Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Calibration of Reported Uncertainty in State-of-Charge Estimation for Lithium Iron Phosphate Cells

Alexandre Barbosa de Lima

ablima@pucsp.br

Pontifícia Universidade Católica de São Paulo

Research Article

Keywords: State of charge, Uncertainty quantification, Calibration, Bayesian filtering, Lithium iron phosphate, Preregistration

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11787350/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Calibration of Reported Uncertainty in State-of-Charge Estimation for Lithium Iron Phosphate Cells

Alexandre Barbosa de Lima[1,2*]

[1*]Faculty of Exact Sciences and Technology, Pontifical Catholic University of São Paulo (PUC-SP), São Paulo, SP, Brazil.
[2]Artificial Intelligence and Quantum Technologies Research Group, Naval War College, Rio de Janeiro, RJ, Brazil.

Corresponding author(s). E-mail(s): ablima@pucsp.br;

Abstract

A battery management system acts on the state of charge (SOC) and on the uncertainty that its estimator reports, but estimators are usually evaluated by point accuracy alone. In lithium iron phosphate (LFP) cells, the flat open-circuit voltage (OCV) and its hysteresis make this uncertainty hard to trust. This paper examines when SOC estimators produce credible intervals with coverage consistent with their nominal levels. A preregistered simulation study with 144 primary and 96 control conditions, and an exact Bayesian reference in the controls, compares a Fisher-information estimator, an iterated extended Kalman filter (IEKF), an unscented Kalman filter (UKF), and a Rao–Blackwellized grid filter on a second-order equivalent-circuit model with hysteresis, whose state the estimators know, measured-current noise, and three voltage sensors. The grid filter was calibrated in aggregate (coverage of 52.4, 80.7 and 94.0% at nominal 50, 80 and 95%), its coverage was within 0.2 percentage points of that of the exact posterior. Intervals reported with 95% probability, which should contain the true SOC 95% of the time, contained it 59.4, 3.1 and 92.5% of the time for the IEKF, the Fisher-information estimator, and the UKF. Exploratory tests implicate fine-scale variations of the interpolated OCV slope used in the IEKF linearization. The grid filter lost tail coverage with a very precise voltage sensor and no process noise in the resistor–capacitor branches. Calibration, and not only point accuracy, should be evaluated.

Keywords: State of charge, Uncertainty quantification, Calibration, Bayesian filtering, Lithium iron phosphate, Preregistration

1 Introduction

The state of charge (SOC) of a lithium-ion cell cannot be measured and must be estimated from current and voltage with a model of the cell (Plett 2016). A battery management system (BMS) uses this estimate to limit power, to schedule balancing, and to report the remaining range. Each of these decisions depends not only on the estimate but also on how wrong it may be: a power limit that is safe for an SOC of 20% may be unsafe if the true SOC can be 12%. Model-based estimators of the Kalman family report this uncertainty as a variance, from which they build intervals around the estimate. Using these intervals for decisions requires calibration, a term used here in its statistical sense and not in the sense of tuning a filter (Gneiting et al. 2007): an interval reported

with 95% probability should contain the true SOC about 95% of the time, and likewise at any other level. The fraction of time in which the true SOC lies inside the reported interval, called coverage, measures it.

Lithium iron phosphate (LFP) cells make this question acute. Over a wide range of SOC, the open-circuit voltage (OCV) changes only slightly with the SOC, and it differs between charge and discharge because of hysteresis (Plett 2016; Zhang et al. 2025). On the plateau, the voltage carries little information about the SOC; near its ends, the OCV curve bends sharply; and when the OCV comes from interpolated tables of measured data, its local slope can also vary at a scale much finer than the uncertainty of the SOC. The first condition makes the true uncertainty large, and the others can make the linearizations used by extended Kalman filters inaccurate. An estimator may therefore be accurate on average and still report variances that are much smaller or much larger than its actual errors. Figure 1 shows these conditions on the OCV curve used in this study.

Comparisons of estimators nevertheless rely mostly on point accuracy, such as the root-mean-square error (RMSE) of the SOC on a few drive cycles (Wang et al. 2021; Silva 2022; Priya et al. 2023; Dinh and Nguyen 2026; Bessa et al. 2026). Observers with guaranteed error bounds give worst-case limits rather than probabilistic statements that can be checked for calibration (Dinh and Nguyen 2026; Bessa et al. 2026), and work on the tuning of the noise covariances (Knox et al. 2024; Previtali and Previdi 2025), including their online adaptation (da Silva et al. 2024), treats calibration as tuning for point accuracy. Consistency tests for filters are well established (Bar-Shalom et al. 2001), and early work on Kalman filtering for BMS already compared the error bounds of extended and sigma-point filters (Plett 2006), but calibration is seldom the primary endpoint of such comparisons. A recent estimator for LFP cells sets the variance of an SOC obtained from a moving-window regression of the OCV from the Cramér–Rao bound given by the Fisher information, and its authors reported it to be accurate and robust to current-sensor bias, voltage quantization, and prolonged operation on the plateau (Shi et al. 2026). A lower bound on the variance need not describe the actual error, and the original study did not examine whether it does.

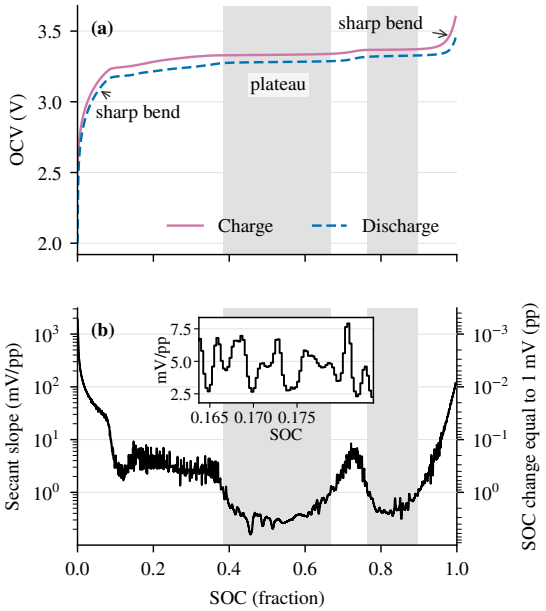

Fig. 1 OCV of the LFP cell used in this study (curve A). (a) Charge (solid purple) and discharge (dashed blue) branches, whose gap is modeled as hysteresis; gray bands mark the plateau (slope below 1 mV per percentage point of SOC) and arrows the sharp bends. (b) Local slope of the mean of the branches (secant over 0.025 percentage points, logarithmic scale): 0.16 mV per percentage point on the plateau, more than 100 at the ends, and between 2.3 and 7.9 within 2 percentage points in the inset (SOC 0.164 to 0.184). The right axis gives the change of SOC that moves the OCV by 1 mV, a typical voltage-sensor noise.

Two obstacles explain why calibration is seldom tested. First, judging a reported variance requires a reference posterior; for real data, the true SOC is itself uncertain, and an independently validated reference posterior is generally unavailable. Second, coverage is a noisy statistic: the error of an SOC estimator persists for long periods, so a single trajectory contributes little more than one independent observation of whether the interval held. A comparison with few trajectories cannot distinguish a calibrated estimator from a miscalibrated one, and a comparison designed after seeing the data can find whatever it looks for.

