energies

MDPI

Article

Data Augmentation for Lithium-Ion Battery State-of-Charge Estimation Across Methods and Models

Alexey A. Druzhinin * and Daniel Lisnovetski

Moscow Center for Advanced Studies, 123592 Moscow, Russia
* Correspondence: drusalexey@gmail.com

Abstract

Accurate state-of-charge (SoC) estimation from voltage, current, and temperature is required for lithium-ion pack energy management. Learning-based estimators need labeled field traces that are costly on a deployed sensing chain, and data augmentation is used to enlarge those sets. Published studies almost always train one synthesizer for one learner, so it remains unknown which operators help which estimator classes, at what intensity, and after how much real labeled volume. In this paper, a factorial of classical, physics-informed, and generative operators is evaluated on twelve SoC estimators (dose $L = 1$–10, labeled volume $n = 1$–5; >5000 configurations) using dual stand–logger unmanned aerial vehicle (UAV) pack flights as the labeled corpus, with root-mean-square error (RMSE) scored on a locked acceptance trajectory. RMSE change versus the matched unaugmented baseline is class-conditional: sequential and convolutional estimators improve by +18% to +32%, LightGBM and XGBoost average ≈−2.4% and −4.9%, and a physics-informed network degrades by ≈22%. Labeled volume accounts for ≈53% of log RMSE variance versus ≈2% for synthesizer identity. Operators that preserve the empirical training distribution (thermo-voltage coupling and point shuffle) outrank mean-shift and TimeGAN. Dose L is a process parameter: thermo-voltage is optimal at $L = 8$ for all eleven classical estimators, and selecting L on the acceptance set inflates reported gain by a factor of ≈2.5 relative to averaging over L. Admissible operators therefore depend on estimator class and on labeled volume n.

Keywords: data augmentation; state of charge; lithium-ion batteries; battery management systems; machine learning; energy storage; unmanned aerial vehicles

1. Introduction

Lithium-ion batteries in stationary storage, electric vehicles, and other mobile platforms require online state-of-charge (SoC) estimates from voltage, current, and temperature. Those estimates must hold under sensor bias, electromagnetic interference, and highly dynamic loads [1–7]. Battery management systems (BMSs) still rely on Coulomb counting. They still rely on open-circuit voltage (OCV) correction after rest. They still rely on extended and unscented Kalman filters built on equivalent-circuit models (ECMs) [8–19]. These observers remain attractive because process and measurement noise can be tuned explicitly, failsafe limits are straightforward to impose, and the structure maps onto embedded BMS firmware. Their accuracy, however, depends on ECM fidelity under temperature transients, sensor bias, and current spikes that laboratory drive cycles under-represent.

Learning-based estimators—gradient-boosted trees and deep sequence models—are competitive on laboratory cycles [20–28]. Labeled field telemetry that matches a deployed sensing chain remains scarce. Dual stand–logger unmanned aerial vehicle (UAV) flights

Check for updates

Academic Editor: King Jet Tseng

Received: 21 September 2026
Revised: 21 October 2026
Accepted: 3 November 2026
Published: 5 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

supply the high-dynamic labeled corpus. Propulsion spectra, logger scale and bias residuals, and scarce synchronized reference SoC are treated as properties of the dataset rather than as laboratory artifacts [29–35]. The study addresses lithium-ion SOC estimation under label scarcity in general; the UAV pack serves as the experimental corpus rather than as the target application.

1.1. Motivation and Problem Statement

Deploying learning-based SoC estimators is limited less by model capacity than by the cost of labeled trajectories with synchronized reference SoC. Domain shift between laboratory cycles and field hardware is the other limit [7,36–38]. Synthetic expansion of the training set is an attractive option [39,40]. The recent literature has emphasized generative adversarial networks (GANs) and TimeGAN variants for battery state estimation [41–52]. Mohanty et al. combined TimeGAN with a BERT-based electric-vehicle (EV) SoC predictor [48], while Channegowda et al. demonstrated deep generative synthesis of battery time series [49].

That literature almost always trains *one synthesizer* for *one learner*. A single pairing can show whether a proposed generator improves a chosen network. It does not map the interactions that control a data pipeline: (i) which transform family should be enabled, (ii) at what intensity, (iii) for which estimator class, and (iv) after how much real labeled volume synthesis remains justified.

Additional labeled traces require calibrated stands, synchronized loggers, and held-out acceptance trajectories. Synthetic samples are essentially free once the pipeline is implemented. If synthesizer identity explained most of the residual error, investment in generative models would be the priority. If labeled volume explained most of the residual error, additional instrumentation would be the priority. A factorial map across methods and models is required before either conclusion can be drawn [4,44,48]. (1) Labeled state-of-charge traces are scarce, and augmentation is used without a map of its effects. (2) Published studies train one synthesizer for one learner. (3) This paper runs a factorial on twelve estimators, eight classical operators plus TimeGAN, $L = 1 \dots 10$, and $n = 1 \dots 5$—more than 5000 trained configurations, scored on one locked acceptance flight. (4) Gain is conditional on estimator class, labeled volume dominates synthesizer identity, and dose is a process parameter. Those three findings are C1, C2, and C4. C3 adds that the operator ranking tracks preservation of the training distribution.

1.2. Contributions

The factorial isolates those interactions for twelve SoC estimators. Dual stand–logger UAV flights supply the labeled corpus. The experimental factors are the augmentation method m, the cumulative dose L, the estimator f, and the real labeled volume n. The study compares existing augmentation operators rather than proposing a new generative architecture. The measured contrasts are as follows: RMSE is the root-mean-square error on the locked acceptance flight. $\Delta\%$ is the relative change of that RMSE against the matched unaugmented twin. Positive $\Delta\%$ means that the error fell. GRU is a gated recurrent unit. TCN is a temporal convolutional network. PINN is a physics-informed neural network.

1. **Estimator-Class Conditional Gain (C1):** Under matched unaugmented baselines, sequential and convolutional estimators improve mean RMSE by +18% to +32%. LightGBM and XGBoost average $\Delta\% \approx -2.4\%$ and -4.9%, respectively. The value -2.4 includes the TimeGAN configurations; without them (Section 6.1), the LightGBM mean is -2.7. A physics-informed neural network (PINN) degrades by ≈22%. The same synthetic corpus that regularizes a GRU or TCN estimator can be neutral or harmful

for a tree ensemble, or for a PINN whose residual constraints are not co-designed with the operator. The same operator is not admissible for every estimator class.

2. **Labeled Volume Dominates Synthesizer Identity (C2):** One-factor variance shares of log RMSE are $n \approx 53\%$, estimator $\approx 7\%$, dose $L \approx 3\%$, and method $\approx 2\%$. Under this sensing chain, collecting an additional labeled flight dominates exchanging generators. Mean unaugmented RMSE is non-monotonic in n for most sequence models, while mean augmentation $\Delta\%$ increases with n. Method identity does not transfer from $n = 1$ to $n = 5$. Only 25% of estimators keep the same best-performing method from $n = 1$ to $n = 5$ (mean within-estimator Kendall $\tau \approx 0.03$). Fixing the $n = 1$ method costs about 13 percentage points of mean relative gain at $n = 5$. Vector shift collapses from a low-n top contender to last place as n grows.
3. **Ranking Tracks Distribution Preservation (C3):** Thermo-voltage (+22.5% mean) and point shuffle (+18%) lead; mean-shifting vector shift and TimeGAN trail. The ranking is monotonic in how closely the operator preserves the empirical training law—a signature of regularization under a fixed epoch budget rather than of information gain. Operators advertised as physics-aware or distribution-matching do not automatically outrank inexpensive preservation transforms under a common acceptance trajectory. TimeGAN occupies a volume-conditional niche: it wins no best-of-L cells at $n \leq 2$, but at $n = 3$ it is best for both tree and both feedforward estimators while remaining dominated for sequential estimators. That niche does not overturn the pooled ranking; it confines TimeGAN to tree and feedforward estimators at intermediate n.
4. **Dose Is a Process Parameter (C4):** At $L = 8$, thermo-voltage is optimal for all eleven classical estimators; early $L = 2$ optima for noise on trees reverse at high dose. Dose must therefore be swept end-to-end before the training configuration is fixed. Selecting the best L on the acceptance flight overstates the mean dose-averaged improvement by a factor of approximately 2.5× relative to averaging over L. Published performance reports must separate validation-selected dose from acceptance-reported RMSE. First-copy ($L = 2$) winners are not high-dose winners: Spearman correlation of method mean $\Delta\%$ at $L = 2$ versus $L \in \{8, 9, 10\}$ is only 0.17, and the same best method appears in only 15% of estimator×n cells. Fixing the $L = 2$ winner costs about 25 percentage points at high dose. Gaussian noise on trees peaks at $L = 2$ (+22%/+34%) and reverses by $L = 10$. Overdose past $L^\star$ is common and class-conditional: about 63% of dose curves peak and then degrade by >5% toward $L_{\max}$ (36% by >20%). PINN and sequential estimators under $n = 5$ are the settings most sensitive to overdose; trees are more overdose-tolerant.

The remaining sections proceed as follows: Section 2 reviews related work and positions the contribution with a structured comparison. Section 3 formalizes the estimation problem and dataset. Section 4 presents the factorial augmentation framework and operator equations. Section 5 details the experimental protocol and estimators. Section 6 reports results corresponding to C1–C4. Section 7 discusses mechanisms, battery-management implications, and threats to validity. Section 8 translates the findings into engineering guidelines. Section 9 states limitations. Section 10 concludes the paper.

2. Related Work and Positioning

2.1. Model-Based and Embedded SoC Estimation

Online SoC on embedded battery management system (BMS) hardware is obtained from Coulomb counting, open-circuit voltage (OCV) after rest, and equivalent-circuit model (ECM) observers [4]. Reviews place those observers beside learning-based estimators [5,6,22,23,27,53,54]. Adaptive dual and forgetting-factor Kalman schemes remain strong baselines under dynamic loading [11–14,19].

An extended Kalman filter is not retuned within the factorial; instead, Section 6.6 scores Coulomb counting and the memoryless reduction of (1) and (2) on the locked acceptance flight. It uses the stack's error law and circuit structure as constraints on any input-space synthesizer. Sensor bias and variance propagate into SoC error in closed form [7,18]. Operators that emulate calibration residuals follow from that error law. Arbitrary waveform warping does not.

A first-order Thévenin model commonly used in onboard filters is

$$V_t = \mathrm{OCV}(y_t) - R_0 I_t - v_{p,t}, \tag{1}$$

with polarization dynamics

$$\dot{v}_{p,t} = -\frac{1}{\tau} v_{p,t} + \frac{R_p}{\tau} I_t, \tag{2}$$

where $y_t = SoC(t) \in [0, 1]$ [15–17,55]. Hybrid physics–data estimators attach neural correctors to the same structure [29,56]. An operator that ignores (1) and (2) can therefore regularize a sequence model and, on the same corpus, violate a physics-informed residual.

2.2. Data-Driven SoC Across Laboratory, EV, and UAV Corpora

Deep sequence models and tree ensembles map (V, I, T) windows to SoC on laboratory and electric-vehicle (EV) corpora [20,21,24–26,28,57,58]. The sequence family includes long short-term memory (LSTM), gated recurrent unit (GRU), and Transformer models. Unmanned aerial vehicle (UAV) and urban-air-mobility studies apply related pack-level estimators, unscented Kalman filter (UKF) variants, and logger-bias corrections under high-dynamic profiles [29–35]. Laboratory cells, EV packs, and flight traces are therefore regarded here as alternative scarce-label corpora for the same lithium-ion SOC estimation problem. Transfer-learning surveys document domain shift across temperature, aging, and pack inconsistency [36–38]. Those corpora therefore require structured enrichment of the training set before a field root-mean-square error (RMSE) is interpreted as a pack property.

Architecture studies vary the network on a fixed corpus and report RMSE. Labeled volume, synthesizer family, and dose stay fixed. That design cannot map which transform family to enable, at what intensity, for which estimator class, or after how much real labeled volume. Data-pipeline studies fix the training recipe and the acceptance trajectory, and then vary those factors. The twelve-estimator portfolio is a pipeline factor under that frozen recipe and locked trajectory.

2.3. Augmentation for Battery State Estimation

Generative enrichment for SoC and state of health (SoH) includes TS-DCGAN [41], conditional GAN [42], time-series WGAN [43], TimeGAN and its variants [46–48,59], Gaussian schedules with TimesNet [50], and synthetic dataset generation for battery time series [49]. Reviews attribute substantial error reductions to generative enrichment under data shortage [44,45,52]. Classical time-series augmentation (noise, jitter, warping) supplies the non-generative baselines [39,40,60].

The published template trains one synthesizer for one learner. Two pipeline quantities then remain unmeasured. Dose, or intensity, is rarely swept as a process parameter; a single enriched training set is published instead. Downstream evaluation is locked to one network family, so synthesizer quality is confounded with estimator inductive bias. Without an L schedule and without a learner portfolio, an RMSE change cannot be attributed to the operator.

