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.

Hybrid Extended Kalman Filtering and Feedforward Neural Correction for Lithium-Ion Battery State-of-Charge Estimation

Daksh Rastogi
Indian Institute of Technology Roorkee
Mayank Goswami
mayankfph@iitr.ac.in

Indian Institute of Technology Roorkee

Research Article

Keywords: lithium-ion battery, state-of-charge estimation, extended Kalman filter, feedforward neural network, battery management system, residual learning, equivalent-circuit model

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11634907/v1

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

Additional Declarations: The authors declare no competing interests.

Abstract

Estimating the state of charge (SOC) of a lithium-ion battery from terminal measurements is difficult because voltage depends on load, temperature and internal polarization as well as remaining charge. This study evaluates a hybrid estimator coupling a calibrated, temperature-dependent first-order equivalent-circuit extended Kalman filter (EKF) with a feedforward neural network (FNN). Rather than predicting SOC independently, the FNN uses causal measurement history and EKF diagnostics to produce two bounded corrections: a voltage-model offset applied within the EKF observation update and an output-only SOC residual correction. The frozen network was developed using four Panasonic 18650PF recordings and two development-validation recordings. In a diagnostic replay of ten previously examined automotive load-profile recordings at 10°C and 25°C, equal-weighted mean recording-level RMSE against a charge-counter-derived SOC reference decreased from 0.5863 to 0.4516 SOC percentage points, a 22.98% relative reduction. Eight recordings improved, while both UDDS recordings deteriorated. Although the results demonstrate reproducible condition-dependent error correction, the reference relies on an assumed 2.79826 Ah capacity, benchmark recordings were previously inspected, and no independent multi-cell generalization or safety claim is made.

1. Introduction

The state of charge (SOC) is the fraction of a cell's available charge remaining relative to a defined capacity. It is central to battery management systems (BMSs) because it informs operating limits, usable-energy estimates and user-facing battery indicators. SOC cannot be directly measured with a sensor in the manner of voltage or current. In particular, a loaded cell's terminal voltage is affected by ohmic drop, diffusion-related polarization, temperature and past current, which complicates the inversion of measured voltage into SOC [2], [5].

Coulomb counting integrates measured current but accumulates errors from uncertain capacity, sensor offsets and inaccurate initial conditions. Equivalent-circuit models (ECMs) provide computationally tractable relations among SOC, voltage and current, while Kalman filtering fuses dynamic predictions with noisy voltage measurements [2]–[5]. Nonetheless, even a calibrated ECM leaves mismatch because real cells exhibit behavior not fully represented by a low-order model. Machine-learning methods can learn repeatable relationships in that mismatch, but data-driven corrections may overfit or degrade an accurate baseline.

This investigation develops a compact physics-guided, two-output FNN as an augmentation to a calibrated EKF for one Panasonic 18650PF cell. The focus is the measured behavior and reproducibility of a specific trained estimator, rather than an unsupported claim of architectural novelty. The principal question is whether neural correction lowers SOC error across heterogeneous dynamic load recordings, and whether the same correction can fail when the EKF is already accurate.

2. Related work and motivation

Plett established an extended-Kalman-filter framework for battery state and parameter estimation, including battery modeling and the use of dynamic voltage observations [2]–[4]. Hu, Li and Peng compared equivalent-circuit alternatives and emphasized the need to balance model fidelity, complexity and robustness [5]. These studies motivate using an EKF as a physically interpretable estimator rather than replacing it indiscriminately with a fully data-driven mapping.

Neural battery SOC estimators have also been explored extensively. Chemali et al. employed long short-term memory (LSTM) networks for SOC prediction using measured battery signals [6]. Vidal et al. compared recurrent and non-recurrent architectures and demonstrated that filtered signal histories can make feedforward networks effective, including on Panasonic 18650PF measurements [7]. More recent literature surveys discuss supervised-learning, feature engineering and hybrid estimation approaches, along with their validation challenges [8]. These studies establish substantial prior art for neural SOC estimation; neither FNNs nor physics/neural combinations are claimed as new here.

The specific engineering choice examined here is to retain a fixed physics-calibrated EKF and train a network that predicts both pre-update voltage mismatch and an SOC output residual. The first network output enters the filter's observation prediction; the second is deliberately not fed back into the EKF state. This distinction makes it possible to investigate two complementary ways in which learned errors affect the final estimate.