This paper addresses both obstacles in a matched-model simulation study, with one deliberately mismatched sensor condition, to ask when an SOC estimator knows its own error on an LFP cell with hysteresis. It compares the Fisher-information estimator (Shi et al. 2026), hereafter the Fisher estimator, an iterated extended

Kalman filter (IEKF) (Bell and Cathey 1993; Grewal and Andrews 2015; Särkkä and Svensson 2023), an unscented Kalman filter (UKF) (Julier and Uhlmann 2004), and a Rao–Blackwellized grid filter that represents the SOC on an adaptive grid and, conditionally on the SOC path, marginalizes the resistor–capacitor (RC) states analytically (Bucy and Senne 1971; Schön et al. 2005). A numerically converged evaluation of the exact posterior, available when the current is measured without noise, serves as the reference, and a preregistration fixed the hypotheses, margins, sample sizes, and seeds before any confirmatory outcome was generated.

The paper makes three contributions. First, it provides a calibration benchmark for SOC estimators on LFP cells with hysteresis, measured-current noise, and quantized voltage, with an exact posterior in control conditions, multi-level coverage under prior sampling as in simulation-based calibration (Cook et al. 2006; Talts et al. 2018), a block bootstrap for conditions that share random draws, and a preregistered design whose dated amendments document every later change (Section 5). Second, it shows that estimators that share the same model can differ sharply in calibration: the UKF was close to nominal in aggregate and the IEKF was not, and a Rao–Blackwellized grid filter met every preregistered criterion and was within 0.2 percentage points of the exact posterior in the control conditions, although it was not uniformly better than the UKF and lost coverage in one identified regime. Third, it documents miscalibration in common estimators: the Fisher estimator reports variances orders of magnitude smaller than its errors, the 95% intervals of the IEKF cover less than 90% of the time in 107 of the 144 primary conditions, and exploratory tests implicate the slope used in its linearization rather than the preregistered nonlinearity index.

The remainder of the paper is organized as follows. Section 2 presents the cell and sensor models, and Section 3 describes the estimators and the exact reference. Section 4 defines the calibration metrics and the statistical analysis, and Section 5 describes the preregistered experimental design. Section 6 reports and interprets the results, and Section 7 draws the implications for BMS design, states the limitations, and concludes the paper. The supplementary material (Online Resource 1), whose tables, figures and sections carry the prefix S, gives further details.

2 Cell and Sensor Models

This section defines the model that generates the data. The estimators know its parameters, and all except the Fisher estimator (Section 3) also know its structure, so differences in calibration reflect the estimators and not model error, except for the quantized sensor and a neglected noise correlation (Section 2.3).

Notation used in the paper: scalars are italic (z_k), vectors bold lowercase ($\mathbf{x}_k$), and matrices bold uppercase ($\mathbf{F}$); vectors are column vectors, and the superscript $\top$ denotes transposition. The subscript k denotes discrete time, $\mathrm{diag}(\cdot)$ a diagonal matrix, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, and $\sim$ "distributed as". A hat marks an estimate, a tilde a measured quantity, and primes derivatives with respect to the SOC; $\mathrm{sgn}(\cdot)$ is the sign function and $\mathbf{1}\{\cdot\}$ the indicator function.

2.1 Equivalent-circuit model

A second-order equivalent-circuit model (ECM) represents the cell, with a series resistance R_0 and two RC branches with resistances R_1, R_2 and time constants τ_1, τ_2 (Plett 2016) (Fig. 2). The discrete time k corresponds to the instant $t = k\,\Delta t$, where Δt is the sampling interval. With the state $\mathbf{x}_k = (z_k, v_{1,k}, v_{2,k})^\top$, where z_k is the SOC and $v_{j,k}$ is the voltage across branch j, $j = 1, 2$, and with the current i_k, positive in discharge, held constant between $t = k\,\Delta t$ and $t = (k+1)\,\Delta t$ (zero-order hold), the exact discretization of the RC dynamics gives

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{F}\,\mathbf{x}_k + \mathbf{g}\,i_k + \mathbf{w}_k, \\ \mathbf{F} &= \mathrm{diag}(1, a_1, a_2), \quad a_j = e^{-\Delta t/\tau_j}, \end{aligned} \tag{1}$$

with

$$\mathbf{g} = \left(-\tfrac{\Delta t}{3600\,Q_\mathrm{n}},\; R_1(1-a_1),\; R_2(1-a_2) \right)^\top, \tag{2}$$

where a_j is the fraction of the voltage across branch j that remains after one sampling interval with zero current, $R_j(1-a_j)$ is the corresponding input gain, Q_n is the capacity in ampere-hours,

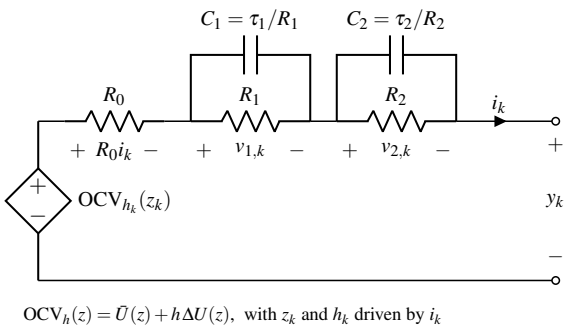

Fig. 2 Second-order equivalent-circuit model with hysteresis. The OCV source $\mathrm{OCV}_{h_k}(z_k)$ depends on the SOC z_k and on the hysteresis state h_k, both driven by the current i_k (positive in discharge); the terminal voltage y_k is the OCV minus the drop $R_0 i_k$ and the voltages $v_{1,k}$ and $v_{2,k}$ across the RC branches, with $C_j = \tau_j/R_j$.

the coulombic efficiency is taken as one, and $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(0, q_{\mathrm{RC}}, q_{\mathrm{RC}}))$ is an optional process noise on the RC branches. The terminal voltage is

$$\begin{aligned} y_k &= \mathrm{OCV}_{h_k}(z_k) - v_{1,k} - v_{2,k} - R_0\, i_k + e_k, \\ \mathrm{OCV}_h(z) &= \bar{U}(z) + h\,\Delta U(z), \end{aligned} \tag{3}$$

where $\bar{U}$ is the mean of the charge and discharge OCV branches, ΔU is half the difference between the charge and the discharge branches, $h_k \in [-1, 1]$ is the hysteresis state (Plett 2004), with $h_k = 1$ on the charge branch and $h_k = -1$ on the discharge branch, and $e_k \sim \mathcal{N}(0, \sigma_v^2)$ is the voltage-sensor noise. The hysteresis state evolves as $h_{k+1} = \lambda_k h_k - (1 - \lambda_k)\,\mathrm{sgn}(i_k)$, with $\lambda_k = e^{-\gamma |i_k| \Delta t/(3600\, Q_{\mathrm{n}})}$ (Plett 2004), driven by the true current and initialized at the value that the source cycle reaches at the start of each segment. The rate $\gamma = 10$ is nominal and was not identified for this cell; with it, a change of 10 percentage points of SOC in one direction moves h_k about 63% of the way to the corresponding branch. All estimators receive the true h_k and use it only to evaluate the OCV map, not to infer the current from its evolution, so the study isolates the uncertainty of the SOC and of the RC voltages from that of the hysteresis state. Table 1 lists the parameter values.