2.4. Gap Analysis and Positioning

The reported gain may be a property of the synthesizer, of the learner, of the dose, or of how much real labeled data is already available. Across laboratory cells, EV cycles, SoH aging sets, and UAV packs, the published template remains one synthesizer × one learner: TimeGAN with a BERT-based EV predictor [48], TS-DCGAN [41], conditional GAN [42], Gaussian schedules with TimesNet [50], and generative synthesis of battery series [49]. Reviews that attribute large error reductions to generative enrichment under shortage inherit that single-estimator protocol [44,45].

Table 1 positions studies by four experimental axes rather than by application domain. The first two axes are the gaps above: (i) dose L swept as a process parameter, and (ii) a multi-estimator portfolio (trees, networks, physics-informed neural network (PINN)). The remaining two complete the pipeline grid: (iii) labeled volume n as a factor, and (iv) classical, physics-informed, and generative operators under one acceptance trajectory. Generative papers occupy the EV and laboratory-cell rows [48,49]. UAV estimators occupy the scarce-flight rows without an augmentation grid [29,30]. The single-estimator template leaves those four axes unmeasured. Dual stand–logger UAV flights supply the labeled corpus of the factorial.

Table 1. Positioning of this method × model SOC-augmentation factorial. Domain names (lab/EV/UAV) denote the labeled corpus used in each study. Aug.: operator family; #est.: number of downstream estimators; L: dose sweep; n: labeled-volume factor (– = not reported or not applicable).

Study	Domain	Data	Aug.	L	#est.	n	Gap vs. This Work
Meng [4]	EV/BMS	online V, I, T	–	–	survey	–	no synthesizer map
Mohanty [48]	EV	lab/EV cycles	TimeGAN	–	1	scarce	one learner; no L/n
Channegowda [49]	lab	synthetic series	DL gen.	–	–	–	not SoC factorial
Hu [41]	lab	cell cycles	TS-DCGAN	–	few	–	no dose/volume axes
Zhang [50]	lab	drive cycles	Gaussian	weak	1	–	single model family
Echabarri [46]	cells	aging/SoH	TimeGAN	–	1	split	SoH; one learner
Cao [29]	UAV	flight/pack	–	–	hybrid	scarce	no aug. grid
Shibl [30]	UAV	UAV BMS	–	–	few	–	no synthesizer study
Seol [47]	lab	aging	TimeGAN	–	few	–	SoH focus
Qiu [42]	lab	cell cycles	cGAN	–	few	–	no multi-estimator/L
This work	**SoC (scarce labels)**	**stand–logger (UAV corpus)**	**8+ ops**	**1–10**	**12**	**1–5**	**method × model map**

3. System Model and Problem Formulation

The estimator maps a causal logger window of current, voltage, and temperature to a remaining-capacity fraction. That fraction is calibrated on the stand channel. Dual stand–logger UAV flights supply the labeled corpus.

3.1. Definition of State of Charge

Let $Q_{\max}$ denote the usable capacity (in ampere-hours) at the present health and temperature operating point. With Coulombic efficiency η_c and current $I(t)$ (amperes),

$$y(t) = SoC(t) = y(t_0) - \frac{1}{3600\, Q_{\max}} \int_{t_0}^{t} \eta_c(\tau)\, I(\tau)\, d\tau. \tag{3}$$

The quantity $y(t)$ is the remaining-capacity fraction on $[0, 1]$ and is the supervision target, not a logger measurement. In discrete time with sampling period Δt,

$$y_k = y_{k-1} - \frac{\eta_c I_k\, \Delta t}{3600\, Q_{\max}}. \tag{4}$$

Reference labels in battery-management practice typically combine (4) with intermittent OCV correction after relaxation [8,10,11]. Stand-channel labels follow that practice: stand current is integrated, OCV-corrected after rest, and aligned to the logger time base.

3.2. Supervised Estimation Problem

Let the measurable logger feature vector be

$$\mathbf{x}_k = \begin{bmatrix} I_k^{\log} & V_k^{\log} & T_k \end{bmatrix}^{\top} \in \mathbb{R}^3. \tag{5}$$

A causal sliding window of length w is defined as

$$\mathbf{X}_k = (\mathbf{x}_{k-w+1}, \ldots, \mathbf{x}_k) \in \mathbb{R}^{3\times w}. \tag{6}$$

The window ends at sample k and contains no future values. A parametric estimator f_θ maps that window to a scalar SoC at the same instant,

$$\hat{y}_k = f_\theta(\mathbf{X}_k). \tag{7}$$

Sequence and convolutional estimators consume $\mathbf{X}_k$ as a $3 \times w$ series. Tree ensembles flatten the same window into a feature vector. No equivalent-circuit state is carried between windows.

Given an augmented training set $\mathcal{D}$, parameters are obtained by empirical risk minimization,

$$\hat{\theta} \in \arg\min_\theta \frac{1}{|\mathcal{D}|} \sum_{(\mathbf{X},y)\in\mathcal{D}} \ell\big(f_\theta(\mathbf{X}), y\big), \tag{8}$$

where ℓ denotes a squared or Huber loss. On a locked hold-out flight $\mathcal{T}$, mean absolute error (MAE) and root-mean-square error (RMSE) are

$$\mathrm{MAE}(\mathcal{T}) = \frac{1}{|\mathcal{T}|} \sum_{k\in\mathcal{T}} |y_k - \hat{y}_k|, \tag{9}$$

$$\mathrm{RMSE}(\mathcal{T}) = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{k\in\mathcal{T}} (y_k - \hat{y}_k)^2}. \tag{10}$$

Both metrics are in fractional SoC units on $[0, 1]$. An RMSE of 0.01 corresponds to one percentage point of SoC. MAE is the mean absolute deviation. RMSE weights large residuals more strongly and is the primary ranking score in later sections. A sustained RMSE of 0.03 on a high-dynamic trajectory is more relevant to pack-energy management than a laboratory RMSE of 0.005 obtained on rest-rich cycles that do not appear in operation. Scores are therefore reported on the locked acceptance flight rather than on reshuffled laboratory folds.

Relative improvement against the matched unaugmented baseline is

$$\Delta\% = 100 \cdot \frac{\mathrm{RMSE}_{L=1} - \mathrm{RMSE}_L}{\mathrm{RMSE}_{L=1}}. \tag{11}$$

A positive $\Delta\%$ is an RMSE improvement: error falls relative to the matched $L = 1$ twin. A negative $\Delta\%$ means that the augmented training set increased the RMSE.

3.3. UAV Stand–Logger Measurement Chain

The corpus comprises seven synchronized flight-stand recordings. Each recording provides stand and logger currents and voltages, temperature, and reference SoC. The UAV pack supplies high-dynamic labeled traces under a dual sensing chain. The stand

approximates a calibrated reference path. The logger approximates an onboard sensing path with scale and bias residuals:

$$V^{\log} = s_V V^{\text{stand}} + b_V, \quad I^{\log} = s_I I^{\text{stand}} + b_I. \tag{12}$$

Even small (s, b) residuals alter the apparent resistance and, therefore, the voltage–SoC map learned by purely data-driven estimators [7,35]. Typical recordings contain in the order of 10^4 samples at $\Delta t = 0.1$ s, pack voltages near 15 V (4S-class operation), and highly dynamic currents. The platform is a DJI-class multirotor with Dongxingwei D2216 1000 KV motors (Dongxingwei, Shenzhen, China), OPTO 30A ESCs (OPTO, Shenzhen, China), and 9.4-inch propellers (DJI, Shenzhen, China). The peak current is about 60 A/890 W. The airframe is rigidly fixed on a test pallet. The pack is a C-NMC 4S1P module with nominal voltage 14.8 V and nameplate capacity 8 Ah. This pack constitutes the labeled corpus; its usable capacity for Coulomb counting is identified in Section 6.6. The logger was built for this stand. It uses a Raspberry Pi Zero 2 W (Raspberry Pi Ltd., Cambridge, UK), an ADS1115 16-bit ADC (Texas Instruments, Dallas, TX, USA), an ACS758 Hall current sensor (Allegro MicroSystems, Manchester, NH, USA), a resistive voltage divider, a DS18B20 temperature sensor (Analog Devices, Wilmington, MA, USA), an LM2596 regulator (Texas Instruments, Dallas, TX, USA), and XT60 connectors (Amass, Changzhou, China). Voltage and current are logged at 10 Hz. Temperature is logged at 1 Hz. The sensor calibration RMSE is below 0.05 V and 0.1 A. Before each flight, the pack was charged on a SKYRC IMAX B6 V2 charger (SkyRC Technology, Shenzhen, China). The constant-current stage was 0.5C (4 A) up to 16.8 V (4.2 V per cell). The constant-voltage stage ended when the current reached 0.05C (0.4 A). The pack then rested for at least 1 h. Flights include take-off, hover, and maneuvering. After landing, the pack rested for at least 2 h. The laboratory reference path is a Neware BTS-60V/100A cycler (Neware Technology, Shenzhen, China). Reference SOC is stand Coulomb counting with OCV correction. The nominal sample step is 0.1 s. Table 2 lists the seven stand-channel flights. The sample count is the number of rows in each recording, and the duration is the span of the time column. Stand current in the table is positive on discharge, as in (4); on flights 1 and 4, its minimum is slightly negative.

Figure 1 shows a single representative recording; the statistics of all seven flights are summarized in Table 2.

Table 2. Seven stand-channel flights. Duration is the span of the time column. Samples is the number of rows in the recording. SoC is the stand reference in percent, as stored. Current is $I = -$Current_A_Stand, positive on discharge as in (4); min and max are the extremes of that series. Temperature is the stand reading. Held out: flight F_1 never enters training. Training: flights F_2, F_3, F_4, F_6, F_7 enter the nested sets. Acceptance: flight F_5 is locked and is never used for fitting.

Flight	Role	t (s)	Samples	SOC (%)		I (A)		T (°C)	
				Min	Max	Min	Max	Min	Max
F_1	held out	1560.0	15,601	20.0	100.2	−0.585	28.89	24.12	26
F_2	training	1149.7	11,498	20.6	98.7	2.415	34.73	23.62	37
F_3	training	1690.0	16,901	20.8	99.5	2.728	59.76	25	38.44
F_4	training	1360.0	13,597	22.2	100.7	−0.318	59.08	23.81	39.88
F_5	acceptance	1150.0	11,500	20.2	99.3	2.601	46.94	24	41.5
F_6	training	1359.9	13,600	19.8	98.4	2.74	50.47	23.31	33.81
F_7	training	1240.0	12,401	20.2	99.1	2.636	32.61	24.38	35.88

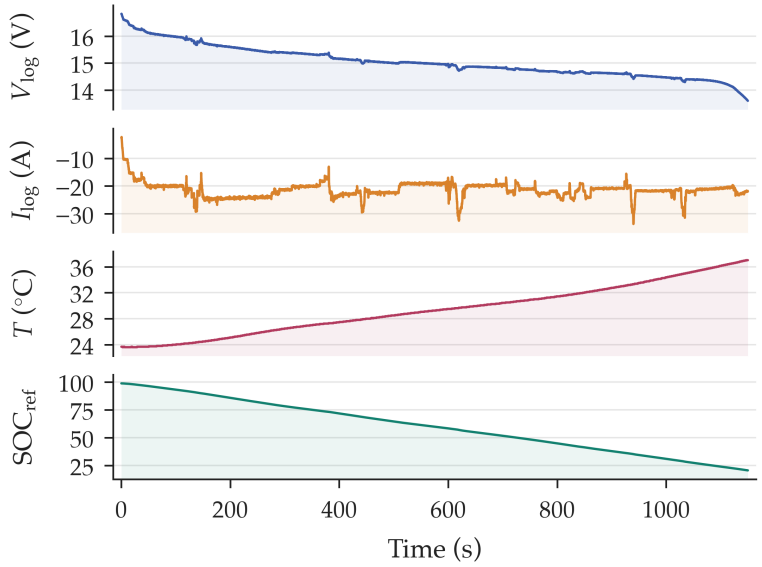

Figure 1. Representative stand–logger trajectory: pack voltage, current, temperature, and reference SOC. A single recording is shown; all seven flights are summarized in Table 2.

Laboratory cell cyclers suppress logger artifacts that a deployed BMS must tolerate. Automotive drive cycles do not reproduce UAV propulsion spectra. Logger-domain features $\mathbf{x}_k$ are the estimator inputs. Stand-calibrated SoC y_k is the supervision signal.