3. Data and experimental protocol

3. Data and experimental protocol

3.1 Panasonic 18650PF recordings

The measurements are from the public Panasonic 18650PF battery dataset released by Kollmeyer [1]. The underlying experiments were conducted using a lithium-ion cell in a temperature-controlled chamber with dynamic vehicle-inspired electrical loading. The study uses nominal 10°C and 25°C recordings and does not claim measurements from other physical cells.

The frozen project designates four recordings for neural training (10°C Cycles 1 and 3; 25°C Cycles 2 and 3) and two Cycle 4 recordings, one per temperature, for development validation. Ten named driving-cycle recordings are used for the reported diagnostic benchmark: HWFET, NN, UDDS and US06 at 10°C; HWFETa, HWFETb, LA92, NN, UDDS and US06 at 25°C. All ten recordings and the two validation recordings come from the same source cell and related measurement campaign. They are separate recording files, not evidence of independently sampled cells.

Table 1
Recording assignments. All benchmark cases had been inspected in earlier project development and must not be described as entirely untouched independent tests.

Assignment	Recordings	Intended use
Training (4)	10°C Cycle 1, Cycle 3; 25°C Cycle 2, Cycle 3	Fit network parameters and input scaler
Development validation (2)	10°C Cycle 4; 25°C Cycle 4	Model selection, early stopping, correction strengths
Diagnostic benchmark (10)	10°C: HWFET, NN, UDDS, US06; 25°C: HWFETa/b, LA92, NN, UDDS, US06	Frozen inference and displayed scores

3.2 Data handling and provisional reference

MAT records are loaded using the preserved reader and validated for sensor consistency. Current is represented as positive during discharge. For one-second replay, the most recent recorded observation at or before each integer-second reporting instant is selected; the algorithm does not interpolate using future values. Each replay begins with an estimated SOC of 0.80, while the derived reference assumes a fully charged origin. This deliberately incorrect initialization creates an initial transient and should not be confused with an actual starting charge measured by an independent instrument.

The reference SOC is reconstructed from the tester's ampere-hour counter with an assumed effective full-charge capacity Qref = 2.79826 Ah:

$z_{ref,k} = 1 + Ah_{counter,k} / 2.79826.$ (1)

The sign convention corresponds to the supplied discharge-counter field. This label is a charge-throughput-derived reference, not an independently measured electrochemical ground truth. Capacity assumptions are a consequential limitation of every reported RMSE, especially for very small residual differences.

4. Physics-based estimator

4.1 One-RC equivalent-circuit formulation

The EKF tracks a two-dimensional state $x = [z, V_p]^T$ comprising SOC z and a lumped polarization voltage Vp. With discharge-positive current I, equivalent capacity Q, ohmic resistance R0, polarization resistance R1 and time constant τ, the model uses:

$z^- = clip(z_{k-1} - \bar{I}_k \Delta t / (3600\, Q), 0, 1),$ (2)

$V_p^- = a_k V_{p,k-1} + (1 - a_k) R1\, \bar{I}_k, \quad a_k = \exp(-\Delta t/\tau).$ (3)

Here $\bar{I}_k$ is the average of current readings across the update. The modeled terminal voltage is:

$V_{model,k} = OCV(z^{-}_k, T_k) - R0(z^{-}_k, T_k)\, I_k - V^{-}_{p,k}$. (4)

OCV, R0, R1 and τ are obtained by interpolation from calibrated SOC- and temperature-dependent tables. The EKF propagates a covariance matrix, linearizes the voltage observation using OCV and resistance derivatives and forms a Kalman-gain-weighted update. An adaptive, bounded voltage-measurement noise term moderates the filter's response to changing voltage innovations. The original physics calibration is kept fixed for baseline and hybrid comparisons.

4.2 EKF measurement correction

At time k, the innovation of the physics-only filter is $v_k = V_{measured,k} - V_{model,k}$. Denoting the observation Jacobian H_k, predicted state covariance P^{-}_k, and observation-noise covariance R_k, the update is represented conventionally by:

$K_k = P^{-}_k H_k^{T} (H_k P^{-}_k H_k^{T} + R_k)^{-1}$, $x_k = x^{-}_k + K_k v_k$. (5)

This formulation explains the hybrid entry point: an ML-proposed offset is added to the modeled voltage before calculating the innovation. The network does not replace the physics state transition, numerical calibration or independent hardware protection functions.

5. Feedforward neural augmentation

5.1 Causal feature construction

Fifteen features are constructed for each second: present current and cell temperature; previous measured voltage, current and temperature; EKF SOC and polarization; previous voltage innovation; fast and slow exponential moving averages of current and voltage; current change; predicted SOC uncertainty; and elapsed discharge time. Voltage features and previous innovations are updated only after the current time-step correction is computed. Consequently, the voltage head does not receive the same time-step measured voltage that it is tasked with correcting.