2.2 OCV curves

The exact posterior is calibrated whatever the OCV curve, so a difference in coverage between two representations of the OCV of the same cell can only come from the approximations made by the estimators. The study therefore uses two curves, both built from the charge and discharge OCV tables of an A123 LFP cell at 25 °C from the CALCE data (CALCE Battery Research Group 2026). Curve A interpolates smoothed versions of these tables with a shape-preserving piecewise-cubic interpolant; its slope ranges from about 0.16 mV per percentage point of SOC on the plateau to more than 100 mV per percentage point at the ends of the domain and still varies at the scale of a fraction of a percentage point (Fig. 1b). Curve B, a stress variant for estimators that rely on the local slope, is built from the unsmoothed tables: a linear term with a slope of 0.05 mV per percentage point is subtracted, the remainder is replaced by its isotonic regression, that is, its least-squares fit under the constraint of being nondecreasing (Barlow et al. 1972; Guntuboyina and Sen 2018), and the linear term is added back, which keeps the curve strictly increasing. The result alternates between flat segments with the floor slope and short steep steps, as in an unsmoothed OCV table: between SOC 0.2 and 0.8, 38% of the points of a grid with a spacing of 0.05 percentage points lie on the floor, with 185 alternations (Fig. S1). Both curves coincide at the ends of their common domain [0, 0.996751].

Table 1 Model parameters (25 °C).

Symbol	Meaning	Value
Q_{n}	capacity	1.0636 Ah
R_0	series resistance	12 mΩ
R_1, τ_1	first RC branch	6 mΩ, 18 s
R_2, τ_2	second RC branch	4 mΩ, 115 s
Δt	sampling interval	1 s
γ	hysteresis rate	10 (nominal)
q_{RC}	RC process noise	0 or 10^{-8} V^2 per step
σ_I	current-sensor noise	0 (control) or 0.01 A

2.3 Current and voltage sensors

A BMS observes the cell through a current sensor, whose reading drives the prediction step of every estimator, and a voltage sensor, whose reading is the measurement. In the control conditions, the estimators receive the true current; otherwise, they receive the measured current $\tilde{\imath}_k = i_k + n_k$, with $n_k \sim \mathcal{N}(0, \sigma_I^2)$ and $\sigma_I = 0.01$ A, white and independent of all other noises, while the true current i_k drives the cell. Replacing i_k by $\tilde{\imath}_k$ in (1)

gives

$$\mathbf{x}_{k+1} = \mathbf{F}\,\mathbf{x}_k + \mathbf{g}\,\tilde{\imath}_k - \mathbf{g}\,n_k + \mathbf{w}_k, \qquad (4)$$

so the current noise enters all three states through the same scalar n_k, and the process covariance seen by the estimators is

$$\mathbf{Q}_x = \sigma_I^2\,\mathbf{g}\mathbf{g}^\top + \mathrm{diag}(0, q_{\mathrm{RC}}, q_{\mathrm{RC}}), \qquad (5)$$

which couples the SOC negatively with both RC voltages and the RC voltages positively with each other. Relative to the model used by the estimators, whose ohmic term contains $\tilde{\imath}_k$, the measurement y_k contains the additional term $+R_0 n_k$. Because n_k is independent of the state predicted from the measurements up to $k-1$, the effective variance of the measurement noise is $\sigma_v^2 + R_0^2\sigma_I^2$, which the estimators use. The same n_k, however, also drives the transition from k to $k+1$, so the measurement noise at k and the process noise of the next transition have the cross-covariance $-R_0\sigma_I^2\,\mathbf{g}$, which none of the estimators includes; the model they filter with is therefore not exact. Because $R_0\sigma_I = 0.12$ mV, this neglected correlation matters most with the less noisy voltage sensor.

For the voltage sensor, the study considers three conditions: S1, Gaussian noise with $\sigma_v = 0.1$ mV; S2, Gaussian noise with $\sigma_v = 1$ mV; and S3, the S2 signal quantized by a 10-bit converter over 5 V with rounding to the nearest step, $\delta = 5/1024$ V. In S3, the estimators add the variance $\delta^2/12$ of a uniform quantization error (Widrow and Kollár 2008); with $\sigma_v \approx 0.2\,\delta$, however, the quantization error is neither uniform nor independent of the signal, and it is correlated in time whenever the voltage stays within one step, so S3 is mismatched by construction.

2.4 Initial state

The generator draws the initial SOC from a Gaussian prior with mean μ_0, the nominal initial SOC of the segment (Section 5), and standard deviation $s_0 \in \{1, 10\}$ percentage points, truncated so that the entire true trajectory stays inside the OCV domain (Table S1). The grid filter and the exact reference use this truncated prior; the other estimators use an untruncated Gaussian with the same mean and variance. The RC voltages start at zero in every segment, and the IEKF, the UKF, the grid filter and the exact reference know this value exactly. The hysteresis state, in contrast, starts at the value reached in the source cycle, so the generator omits the RC transient that the source cycle would leave at the start of a segment; because the generator and the estimators share this initial condition, it introduces no model error.

Table 2 Estimators compared.

Name in the text	Estimator	Reported uncertainty
Fisher estimator	Fisher-information OCV fusion (Shi et al. 2026)	Scalar Gaussian variance
IEKF	Iterated extended Kalman filter	Gaussian covariance
UKF	Unscented Kalman filter	Gaussian covariance
Grid filter	Rao–Blackwellized grid filter	Full posterior of the SOC
Exact reference	Exact posterior (controls only)	Full posterior of the SOC

3 Estimators and Exact Reference

The IEKF, the UKF and the grid filter share the model of Section 2, the measured current, the measurement variance $\sigma_v^2 + R_0^2\sigma_I^2$ (plus $\delta^2/12$ in S3), and the process covariance (5), and neglect the noise correlation of Section 2.3. Within the scalar formulation of its original study, the Fisher estimator uses the same OCV map, capacity, measured current, measurement variance, and true hysteresis state, with the variance $(\sigma_I \Delta t/3600\,Q_{\mathrm{n}})^2$ for coulomb counting. No estimator receives the true SOC, RC voltages, or current (Table 2).

3.1 Fisher-information OCV fusion

The estimator of Shi et al. (2026) fits, over a moving window of 100 s, a regression that is linear in the OCV and in the coefficients of the RC dynamics, with regressors built by zero-order-hold filters of the current. The OCV estimate is inverted with the hysteresis-weighted OCV map on the SOC interval $[0.05, 0.95]$ into an SOC measurement whose variance follows the Fisher-information bound of the original paper (its Eqs. 16–21), and a scalar Kalman filter fuses this

measurement with coulomb counting (its Eqs. 22–26). The update is skipped when the inversion fails or the predicted SOC lies outside that interval, and before the first window is filled, at about 105 s, the estimator relies on coulomb counting alone. Its intervals are central Gaussian intervals built from the fused variance.

3.2 Iterated extended Kalman filter

The IEKF performs the measurement update as Gauss–Newton iterations on the posterior mode (Bell and Cathey 1993). At each iteration, the measurement Jacobian is $(U_h'(z), -1, -1)$, where U_h' is the secant slope of the interpolated OCV map over the grid interval of 0.025 percentage points that contains the current iterate, consistent with the interpolation used to predict the voltage. Iterations stop when the change in SOC is below 10^{-10} or after 10 iterations, and the covariance is updated in Joseph form (Grewal and Andrews 2015). The OCV table is held constant outside the domain of the curve, and the iterates are clipped to $[0, 1]$. A preregistered sensitivity analysis repeats the IEKF with a limit of 50 iterations.

3.3 Unscented Kalman filter

The UKF uses the scaled unscented transform (Julier and Uhlmann 2004) with $\alpha = 1$, $\beta = 2$, and $\kappa = 0$ for the three-dimensional state. Sigma points outside the domain of the OCV curve are evaluated by extrapolation of the interpolant. A sensitivity analysis repeats the UKF with $\alpha = 10^{-3}$, which places the sigma points much closer to the mean. The Gaussian intervals of the Fisher estimator, the IEKF and the UKF are not truncated to the SOC domain.