3.4. Real-Flight Volume Factor

Let $\mathcal{R}_n$ denote the nested set of real training recordings, distinct from the sample set $\mathcal{D}$ and from the flight labels F_i. Nested training sets are defined with a frozen acceptance flight:

$$\begin{aligned} n = 1 &: \ \mathcal{R}_1 = \{F_2\}, \\ n = 2 &: \ \mathcal{R}_2 = \{F_2, F_3\}, \\ n = 3 &: \ \mathcal{R}_3 = \{F_2, F_3, F_4\}, \\ n = 4 &: \ \mathcal{R}_4 = \{F_2, F_3, F_4, F_6\}, \\ n = 5 &: \ \mathcal{R}_5 = \{F_2, F_3, F_4, F_6, F_7\}, \end{aligned} \tag{13}$$

while the acceptance set remains $\mathcal{T} = \{F_5\}$ for all experiments. The integer n is the number of real labeled training recordings (flights) rather than the number of samples. The construction is nested: $\mathcal{R}_1 \subset \mathcal{R}_2 \subset \cdots \subset \mathcal{R}_5$. Flights are added in logger index order, skipping F_5 because it is the locked acceptance flight, so the sequence is F_2, F_3, F_4, F_6, F_7; hence, F_7 enters only at $n = 5$. Flight F_1 is withheld from every training set and is the flight used for the separate cross-flight evaluation.

4. Factorial Augmentation Framework

Synthetic traces are generated by maps $\mathcal{T}_{m,\ell}$ under a shared physical-unit protocol and a cumulative concatenation of intensity layers. The experimental factors are the operator m, dose L, estimator f, and labeled volume n. The scaler and every operator see only $\mathcal{R}_n$; the acceptance flight $\mathcal{T}$ is excluded from fitting and from synthesis.

4.1. Cumulative Dose Schedule

Let s be a per-channel min–max scaler fitted on the real training series of $\mathcal{R}_n$ only, and write $\mathbf{x}^{\mathrm{n}} = s(\mathbf{x})$ for the normalized features used by f_θ. Let $\mathcal{D}_{\mathrm{real}}$ denote that scaled real training series (later windowed to length w for sequence models). Dose $L = 1$ uses $\mathcal{D}_{\mathrm{real}}$ exclusively and is the unaugmented baseline of (11).

Each operator m at intensity index ℓ is a map

$$\mathcal{T}_{m,\ell} : (\mathbf{x}, y) \longmapsto (\tilde{\mathbf{x}}, \tilde{y}) \tag{14}$$

applied to the *base* real series, not to previously synthesized layers. The synthetic layer at index ℓ is the image

$$\mathcal{A}_{m,\ell} = \{\mathcal{T}_{m,\ell}(\mathbf{x}, y) : (\mathbf{x}, y) \in \mathcal{D}_{\text{real}}\}. \tag{15}$$

For method m and dose $k \geq 2$, the training set is the cumulative concatenation

$$\mathcal{D}_{L_k}^{(m)} = \mathcal{D}_{\text{real}} \cup \bigcup_{j=2}^{k} \mathcal{A}_{m,j}, \tag{16}$$

as shown in Figure 2. Because each layer is a full-length transform of the same real series,

$$|\mathcal{D}_{L_k}^{(m)}| = k\,|\mathcal{D}_{\text{real}}|. \tag{17}$$

Sliding windows of length w are cut from this concatenation after the layers are stacked. Sequence models therefore receive both real windows and windows that straddle synthetic orderings. Validation traces remain purely real, so early stopping cannot lock onto synthetic artifacts.

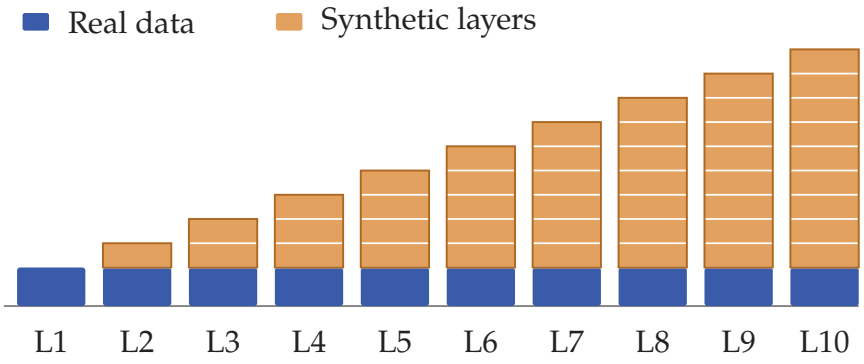

Figure 2. Cumulative training composition versus dose $L = 1 \ldots 10$. Blue: real flight windows; orange: synthetic layers.

All operators except point shuffle act in physical units and return through the same clip–rescale map:

$$\mathbf{x} = s^{-1}(\mathbf{x}^{\text{n}}), \qquad \tilde{\mathbf{x}}^{\text{n}} = s\big(\Pi(\tilde{\mathbf{x}})\big), \tag{18}$$

where Π clips each channel to the stand envelope used in the campaign:

$$V \in [10, 17.6]\ \text{V}, \quad I \in [0, 100]\ \text{A}, \quad T \in [0, 60]\ ^\circ\text{C}. \tag{19}$$

Point shuffle permutes already-scaled samples and, therefore, skips (18). The default label rule is $\tilde{y}_k = y_k$; mixup is the exception that forms a convex combination of two labels.

4.2. Intensity Schedule as a Process Parameter

Dose L is a process parameter of (16): it jointly sets perturbation amplitude and synthetic cardinality. For noise-type operators, channel-wise amplitudes grow approximately linearly with the intensity index ℓ:

$$\alpha_c^{(\ell)} = \alpha_c^{(2)} + \frac{\ell - 2}{8}\big(\alpha_c^{(10)} - \alpha_c^{(2)}\big), \qquad c \in \{V, I\}, \tag{20}$$

while (17) makes $|\mathcal{D}_L|$ grow with L. Segment operators (thermo-voltage, mixup, dropout, RC) use the same index ℓ to raise both the perturbation amplitude and the coverage:

$$\rho^{(\ell)} = |\mathcal{S}^{(\ell)}|/N, \tag{21}$$

the fraction of samples that receive a nonzero edit. L therefore confounds perturbation strength and synthetic sample budget.

The experimental design is a four-way factorial:

$$(m, L, f, n) \in \mathcal{M} \times \{1, \ldots, 10\} \times \mathcal{F} \times \{1, \ldots, 5\}, \tag{22}$$

with response $R(m, L, f, n) = \text{RMSE}$ on the locked acceptance flight. The matched $L = 1$ cell of (11) is the unaugmented reference for every (f, n).

The operators fall into three families (Figure 3): distribution-preserving near-duplicates (thermo-voltage, point shuffle, zero-mean Gaussian), mean-shifting maps (vector shift, Gaussian-plus-bias, channel dropout), and physics or generative maps (RC polarization, TimeGAN, SoC-aware mixup).

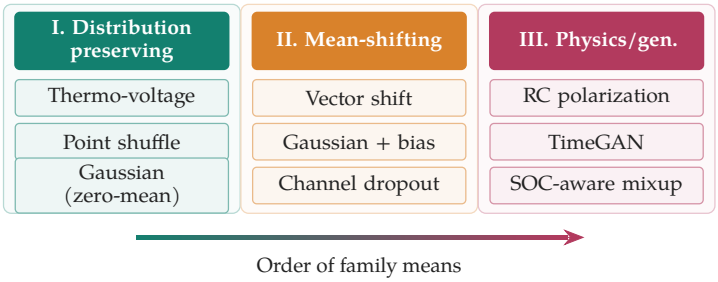

Figure 3. Columns group the operators by family. The arrow orders the three families by the unweighted mean of their three pooled Δ% entries in Section 6.3: distribution-preserving +18.4, mean-shifting +10.9, and physics/generative +9.4. Within a column, the operators are listed without ranking. Channel dropout (+14.2) lies in the mean-shifting column and exceeds every physics/generative operator; the arrow therefore orders family means rather than individual operators. The ranking of individual operators is given in Section 6.3.

4.3. Gaussian Noise and Bias

Zero-mean Gaussian noise is added independently to voltage and current; temperature is left untouched. At dose ℓ,

$$\begin{aligned} \tilde{V}_k &= V_k + \alpha_V^{(\ell)} \sigma_V z_k^V, \\ \tilde{I}_k &= I_k + \alpha_I^{(\ell)} \sigma_I z_k^I, \end{aligned} \qquad z_k^V, z_k^I \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \tag{23}$$

where σ_V and σ_I are the training-set standard deviations of the physical channels and $\tilde{y}_k = y_k$. The schedule (20) runs α_V from 3×10^{-4} at $L = 2$ to 2.7×10^{-3} at $L = 10$, and α_I from 5×10^{-3} to 4.5×10^{-2}. The map injects logger-scale jitter and leaves the mean of (V, I) unchanged.

Gaussian-plus-bias applies the same noise and then a deterministic offset that mimics a calibration residual [7,18]:

$$\tilde{V}_k \leftarrow \tilde{V}_k + \beta_V^{(\ell)}, \qquad \tilde{I}_k \leftarrow \tilde{I}_k + \beta_I^{(\ell)}, \tag{24}$$

with

$$\beta_V^{(\ell)} = \frac{p_V^{(\ell)}}{100} \bar{V}, \qquad \beta_I^{(\ell)} = \frac{p_I^{(\ell)}}{100} \bar{I} \tag{25}$$

and $(\bar{V}, \bar{I})$ the campaign means (15 V, 18.6 A). The percentages $p_V^{(\ell)}$ and $p_I^{(\ell)}$ alternate in sign along L, so successive layers displace the empirical mean in opposite directions.

4.4. Vector Shift

A discrete grid of physical offsets is applied uniformly to every sample of the base series:

$$\tilde{\mathbf{x}}_k = \mathbf{x}_k + \boldsymbol{\delta}^{(\ell)}, \qquad \boldsymbol{\delta}^{(\ell)} = \begin{bmatrix} \Delta I^{(\ell)} & \Delta V^{(\ell)} & \Delta T^{(\ell)} \end{bmatrix}^\top. \tag{26}$$

$L = 1$ is the zero offset; $L = 2, \dots, 10$ index nine grid nodes with $|\Delta V| = 75\,\mathrm{mV}$, $|\Delta I| \approx 0.09\,\mathrm{A}$, and $\Delta T \in \{-3, 0, +3\}\,\mathrm{K}$. The waveform is copied; only the empirical mean of (I, V, T) moves. Labels are copied. Unlike (23), the map injects no additional variance.

4.5. Point Shuffle

Let $b_\ell \in \{5, 10, \dots, 50\}$ be a block length that grows with dose. The integer index set $\{1, \dots, N\}$ is partitioned into contiguous blocks of length b_ℓ; inside each block, a uniform random permutation π_ℓ is drawn and applied jointly to features and labels:

$$(\tilde{\mathbf{x}}_k, \tilde{y}_k) = \left(\mathbf{x}_{\pi_\ell(k)}, y_{\pi_\ell(k)}\right). \tag{27}$$

Every pair $(\mathbf{x}, y)$ is therefore preserved, and the one-dimensional marginals of (I, V, T, y) are unchanged. The map edits short-horizon temporal order only. Sequence models that read windows of length w then see locally scrambled dynamics; tree ensembles that flatten the same window see a permutation of already-used coordinates.

4.6. SoC-Aware Mixup

Two windows of length $w_{\mathrm{mix}} = 48$ (about 4.8 s at $\Delta t = 0.1\,\mathrm{s}$) are mixed only when they lie in a joint SoC–current neighborhood. For a source window i with mid-window label $\bar{y}^{(i)}$ and mean current $\bar{I}^{(i)}$, admissible partners j satisfy

$$|\bar{y}^{(i)} - \bar{y}^{(j)}| < \varepsilon_y^{(\ell)}, \qquad |\bar{I}^{(i)} - \bar{I}^{(j)}| < \varepsilon_I^{(\ell)} \sigma_I. \tag{28}$$

The tolerances $\varepsilon_y^{(\ell)}$ and $\varepsilon_I^{(\ell)}$ widen with dose (from about 0.02 and 0.12 at $L = 2$ to 0.08 and 0.35 at $L = 10$). A mixing weight is drawn as $\lambda \sim \mathrm{Beta}(2, 2)$ and

$$\begin{aligned} \tilde{\mathbf{X}} &= \lambda \mathbf{X}^{(i)} + (1 - \lambda)\mathbf{X}^{(j)}, \\ \tilde{y} &= \mathrm{clip}_{[0,1]}\left(\lambda y^{(i)} + (1 - \lambda) y^{(j)}\right). \end{aligned} \tag{29}$$

This is the only operator that edits the label. The neighborhood (28) excludes pairs that mix a rest window with a propulsion spike. Coverage $\rho^{(\ell)}$ in (21) grows from about 6% to 28% of the series.

4.7. Channel Dropout

A channel c (voltage, current, or temperature) and a span $\mathcal{S}$ of length at most a dose-dependent fraction of the series are drawn uniformly, and the selected samples are zeroed:

$$\tilde{x}_{k,c} = \begin{cases} 0, & k \in \mathcal{S}, \\ x_{k,c}, & \text{otherwise}, \end{cases} \qquad \tilde{y}_k = y_k. \tag{30}$$

The number of spans grows from 2 at $L = 2$ to 18 at $L = 10$; the maximum span fraction grows from 0.03 to 0.12. The map zeroes a channel on $\mathcal{S}$ while copying the SoC label. Unlike (23), the edit is sparse in time and total in amplitude on the chosen channel.