For each predicted time step, a causal history of the latest 64 feature vectors is normalized using statistics learned from the training recordings alone. Four 15-dimensional summaries are concatenated: the latest vector, mean of the latest eight, mean of the latest 32, and mean of all 64. This produces a 60-dimensional input, so temporal context is supplied without recurrent architecture.

5.2 Network design and training objective

The selected feedforward network uses fully connected layers 60 → 48 → 24 → 2, ReLU hidden activations and a final tanh activation. The two outputs correspond to normalized voltage innovation and normalized SOC residual, each bounded during supervised training. The network contains 4,154 trainable parameters and uses Smooth L1 (Huber-type) loss for the two target channels, AdamW optimization (learning rate 0.001, weight decay 0.001) and early stopping on development-validation loss. The

training-script seed is 351; the artifact contains fixed trained weights, a feature scaler and selection metadata.

y_t = FNN([f_t; $mean_8(f)$; $mean_{32}(f)$; $mean_{64}(f)$]), $y_t \in [-1,1]^2$. (6)

The voltage head was trained against the clipped baseline EKF innovation scaled by 0.025 V, while the SOC head was trained against the clipped charge-counter-reference minus baseline EKF SOC, scaled by 0.04 SOC fraction. During inference the learned outputs are intentionally conservative, using development-selected multipliers $\alpha V = 0.5$ and $\beta z = 0.5$.

5.3 Coupling into the EKF

The first output modifies the EKF-predicted voltage, influencing the innovation and its state update. The second output is a separately bounded, output-only SOC correction:

$V_{hybrid\text{-}pred,t} = V_{model,t} + \alpha V \times 0.025 \times y_{t,voltage}$, (7)

$z_{final,t} = clip(z_{hybrid\text{-}EKF,t} + \beta z \times 0.04 \times y_{t,SOC}, 0, 1)$. (8)

The final SOC residual correction does not feed back into the EKF state at the next second. Learned corrections are suppressed for the first 60 seconds after initialization and when measured temperature falls outside approximately 9–31°C. These bounds are safeguards against unsupported corrections, but should not be interpreted as a quantified probabilistic confidence measure or a formal safety guarantee.

6. Evaluation and reproducibility

The trained model and EKF calibration are frozen for the reported experiment. The provided reproduction script verifies file integrity, reads original MAT recordings, executes EKF and FNN inference, computes errors, and generates the figures anew. Previously stored results are not used as a source of predicted SOC values. Feature calculation is causal, and normalization statistics were fitted using designated training recordings. The study is a model-development evaluation, not a fully prospective validation protocol; benchmark profiles had been viewed during earlier iterations of the project.

Let $e_t = 100(\hat{z}_t - z_{ref,t})$ be the estimation error in SOC percentage points (pp). For an evaluated recording of N samples:

$RMSE = \sqrt{[(1/N) \Sigma_t e_t^2]}$, $MAE = (1/N) \Sigma_t |e_t|$. (9)

The headline mean is the arithmetic mean of the ten per-recording RMSE values, granting each named recording equal weight rather than weighting by its number of sensor samples. Relative reduction is 100 × (mean_RMSE_EKF − mean_RMSE_hybrid) / mean_RMSE_EKF. Peak absolute error is assessed separately. No confidence interval or cross-cell generalization is implied by the descriptive averages.

7. Results

Table 2
Full diagnostic benchmark for the 80%-initialized estimator. Percentage values represent relative change in recording RMSE; negative values indicate degradation.

Recording	EKF (pp)	Hybrid (pp)	Δ RMSE	Result
10C HWFET	1.2226	0.7562	+ 38.1%	Improved
10C NN	0.6112	0.3373	+ 44.8%	Improved
10C UDDS	0.1424	0.2603	-82.8%	Worse
10C US06	0.7900	0.5466	+ 30.8%	Improved
25C HWFTa	0.6962	0.5557	+ 20.2%	Improved
25C HWFTb	1.0720	0.9726	+ 9.3%	Improved
25C LA92	0.3988	0.2990	+ 25.0%	Improved
25C NN	0.3951	0.2907	+ 26.4%	Improved
25C UDDS	0.1684	0.2431	-44.4%	Worse
25C US06	0.3667	0.2546	+ 30.6%	Improved

The equal-weighted mean recording RMSE decreased from 0.586341 pp for the adaptive EKF to 0.451611 pp for the hybrid, a 22.98% relative reduction. Eight recordings improved and two worsened. These are reductions in prediction disagreement with the specified derived reference, not in physical stored energy. The mean 25°C case RMSE decreased from 0.5162 to 0.4359 pp, while the mean 10°C case RMSE decreased from 0.6916 to 0.4751 pp.