3.4 Rao–Blackwellized grid filter

Given a path of the SOC, the model (1)–(3) is linear and Gaussian in the RC voltages. The grid filter exploits this structure (Schön et al. 2005): it represents the marginal posterior of the SOC by weights on an adaptive grid of nodes (Bucy and Senne 1971) and attaches to each node the conditional mean and covariance of the RC voltages, which a Kalman filter propagates and updates. The prediction convolves the SOC weights with the Gaussian SOC increment implied by (5), conditions the RC moments on that increment so that the cross-covariance of (5) is preserved, and combines at each node the Gaussians arriving from all source nodes by moment matching; the update multiplies each weight by the marginal likelihood of the measurement under the conditional Kalman filter of its node. The grid spacing follows the predicted standard deviation of the SOC and is refined when a narrow likelihood falls between nodes, so that at the end of every step the resolution ratio R, the posterior standard deviation over the local spacing, is at least 10, with at least 20 nodes holding 99% of the mass, as registered; Section S1 gives the numerical rules. Credible intervals are read from the discrete cumulative distribution, without a Gaussian assumption. With noiseless current, the grid filter targets the exact posterior up to discretization errors; with current noise, carrying a single Gaussian for the RC voltages at each node is an approximation, because paths that reach the same node have different RC histories.

3.5 Exact reference

When the current is measured without noise, $z_k = z_0 - \sum_{m<k} \Delta t\, i_m/(3600\, Q_\mathrm{n})$ is a deterministic shift of the initial SOC, so the posterior of z_k is the posterior of z_0 mapped by that shift. For each value of z_0 on a grid, a Kalman filter for the RC voltages yields the exact marginal likelihood of all measurements, also when $q_\mathrm{RC} > 0$, and normalizing its product with the truncated prior gives the exact posterior up to quadrature error. The grid is refined until the posterior quantiles converge, and an independent evaluation on a much finer uniform grid confirmed the result (Section S1). This exact reference is available only in control conditions with Gaussian voltage noise and is not an online estimator.

4 Evaluation of Calibration

This section defines the coverage of the reported intervals (Section 4.1), explains why it must be aggregated over many trajectories (Section 4.2) and how its confidence intervals are obtained (Section 4.3), and defines the nonlinearity index (Section 4.4) and the secondary metrics (Section 4.5).

4.1 Coverage

For an estimator that reports a central credible interval $I_k(\gamma)$ with nominal level γ, that is, the interval between the quantiles $(1-\gamma)/2$ and $(1+\gamma)/2$ of the posterior distribution of z_k reported by the estimator, to which the estimator assigns probability γ (Gelman et al. 2013), the time coverage of realization r is

$$c_r(\gamma) = \frac{1}{K}\sum_{k=1}^{K} \mathbf{1}\{z_k \in I_k(\gamma)\}, \tag{6}$$

where K is the number of samples in the segment. The coverage of a condition is the unweighted mean of $c_r(\gamma)$ over its n realizations, and the aggregate coverage is the unweighted mean over conditions. The study evaluates coverage at $\gamma \in \{0.50, 0.80, 0.95\}$, because an estimator can match the 95% level with intervals of the wrong shape.

If the true SOC is drawn from the prior that the estimator uses and the estimator computes the exact posterior, the expected coverage at every instant equals γ, whatever the profile or curve (Cook et al. 2006; Talts et al. 2018; Gneiting et al. 2007). The generator therefore samples the initial SOC from the prior in every primary condition, as in simulation-based calibration, so that calibration can be tested condition by condition. Calibration in this sense is averaged over the prior and does not guarantee nominal coverage for a particular initial error.

4.2 Why many trajectories are needed

Errors in the SOC persist over the whole segment, so $c_r(\gamma)$ tends to be close to 0 or to 1. In pilot runs, the standard deviation of $c_r(0.5)$ between realizations was 0.17 to 0.32, so that 20 realizations give a standard error of 4 to 7 percentage points (pp) for a single condition, and a single condition would need about 45 to 160 realizations for the 95% interval of an exact estimator to fall within ±5 pp of nominal. The confirmatory criteria are therefore defined on the aggregate coverage (Section 5.5).

4.3 Interval estimation

Conditions that share a profile, a segment, and an OCV curve share their random draws (Section 5.3), so their outcomes are dependent. A block bootstrap (Davison and Hinkley 1997) resamples realization indices within each of these 12 groups, carrying all conditions of a group together, with 10 000 resamples, and reports percentile intervals. Equivalence to the nominal level is declared when the entire 95% interval lies within the margin, which corresponds to two one-sided tests at the 2.5% level each (Schuirmann 1987) and is more conservative than the usual procedure with a 90% interval. Each hypothesis is decided separately, without adjustment for multiple comparisons.

4.4 Nonlinearity index

The preregistered explanation of the behavior of the IEKF uses the index

$$\kappa_k = \sqrt{P_{z,k}}\,\frac{|U_h''(\hat{z}_k)|}{|U_h'(\hat{z}_k)|}, \tag{7}$$

evaluated at the IEKF estimate $\hat{z}_k$, where $P_{z,k}$ is its reported SOC variance and U_h', U_h'' are the derivatives of the hysteresis-weighted piecewise-cubic interpolants of the two branches (on curve B, of the interpolated staircase plus the floor slope). The index compares the change of the slope over one standard deviation with the slope itself, and its median over samples and realizations characterizes each condition. Because $P_{z,k}$ is reported by the IEKF itself, the index is not independent of the calibration that it is meant to explain (Section 6.3).

4.5 Other metrics

The study also reports the SOC RMSE, the normalized error $(\hat{z}_k - z_k)/\sqrt{P_{z,k}}$ (descriptively, because the posterior of the grid filter need not be Gaussian; for the grid filter and the reference, $\hat{z}_k$ and $P_{z,k}$ are the posterior mean and variance computed on the grid), the number of IEKF updates that reach the iteration limit, and the wall-clock cost per update.

5 Preregistered Experimental Design

This section describes the design of the confirmatory study, which was fixed before any

of its outcomes was generated: the preregistration (Section 5.1), the factors (Section 5.2), the random draws (Section 5.3), the hypotheses (Section 5.4), the reference check and the power analysis (Section 5.5), and the cost benchmark (Section 5.6).

5.1 Preregistration

The protocol was committed to a private version-controlled repository before the software for the confirmatory run was written. Seven amendments, all made before the confirmatory run started and before any of its outcomes was examined, corrected an input table, added the sensitivity analysis of the IEKF, introduced the numerical safeguards and the likelihood-driven refinement of the grid filter after a first execution stopped on a numerical overflow, replaced an inaccurate implementation of the exact reference, and kept the grid filter unchanged with documented local deviations from the exact posterior (Table S3). The interrupted execution was archived without examination of its outcomes, the confirmatory run restarted from zero with the registered seeds, and software tests used a seed reserved for that purpose. A public package (Lima 2026) contains the preregistration, the amendments and the code, with the SHA-256 hash of each file, the commit of the private history in which it was recorded, and its agreement with the hashes recorded during the run; the full history is available from the author on request.