4.8. Thermo-Voltage Coupling

On randomly placed segments of length 120 samples (12 s), temperature and voltage are moved together along a first-order OCV temperature coefficient,

$$\begin{bmatrix} \tilde{T}_k \\ \tilde{V}_k \end{bmatrix} = \begin{bmatrix} T_k \\ V_k \end{bmatrix} + \delta T \begin{bmatrix} 1 \\ \kappa \end{bmatrix}, \qquad k \in \mathcal{S}, \tag{31}$$

with $\kappa = 2.5\,\mathrm{mV\,K^{-1}}$ and

$$\delta T = s\, \delta T_{\max}^{(\ell)} u, \qquad s \sim \mathrm{Unif}\{-1, +1\}, \quad u \sim \mathrm{Unif}[0.3, 1]. \tag{32}$$

The cap $\delta T_{\max}^{(\ell)}$ grows from 0.6 K at $L = 2$ to 3 K at $L = 10$; coverage $\rho^{(\ell)}$ grows from about 4% to 18%. Current and labels are copied. Only a minority of samples move, and they move by millivolts, so the empirical law of $\mathcal{D}_{\text{real}}$ is nearly preserved.

4.9. RC Polarization Synthesis

A first-order Thévenin polarization voltage is integrated on randomly placed segments and subtracted from terminal voltage [16,17]:

$$\begin{aligned} v_{p,k} &= \mathrm{e}^{-\Delta t/\tau^{(\ell)}} v_{p,k-1} + R_p^{(\ell)}\left(1 - \mathrm{e}^{-\Delta t/\tau^{(\ell)}}\right) I_{k-1}, \\ \tilde{V}_k &= V_k - v_{p,k}, \qquad k \in \mathcal{S}, \end{aligned} \tag{33}$$

with $\Delta t = 0.1$ s. Dose raises both circuit parameters: $R_p^{(\ell)}$ from 6 mΩ to 25 mΩ and $\tau^{(\ell)}$ from 6 s to 30 s; segment coverage grows from about 4% to 16%. Labels and (I, T) are copied. The map writes equivalent-circuit dynamics into V under the concatenation schedule (16).

4.10. TimeGAN Baseline

A TimeGAN-style generative baseline [46,48,59] is included under the same cumulative concatenation at matched L, matched f, and matched n. Let E, R, G, and D denote the embedder, recovery, generator, and discriminator of Yoon et al. Training on windows from $\mathcal{R}_n$ (never $\mathcal{T}$) uses a reconstruction loss on the real window $\mathbf{X}$,

$$\mathcal{L}_{\text{rec}} = \mathbb{E}\left\|\mathbf{X} - R(E(\mathbf{X}))\right\|_2^2, \tag{34}$$

together with the supervised latent loss and an adversarial term on G versus D. Once trained for a given n, the sampler is frozen. Synthetic layers $\mathcal{A}_{\text{TG},j}$ are independent draws from that frozen G, then stacked exactly as in (16). For this baseline, L is a cardinality schedule: how many generated copies are concatenated. Unlike the Gaussian-noise schedule (20), it leaves the generator intensity unchanged.

5. Experimental Setup

The reconstruction protocol for the factorial in (22) fixes the operator m, dose L, estimator f, and labeled volume n. Unmanned aerial vehicle (UAV) stand–logger flights supply the labeled corpus [29,30]. Nested training sets $\mathcal{R}_n$ and the locked acceptance trajectory $\mathcal{T} = \{F_5\}$ (Logger 5) follow (13).

5.1. Protocol

Figure 4 maps nested real flights onto locked Logger

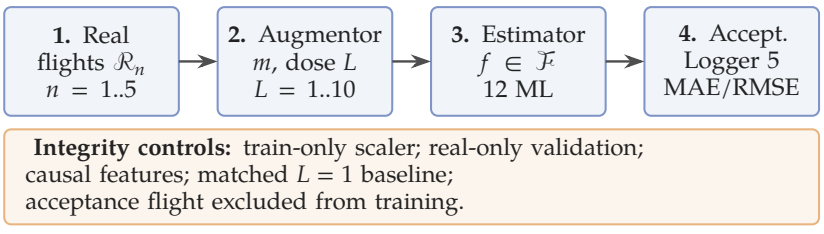

Figure 4. Experimental protocol and integrity controls for the factorial SOC-augmentation grid on the UAV stand–logger corpus.

For each tuple (m, L, f, n), the following steps are executed:

1. Fit a feature scaler on training flights $\mathcal{R}_n$ only.

2. Form a block-wise real validation split (no synthetic samples).
3. Construct $\mathcal{D}_L^{(m)}$ according to (16) when $L \geq 2$.
4. Train estimator f with early stopping on validation RMSE under a shared epoch/batch recipe.
5. Evaluate MAE and RMSE on the locked acceptance flight via (9) and (10).

The classical grid comprises eleven estimators × seven operators × ten doses per n. Extended branches add a physics-informed neural network (PINN), RC polarization, and TimeGAN [48,59]. The experimental corpus exceeds five thousand trained configurations.

Each augmented cell is compared with its matched $L = 1$ twin for the same (f, n) and, where applicable, the same method branch. Relative improvement Δ% uses that twin as the baseline. The sign convention of (11) is that positive Δ% is an RMSE improvement. The physics-informed network is scored as an extended twelfth branch. Method-mean tables and best-(m, L) cells omit a PINN $\min_L$ cell. Seed-sensitive recurrent repeats are reported with the corresponding cells.

5.2. Downstream Estimators

The portfolio contains twelve SoC estimators (Table 3):

- Gradient boosting: XGBoost, LightGBM;
- Feedforward network (FNN) [20];
- Convolutional neural network (CNN) and temporal convolutional network (TCN);
- Hybrid sequence stacks: CNN–LSTM and a distinct hybrid CNN–LSTM model [21];
- Recurrent networks: gated recurrent unit (GRU), long short-term memory (LSTM), and bidirectional LSTM (BiLSTM) [24];
- Nonlinear autoregressive model with exogenous inputs (NARX);
- Physics-informed SoC network as an extended branch [56].

Table 3. Estimator portfolio used in the factorial grid (twelve estimators). Conv.: convolutional; Sys. ID: system identification; PINN: extended branch. Hybrid CNN–LSTM is distinct from CNN–LSTM.

Class	Estimators	Input View
Trees	LightGBM, XGBoost	flat features
FNN	feedforward network	flat features
Conv.	CNN, TCN	temporal
Hybrid	CNN–LSTM; hybrid CNN–LSTM	temporal
RNN	GRU, LSTM, BiLSTM	sequences
Sys. ID	NARX	ARX windows
PINN	physics-informed	residual + data

CNN–LSTM and hybrid CNN–LSTM are two distinct models. The PINN is the twelfth estimator and is not pooled with the eleven classical estimators when method-mean RMSE is reported. Hyperparameters follow a shared early-stopping rule, a train-only scaler, and a capped epoch budget. Cells are not retuned individually.

5.3. Integrity Controls and Acceptance Criterion

Scaler parameters are estimated on the real training flights $\mathcal{R}_n$ only. Logger 5 never enters training or augmentation: $\mathcal{T} = \{F_5\}$ is the locked acceptance flight for every cell. Validation remains purely real, so early stopping cannot exploit synthetic artifacts. Rolling features are causal (no future leakage). An augmented cell is treated as an improvement only if its RMSE on the locked acceptance flight is lower than that of the matched $L = 1$ twin.

5.4. Statistical Reporting

The primary metrics are mean absolute error (MAE) and root-mean-square error (RMSE) in fractional SoC units on $[0,1]$, as defined in (9) and (10). Relative improvement $\Delta\%$ is defined in (11) (positive = RMSE improvement). For variance attribution, let $z = \log \mathrm{RMSE}$, and let F denote a factor. The one-factor explained-variance share is

$$\eta_F^2 = \frac{\mathrm{Var}\big(\mathbb{E}[z \mid F]\big)}{\mathrm{Var}(z)}, \tag{35}$$

estimated empirically over the experimental corpus (>5000 configurations). The shares below are computed on 5130 cells: the 4800-cell classical grid plus 330 TimeGAN cells. A percentile bootstrap with 800 replicates resamples dose–method cells inside each (estimator, n) group. These intervals reflect variation across configurations rather than across random seeds. The 95% intervals are n 51.4–55.4%, estimator 6.7–8.1%, L 2.5–4.1%, and method 1.4–2.8%, around the reported shares of about 53%, 7%, 3%, and 2%. Because each configuration was trained once, seed-level error bars are not available; the band in the dose figure shows the standard deviation across the twelve estimators. Method ranking uses Friedman tests across blocks and pairwise Wilcoxon signed-rank contrasts where applicable. Effect sizes and selection rules accompany the rank tests.

6. Results

The factorial cell is the tuple (m, L, f, n) with response RMSE on the locked acceptance flight (10). Relative improvement $\Delta\%$ is taken against the matched unaugmented twin (11). The next four subsections are C1, C2, C3, and C4, in that order.

6.1. Estimator-Class Conditional Gain

C1: The mean RMSE change versus the matched unaugmented baseline depends on the estimator class. Figure 5 shows the mean $\Delta\%$ of each estimator, averaged over operators, doses, and labeled volumes.

Sequential and convolutional estimators gain +18% to +32% RMSE relative to their matched $L = 1$ baselines. By contrast, the mean is −2.7 for LightGBM, −4.9 for XGBoost, and −22.4 for the PINN. The hybrid CNN–LSTM (+7.3) is the only sequential or convolutional estimator outside the +18 to +32 range. The mean gain is therefore estimator-class conditional (C1): the same synthetic corpus that stabilizes a GRU or TCN can be near-neutral for LightGBM and harmful for a PINN.

Sequence models under scarce flights benefit from mild input-space regularization that enlarges the set of temporally coherent windows without destroying the $(V, I, T) \mapsto y$ map. Tree ensembles already average over partitions of flat features, so near-duplicates add little, and mean-shifting maps can send windows into poorly supported leaves. Physics residuals that couple (V, I, T, y) are disrupted when the inputs are warped without co-designing the residual [56].

Figure 6 repeats the split at operator resolution ($n = 1$, mean $\Delta\%$ over $L \geq 2$). Sequential and convolutional rows stay predominantly green on distribution-preserving operators; LightGBM and XGBoost rows are near zero or red on several aggressive transforms. Thermovoltage on a GRU and vector shift on LightGBM are therefore separate configuration choices. C1 Summary: The same synthetic corpus can improve a sequence or convolutional estimator and harm a tree ensemble or a PINN.

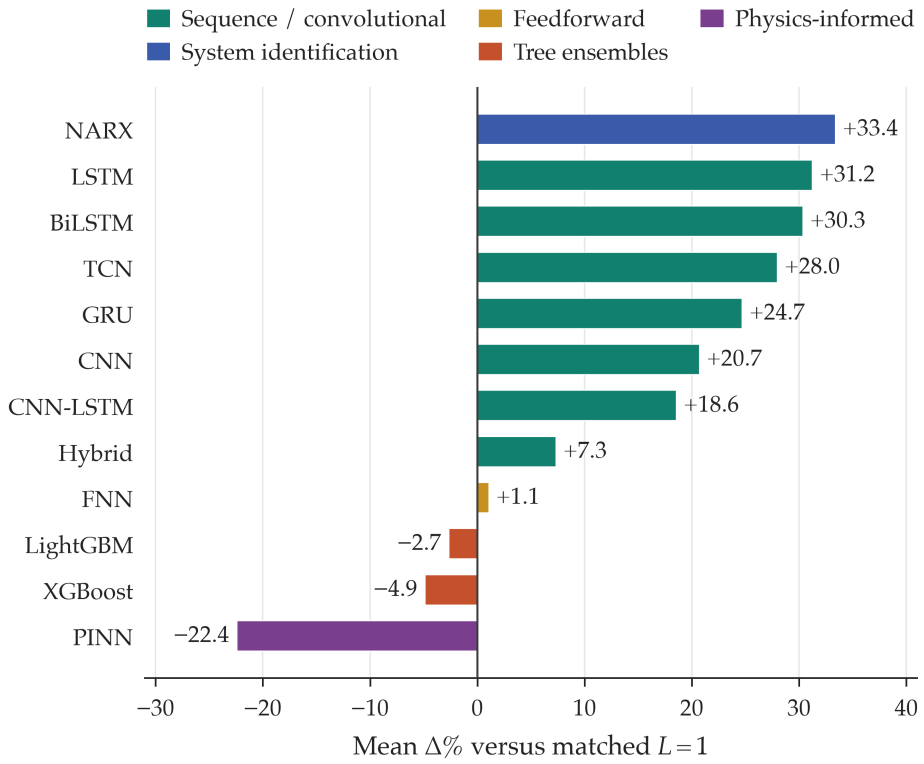

Figure 5. Each bar is the unweighted mean of per-cell Δ% over eight operators (thermo-voltage, point shuffle, Gaussian, Gaussian+bias, vector shift, channel dropout, SOC mixup, and RC polarization), doses $L = 2 \ldots 10$, and $n = 1 \ldots 5$ (360 cells). $L = 1$ is the matched baseline and is not inside the mean. TimeGAN is not included. Positive Δ% is an RMSE reduction. Hybrid denotes hybrid CNN–LSTM; PINN denotes PINN- SOC.