7.1 Beneficial corrections

Among all recordings, the greatest relative reduction occurred on the 10°C NN profile: RMSE decreased from 0.6112 to 0.3373 pp (44.8%). The largest absolute reduction occurred on 10°C HWFET, from 1.2226 to 0.7562 pp (0.4664 pp, or 38.1% relative). Both US06 recordings also improved (10°C: 0.7900 to 0.5466 pp; 25°C: 0.3667 to 0.2546 pp), indicating the model did not only help smooth highway-like loads. These effects are recording-level observations and should not be generalized as guarantees for all highway or aggressive driving conditions.

7.2 Failure on already accurate EKF trajectories

The two UDDS recordings show the opposite outcome. At 10°C, RMSE increases from 0.1424 to 0.2603 pp (82.8% relative increase), and at 25°C from 0.1684 to 0.2431 pp (44.4% increase). Although these absolute RMSEs remain below 0.3 pp, the learned correction reduces the quality of a baseline that already tracks the derived reference closely. This is a critical model-design limitation and argues against unconditional correction in a practical BMS.

7.3 Maximum error and reference sensitivity

In the replayed recordings, maximum absolute errors reported for EKF and hybrid are identical because the dominant error occurs during the initial, correction-disabled portion of the trajectories. The improvement in mean RMSE must therefore not be misrepresented as a reduction in peak error or improved initial SOC acquisition. A separate exploratory capacity-sensitivity study in the project history found that changing assumed reference capacity can change comparative model rankings; this manuscript uses the frozen 2.79826 Ah reference consistently and does not claim independently validated absolute SOC.

8. Discussion

The empirical results support a narrow but useful conclusion: a modest FNN with causal measurement context can improve an already calibrated EKF's average agreement with a derived SOC reference on several dynamic load histories. In effect, the voltage head changes how the filter interprets model mismatch, while the SOC head corrects a residual on the final reported output. These two paths offer expressive correction without replacing the underlying equivalent-circuit state model.

The failure pattern is as important as the mean. On UDDS the physics-based estimator performs so well that additional correction becomes unnecessary or harmful. A model trained primarily to reduce average residuals can fail to identify this situation. The 64-step causal history captures short-term dynamics but not all uncertainties in capacity, OCV hysteresis, temperature variation or acquisition provenance. Moreover, observations from one laboratory cell are highly correlated and do not represent a statistical sample of EV battery packs.

For a practical BMS, improved SOC estimation may help energy indication or control-state awareness, but deployment requires additional calibration, embedded timing measurements, fault handling, independent cell populations and temperature/ageing tests. No enhanced range, charge capacity, safety margin or operating lifetime has been measured in this project. The hybrid model is a research demonstrator and must not replace independent cell voltage, current, temperature or protection limits.

9. Limitations and future work

Four limitations govern interpretation. First, SOC labels are reconstructed from a tester counter with assumed full-charge capacity, so low-RMSE results partly depend on that modeling decision. Second, the benchmark recordings were previously examined during project development and therefore do not establish fully untouched prospective generalization. Third, the training and benchmark data arise from one cell and only two supported temperatures, precluding claims about cross-cell performance, ageing or extrapolation. Fourth, the 60-second warm-up and two UDDS failures show that the network cannot guarantee improvements in all regimes or correct large initial state errors.

A follow-up study should characterize capacity and SOC reference more independently, freeze evaluation conditions before acquisition, include multiple aged and fresh cells, evaluate across temperatures outside those used for training and learn a validation-tested reliability policy that can decline correction when expected benefit is uncertain. Such a policy should be judged using entirely new experiments rather than additional tuning to these ten previously viewed recordings.

10. Conclusion

A frozen feedforward neural model was integrated with a temperature-aware EKF for Panasonic 18650PF SOC estimation and evaluated using reproducible one-second replay. Compared with the physics-only EKF under a specified derived SOC reference, the hybrid reduced equal-weighted mean recording-level RMSE by 22.98% (0.5863 to 0.4516 pp) and improved eight of ten observed benchmark recordings. Two UDDS recordings deteriorated, and peak errors were unchanged. The principal contribution is a transparent engineering case study in combining causal neural residual estimation with model-based state estimation, including explicit documentation of where the correction helps and where it harms. Broader generalization and deployment readiness remain open questions.

Declarations

Conflict of interest.