5.2 Factors

Segments of three driving profiles of the CALCE A123 data (CALCE Battery Research Group 2026), DST, US06 and FUDS, drive the model (Fig. S2). Each profile contributes a crossing segment, from SOC 0.88 down to 0.20, which covers the plateau and its lower bend, and a plateau segment, from 0.70 to 0.30; the upper end is the nominal initial SOC μ_0. Each source interval, of 2928 to 5284 s, is rescaled in time to 3600 s, and its current is multiplied by the same factor, from 0.81 to 1.47, to preserve the charge. The primary design crosses segment (crossing, plateau), OCV curve (A, B), profile (DST, US06, FUDS), voltage sensor (S1, S2, S3), prior standard deviation of the initial SOC (1 and 10 percentage points), and RC process noise $q_{\mathrm{RC}} \in \{0, 10^{-8}\}$ V^2 per step, matched between generator and estimators, with measured-current noise $\sigma_I = 0.01$ A and no current-sensor bias in every primary condition, for a total of 144 primary conditions with $n = 50$ realizations each. Ninety-six control conditions repeat the S1 and S2 conditions with noiseless current, so that the exact reference applies. A robustness set of 216 conditions with fixed initial errors of 0 and ± 10 percentage points and $n = 20$ realizations is reported separately and enters no hypothesis.

5.3 Random draws

Each realization uses four independent streams (measured-current noise, standardized voltage noise, RC process noise, and the uniform variate that is mapped to the initial SOC), derived from a registered master seed and from the codes of profile, segment, curve, and realization. No stream depends on the voltage-sensor condition, the prior, or q_{RC}. The same standardized voltage noise is therefore scaled for the Gaussian sensors S1 and S2 and quantized for S3 (Section 2.3), and all estimators see identical data within a realization. This pairing makes comparisons between estimators and between conditions sharper, and it is the reason for the block bootstrap of Section 4.3.

5.4 Hypotheses

Table 3 lists the confirmatory and secondary hypotheses with their decisions. The calibration hypotheses for the grid filter use only the 96 primary conditions with Gaussian voltage noise (S1 and S2), a restriction fixed on theoretical grounds before any S3 outcome, because S3 is mismatched by construction; its coverage in S3 is reported descriptively. Each hypothesis is classified as supported, not supported, or inconclusive.

5.5 Reference check and power

Before the analysis of any primary outcome, the aggregate coverage of the exact reference over the 96 control conditions had to lie within nominal ± 5 pp at all three levels; control conditions whose 95% interval lies entirely outside this band are flagged without changing any denominator. A power analysis on pilot trajectories recentered to nominal coverage (Section S2) set $n = 50$: for a simulated estimator whose coverage was

at or slightly below nominal (49.98, 79.59 and 94.25%), and with perfectly correlated outcomes within each group of conditions that share random draws, the probability of supporting H1 was at least 0.977, and that of passing the reference check, 0.978.

5.6 Cost benchmark

The cost of the grid filter was measured on an Apple M3 Pro laptop on a fixed subset of 12 trajectories, with one warm-up run and three timed repetitions in isolated processes, excluding data loading, simulation, and input and output. All estimators are implemented in Python 3.11 with NumPy and SciPy.

6 Results

This section reports the outcomes of the 456 conditions of the design, all of which were completed and analyzed with the analysis code written before the run, without changes; Table 3 summarizes the decision on each hypothesis. Exploratory analyses, run after the confirmatory study with new random seeds and without changing any confirmatory output, help to interpret three of the findings and are identified as such.

6.1 Validity of the reference and of the grid filter

Over the 96 control conditions, the exact reference covered the true SOC 49.8, 80.0 and 94.9% of the time at the nominal levels of 50, 80 and 95% (Table 3), so the reference check passed and the primary analysis proceeded. Fig. S4 compares the grid filter with the reference condition by condition. Even the exact reference does not reach the nominal level in every condition: with 50 realizations, the coverage of a single condition varies by several percentage points around the nominal level (Section 4.2), from about 88 to 99% at 95%. Because both methods process the same data, the grid filter follows this variation, and almost all points lie on the identity line: it reproduces the coverage of the exact reference condition by condition, and not only in aggregate. Over all control conditions, its coverage exceeded that of the reference by +0.13, +0.11 and +0.11 percentage points at the three levels, within the margin of ±1 percentage point of HA2-ref.

Five of the 4800 reference trajectories, all on curve B, plateau segment, S1, prior of 10 percentage points and $q_{\mathrm{RC}} = 10^{-8}$ V^2, did not meet the quantile-convergence criterion after eight halvings of the grid. As registered, they were counted with zero coverage, a penalty for numerically inconclusive trajectories rather than an observation that valid intervals missed the true SOC, and they produce the points above the identity line in Fig. S4. Without them, in an exploratory sensitivity analysis, the difference between the grid filter and the reference was +0.09 [0.04, 0.15], +0.05 [0.01, 0.10] and +0.03 [0.01, 0.05] percentage points, so most of the registered difference at 95% came from these trajectories.

Agreement in coverage does not imply that the two posteriors agree at every instant: in exploratory comparisons, the grid filter deviated locally from the exact posterior by up to 4.1 posterior standard deviations (Section S1). No step of the grid filter or of the reference violated the registered resolution criterion (minimum resolution ratios of 17.3 and 24.1), and the grid filter never failed numerically.

6.2 Calibration of the grid filter (H1)

Over the 144 primary conditions, the 95% coverage was 3.1% for the Fisher estimator, 59.4% for the IEKF, 92.5% for the UKF and 94.0% for the grid filter, with a mean trajectory RMSE of 2.61, 1.56, 0.130 and 0.130 percentage points (Table S6). The grid filter covered the true SOC at 52.4, 80.7 and 94.0% in the 96 primary conditions with Gaussian voltage noise (S1 and S2), within the preregistered margin of 5 percentage points at every level (H1). In the control conditions, where its target is the exact posterior, its coverage differed from that of the reference by less than 0.2 percentage points at every level (HA2-ref). In the 48 conditions with the quantized sensor S3, reported descriptively, its coverage was 49.6, 78.7 and 94.0%. Figure 3 summarizes the calibration of all estimators by OCV curve: a point below the grey line means that the intervals of the estimator contain the true SOC less often than they claim. The grid filter and the UKF lie close to the line on both curves; the IEKF lies below it

Table 3 Preregistered hypotheses and decisions. Coverage in percent at 50, 80 and 95%; differences in percentage points (pp); intervals are 95% block-bootstrap intervals. The labels of the preregistration are given in parentheses. Primary: the 144 primary conditions; controls: the 96 control conditions.