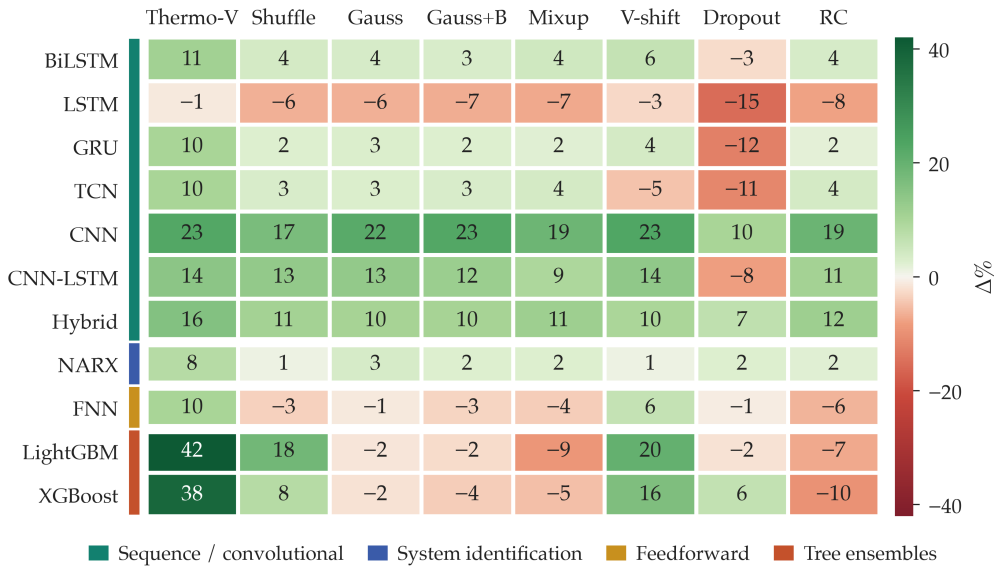

Figure 6. Each integer is the unweighted mean of per-cell Δ% over $L = 2 \ldots 10$ at $n = 1$ (nine doses), rounded to the nearest integer. The map covers the eleven classical estimators and eight operators (the PINN and TimeGAN are excluded); green denotes an RMSE reduction.

6.2. Labeled Volume Dominates Synthesizer Identity

C2: Labeled volume accounts for about half of log RMSE variance, and synthesizer identity for about two percent. Figure 7 reports one-factor η^2 shares of log RMSE from (35) on the experimental corpus: $\eta^2_n \approx 53\%$, $\eta^2_{\text{model}} \approx 7\%$, $\eta^2_L \approx 3\%$, and $\eta^2_{\text{method}} \approx 2\%$. The joint

share of estimator and n exceeds 68%. Figure 8 shows unaugmented acceptance RMSE against labeled volume n. The RMSE does not decrease monotonically for every estimator; nevertheless, labeled volume remains the largest one-factor share of log RMSE, and its effect is distinct from that of the synthesizer.

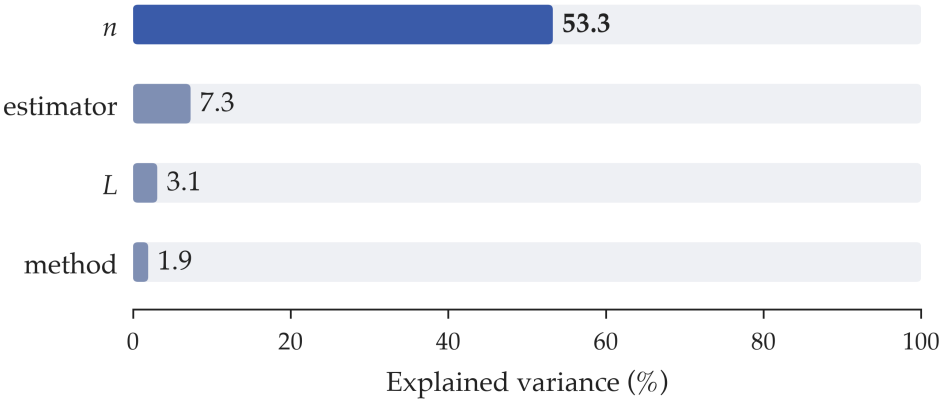

Figure 7. One-factor shares of log RMSE on 5130 cells; bootstrap confidence intervals are given in the Statistical Reporting subsection.

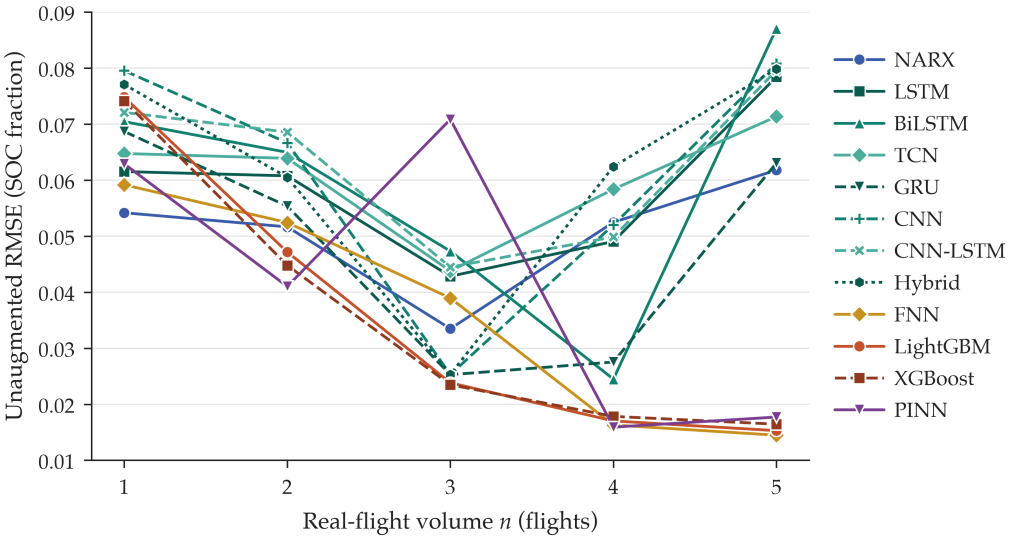

Figure 8. Unaugmented RMSE on the acceptance flight versus n; each point corresponds to a single training run.

Labeled volume ($n \approx 53\%$) outranks method identity ($\approx 2\%$) in the log RMSE partition (C2). Mean $\Delta\%$ versus $L = 1$ still rises with dose, from approximately +4% at $L = 2$ to approximately +18% at $L = 10$, and about 74% of cells improve monotonically in L (the Spearman association between L and gain is positive but modest). Synthesis therefore acts as a sample-budget and regularization term under a capped epoch recipe (16); it does not replace an additional calibrated flight.

The volume axis is not i.i.d. expansion of one law. Only three of twelve estimators (LightGBM, XGBoost, FNN) improve monotonically as n grows from 1 to 5 at $L = 1$; most sequence models show non-monotonic unaugmented RMSE when Loggers 6–7 enter training. At the same time, corpus-mean augmentation $\Delta\%$ rises from roughly +2% at $n = 1$ to more than +30% at $n = 5$. When added flights shift the mixture relative to the acceptance logger, mild synthesizers densify the older mode and reduce variance; they do not create missing physics. The correlation between poor unaugmented baselines and large augmentation gains is positive ($r \approx 0.5$), which matches that reading.

Method identity is not stable in n. Only 25% of estimators keep the same best-performing method from $n = 1$ to $n = 5$; the mean within-estimator Kendall τ of method ranks is approximately 0.03 (near chance). The share of estimators for which thermo-voltage is the best operator falls from 75% at $n = 1$ to 25% at $n = 5$. Vector shift drops from a low-n contender to last place, while channel dropout rises. Fixing each model's $n = 1$ best operator and re-scoring at $n = 5$ costs about 13 percentage points of mean relative gain. A synthesizer profile is therefore keyed to (estimator class, n). C2 Summary: an additional labeled flight dominates exchanging generators, and the operator that wins at $n = 1$ need not win at $n = 5$.

6.3. Ranking Tracks Distribution Preservation

C3: The pooled operator ranking follows how closely the operator preserves the empirical training law. Figure 9 ranks methods by mean Δ% versus $L = 1$. Thermo-voltage leads (approximately +22.5%), followed by point shuffle ($\approx$ +18%) and the noise family; mean-shifting vector shift and TimeGAN trail ($\approx$ +6% to +8%). Thermo-voltage moves only a minority of samples by coordinated millivolt/Kelvin edits (31); point shuffle permutes order and leaves values unchanged (27). Pooled ranking therefore tracks how closely an operator preserves the empirical law of $\mathcal{D}_{\text{real}}$ (C3). Table 4 reports, at $n = 1$, the unweighted mean RMSE over $L = 1 \ldots 10$ (eleven classical estimators) and the relative change in that mean from $L = 1$ to $L = 10$. Table 5 isolates the first synthetic layer, where cardinality (17) has not yet accumulated. At $L = 2$, point shuffle leads (+12.4%), TimeGAN is essentially neutral (+0.1%), and channel dropout is already harmful (−4.7%).

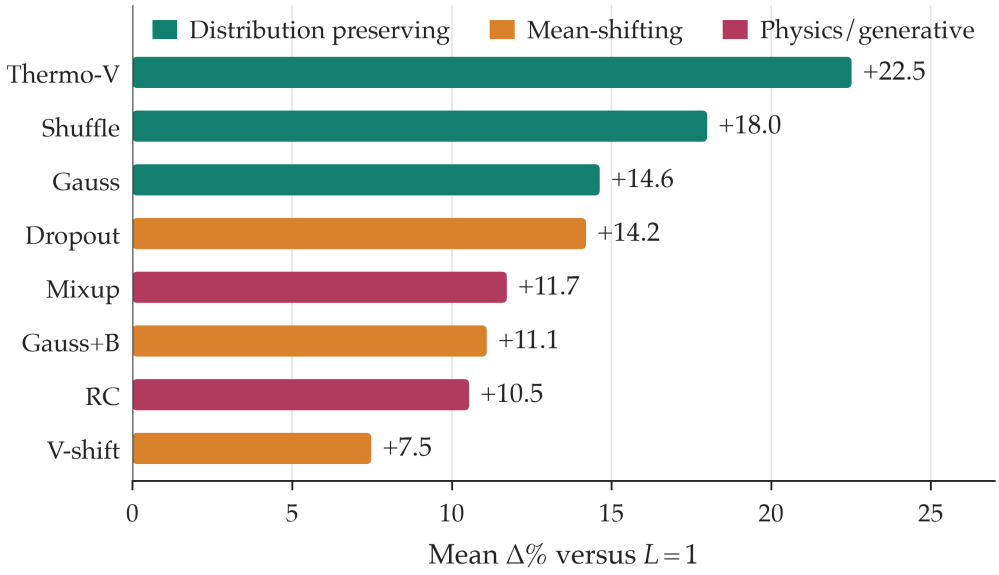

Figure 9. Each bar is the pooled mean Δ% of Table 6: the unweighted mean over $L = 2 \ldots 10$, twelve estimators, and $n = 1 \ldots 5$ (540 cells). TimeGAN is not shown; its table entry (+5.9) is averaged over a smaller subset: eleven estimators and $n = 1, 2, 3$ (297 cells).

Pairwise signed-rank tests prefer thermo-voltage over the remaining operators; TimeGAN is distinguishable mainly from thermo-voltage, not from every classical baseline. On the eight classical operators, Friedman tests of method ranks inside blocks (estimator, L, n) for $L \geq 2$ use 540 blocks and $k = 8$ methods. The statistic is $\chi^2(7) = 218.3$, $p = 1.5 \times 10^{-43}$, and Kendall $W = 0.058$. The small W matches C2: method identity is a detectable but small share of log RMSE. Thermo-voltage has the best mean rank, at 3.29, where rank 1 is the lowest RMSE. Wilcoxon signed-rank tests compare its RMSE with each of the other seven operators on the same 540 blocks. Every Holm-adjusted p is below 10^{-9}. The rank-

biserial correlations run from +0.22 against point shuffle to +0.45 against SOC-aware mixup. TimeGAN is excluded from the Friedman test because it was trained only for eleven estimators and $n \leq 3$. The Wilcoxon tests use the 540 blocks as paired samples.