The author reports no known competing financial interests related to this work; this statement requires final author confirmation before posting.

Author contributions

and acknowledgements. The student author and any additional contributors should review and approve accurate contributions, acknowledgements, and author order before preprint submission. Faculty supervision is not automatically equivalent to coauthorship.

Data availability

, code and declarations

Data availability. Original cell measurements were obtained from the public Panasonic 18650PF dataset by Kollmeyer [1]. The archived course package contains the subset used for training, development validation and diagnostic reproduction, subject to the dataset’s stated attribution terms.

Code availability. A frozen Python environment specification, trained network artifact, EKF calibration, dataset manifest and reproduction script accompany the course submission. A public code URL is

intentionally omitted until an actual release is authorized.

References

1. , [2], Hu X, Li S, Peng H (2012) A comparative study of equivalent circuit models for Li-ion batteries. J Power Sources 198:359–367. 10.1016/j.jpowsour.2011.10.013. [3], [4], [5], [6], [7] [1]
2. Plett GL (2004) Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation. J Power Sources 134(2):277–292. 10.1016/j.jpowsour.2004.02.033
3. Plett GL (2004) Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 2. Modeling and identification. J Power Sources 134(2):262–276. 10.1016/j.jpowsour.2004.02.032
4. Plett GL (2004) Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1. Background. J Power Sources 134(2):252–261. 10.1016/j.jpowsour.2004.02.031
5. Chemali E, Kollmeyer PJ, Preindl M, Ahmed R, Emadi A (2018) Long Short-Term Memory Networks for Accurate State-of-Charge Estimation of Li-ion Batteries. IEEE Trans Industr Electron 65(8):6730–6739. 10.1109/TIE.2017.2787586
6. Vidal C, Malysz P, Naguib M, Emadi A, Kollmeyer PJ (2022) Estimating battery state of charge using recurrent and non-recurrent neural networks. J Energy Storage 47:103660. 10.1016/j.est.2021.103660
7. Fadlaoui O, Masaif N Machine Learning-Based Estimation of State of Charge and State of Health in Lithium-Ion Batteries for Electric Vehicles, in (2025) *International Conference on Circuit, Systems and Communication (ICCSC)*, 2025. 10.1109/ICCSC66714.2025.11135292

