			
$\Delta D_{s,p}$	$\Delta\varepsilon_{s,p}$	$\alpha_{D_{s,p}}$	Input Profile	Data Selection Framework (Adaptive Rating)	Univariate Approach ($\varepsilon_{s,p}$)	Bivariate Approach ($D_{s,p}, \varepsilon_{s,p}$)	Trivariate Approach ($D_{s,p}, \varepsilon_{s,p}, \Delta V$)
20%	10%	−3/8	FUDS	3.6%	12.7%	−23.3%	−32.4%
			UDDS	−1.5%	7.5%	−26.7%	−35.7%
			US06	3.5%	14.5%	−22.0%	−35.8%
			DST	4.7%	15.7%	−22.3%	−35.4%
−20%	10%	1/12	FUDS	0.8%	1.8%	−23.3%	−32.4%
			UDDS	−3.0%	−3.0%	−26.7%	−35.7%
			US06	0.6%	3.3%	−22.0%	−35.8%
			DST	1.5%	4.3%	−22.3%	−35.4%

5.2. Experimental validation

The experimental validation of the adaptive rating formula is identical to that of the preliminary rating formula in Section 4.2.2, except that the *a posteriori* rating evaluation is performed with the adaptive rating formula in Eq. (16). Fig. 7 shows the *a posteriori* rating-error correlations for the same examples in Fig. 5, where the data segments are simply rearranged along the horizontal-axis according to the new rating values. Fig. 7 also includes correlations for $\varepsilon_{s,p}$ under different current profiles and parameter uncertainties. The α_{ϕ} weight was selected to be 1/12, −1/3, −3/8, and 1/12 for the four uncertain parameter sets in Table 1, corresponding to arbitrary parameter uncertainty guesses ($\Delta\hat{\phi}$) of 10%, −30%, −30%, and 10% of the true parameter values, respectively, as $\alpha_{\phi}\hat{\phi} = \Delta\hat{\phi}$. To improve the numerical stability of the rating, the denominator threshold was selected to be 0.1 for $D_{s,p}$ estimation and 1.7 for $\varepsilon_{s,p}$ estimation. These values were hand-tuned for each target parameter, though we intend to develop a method for automatically determining them in future work.

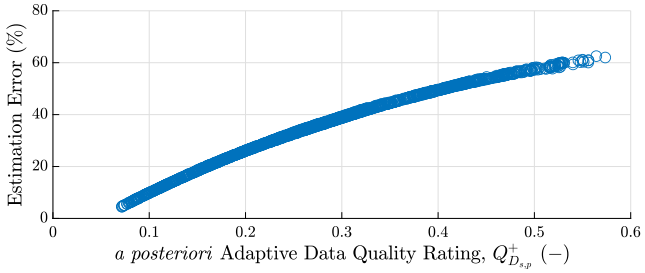
It can be seen that each rating-error correlation in Fig. 7 is strongly monotonic, indicating the adequacy of using it for data selection, as the highest-quality data segments are associated with the smallest ratings. The rating-error correlations were significantly improved over those using the preliminary rating formula shown in Fig. 5, which attests to the effectiveness of the model uncertainty approximation for enhancing the reliability of the rating formula. For each case, the final estimate was returned by averaging the estimates from the five data segments with the smallest *a posteriori* ratings. The $D_{s,p}$ and $\varepsilon_{s,p}$ estimation results are summarized in Tables 6 and 7 alongside results from the conventional approaches of univariate, bivariate, and trivariate estimation without data selection.

In comparison with the results under the preliminary rating formula (Tables 4 and 5), the adaptive rating improved the estimation accuracy for most cases, with several error reductions of one order of magnitude. The adaptive rating also delivered tighter error variation than the preliminary rating, i.e., [−4.8%, 37.6%] vs. [−28.4%, 52.6%] for $D_{s,p}$ and [−3.0%, 4.7%] vs. [−9.8%, −1.3%] for $\varepsilon_{s,p}$. Estimation accuracy was notably improved in $\varepsilon_{s,p}$ for the case of $\Delta D_{s,p}$: −20% and $\Delta\varepsilon_{s,p}$: 10%, with errors consistently less than or equal to those of the three conventional approaches. In summary, the data selection framework outperformed or matched the conventional approaches in every case, and yielded estimation errors that were at least one order of magnitude smaller in 28 out of the 48 cases. Thus, the adaptive rating formula has been validated as an effective means for improving the performance of the data selection framework.