Table 4. Mean RMSE is the unweighted mean of cell RMSE over $L = 1 \ldots 10$ and the eleven classical estimators at $n = 1$ (110 cells per operator). The Δ% column is the relative change in that eleven-estimator mean RMSE at $L = 10$ versus the same mean at $L = 1$; it is computed from the averaged RMSE rather than by averaging the eleven per-estimator percentages. PINN is the twelfth estimator (Figure 5) and is omitted here: it is an extended branch with negative mean Δ%. Results for RC polarization and TimeGAN are reported in Tables 5 and 6. Δ% follows (11) (positive = RMSE improvement).

Method	Mean RMSE	Δ% at $L = 10$
Thermo-voltage	0.058	+28.1
Vector shift	0.063	+10.0
Point shuffle	0.065	+7.4
Gaussian	0.066	+7.5
Gaussian+bias	0.066	+5.6
SoC mixup	0.067	+1.6
Channel dropout	0.070	+8.3

Table 5. Unweighted mean of per-cell Δ% at $L = 2$ versus the matched $L = 1$ cell. For every operator except TimeGAN the mean uses twelve estimators, including the PINN, and $n = 1 \ldots 5$ (60 cells). Gaussian and Gaussian+bias have equal means (+3.6). TimeGAN uses the same mean over the eleven non-PINN estimators and $n = 1, 2, 3$ only (33 cells); it was not trained at $n = 4$ and $n = 5$; Δ% follows (11) (positive = RMSE improvement).

Method	Mean Δ% at $L = 2$
Point shuffle	+12.4
SoC mixup	+7.7
Thermo-voltage	+5.9
Vector shift	+4.7
Gaussian	+3.6
Gaussian+bias	+3.6
RC polarization	+1.0
TimeGAN	+0.1
Channel dropout	−4.7

Two additional physics operators, tested only in preliminary runs and not in the main grid, point the same way. Ohmic IR-drop that moves terminal voltage more aggressively tends to raise RMSE; Coulomb-consistent current rescaling with relabeled SoC underperforms near-duplicate thermo-voltage on matched tree models. TimeGAN wins no best-of-L cell at $n \leq 2$, is best at $n = 3$ for both tree and both feedforward models, and remains dominated for sequential estimators. That cell does not overturn the pooled ranking; it bounds TimeGAN to intermediate-volume tree/feedforward models.

Table 6 scores the same operators by distribution preservation, physics claim, mean Δ%, and whether it is recommended as a first choice. The highest-gain maps are those that perturb least while expanding $|\mathcal{D}|$ (see also Figure 3).

The order is not the same at every n. At $n = 1$, thermo-voltage leads and vector shift is second; at $n = 3$, TimeGAN leads on this mean; at $n = 5$, thermo-voltage leads again and vector shift is last. This dependence on n reflects the volume effect (C2), while the pooled ranking in Table 6 remains the same. C3 Summary: Thermo-voltage and point shuffle lead that pooled mean, and TimeGAN does not.

Table 6. Compact operator matrix. The $\overline{\Delta\%}$ column is the unweighted mean of per-cell $\Delta\%$ over $L = 2 \dots 10$ (the $L = 1$ baseline is not included). Twelve estimators and $n = 1 \dots 5$ give 540 cells for every operator except TimeGAN. TimeGAN uses eleven estimators, $n = 1, 2, 3$, and $L = 2 \dots 10$ (297 cells). Pres. $\mathbb{P}$: preserves the empirical training law; Phys.: physics-informed construction; 1st?: recommended first-choice synthesizer (✓ = yes; ∼ = partial; – = no). The same mean sliced by n is Table 7.

Operator	Pres. $\mathbb{P}$	Phys.	$\overline{\Delta\%}$	1st?
Thermo-voltage	∼	✓	+22.5	✓
Point shuffle	✓	–	+18.0	✓
Gaussian	∼	–	+14.6	∼
Channel dropout	–	–	+14.2	–
SoC mixup	∼	–	+11.7	∼
Gaussian+bias	–	–	+11.1	∼
RC polarization	∼	✓	+10.5	∼
Vector shift	–	–	+7.5	–
TimeGAN	–	–	+5.9	–

Table 7. Operator mean $\Delta\%$ by labeled volume n, using the pool of Table 6 sliced by n. For each n the classical and RC entries average twelve estimators over $L = 2 \dots 10$ (108 cells). TimeGAN averages eleven estimators over $L = 2 \dots 10$ (99 cells) and was trained only at $n = 1, 2, 3$. The pooled column equals Table 6. An em dash means that the cell was not trained. Green is the largest mean in that column. A pale cell is a negative mean, so RMSE rose.

Operator	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	Pooled
Thermo-voltage	12.7	20.2	17.4	18.0	44.3	22.5
Point shuffle	4.6	10.4	11.7	21.1	42.1	18.0
Gaussian	1.3	7.8	12.5	14.5	37.1	14.6
Channel dropout	−5.6	5.7	12.3	22.3	36.4	14.2
SoC mixup	−1.6	4.4	7.4	15.7	32.7	11.7
Gaussian+bias	−1.4	0.9	10.6	9.0	36.5	11.1
RC polarization	−2.6	0.2	2.2	18.4	34.5	10.5
Vector shift	10.1	7.6	−4.8	3.3	21.1	7.5
TimeGAN	−1.4	−4.5	23.6	–	–	5.9

6.4. Dose Is a Process Parameter

C4: Dose L sets both perturbation amplitude and the number of synthetic copies, so one reported dose does not identify the operator. Dose L jointly raises the perturbation amplitude (20) and synthetic cardinality (17). Figure 10 shows the mean dose–response at $n = 1$.

Per-dose winners move from vector shift at mid L to unanimous thermo-voltage at $L = 8$ on all eleven classical estimators. On trees, Gaussian-type operators can look strong at $L = 2$ and then inflate error toward $L = 10$. Near-duplicate maps enlarge $|\mathcal{D}|$ under a capped epoch budget without breaking the SoC manifold; non-physical dropout already harms at low dose. A single reported L does not summarize the dose curve (C4).

The first synthetic copy is a poor predictor of the high-dose identity. At $L = 2$, point shuffle leads (+12.4%) and channel dropout harms (−4.7%); at $L \in \{8, 9, 10\}$, thermo-voltage leads (+31%) and dropout recovers to second place. Spearman correlation between these regimes is 0.17; only 15% of estimator × n cells keep the same best method. Fixing the $L = 2$ winner costs about 25 percentage points at high dose. On LightGBM/XGBoost, Gaussian noise peaks at $L = 2$ (+22%/+34%) and reverses by $L = 10$, while thermo-voltage improves toward high dose. $L = 2$ helps to detect strongly harmful transforms; it is not used as the final dose.

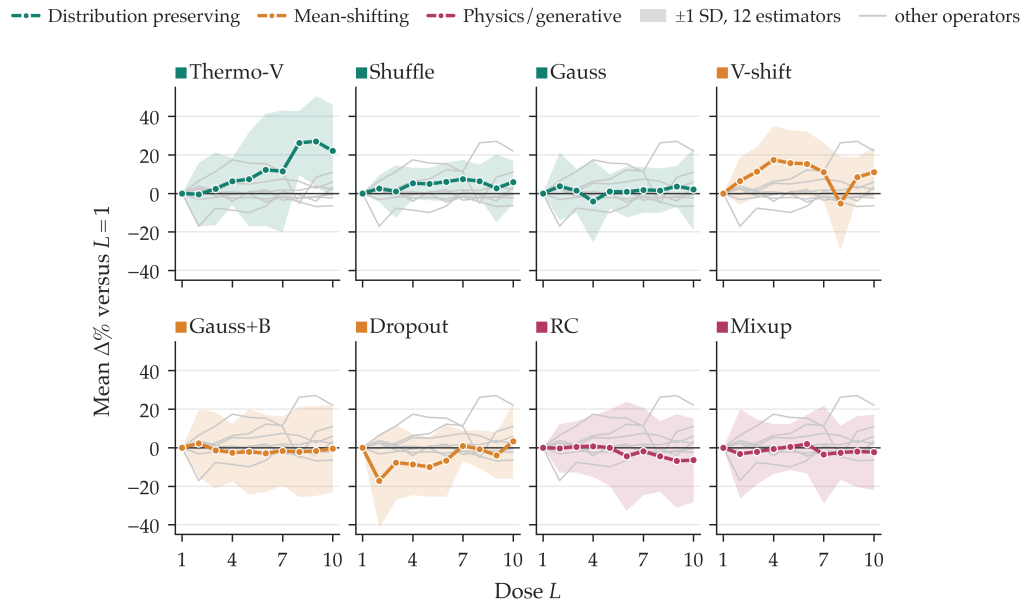

Figure 10. Each point is the unweighted mean, over the twelve estimators including the PINN, of per-estimator Δ% at that dose and at $n = 1$. $L = 1$ is zero by definition, and TimeGAN is not shown. On this mean, thermo-voltage is +26.2 at $L = 8$, +27.0 at $L = 9$, and +22.0 at $L = 10$; vector shift is +17.4 at $L = 4$ and −5.2 at $L = 8$; channel dropout is −17.1 at $L = 2$. The shaded band is plus or minus one standard deviation across the twelve estimators. It therefore describes the spread across estimators rather than a confidence or seed interval.

Overdose past the validation minimizer $L^\star$ is common and class-conditional: about 63% of dose curves peak and then degrade by >5% toward $L_{\max}$ (36% by >20%). PINN and sequential estimators at $n = 5$ are the settings most sensitive to overdose; trees tolerate overdose better.

Reporting $\max_L \Delta\%$ on the acceptance flight inflates the mean reported improvement by approximately 2.5× relative to averaging over L. High-dose optima still exist; a dose $L^\star$ selected on the acceptance flight is not an unbiased estimate of field RMSE. The reporting map is

$$L^{\text{val}} = \arg\min_L \text{RMSE}_{\text{val}}(m, f, n, L), \quad \text{report RMSE}_{\text{acc}}\left(m, f, n, L^{\text{val}}\right). \tag{36}$$

Dose is chosen on real validation; acceptance RMSE is written once. Tables that publish only $\min_L \text{RMSE}_{\text{acc}}$ are optimistic upper bounds. C4 Summary: The final dose is chosen on real validation, and the acceptance flight is used only to report RMSE.

6.5. Absolute Reference Cells

Table 8 reports the best (m, L) pair at $n = 1$ on the locked acceptance flight. LightGBM with thermo-voltage at $L = 10$ reaches RMSE 0.030 (+59% versus its $L = 1$ twin). Convolutional stacks often prefer mid-dose vector shift to maximum thermo-voltage dose.

As n grows, the absolute RMSE of the strong families falls (best LightGBM+thermo-voltage cells from approximately 0.030 at $n = 1$ toward approximately 0.014 at $n = 5$). Extreme minima at intermediate n should be read with seed-sensitivity in mind; an acceptance result should be reported as an interval, not only as the best cell.

Table 8. Best (m, L) at $n = 1$ on the locked acceptance flight (eleven classical estimators). PINN is omitted: a single $\min_L$ cell is not a representative deployment estimate for that class (mean $\Delta\% \approx -22\%$, Figure 5). RMSE in SoC fraction on $[0, 1]$; $\Delta\%$ follows (11) (positive = improvement). TV: thermo-voltage; VS: vector shift; Hybrid: hybrid CNN–LSTM.

Est.	RMSE $L = 1$	Best (m, L)	Best RMSE	$\Delta\%$
LightGBM	0.075	TV, $L = 10$	0.030	+59
XGBoost	0.074	TV, $L = 9$	0.032	+57
CNN	0.080	VS, $L = 6$	0.047	+40
CNN–LSTM	0.072	VS, $L = 6$	0.048	+34
TCN	0.065	TV, $L = 8$	0.044	+32
Hybrid	0.077	TV, $L = 10$	0.054	+30
GRU	0.069	TV, $L = 9$	0.050	+27
BiLSTM	0.071	TV, $L = 9$	0.053	+24
FNN	0.059	TV, $L = 6$	0.045	+24
NARX	0.054	TV, $L = 10$	0.042	+23
LSTM	0.062	VS, $L = 6$	0.052	+16

6.6. Model-Based Reference on the Acceptance Flight

Two references use the fractional RMSE of (10) on F_5. Both the capacity and the open-circuit-voltage curve are estimated without using F_5 or F_1.

On F_2, F_3, F_4, F_6, F_7, the logger current is integrated with the discharge-positive sign of (4). The usable capacity is the least-squares value against the stand label, with the initial state equal to that label at the first sample. The estimate is $Q_{\max} = 8.790$ Ah.

Frozen at that capacity, the same recursion on F_5 starts from the stand label at the first sample and then uses only the logger current. Its RMSE is 3.6×10^{-4}. For comparison, the Coulomb trace recorded with the flight, which is independent of this fit, has an RMSE of 0.0158 against the same stand label.