Figures

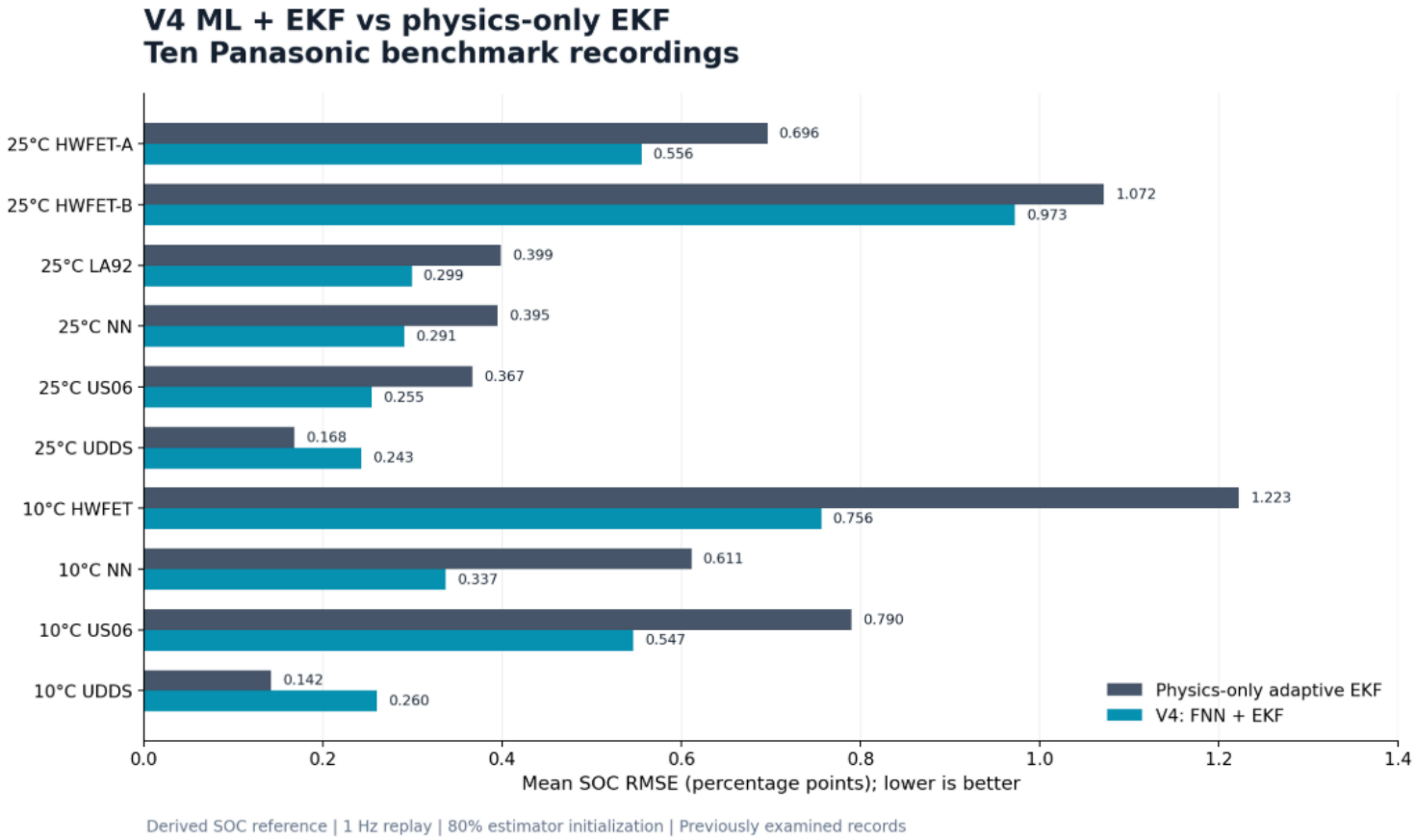


Figure 1

Recording-by-recording SOC RMSE for the frozen hybrid and identical physics EKF baseline. Lower bars indicate better agreement with the derived SOC reference.