Hypothesis	Statement	Result	Decision
Reference check	Coverage of the exact reference within nominal ±5 pp at all levels (controls)	49.8 [48.6, 51.1], 80.0 [78.9, 81.0], 94.9 [94.3, 95.5]; no flagged control	passed
H1 (HA2′)	Aggregate coverage of the grid filter within nominal ±5 pp at all levels (96 primary, S1 and S2)	52.4 [51.5, 53.4], 80.7 [79.9, 81.5], 94.0 [93.5, 94.5]	supported
H2 (HB′)	IEKF 95% coverage below 80% in at least 75% of the upper-tercile conditions of κ (primary)	29 of 48 (60%)	not supported
H3 (HC)	RMSE of the grid filter not larger than that of the IEKF in at least 116 conditions (primary)	121 of 144	supported
H4 (HD)	95th percentile of the cost of the grid filter per update below 2 ms (cost benchmark)	mean 0.57 ms, 95th percentile 1.07 ms	supported
H5 (HE)	95% coverage of the Fisher estimator below 50% in at least 116 conditions (primary)	144 of 144	supported
HA2-ref	Grid filter minus exact-reference coverage within ±1 pp (controls)	+0.13 [0.06, 0.19], +0.11 [0.04, 0.18], +0.11 [0.04, 0.20] pp	supported
HA1	Lower bound of the aggregate 95% coverage of the grid filter at least 90% (primary)	93.5%	supported
HB-cal	Aggregate equivalence of the IEKF to nominal, by curve (primary)	at 95%: 77.7 [76.4, 79.0] on curve A, 41.2 [39.3, 43.2] on curve B	not supported
HC-sup	RMSE advantage of the grid filter in the upper tercile of κ (primary)	mean log RMSE ratio, grid filter/IEKF: −0.98 [−1.02, −0.64]	supported

on curve A and far below it on curve B, where its 95% intervals contained the true SOC only 41.2% of the time; and the intervals of the Fisher estimator almost never contained it. In the control trajectory of Fig. S3, chosen by a rule fixed before plotting, the 95% interval of the exact reference narrows from several percentage points to a few hundredths; the intervals of the grid filter and of the UKF follow it, that of the IEKF is narrower, and after 100 s that of the Fisher estimator is narrower than 0.01 percentage points while its error stays between about 0.1 and 1 percentage point.

The aggregate hides a regime in which the grid filter is not calibrated. A permutation test of heterogeneity over the 144 primary conditions, registered as descriptive, rejected equal coverage at every level ($p < 10^{-4}$); Fig. 4 shows the 95% coverage of every primary condition, and Table S2 lists the coverage of the grid filter by stratum. With the most precise voltage sensor (S1) and no RC process noise, the 95% coverage of the grid filter fell to 86.3% on the crossing segments, while its 50% coverage rose to 55.6% (60.0% on the plateau segments): the 50% intervals are too wide and the 95% intervals too narrow, which indicates errors more concentrated near zero and with heavier tails than the reported posterior. Every other combination of sensor and RC noise stayed within about 2.5 percentage points of nominal at 95%.

With current noise, the grid filter carries a single Gaussian for the RC voltages at each node, an approximation expected to matter most when a precise voltage sensor constrains the SOC tightly and no RC process noise absorbs the mismatch between paths. In an exploratory diagnosis of this regime (80 trajectories), a particle filter with 20 000 particles covered the true SOC at 46.6, 76.4 and 93.2%, and the grid filter at 52.5, 75.7 and 85.4%, so the loss of coverage comes from the grid filter and not from the problem itself. Doubling the grid resolution did not help (84.3% at 95%), while a minimal RC process noise (10^{-10} V^2) in the filter alone raised its 95% coverage to 97.2%, which is consistent with a degenerate conditional covariance of the RC voltages. The neglected noise correlation, largest in this regime, is a less likely cause, because the UKF, which neglects it too, over-covered (98.1% at 95%). The single Gaussian per node is therefore the most likely cause, but the diagnosis does not establish it.

6.3 Calibration of the IEKF (H2)

The IEKF was miscalibrated in most conditions: its 95% coverage was below 90% in 107 and below 80% in 91 of the 144 primary conditions, and its 95% intervals contained the true SOC 77.7% of the time on curve A and 41.2% on curve B. H2 predicted that the failures would concentrate where the nonlinearity index κ is large, and it was not supported: the 95% coverage was below 80% in 29 of the 48 conditions of the upper tercile of κ, short

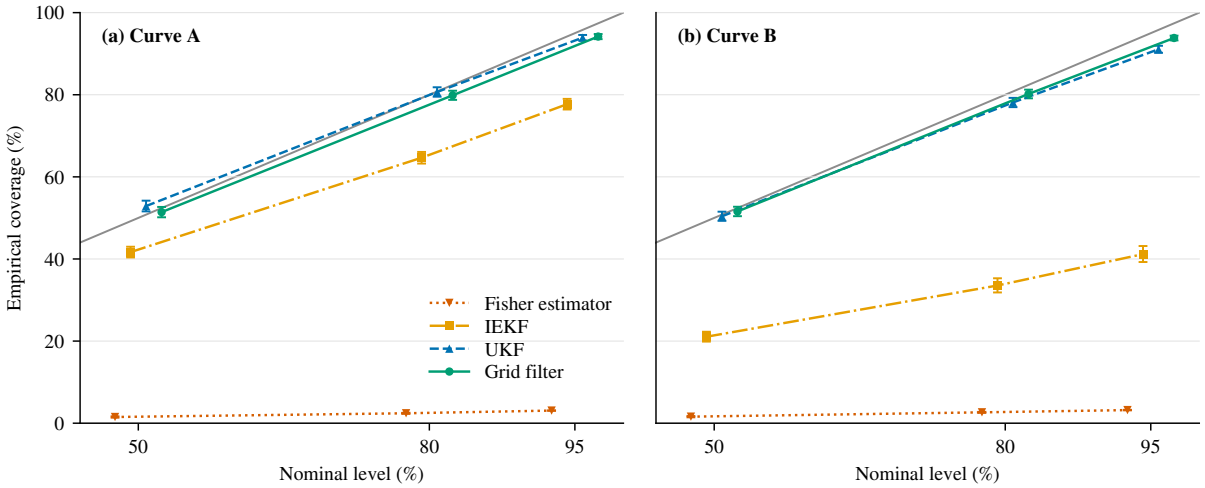

Fig. 3 Aggregate coverage against nominal level over the 72 primary conditions of (a) curve A and (b) curve B. The grey line marks perfect calibration; points below it indicate intervals that are too narrow, that is, an estimator that is more confident than its errors justify. Bars are 95% block-bootstrap intervals; the curve-specific intervals of the Fisher estimator, the UKF and the grid filter were computed after the confirmatory analysis with the registered procedure.

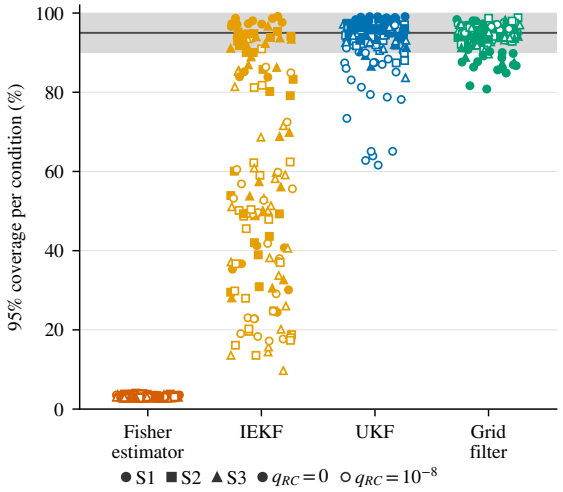

Fig. 4 95% coverage of each of the 144 primary conditions, by estimator. Marker shape indicates the voltage sensor; filled and open markers indicate $q_{\mathrm{RC}} = 0$ and 10^{-8} V^2. The shaded band marks 90 to 100%.

of the required 36. The failure was not concentrated in that tercile. The 95% coverage was 56.3, 58.6 and 63.4% in the lower, middle and upper terciles, and the lower tercile had the largest RMSE. With a prior of 10 percentage points, the criterion held in 20 of 24 upper-tercile conditions, and with a prior of 1 percentage point in 9 of 24, which shows that the tercile of κ follows the variance reported by the IEKF.