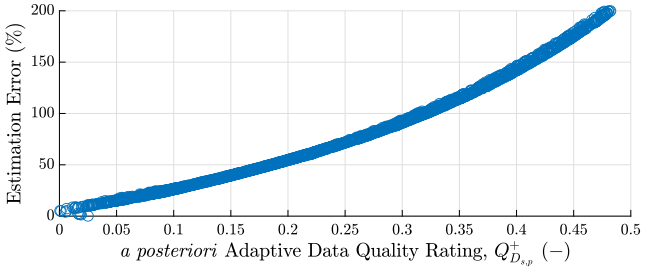
6. Conclusions

In this paper, an uncertainty-aware data selection framework was proposed and demonstrated for the accurate estimation of health-related Li-ion battery parameters. The foundation of the framework is the proposed data quality rating, which is a metric for predicting the extent to which a selected data segment will propagate system uncertainties to the estimation result. The data selection and estimation procedures were integrated through *a priori* and *a posteriori* evaluations of data segment quality, and the *a posteriori* data quality rating was enhanced with a novel approximation of model/measurement uncertainty that considers the influence of unmodeled dynamics and/or sensor noise. Two health-related electrochemical parameters, $D_{s,p}$ and $\varepsilon_{s,p}$, were separately estimated in simulation and experiment to evaluate the performance of the framework. Excellent estimation accuracy was achieved, even in the difficult case of estimating the weakly sensitive $D_{s,p}$ under uncertainty in the strongly sensitive $\varepsilon_{s,p}$. Additionally, model/measurement uncertainty was generally observed to reduce estimation accuracy, though its influence was mitigated through successful integration of the adaptive model/measurement uncertainty approximation into the rating formula. The framework yielded experimental estimation error reductions of one order of magnitude when compared with the conventional approaches of univariate and multivariate estimation without data selection. It was thus validated as an effective means for improving parameter estimation accuracy when control authority is not available over the data.

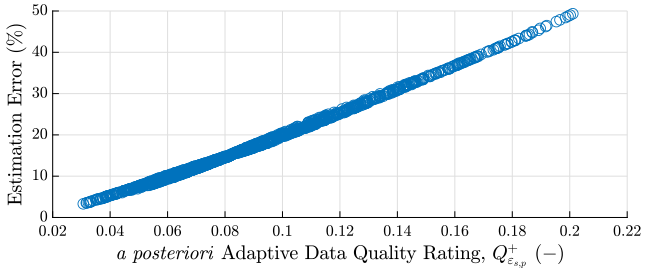
The data selection framework is significant because it has the potential to change the paradigm of both online and offline parameter estimation. Throughout this work, we showed that drive-cycle data



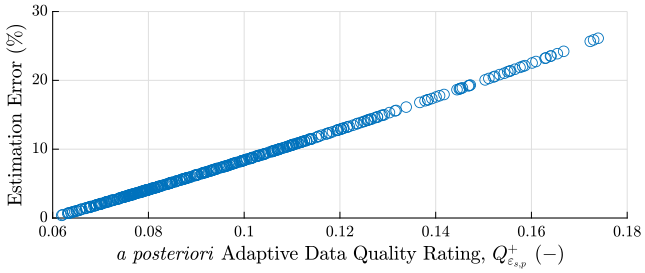
(a) Estimation of $D_{s,p}$ under FUDS profile, $\Delta D_{s,p}$: -20%, $\Delta\varepsilon_{s,p}$: -20%, and $\alpha_{\varepsilon_{s,p}}$: 1/12.



(b) Estimation of $D_{s,p}$ under UDDS profile, $\Delta D_{s,p}$: 20%, $\Delta\varepsilon_{s,p}$: 10%, and $\alpha_{\varepsilon_{s,p}}$: -1/3.