The polarization voltage in (2) is set to zero for a second reference. The open-circuit-voltage curve is the mean stand open-circuit voltage in equal SoC bins on the same five flights. $R_0 = 0.0307\,\Omega$ is the median of $(\mathrm{OCV} - V)/I$ on those flights for $|I| > 8$ A. Each logger sample on F_5 is inverted through (1). The RMSE is 0.097.

A scalar voltage correction to the Coulomb update was also tuned on F_2; since the training error was minimal at zero gain, no correction was applied on F_5. The process and measurement covariances of an extended Kalman filter were not tuned for this reference.

The LightGBM thermo-voltage cell at $n = 1$ and $L = 10$ has RMSE 0.030 (Table 8). The best learning-based configuration at $n = 1$ has RMSE 0.018. Both lie above the calibrated Coulomb count. The memoryless inversion lies above the unaugmented eleven-estimator mean at $n = 1$, which is 0.069. The learning-based estimator therefore remains a correction on Coulomb counting (38), not a replacement for it.

6.7. Negative Results and Unfavorable Settings

Channel dropout frequently degrades low-dose performance (Table 5). Several LSTM configurations keep $L = 1$ as the optimum for noise-like operators. The physics-informed network loses accuracy under generic input-space distortion on average. Unaugmented $L = 1$ repeats on identical real data can diverge for a minority of recurrent models, so seed intervals should be reported with the results.

6.8. Synthesis

Estimator class, labeled volume n, operator rank, and the L sweep jointly constrain which synthesizer is admissible. Section 8 turns those constraints into a practical selection procedure (Figure 11).

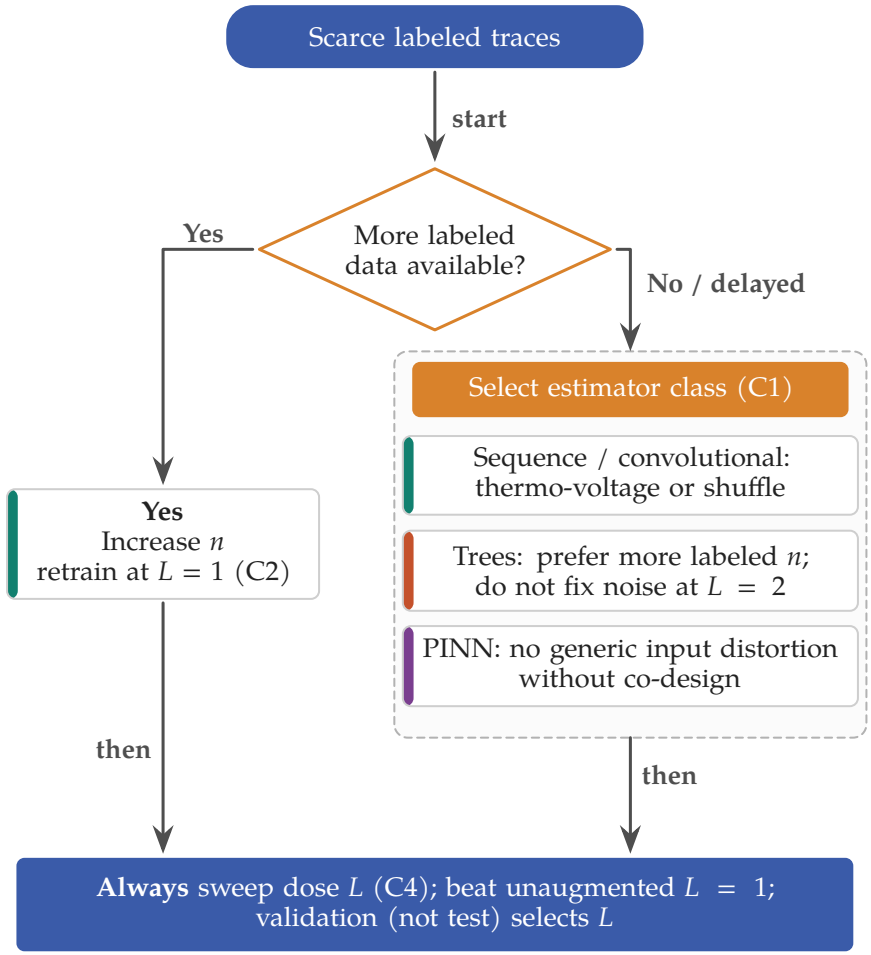

Figure 11. Decision map for synthesizer selection under scarce SOC labels.

7. Discussion

7.1. Mechanistic Interpretation

The method ranking in Section 6 matches regularization under a fixed epoch budget, not information gain. Near-duplicate operators increase $|\mathcal{D}|$ while approximately preserving $\mathbb{P}(\mathbf{x}, y)$. This reduces estimator variance for a capped training schedule. Mean-shifting maps and generative replacement move the probability mass away from the empirical law of the stand–logger pack. They compete with the scarce labeled evidence rather than densifying it.

Point shuffle changes no measured values, yet it ranks near the top, because it densifies already-observed windows. Thermo-voltage succeeds with millivolt/Kelvin-scale coupling applied to a minority of samples. The edit is physically motivated, but the accuracy gain is still that of a mild perturbation that leaves the training law nearly intact. TimeGAN, trained as a learned sampler under the same locked acceptance protocol, underperforms those inexpensive classical operators [46,48,59].

Formally, let $\hat{\mathbb{P}}_L$ denote the empirical law of the dose-L training set. Operators that keep

$$\mathrm{TV}(\hat{\mathbb{P}}_L, \hat{\mathbb{P}}_1) \tag{37}$$

small while increasing $|\mathcal{D}_L|$ behave like bagging or other variance-reduction regularizers. Operators that enlarge total variation without adding new labeled physics tend to act as covariate shift. The observed ranking is consistent with minimizing (37) under a fixed epoch budget, not with maximizing the advertised physical fidelity of the operator. The same confound explains why L is not a purity index of "more physics." Dose jointly raises perturbation amplitude (20) and synthetic cardinality (17), so a high-L improvement can be extra training on near-duplicates rather than a better physical model of the pack.

On this stand–logger corpus and training recipe, distribution-preserving enrichment is the more productive first stage. Generative models remain a valid later stage of battery learning. A generative stack is justified after labeled volume n and mild operators are exhausted, not before [44,49]. That ordering is compatible with generative adversarial network (GAN) and TimeGAN studies that improve a chosen learner on laboratory or electric-vehicle cycles. Those papers test whether a generator can help one network. They do not map which operator a battery management system should enable for which estimator and after how much labeled volume. When $\hat{\mathbb{P}}_1$ itself drifts because new flights enter the training pool, mild synthesizers can temporarily compensate by densifying the older mode of the mixture. They do not create information about the drifted law.

7.2. Comparison with Prior Augmentation Studies

Prior augmentation studies ask whether one generator helps one learner, or whether synthetic series match a battery corpus. This factorial asks which operator helps which estimator class, at which dose, and after how much labeled volume, on one locked acceptance flight.

The contrast is TimeGAN with one EV predictor (Mohanty) [48], TS-DCGAN (Hu) [41], conditional GAN (Qiu) [42], time-series WGAN (Gu) [43], Gaussian schedules with TimesNet (Zhang) [50], and generative synthesis (Channegowda) [49]. SoH targets are a different quantity: Echabarri, Seol, and Yao [46,47,51]. Reviews (Kabir, Sun) inherit the single-estimator protocol [44,45].

On this acceptance flight, the pooled ranking does not put that generator class first. Thermo-voltage (+22.5) and point shuffle (+18.0) lead the 540-cell mean. TimeGAN is +5.9 on eleven estimators at $n = 1, 2, 3$ (297 cells), not on the 540-cell mean, and is confined to tree and feedforward models at intermediate n. Moreover, the same synthetic corpus can help sequence models and harm trees or a PINN (C1), labeled volume outweighs the choice of method (C2), and dose acts as a process parameter (C4); these findings refer to the present stand–logger corpus.

Because laboratory cycles, EV cycles, and aging datasets differ from the present stand–logger pack, gains reported on them are not directly comparable with the ranking obtained here. The evaluation is a supervised window-to-SOC regression: each configuration is scored by RMSE on one acceptance flight against its matched unaugmented twin, without a forecasting horizon or a reward function. In all cases, the learning-based estimator acts as a correction to Coulomb counting rather than a replacement for it.

7.3. Implications for Battery Management Systems

Labeled SoC traces for unmanned aerial vehicle (UAV) packs, and for other energy systems whose reference capacity is expensive to synchronize, require calibrated stands and held-out acceptance flights. Classical synthesizers are cheap once the pipeline exists. The cost asymmetry does not license one operator for every estimator on the pack. The same synthetic corpus that regularizes a sequence or convolutional estimator can be neutral or harmful for a tree ensemble or a physics-informed residual whose constraints are not co-designed with the operator. That class split (C1) follows how each estimator uses (V, I, T):

sequence and convolutional estimators absorb mild near-duplicates; tree partitions and unmatched physics residuals do not. On this sensing chain, labeled volume n outranks synthesizer identity (C2). Additional labeled trajectories outrank exchanging generators. Non-monotonic $L = 1$ baselines after adding loggers are a domain-shift alarm, not proof that more data hurts learning. A synthesizer profile keyed to one n need not survive the next.

Thermo-voltage and point shuffle preserve the empirical training law. Until the gains from these two operators are exhausted, mean-shift and GAN enrichment add covariate shift rather than new pack physics (C3). TimeGAN remains a documented niche for tree and feedforward models at intermediate n. Root-mean-square error (RMSE) versus L should be reported as part of battery management system (BMS) validation [4,12,14]. The dose axis (C4) confounds intensity with cardinality.

Learning-based estimators should correct—not replace— Coulomb counting, open-circuit voltage (OCV) correction, and Kalman-filter estimates [11,29,32]:

$$\hat{y}^{\mathrm{BMS}} = \hat{y}^{\mathrm{filter}} + \mathrm{sat}\Big(g_\phi\big(\mathbf{X}, \hat{y}^{\mathrm{filter}}\big)\Big), \tag{38}$$

where g_ϕ is the learning corrector and $\mathrm{sat}(\cdot)$ limits the magnitude of the correction.

8. Engineering Guidelines for BMS Deployment

Learning-based SoC estimators on this stand–logger corpus are configured in a fixed order. Labeled volume is decided first. The operator is then selected according to estimator class. Dose is swept last, against an unaugmented $L = 1$ baseline (Figure 11).

Labeled SoC traces on this sensing chain require a calibrated stand and a held-out acceptance flight. Figure 7 assigns most residual log RMSE variance to n rather than to the operator (C2). Under this sensing chain, an additional labeled flight reduces residual error more than exchanging generators [44,48].

When a further trajectory cannot be collected, synthesis is applied to the training flights already in hand. The operator is bound to the estimator already chosen for the pack (C1). Sequential and convolutional estimators use thermo-voltage coupling or mild point shuffle at mid-to-high dose. Tree ensembles still gain more from additional real flights, and from a complete dose curve, than from a noise operator frozen at $L = 2$. A physics-informed neural network (PINN) is not subjected to generic input-space distortion unless the residual is co-designed with the operator. Within each class, the order of preference tracks preservation of the empirical training law (C3). Thermo-voltage and point shuffle are used first. Mean-shifting operators and TimeGAN are considered only after the gains from these two operators are exhausted. TimeGAN remains a niche for tree and feedforward models at intermediate n.

Dose L is swept as a process parameter (C4) [4,12]. Validation plots RMSE against L. The first-copy setting $L = 2$ helps to detect strongly harmful transforms. Its winner is a poor proxy for the high-dose identity. Overdose past the validation minimizer $L^\star$ is common. PINN and sequence models under large n require an earlier stop than trees. An augmented estimator is retained only when it beats its matched unaugmented twin on the locked acceptance flight. The index L is selected on real validation (36). The value $\max_L \Delta\%$ selected on the acceptance flight is an upper bound, not a field RMSE. If added flights raise the $L = 1$ baseline on the acceptance logger, the training mixture has drifted relative to $\mathcal{T}$. Mild synthesizers can densify the older mode for a time. The flight mix is reweighted before generative complexity is increased. Operator identity is re-selected for (class, n), not inherited from the $n = 1$ best operator. The learning-based estimator remains a bounded corrector on Coulomb counting, open-circuit voltage (OCV) correction, and Kalman-filter estimates (38). Reported quantities are unaugmented RMSE, validation-

selected-dose RMSE, method, n, seed intervals, and whether L was chosen on validation or on acceptance data.

9. Limitations

The factorial ranking is tied to one stand chemistry and one logger hardware pair. Transfer across packs, temperatures, and sensing chains remains open [36,37].