The iteration limit does not explain the failure: only 0.8% of the IEKF updates reached the limit of 10 iterations, and a limit of 50 changed its coverage by at most 1.1 percentage points on either curve. The UKF, with the same model, noise covariances, and Gaussian prior as the IEKF, covered the true SOC 92.5% of the time at 95%, but with its sigma points close to the mean ($\alpha = 10^{-3}$, a preregistered sensitivity analysis) its 95% coverage fell to 53.9% (75.1% on curve A and 32.7% on curve B), close to that of the IEKF; 101 of its 7200 trajectories failed numerically and were counted with zero coverage.

The failure of H2 says more about the index than about the IEKF: κ is built from the variance reported by the IEKF, which shrinks when the filter becomes overconfident (Jazwinski 1970), so the trajectories in which it fails most can be classified as nearly linear, as the largest RMSE and lowest coverage of the lower tercile suggest. An index based on the exact posterior or on the true SOC, and on the secant slope that the IEKF actually uses, would avoid this problem.

Exploratory tests point to a different mechanism (Fig. 5). Because the IEKF linearizes the OCV map through its secant over an interval of 0.025 percentage points, its slope, and with it the gain and the variance update, inherits the fine-scale variation of the slope of the interpolated map (Fig. 5a). In six conditions on curve A, a smoothed

curve in both the generator and the filter, which changed the OCV by at most 1.6 mV over the visited range, raised the 95% coverage of the IEKF by 32.9 percentage points [26.2, 39.7], to 92.9%; the smoothed slope used only in its Jacobian raised it by 31.0 percentage points [24.1, 38.1], about 94% of the full effect, in all six conditions. The UKF and the grid filter changed by less than 1 percentage point. The UKF, whose sigma points lie about $\sqrt{3}$ standard deviations from the mean, effectively averages the slope over many grid intervals. The lower coverage of the IEKF on curve B, whose plateau is a staircase, is consistent with this mechanism. Because these tests are exploratory, they identify the slope used in the linearization as the main factor but do not exclude others.

6.4 Accuracy and cost (H3 and H4)

The RMSE of the grid filter was not larger than that of the IEKF in 121 of the 144 conditions, and its mean RMSE was 0.13 percentage points against 1.56 for the IEKF. The preregistered aggregate statistic, the mean over conditions of the natural logarithm of the ratio of the RMSE of the two filters, was −1.523, so that, at the level of conditions, the RMSE of the grid filter was about one fifth of that of the IEKF; for a single trajectory, the typical ratio was about 0.75, because a few IEKF trajectories with very large errors dominate the ratio between conditions (Section S5). In the cost benchmark, the grid filter took 0.57 ms per update on average and 1.07 ms at the 95th percentile, below the limit of 2 ms.

6.5 Calibration of the Fisher estimator (H5)

The 95% intervals of the Fisher estimator contained the true SOC 3.1% of the time over all conditions, and its coverage was below 50% in all 144 conditions. After the first 100 s, when the regression window is first filled and the estimator stops relying on the prior, the coverage fell to 0.17, 0.30 and 0.47% at the three levels. Its normalized errors were of the order of hundreds of reported standard deviations, while its mean RMSE (2.61 pp) was within a factor of two of that of the IEKF. A likely reason is that the Fisher information of the windowed regression measures how much the voltage noise moves the estimated

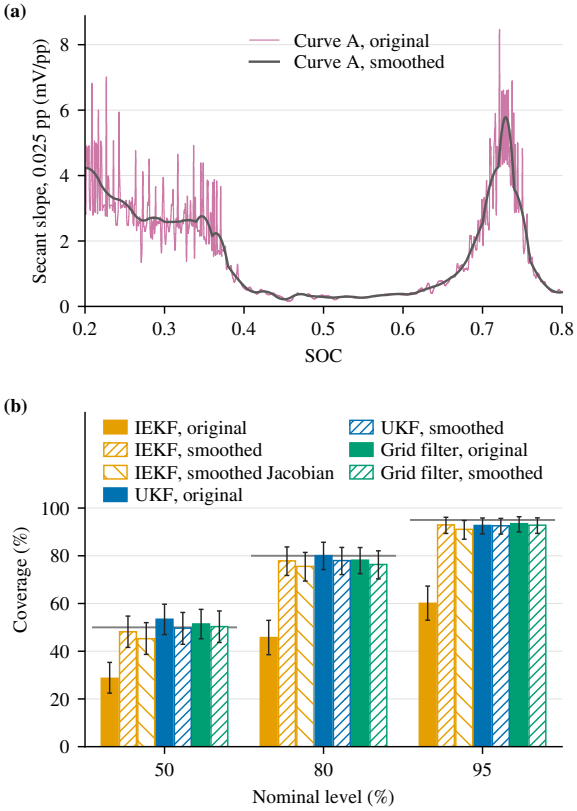

Fig. 5 Exploratory test of the IEKF failure. (a) Secant slope of curve A over intervals of 0.025 percentage points on the plateau, original and smoothed. (b) Coverage of the IEKF with the original curve, with the smoothed curve in generator and filter, and with the smoothed slope only in its Jacobian, and coverage of the UKF and of the grid filter with the original and smoothed curves; six conditions (three profiles and two segments; S2, prior of 1 percentage point, $q_{\mathrm{RC}} = 10^{-8}$ V^2), 120 trajectories per arm, new random seeds. The three IEKF arms share one set of seeds and the UKF and the grid filter arms another. Bars are 95% percentile intervals over trajectories; horizontal lines mark the nominal levels.

OCV, but not the bias of a regression that treats the OCV and the hysteresis state as constant within the window while both change; pilot runs, with coverage of the same order with zero voltage noise and with 1 mV of noise, were consistent with this explanation, which the confirmatory design does not test.

6.6 Robustness to fixed initial errors

In the 216 robustness conditions, the initial SOC differs from the prior mean by a fixed error, so coverage need not be nominal (Section 4.1). With initial errors of 0, +10 and −10 percentage points,

the 95% coverage of the grid filter stayed between 94.2 and 94.4%, that of the UKF fell to 81.3% with +10 percentage points, that of the IEKF to about 41% with either nonzero error, and that of the Fisher estimator stayed near 3% (Table S7).

7 Conclusion

An accurate SOC estimator does not necessarily know its own error. In a preregistered simulation study on an LFP model with hysteresis, filters that share the same model differed sharply in calibration: intervals reported with 95% probability, which should contain the true SOC 95% of the time, contained it 59.4 and 3.1% of the time for the IEKF and the Fisher estimator, respectively, 92.5% for the UKF, and 94.0% for the Rao–Blackwellized grid filter. The grid filter matched the coverage of the exact reference within 0.2 percentage points when the current was measured without noise, although local discrepancies between the two posteriors remained, and exploratory tests implicate the fine-scale variation of the interpolated OCV slope used in the linearization of the IEKF.

An evaluation by RMSE alone would miss these differences, which matter for a BMS that uses the reported variance, for example to set power limits or to decide when to trust coulomb counting. The intervals of the IEKF and of the Fisher estimator should not be used for such decisions in the conditions studied here. The UKF is a reasonable default, close to nominal in aggregate at about one tenth of the cost of the grid filter in the present implementation (Section S4), although its coverage fell to 81.3% with an initial error of +10 percentage points. When the tails matter, the grid filter gave the coverage closest to nominal and kept it with fixed initial errors, but its intervals should not be trusted with a very precise sensor and no RC process noise until the cause of that failure is resolved.

The study has limitations. It evaluates a matched-model setting with OCV curves from a single cell at 25 °C, in which all estimators receive the true hysteresis state; errors in the model, in the capacity, or in the OCV curve, and bias of the current sensor, are not part of the design. The Fisher estimator was implemented from the published equations, and the cost of the grid filter was measured on a laptop, not on BMS hardware. Finally, an equivalent validation with measured data is harder because the true SOC is not directly observed.