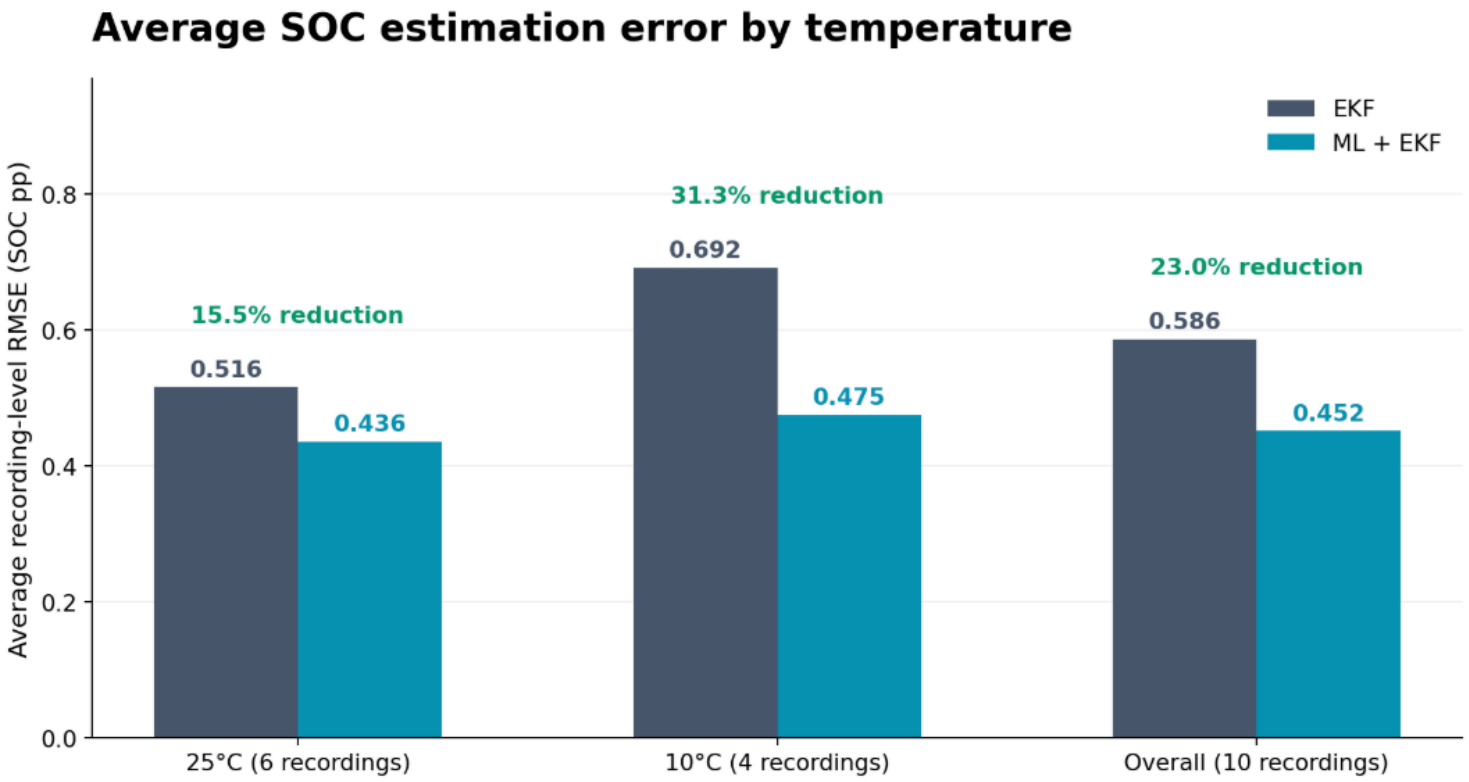


Figure 2

Temperature-stratified equal-weighted mean recording RMSE for the same ten diagnostic benchmark cases.

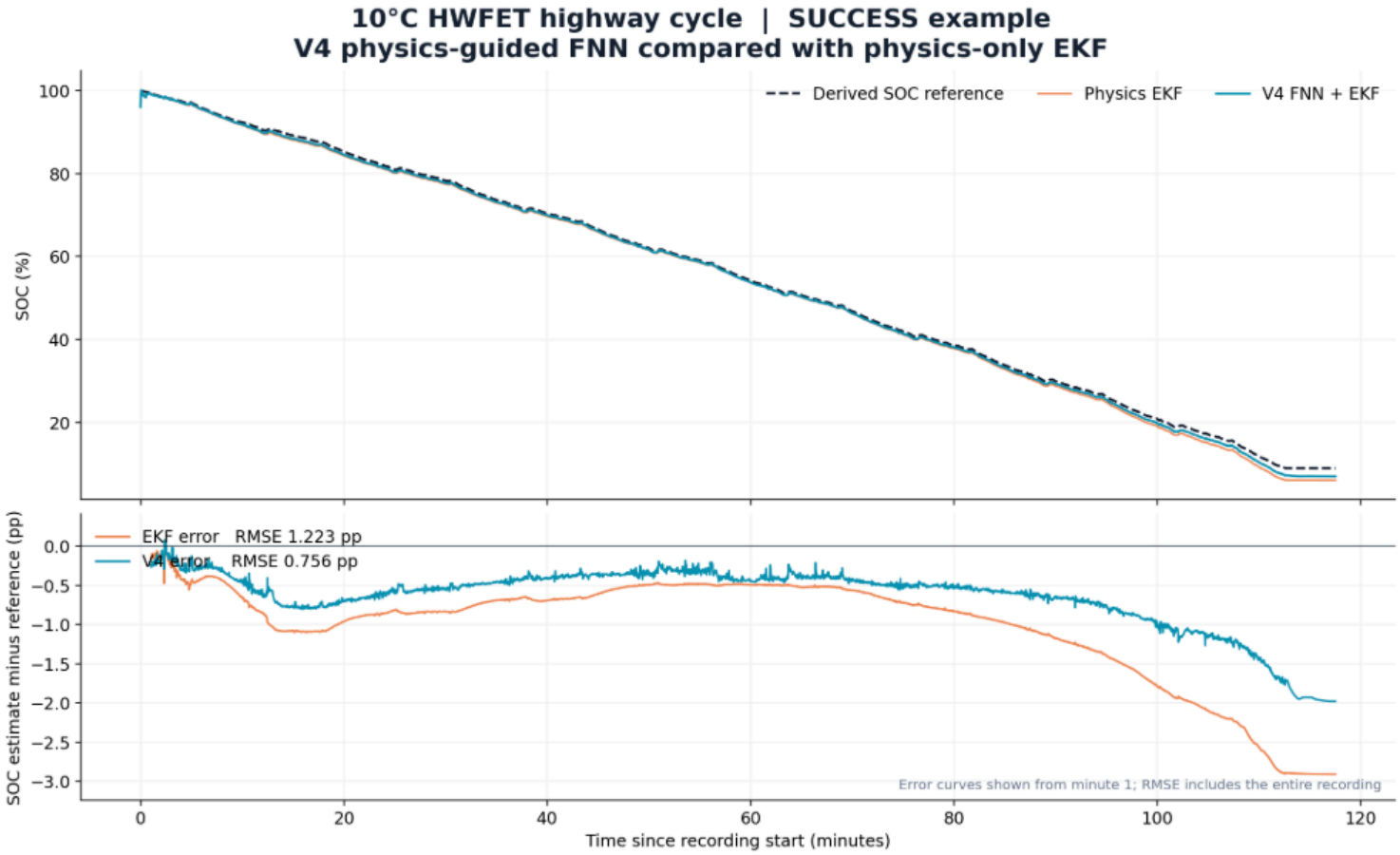


Figure 3

Example of reduced SOC estimation error for the 10 °C HWFET recording. The lower panel displays signed error relative to the Ah-derived reference.