(c) Estimation of $\varepsilon_{s,p}$ under FUDS profile, $\Delta D_{s,p}$: 20%, $\Delta\varepsilon_{s,p}$: 10%, and $\alpha_{D_{s,p}}$: -3/8.



(d) Estimation of $\varepsilon_{s,p}$ under US06 profile, $\Delta D_{s,p}$: -20%, $\Delta\varepsilon_{s,p}$: 10%, and $\alpha_{D_{s,p}}$: 1/12.

Fig. 7. Experimental *a posteriori* rating-error correlations for $D_{s,p}$ and $\epsilon_{s,p}$ using adaptive rating formula under different input profiles and parameter uncertainties.

(e.g., FUDS, UDDS, US06, and DST), which are frequently used for battery SOH estimation as the only available data during battery operation, may not provide accurate estimation results. This is mainly due to the existence of large portions of low-quality data (low sensitivity and high uncertainty) in the cycle. The contribution of this paper is to propose a data selection scheme that can extract the high-quality data segments from random operational data to improve estimation accuracy. In online estimation, high-quality data segments can be selected from a window of a random data stream, which facilitates the optimal use of the incoming data when it cannot be controlled/designed. In offline estimation, high-quality data segments can be selected from an existing database of measurements. This may eliminate or reduce the time, equipment, labor, and expertise required to design and conduct experiments, which would otherwise be necessary to generate suitable data sets when accurate estimation results are critical. For example, consider the emerging field of battery repurposing, in which retired electric vehicle batteries must undergo a lengthy health testing and diagnosis procedure before being appropriately repackaged for stationary applications. Through the data selection framework, an existing high-quality data segment from the vehicle BMS might be extracted and used to quickly and accurately estimate the health-related battery parameters, as was demonstrated in the verification and validation studies in Sections 4 and 5, circumventing the need for lengthy experimentation.

CRediT authorship contribution statement

Jackson Fogelquist: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing – original draft, Visualization. **Xinfan Lin:** Conceptualization, Methodology, Validation, Resources, Writing – original draft, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We appreciate the funding support from the NSF CAREER Program, United States (Grant No. 2046292) and the NASA HOME Space Technology Research Institute, United States (Grant No. 80NSSC19K1052).

References

[1] Chaturvedi NA, Klein R, Christensen J, Ahmed J, Kojic A. Algorithms for advanced battery-management systems. IEEE Control Syst Mag 2010;30(3):49–68. http://dx.doi.org/10.1109/MCS.2010.936293.

[2] Xiong R, Li L, Tian J. Towards a smarter battery management system: A critical review on battery state of health monitoring methods. J Power Sources 2018;405:18–29. http://dx.doi.org/10.1016/j.jpowsour.2018.10.019.

[3] Lin X, Kim Y, Mohan S, Siegel JB, Stefanopoulou AG. Modeling and estimation for advanced battery management. In: Annual review of control, robotics, and autonomous systems. Vol. 2. Annual Reviews; 2019, p. 393–426. http://dx.doi.org/10.1146/annurev-control-053018-023643.

[4] Cramér H. Mathematical methods of statistics (PMS-9). Princeton University Press; 1999, URL http://www.jstor.org/stable/j.ctt1bpm9r4.

[5] Cover TM, Thomas JA. Elements of information theory. second ed.. Hoboken, NJ: John Wiley & Sons, Inc; 2006, p. 392–9.

[6] Fedorov VV. Theory of optimal experiments. New York, NY: Academic Press, Inc; 1972, p. 27–30.

[7] Pronzato L, Walter E. Robust experiment design via stochastic approximation. Math Biosci 1985;75(1):103–20. http://dx.doi.org/10.1016/0025-5564(85)90068-9, URL https://www.sciencedirect.com/science/article/pii/0025556485900689.

[8] Emery AF, Nenarokomov AV. Optimal experiment design. Meas Sci Technol 1998;9(6):864–76. http://dx.doi.org/10.1088/0957-0233/9/6/003.