Future work will investigate the regime in which the grid filter loses calibration, estimate the hysteresis state jointly with the SOC, test the estimators under model error and current-sensor bias, evaluate the cost on BMS hardware, and validate the conclusions with measured data, for example through the consistency of the voltage innovations. Meanwhile, the practical message for BMS design is that a reported variance is a claim to be checked, not assumed: in the conditions studied here, the variance of the UKF could be used with modest caution, whereas that of the IEKF should not be used before its coverage has been verified with the OCV map actually deployed, especially when that map is an interpolated table.

Declarations

- **Competing interests:** The author declares no competing interests.
- **Data availability:** The CALCE data used to build the profiles and the OCV curves are publicly available (CALCE Battery Research Group 2026).
- **Code availability:** The code, the preregistration, and the amendments are available at https://github.com/alexandreblima/LFP-SOC-Calibration and archived at Zenodo (Lima 2026).

References

Bar-Shalom Y, Li XR, Kirubarajan T (2001) Estimation with Applications to Tracking and Navigation. Wiley, New York, NY, USA, https://doi.org/10.1002/0471221279

Barlow RE, Bartholomew DJ, Bremner JM, et al (1972) Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression. J. Wiley

Bell BM, Cathey FW (1993) The iterated Kalman filter update as a Gauss–Newton method. IEEE Transactions on Automatic Control 38(2):294–297. https://doi.org/10.1109/9.250476

Bessa IV, Coutinho D, Robalo TS (2026) Robust nonlinear observer design with learning applied to SOC estimation in Li-Ion batteries. Journal of Control, Automation and Electrical Systems 37:297–316. https://doi.org/10.1007/s40313-026-01248-y

Bucy RS, Senne KD (1971) Digital synthesis of non-linear filters. Automatica 7(3):287–298. https://doi.org/10.1016/0005-1098(71)90121-X

CALCE Battery Research Group (2026) Battery data. Center for Advanced Life Cycle Engineering, University of Maryland, College Park, MD, USA, URL https://calce.umd.edu/battery-data, accessed: Nov. 1, 2026

Cook SR, Gelman A, Rubin DB (2006) Validation of software for Bayesian models using posterior quantiles. Journal of Computational and Graphical Statistics 15(3):675–692. https://doi.org/10.1198/106186006X136976

Davison AC, Hinkley DV (1997) Bootstrap Methods and their Application. Cambridge University Press, Cambridge, UK, https://doi.org/10.1017/CBO9780511802843

Dinh NQ, Nguyen KT (2026) The State of Charge Estimation Method for Lithium Battery Using Attractive Ellipsoid Super-Twisting Sliding Mode Observer. Journal of Control, Automation and Electrical Systems 37:1159–1177. https://doi.org/10.1007/s40313-026-01291-9

Gelman A, Carlin JB, Stern HS, et al (2013) Bayesian Data Analysis, 3rd edn. CRC Press, Boca Raton, FL

Gneiting T, Balabdaoui F, Raftery AE (2007) Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69(2):243–268. https://doi.org/10.1111/j.1467-9868.2007.00587.x

Grewal MS, Andrews AP (2015) Kalman Filtering: Theory and Practice Using MATLAB, 4th edn. John Wiley & Sons, Hoboken, New Jersey

Guntuboyina A, Sen B (2018) Nonparametric shape-restricted regression. Statistical Science 33(4):568–594. https://doi.org/10.1214/18-STS665

Jazwinski AH (1970) Stochastic Processes and Filtering Theory, Mathematics in Science and Engineering, vol 64. Academic Press, New York

Julier SJ, Uhlmann JK (2004) Unscented filtering and nonlinear estimation. Proceedings of the IEEE 92(3):401–422. https://doi.org/10.1109/JPROC.2003.823141

Knox J, Blyth M, Hales A (2024) Advancing state estimation for lithium-ion batteries with hysteresis through systematic extended Kalman filter tuning. Scientific Reports 14:12472. https://doi.org/10.1038/s41598-024-61596-0

Lima AB (2026) LFP-SOC-Calibration: Reproducibility package for the calibration of reported uncertainty in state-of-charge estimation of LFP cells. Zenodo, version v1.0-submission, https://doi.org/10.5281/zenodo.24326741

Plett GL (2004) Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 2. modeling and identification. Journal of Power Sources 134(2):262–276. https://doi.org/10.1016/j.jpowsour.2004.02.032

Plett GL (2006) Sigma-point Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1: Introduction and state estimation. Journal of Power Sources 161(2):1356–1368. https://doi.org/10.1016/j.jpowsour.2006.06.003

Plett GL (2016) Battery Management Systems, Volume II: Equivalent-Circuit Methods. Artech House, Norwood, MA, USA

Previtali D, Previdi F (2025) Optimal calibration of Kalman filters for state of charge estimation of lithium-ion cells. Journal of Energy Storage 139:118753. https://doi.org/10.1016/j.est.2025.118753

Priya RP, R S, Sakile R (2023) State of charge estimation of lithium-ion battery based on

extended Kalman filter and unscented Kalman filter techniques. Energy Storage 5(3):e408. https://doi.org/10.1002/est2.408

Schön T, Gustafsson F, Nordlund PJ (2005) Marginalized particle filters for mixed linear/nonlinear state-space models. IEEE Transactions on Signal Processing 53(7):2279–2289. https://doi.org/10.1109/TSP.2005.849151

Schuirmann DJ (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics 15(6):657–680. https://doi.org/10.1007/BF01068419

Shi J, Jiang S, Tao S, et al (2026) An adaptive estimation approach based on Fisher information to overcome the flat voltage plateau challenges of SOC estimation in LFP batteries. Energy and AI 24:100693. https://doi.org/10.1016/j.egyai.2026.100693

Silva CT (2022) Sistema de gerenciamento de baterias de lítio com estimação de estados com filtro de Kalman estendido adaptativo. PhD thesis, Escola Politécnica, Universidade de São Paulo, São Paulo, https://doi.org/10.11606/T.3.2022.tde-04042023-080719

da Silva CT, Dias BMdA, Araújo RE, et al (2024) A Practical Methodology for Real-Time Adjustment of Kalman Filter Process Noise for Lithium Battery State-of-Charge Estimation. Batteries 10(7). https://doi.org/10.3390/batteries10070233, URL https://www.mdpi.com/2313-0105/10/7/233

Särkkä S, Svensson L (2023) Bayesian Filtering and Smoothing, 2nd edn. Cambridge University Press, https://doi.org/10.1017/9781108917407

Talts S, Betancourt M, Simpson D, et al (2018) Validating Bayesian inference algorithms with simulation-based calibration. https://doi.org/10.48550/arXiv.1804.06788

Wang Z, Gladwin DT, Smith MJ, et al (2021) Practical state estimation using Kalman filter methods for large-scale battery systems. Applied Energy 294:117022. https://doi.org/10.1016/j.apenergy.2021.117022

Widrow B, Kollár I (2008) Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications. Cambridge University Press, Cambridge, https://doi.org/10.1017/CBO9780511754661

Zhang K, Chen C, Er L, et al (2025) Robust state-of-charge estimation for $LiFePO_4$ lithium-ion batteries with pronounced voltage plateau regions. Applied Energy 401:126755. https://doi.org/10.1016/j.apenergy.2025.126755

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OnlineResource1.pdf