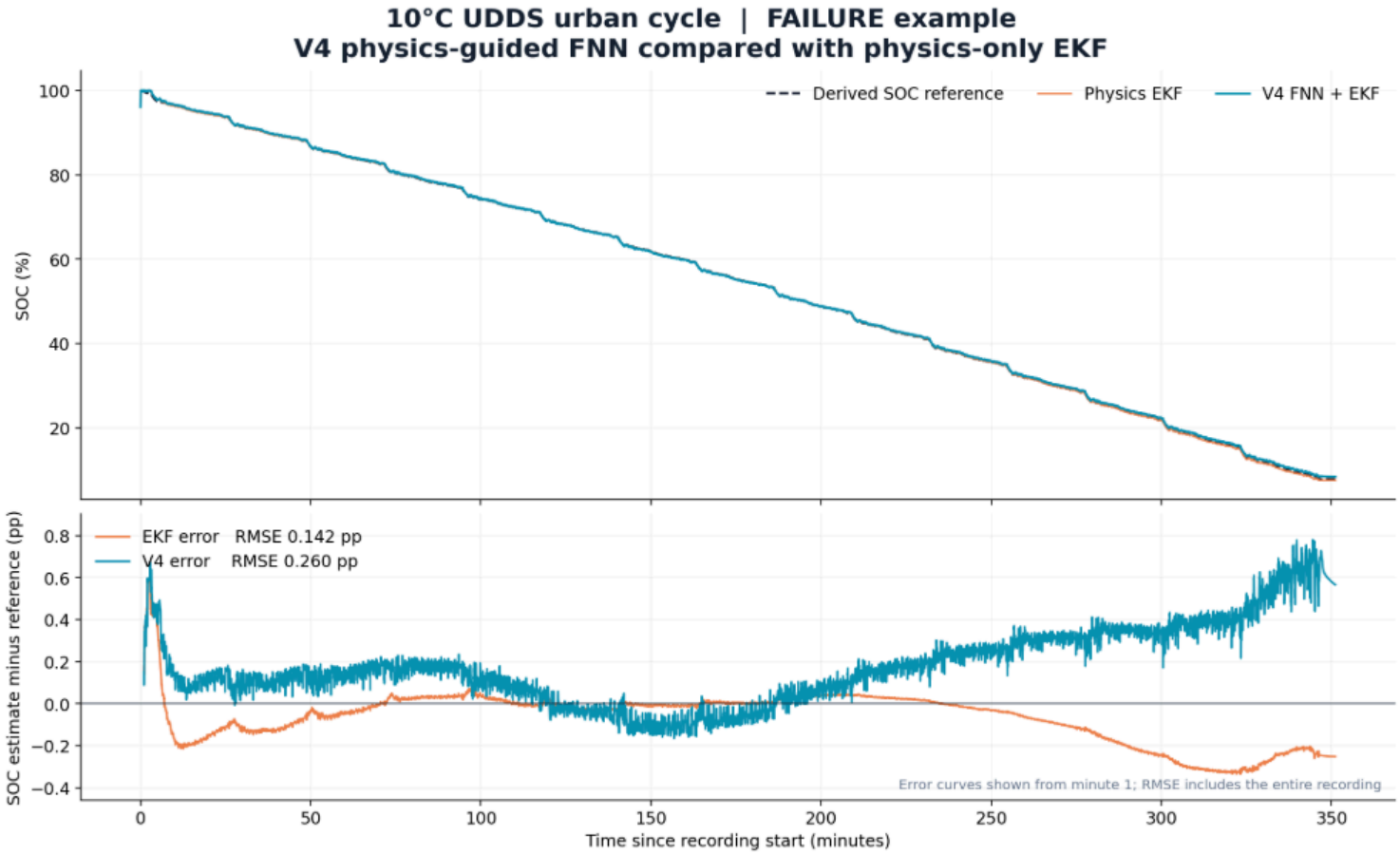


Figure 4

UDDS failure example. The EKF is close to the derived reference and the learned correction increases signed SOC error over parts of the discharge.