[9] Lin X. Analytic analysis of the data-dependent estimation accuracy of battery equivalent circuit dynamics. IEEE Control Syst Lett 2017;1(2):304–9. http://dx.doi.org/10.1109/LCSYS.2017.2715821.

[10] Song Z, Hofmann H, Lin X, Han X, Hou J. Parameter identification of lithium-ion battery pack for different applications based on Cramer-Rao bound analysis and experimental study. Appl Energy 2018;231:1307–18. http://dx.doi.org/10.1016/j.apenergy.2018.09.126, URL https://www.sciencedirect.com/science/article/pii/S0306261918314375.

[11] Forman JC, Moura SJ, Stein JL, Fathy HK. Optimal experimental design for modeling battery degradation. In: Dynamic systems and control conference. Vol. 1. 2012, p. 309–18. http://dx.doi.org/10.1115/DSCC2012-MOVIC2012-8751, arXiv:https://asmedigitalcollection.asme.org/DSCC/proceedings-pdf/DSCC2012-MOVIC2012/45295/309/4446721/309_1.pdf.

[12] Rothenberger MJ, Docimo DJ, Ghanaatpishe M, Fathy HK. Genetic optimization and experimental validation of a test cycle that maximizes parameter identifiability for a Li-ion equivalent-circuit battery model. J Energy Storage 2015;4:156–66. http://dx.doi.org/10.1016/j.est.2015.10.004, URL https://www.sciencedirect.com/science/article/pii/S2352152X15300232.

[13] Park S, Kato D, Gima Z, Klein R, Moura S. Optimal experimental design for parameterization of an electrochemical lithium-ion battery model. J Electrochem Soc 2018;165(7):A1309–23. http://dx.doi.org/10.1149/2.0421807jes.

[14] Lai Q, Ahn HJ, Kim G, Joe WT, Lin X. Optimization of current excitation for identification of battery electrochemical parameters based on analytic sensitivity expression. In: 2020 American control conference. 2020, p. 346–51. http://dx.doi.org/10.23919/ACC45564.2020.9147575.

[15] Hu C, Youn BD, Chung J. A multiscale framework with extended Kalman filter for lithium-ion battery SOC and capacity estimation. Appl Energy 2012;92:694–704. http://dx.doi.org/10.1016/j.apenergy.2011.08.002, URL https://www.sciencedirect.com/science/article/pii/S0306261911004971.

[16] Xiong R, Sun F, Chen Z, He H. A data-driven multi-scale extended Kalman filtering based parameter and state estimation approach of lithium-ion polymer battery in electric vehicles. Appl Energy 2014;113:463–76. http://dx.doi.org/10.1016/j.apenergy.2013.07.061, URL https://www.sciencedirect.com/science/article/pii/S0306261913006284.

[17] Chen C, Xiong R, Shen W. A lithium-ion battery-in-the-loop approach to test and validate multiscale dual H infinity filters for state-of-charge and capacity estimation. IEEE Trans Power Electron 2018;33(1):332–42. http://dx.doi.org/10.1109/TPEL.2017.2670081.

[18] Hua Y, Cordoba-Arenas A, Warner N, Rizzoni G. A multi time-scale state-of-charge and state-of-health estimation framework using nonlinear predictive filter for lithium-ion battery pack with passive balance control. J Power Sources 2015;280:293–312. http://dx.doi.org/10.1016/j.jpowsour.2015.01.112, URL https://www.sciencedirect.com/science/article/pii/S0378775315001287.

[19] Veraart J, Sijbers J, Sunaert S, Leemans A, Jeurissen B. Weighted linear least squares estimation of diffusion MRI parameters: Strengths, limitations, and pitfalls. NeuroImage 2013;81:335–46. http://dx.doi.org/10.1016/j.neuroimage.2013.05.028, URL https://www.sciencedirect.com/science/article/pii/S1053811913005223.

[20] Li Y, Wang X. Conditional extended Kalman filter for battery model parameter identification. In: Dynamic systems and control conference. Vol. 2. 2014, p. 5820–6. http://dx.doi.org/10.1115/DSCC2014-5820, arXiv:https://asmedigitalcollection.asme.org/DSCC/proceedings-pdf/DSCC2014/46193/V002T36A001/4445305/v002t36a001-dscc2014-5820.pdf. V002T36A001.

[21] Zhang Y, Liu H, Chen Y, Wang Q. Selection method of measurement data for the parameters estimation of transmission line. In: 2018 2nd IEEE conference on energy internet and energy system integration. 2018, p. 1–5. http://dx.doi.org/10.1109/EI2.2018.8582624.

[22] Li C, Zhang Y, Zhang H, Wu Q, Terzija V. Measurement-based transmission line parameter estimation with adaptive data selection scheme. IEEE Trans Smart Grid 2018;9(6):5764–73. http://dx.doi.org/10.1109/TSG.2017.2696619.

[23] Lin X. A data selection strategy for real-time estimation of battery parameters. In: 2018 American control conference. 2018, p. 2276–81. http://dx.doi.org/10.23919/ACC.2018.8431747.

[24] Gima ZT, Kato D, Klein R, Moura SJ. Analysis of online parameter estimation for electrochemical Li-ion battery models via reduced sensitivity equations. In: 2020 American control conference. 2020, p. 373–8. http://dx.doi.org/10.23919/ACC45564.2020.9147260.

[25] Lin X, Stefanopoulou A, Laskowsky P, Freudenberg J, Li Y, Anderson RD. State of charge estimation error due to parameter mismatch in a generalized explicit lithium ion battery model. In: Dynamic systems and control conference. Vol. 1. 2011, p. 393–400. http://dx.doi.org/10.1115/DSCC2011-6193, arXiv:https://asmedigitalcollection.asme.org/DSCC/proceedings-pdf/DSCC2011/54754/393/2766969/393_1.pdf.

[26] Mishra PP, Garg M, Mendoza S, Liu J, Rahn CD, Fathy HK. How does model reduction affect lithium-ion battery state of charge estimation errors? Theory and experiments. J Electrochem Soc 2016;164(2):A237–51. http://dx.doi.org/10.1149/2.0751702jes.

[27] Li Z, Huang J, Liaw BY, Zhang J. On state-of-charge determination for lithium-ion batteries. J Power Sources 2017;348:281–301. http://dx.doi.org/10.1016/j.jpowsour.2017.03.001, URL https://www.sciencedirect.com/science/article/pii/S0378775317302859.

[28] Mendoza S, Liu J, Mishra P, Fathy H. On the relative contributions of bias and noise to lithium-ion battery state of charge estimation errors. J Energy Storage 2017;11:86–92. http://dx.doi.org/10.1016/j.est.2017.01.006, URL https://www.sciencedirect.com/science/article/pii/S2352152X16301633.

[29] Lin X. Analytic derivation of battery SOC estimation error under sensor noises. IFAC-PapersOnLine 2017;50(1):2175–80. http://dx.doi.org/10.1016/j.ifacol.2017.08.277, URL https://www.sciencedirect.com/science/article/pii/S2405896317305943. 20th IFAC World Congress.

[30] Lin X. Theoretical analysis of battery SOC estimation errors under sensor bias and variance. IEEE Trans Ind Electron 2018;65:7138–48. http://dx.doi.org/10.1109/TIE.2018.2795521.

[31] Fogelquist J, Lai Q, Lin X. On the error of Li-ion battery parameter estimation subject to system uncertainties. J Electrochem Soc 2023;170(3):030510. http://dx.doi.org/10.1149/1945-7111/acbc9c.

[32] Fogelquist J, Lin X. Uncertainty-aware data selection framework for parameter estimation with application to Li-ion battery. In: 2022 American control conference. IEEE; 2022, p. 384–91. http://dx.doi.org/10.23919/ACC53348.2022.9867243.

[33] Lu L, Han X, Li J, Hua J, Ouyang M. A review on the key issues for lithium-ion battery management in electric vehicles. J Power Sources 2013;226:272–88. http://dx.doi.org/10.1016/j.jpowsour.2012.10.060, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775312016163.

[34] Ramadass P, Haran B, White R, Popov BN. Mathematical modeling of the capacity fade of Li-ion cells. J Power Sources 2003;123(2):230–40. http://dx.doi.org/10.1016/S0378-7753(03)00531-7, URL https://www.sciencedirect.com/science/article/pii/S0378775303005317.

[35] Wang AA, O'Kane SEJ, Brosa Planella F, Houx JL, O'Regan K, Zyskin M, et al. Review of parameterisation and a novel database (LiionDB) for continuum Li-ion battery models. Progr Energy 2022;4(3):032004. http://dx.doi.org/10.1088/2516-1083/ac692c, URL https://iopscience.iop.org/article/10.1088/2516-1083/ac692c.

[36] Wojtala ME, Planella FB, Zulke AA, Hoster HE, Howey DA. Investigating changes in transport, kinetics and heat generation over NCA/Gr-SiOx battery lifetime. In: 2022 Vehicle power and propulsion conference. IEEE; 2022, p. 1–6. http://dx.doi.org/10.1109/VPPC55846.2022.10003425.

[37] Capron O, Gopalakrishnan R, Jaguemont J, Van Den Bossche P, Omar N, Van Mierlo J. On the ageing of high energy lithium-ion batteries—Comprehensive electrochemical diffusivity studies of harvested nickel manganese cobalt electrodes. Materials 2018;11(2):176. http://dx.doi.org/10.3390/ma11020176, URL https://www.mdpi.com/1996-1944/11/2/176.

[38] Zhou X, Huang J, Pan Z, Ouyang M. Impedance characterization of lithium-ion batteries aging under high-temperature cycling: Importance of electrolyte-phase diffusion. J Power Sources 2019;426:216–22. http://dx.doi.org/10.1016/j.jpowsour.2019.04.040, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775319304513.

[39] Channagiri SA, Nagpure SC, Babu S, Noble GJ, Hart RT. Porosity and phase fraction evolution with aging in lithium iron phosphate battery cathodes. J Power Sources 2013;243:750–7. http://dx.doi.org/10.1016/j.jpowsour.2013.06.023, URL https://www.sciencedirect.com/science/article/pii/S0378775313010173.

[40] Dong G, Wei J. A physics-based aging model for lithium-ion battery with coupled chemical/mechanical degradation mechanisms. Electrochim Acta 2021;139133. http://dx.doi.org/10.1016/j.electacta.2021.139133, URL https://www.sciencedirect.com/science/article/pii/S0013468621014237.

[41] Birkl CR, Roberts MR, McTurk E, Bruce PG, Howey DA. Degradation diagnostics for lithium ion cells. J Power Sources 2017;341:373–86. http://dx.doi.org/10.1016/j.jpowsour.2016.12.011, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775316316998.

[42] Han X, Lu L, Zheng Y, Feng X, Li Z, Li J, et al. A review on the key issues of the lithium ion battery degradation among the whole life cycle. eTransportation 2019;1:100005. http://dx.doi.org/10.1016/j.etran.2019.100005, URL https://linkinghub.elsevier.com/retrieve/pii/S2590116819300050.

[43] Fang R, Miao C, Nie Y, Wang D, Xiao W, Xu M, et al. Degradation mechanism and performance enhancement strategies of $LiNi_xCo_yAl_{1-x-y}O_2$ ($x \geq 0.8$) cathodes for rechargeable lithium-ion batteries: A review. Ionics 2020;26(7):3199–214. http://dx.doi.org/10.1007/s11581-020-03569-7, URL https://link.springer.com/10.1007/s11581-020-03569-7.

[44] Prasad GK, Rahn CD. Model based identification of aging parameters in lithium ion batteries. J Power Sources 2013;232:79–85. http://dx.doi.org/10.1016/j.jpowsour.2013.01.041, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775313000700.

[45] Schmidt AP, Bitzer M, Imre ÁW, Guzzella L. Model-based distinction and quantification of capacity loss and rate capability fade in Li-ion batteries. J Power Sources 2010;195(22):7634–8. http://dx.doi.org/10.1016/j.jpowsour.2010.06.011, URL https://www.sciencedirect.com/science/article/pii/S0378775310009948.

[46] Ramadesigan V, Chen K, Burns NA, Boovaragavan V, Braatz RD, Subramanian VR. Parameter estimation and capacity fade analysis of lithium-ion batteries using reformulated models. J Electrochem Soc 2011;158(9):A1048. http://dx.doi.org/10.1149/1.3609926, URL https://iopscience.iop.org/article/10.1149/1.3609926.

[47] Lai Q, Joe WT, Kim G, Lin X. Data optimization for parameter estimation under system uncertainties with application to Li-ion battery. In: 2021 American control conference. 2021, p. 4408–13. http://dx.doi.org/10.23919/ACC50511.2021.9483048.

[48] Lai Q, Ahn HJ, Kim Y, Kim YN, Lin X. New data optimization framework for parameter estimation under uncertainties with application to lithium-ion battery. Appl Energy 2021;295:117034. http://dx.doi.org/10.1016/j.

apenergy.2021.117034, URL https://www.sciencedirect.com/science/article/pii/S0306261921004955.

[49] Moura SJ, Argomedo FB, Klein R, Mirtabatabaei A, Krstic M. Battery state estimation for a single particle model with electrolyte dynamics. IEEE Trans Control Syst Technol 2017;25(2):453–68. http://dx.doi.org/10.1109/TCST.2016.2571663, URL http://ieeexplore.ieee.org/document/7489035/.

[50] Lai Q, Jangra S, Ahn HJ, Kim G, Joe WT, Lin X. Analytical derivation and analysis of parameter sensitivity for battery electrochemical dynamics. J Power Sources 2020;472:228–338. http://dx.doi.org/10.1016/j.jpowsour.2020.228338, URL https://www.sciencedirect.com/science/article/pii/S037877532030642X.

[51] Doyle M, Fuller TF, Newman J. Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. J Electrochem Soc 1993;140(6):1526–33. http://dx.doi.org/10.1149/1.2221597, URL https://iopscience.iop.org/article/10.1149/1.2221597.

[52] Forman JC, Bashash S, Stein JL, Fathy HK. Reduction of an electrochemistry-based Li-ion battery model via quasi-linearization and Padé approximation. J Electrochem Soc 2011;158(2):A93. http://dx.doi.org/10.1149/1.3519059, URL https://iopscience.iop.org/article/10.1149/1.3519059.

[53] Rodríguez A, Plett GL, Trimboli MS. Comparing four model-order reduction techniques, applied to lithium-ion battery-cell internal electrochemical transfer functions. eTransportation 2019;1:100009. http://dx.doi.org/10.1016/j.etran.2019.100009, URL https://linkinghub.elsevier.com/retrieve/pii/S2590116819300098.

[54] Wildfeuer L, Lienkamp M. Quantifiability of inherent cell-to-cell variations of commercial lithium-ion batteries. eTransportation 2021;9:100129. http://dx.doi.org/10.1016/j.etran.2021.100129, URL https://linkinghub.elsevier.com/retrieve/pii/S2590116821000278.

[55] Song Z, Yang X-G, Yang N, Delgado FP, Hofmann H, Sun J. A study of cell-to-cell variation of capacity in parallel-connected lithium-ion battery cells. eTransportation 2021;7:100091. http://dx.doi.org/10.1016/j.etran.2020.100091, URL https://linkinghub.elsevier.com/retrieve/pii/S2590116820300497.

[56] Wei Y, Wang S, Han X, Lu L, Li W, Zhang F, et al. Toward more realistic microgrid optimization: Experiment and high-efficient model of Li-ion battery degradation under dynamic conditions. eTransportation 2022;14:100200. http://dx.doi.org/10.1016/j.etran.2022.100200, URL https://linkinghub.elsevier.com/retrieve/pii/S2590116822000455.

[57] Scharf L, McWhorter L. Geometry of the Cramer-Rao bound. In: IEEE sixth SP workshop on statistical signal and array processing. 1992, p. 5–8. http://dx.doi.org/10.1109/SSAP.1992.246835.

[58] Lin X. On the analytic accuracy of battery SOC, capacity and resistance estimation. In: 2016 American control conference. 2016, p. 4006–11. http://dx.doi.org/10.1109/ACC.2016.7525539.

[59] Zhang L, Lyu C, Hinds G, Wang L, Luo W, Zheng J, et al. Parameter sensitivity analysis of cylindrical LiFePO4 battery performance using multi-physics modeling. J Electrochem Soc 2014;161(5):A762–76. http://dx.doi.org/10.1149/2.048405jes.

[60] Chen C-H, Brosa Planella F, O'Regan K, Gastol D, Widanage WD, Kendrick E. Development of experimental techniques for parameterization of multi-scale lithium-ion battery models. J Electrochem Soc 2020;167(8):080534. http://dx.doi.org/10.1149/1945-7111/ab9050, URL https://iopscience.iop.org/article/10.1149/1945-7111/ab9050.

[61] Forman JC, Moura SJ, Stein JL, Fathy HK. Genetic identification and Fisher identifiability analysis of the Doyle–Fuller–Newman model from experimental cycling of a LiFePO4 cell. J Power Sources 2012;210:263–75. http://dx.doi.org/10.1016/j.jpowsour.2012.03.009, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775312006088.

[62] Khalik Z, Donkers M, Sturm J, Bergveld H. Parameter estimation of the Doyle–Fuller–Newman model for lithium-ion batteries by parameter normalization, grouping, and sensitivity analysis. J Power Sources 2021;499:229901. http://dx.doi.org/10.1016/j.jpowsour.2021.229901, URL https://linkinghub.elsevier.com/retrieve/pii/S0378775321004341.

[63] Li W, Demir I, Cao D, Jöst D, Ringbeck F, Junker M, et al. Data-driven systematic parameter identification of an electrochemical model for lithium-ion batteries with artificial intelligence. Energy Storage Mater 2022;44:557–70. http://dx.doi.org/10.1016/j.ensm.2021.10.023, URL https://linkinghub.elsevier.com/retrieve/pii/S2405829721004864.

[64] Lai Q, Fogelquist JB, Lin X. System identification of battery single particle model parameters using new data optimization approach. In: 2022 American control conference. Atlanta, GA, USA: IEEE; 2022, p. 376–83. http://dx.doi.org/10.23919/ACC53348.2022.9867365, URL https://ieeexplore.ieee.org/document/9867365/.

[65] You G-w, Park S, Oh D. Real-time state-of-health estimation for electric vehicle batteries: A data-driven approach. Appl Energy 2016;176:92–103. http://dx.doi.org/10.1016/j.apenergy.2016.05.051, URL https://linkinghub.elsevier.com/retrieve/pii/S0306261916306456.

[66] Berliner MD, Zhao H, Das S, Forsuelo M, Jiang B, Chueh WH, et al. Nonlinear identifiability analysis of the porous electrode theory model of lithium-ion batteries. J Electrochem Soc 2021;168(9):090546. http://dx.doi.org/10.1149/1945-7111/ac26b1, URL https://iopscience.iop.org/article/10.1149/1945-7111/ac26b1.

[67] Park S, Zhang D, Moura S. Hybrid electrochemical modeling with recurrent neural networks for Li-ion batteries. In: 2017 American control conference. 2017, p. 3777–82. http://dx.doi.org/10.23919/ACC.2017.7963533, ISSN: 2378-5861.

[68] Tu H, Moura S, Fang H. Integrating electrochemical modeling with machine learning for lithium-ion batteries. In: 2021 American control conference. IEEE; 2021, p. 4401–7. http://dx.doi.org/10.23919/ACC50511.2021.9482997.

[69] Tu H, Moura S, Wang Y, Fang H. Integrating physics-based modeling with machine learning for lithium-ion batteries. Appl Energy 2023;329:120289. http://dx.doi.org/10.1016/j.apenergy.2022.120289, URL https://www.sciencedirect.com/science/article/pii/S030626192201546X.

[70] Xi Z, Dahmardeh M, Xia B, Fu Y, Mi C. Learning of battery model bias for effective state of charge estimation of lithium-ion batteries. IEEE Trans Veh Technol 2019;68(9):8613–28. http://dx.doi.org/10.1109/TVT.2019.2929197.