The rank tests and the η^2 shares are computed on the locked acceptance flight F_5. They do not by themselves measure transfer to another flight. Flight F_1 never enters training. The trained checkpoints are replayed on it, with no refit. The mean RMSE on F_1 falls from 0.192 at $n = 1$ to 0.088 at $n = 5$. At $n = 3$, the mean of the 770 per-configuration RMSE ratios, F_1 over F_5, is 5.24. The ratio of the two mean RMSEs is $0.1532/0.0308 = 4.97$. The Spearman rank correlation of the 770 configurations between F_1 and F_5 is 0.606 at $n = 3$, the highest of $n = 1 \ldots 5$. Only 3 of the 10 best F_5 configurations at $n = 3$ lie in the 50 best F_1 configurations. The grand mean over all replayed flights and all available n does not overturn the pooled method order: thermo-voltage has the lowest mean RMSE, at 0.058. The winning cell does not transfer: at $n = 5$, the lowest F_1 cell is GRU with channel dropout at $L = 8$, RMSE 0.0082. This check uses one withheld flight recorded with the same pack and logger; it therefore does not test transfer to another chemistry, and TimeGAN was not included in the replay.

A minority of recurrent configurations show run-to-run variability on identical real data. Multi-seed intervals should be included in any deployment specification.

Best-cell tables inflate when L is chosen on the locked acceptance flight rather than on real validation (36). BMS dose selection must follow that validation rule. The scalar L confounds perturbation intensity with synthetic cardinality (16), (17), and (20). Each increment both strengthens the transform and concatenates another full-length copy of the training series. Fixed-cardinality intensity sweeps and duplicate-only cardinality controls would orthogonalize those axes before a high-dose win is attributed to amplitude rather than to sample count. Nested flight addition is not guaranteed to be i.i.d. expansion of the same law. Successive flights can mix operating modes. Flight weighting and domain-adversarial checks are future controls on the volume axis.

Stricter physics operators (ohmic drop; Coulomb-consistent current rescaling) need a completed factorial before they can be ranked beyond the present preliminary evidence. NARX models that consume autoregressive SoC context should be read cautiously against open-loop sequence models. They remain a system-identification reference, not a drop-in replacement for a BMS observer.

10. Conclusions

The method × model factorial isolates data-augmentation effects for lithium-ion SoC estimation across twelve estimators. Dual stand–logger unmanned aerial vehicle (UAV) flights supply the labeled corpus. Root-mean-square error (RMSE) is scored on a locked acceptance trajectory for more than five thousand trained configurations. Mean RMSE change versus the matched unaugmented baseline is class-conditional (C1). Sequential and convolutional estimators gain +18% to +32% mean RMSE. LightGBM and XGBoost average ≈−2.4% and −4.9%. A physics-informed network (PINN) loses ≈22%. The same synthetic corpus that regularizes a sequence model can be neutral or harmful for a tree ensemble or a PINN.

One-factor shares of log RMSE variance are $\eta^2_n \approx 53\%$ versus $\eta^2_{\text{method}} \approx 2\%$ (C2). Thermo-voltage and point shuffle outrank mean-shift and TimeGAN (C3) because they stay closer to the empirical training law. Dose L is a process parameter (C4): it jointly sets perturbation amplitude and synthetic cardinality. Thermo-voltage is unanimously optimal at

$L = 8$. Selecting L on the acceptance set inflates reported gain by approximately 2.5× relative to dose-averaged improvement. In deployment, L should be chosen on real validation.

On this sensing chain, the admissible operator is therefore a function of estimator class, labeled volume n, and a validation-selected dose. LightGBM with thermo-voltage at $L = 10$ reaches RMSE 0.030 (+59% versus $L = 1$ at $n = 1$). Figure 11 and the hybrid residual (38) place the selected estimator as a bounded corrector on top of Coulomb counting, OCV correction, and Kalman filtering. Future work could include multi-seed intervals, orthogonalizing intensity from synthetic cardinality, cross-pack transfer, and joint SoC/SoH targets.

Author Contributions: Conceptualization, A.A.D.; methodology, A.A.D.; software, A.A.D.; validation, A.A.D.; formal analysis, A.A.D.; investigation, A.A.D.; resources, A.A.D.; data curation, A.A.D.; writing—original draft preparation, A.A.D.; writing—review and editing, A.A.D. and D.L.; visualization, A.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Science and Higher Education of the Russian Federation (Goszadaniye), No. FSMG-2024-0046.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Supporting numerical tables accompanying this study are provided with the manuscript sources. The UAV stand–logger labeled corpus is available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BMS	Battery management system
CNN	Convolutional neural network
EV	Electric vehicle
FNN	Feedforward network
GAN	Generative adversarial network
GRU	Gated recurrent unit
LSTM	Long short-term memory
MAE	Mean absolute error
NARX	Nonlinear autoregressive model with exogenous inputs
OCV	Open-circuit voltage
PINN	Physics-informed neural network
RMSE	Root-mean-square error
RNN	Recurrent neural network
SOC	State of charge
TCN	Temporal convolutional network
UAV	Unmanned aerial vehicle

References

1. Scrosati, B.; Garche, J. Lithium Batteries: Status, Prospects and Future. *J. Power Sources* **2010**, *195*, 2419–2430. [CrossRef]
2. 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*, 272–288. [CrossRef]
3. Wei, Z.; Zhao, J.; He, H.; Ding, G.; Cui, H.; Liu, L. Future Smart Battery and Management: Advanced Sensing from External to Embedded Multi-Dimensional Measurement. *J. Power Sources* **2021**, *489*, 229462. [CrossRef]
4. Meng, J.; Ricco, M.; Luo, G.; Swierczynski, M.; Stroe, D.I.; Stroe, A.I.; Teodorescu, R. An Overview and Comparison of Online Implementable SOC Estimation Methods for Lithium-Ion Battery. *IEEE Trans. Ind. Appl.* **2018**, *54*, 1583–1591. [CrossRef]

5. Hannan, M.A.; Lipu, M.S.H.; Hussain, A.; Mohamed, A. A Review of Lithium-Ion Battery State of Charge Estimation and Management System in Electric Vehicle Applications: Challenges and Recommendations. *Renew. Sustain. Energy Rev.* **2017**, *78*, 834–854. [CrossRef]
6. Xiong, R.; Cao, J.; Yu, Q.; He, H.; Sun, F. Critical Review on the Battery State of Charge Estimation Methods for Electric Vehicles. *IEEE Access* **2018**, *6*, 1832–1843. [CrossRef]
7. Lin, X. Theoretical Analysis of Battery SOC Estimation Errors Under Sensor Bias and Variance. *IEEE Trans. Ind. Electron.* **2018**, *65*, 7138–7148. [CrossRef]
8. Piller, S.; Perrin, M.; Jossen, A. Methods for State-of-Charge Determination and Their Applications. *J. Power Sources* **2001**, *96*, 113–120. [CrossRef]
9. Movassagh, K.; Raihan, A.; Balasingam, B.; Pattipati, K. A Critical Look at Coulomb Counting Approach for State of Charge Estimation in Batteries. *Energies* **2021**, *14*, 4074. [CrossRef]
10. He, H.; Zhang, X.; Xiong, R.; Xu, Y.; Guo, H. Online Model-Based Estimation of State-of-Charge and Open-Circuit Voltage of Lithium-Ion Batteries in Electric Vehicles. *Energy* **2012**, *39*, 310–318. [CrossRef]
11. Fahmy, H.M.; Swief, R.A.; Hasanien, H.M.; Alharbi, M.; Maldonado, J.L.; Jurado, F. Hybrid State of Charge Estimation of Lithium-Ion Battery Using the Coulomb Counting Method and an Adaptive Unscented Kalman Filter. *Energies* **2023**, *16*, 5558. [CrossRef]
12. Shrivastava, P.; Soon, T.K.; Idris, M.Y.I.; Mekhilef, S.; Adnan, S.B.R.S. Combined State of Charge and State of Energy Estimation of Lithium-Ion Battery Using Dual Forgetting Factor-Based Adaptive Extended Kalman Filter for Electric Vehicle Applications. *IEEE Trans. Veh. Technol.* **2021**, *70*, 1200–1215. [CrossRef]
13. Wadi, A.; Abdel-Hafez, M.; Hussein, A.; Alkhawaja, F. Alleviating Dynamic Model Uncertainty Effects for Improved Battery SOC Estimation of EVs in Highly Dynamic Environments. *IEEE Trans. Veh. Technol.* **2021**, *70*, 6554–6566. [CrossRef]
14. Maletić, F.; Deur, J.; Erceg, I. A Multitimescale Kalman Filter-Based Estimator of Li-Ion Battery Parameters Including Adaptive Coupling of State-of-Charge and Capacity Estimation. *IEEE Trans. Control Syst. Technol.* **2023**, *31*, 692–706. [CrossRef]
15. Chen, M.; Rincon-Mora, G.A. Accurate Electrical Battery Model Capable of Predicting Runtime and I–V Performance. *IEEE Trans. Energy Convers.* **2006**, *21*, 504–511. [CrossRef]
16. Hu, X.; Li, S.; Peng, H. A Comparative Study of Equivalent Circuit Models for Li-Ion Batteries. *J. Power Sources* **2012**, *198*, 359–367. [CrossRef]
17. Naseri, F.; Schaltz, E.; Stroe, D.I.; Gismero, A.; Farjah, E. An Enhanced Equivalent Circuit Model with Real-Time Parameter Identification for Battery State-of-Charge Estimation. *IEEE Trans. Ind. Electron.* **2022**, *69*, 3743–3751. [CrossRef]
18. Zhao, S.; Duncan, S.R.; Howey, D.A. Observability Analysis and State Estimation of Lithium-Ion Batteries in the Presence of Sensor Biases. *IEEE Trans. Control Syst. Technol.* **2017**, *25*, 326–333. [CrossRef]
19. Rout, S.; Das, S. Online State-of-Charge Estimation of Lithium-Ion Battery Using a Fault Tolerant and Noise Immune Threefold Modified Adaptive Extended Kalman Filter. *IEEE Trans. Transp. Electrif.* **2024**, *10*, 9366–9380. [CrossRef]
20. Chemali, E.; Kollmeyer, P.J.; Preindl, M.; Emadi, A. State-of-Charge Estimation of Li-Ion Batteries Using Deep Neural Networks: A Machine Learning Approach. *J. Power Sources* **2018**, *400*, 242–255. [CrossRef]
21. Yang, F.; Song, X.; Xu, F.; Tsui, K.L. State-of-Charge Estimation of Lithium-Ion Batteries via Long Short-Term Memory Network. *IEEE Access* **2019**, *7*, 53792–53799. [CrossRef]
22. Vidal, C.; Malysz, P.; Kollmeyer, P.; Emadi, A. Machine Learning Applied to Electrified Vehicle Battery State of Charge and State of Health Estimation: State-of-the-Art. *IEEE Access* **2020**, *8*, 52796–52814. [CrossRef]
23. How, D.N.T.; Hannan, M.A.; Lipu, M.S.H.; Ker, P.J. State of Charge Estimation for Lithium-Ion Batteries Using Model-Based and Data-Driven Methods: A Review. *IEEE Access* **2019**, *7*, 136116–136136. [CrossRef]
24. Hannan, M.A.; How, D.N.T.; Mansor, M.B.; Hossain Lipu, M.S.; Ker, P.; Muttaqi, K. State-of-Charge Estimation of Li-ion Battery Using Gated Recurrent Unit with One-Cycle Learning Rate Policy. *IEEE Trans. Ind. Appl.* **2021**, *57*, 2964–2971. [CrossRef]
25. Hannan, M.A.; How, D.N.T.; Hossain Lipu, M.S.; Mansor, M.; Ker, P.J.; Dong, Z.Y.; Sahari, K.S.M.; Tiong, S.K.; Muttaqi, K.M.; Indra Mahlia, T.M.; et al. Deep Learning Approach Towards Accurate State of Charge Estimation for Lithium-Ion Batteries Using Self-Supervised Transformer Model. *Sci. Rep.* **2021**, *11*, 19541. [CrossRef] [PubMed]
26. Bao, Z.; Nie, J.; Lin, H.; Gao, K.; He, Z.; Gao, M. TTSNet: State-of-Charge Estimation of Li-Ion Battery in Electrical Vehicles with Temporal Transformer-Based Sequence Network. *IEEE Trans. Veh. Technol.* **2024**, *73*, 7838–7851. [CrossRef]
27. Hossain Lipu, M.S.; Hannan, M.A.; Hussain, A.; Ayob, A.; Saad, M.H.M.; Karim, T.F.; How, D.N.T. Data-Driven State of Charge Estimation of Lithium-Ion Batteries: Algorithms, Implementation Factors, Limitations and Future Trends. *J. Clean. Prod.* **2020**, *277*, 124110. [CrossRef]
28. Sharma, S.; Panigrahi, B.K. Lithium-Ion Battery State-of-Charge and State-of-Energy Simultaneous Estimation via Sparse-Quasi Recurrent Neural Networks (S-QRNN). *IEEE Trans. Ind. Appl.* **2025**, *61*, 774–783. [CrossRef]
29. Cao, X.; Liu, L. A Multi-Timescale Method for State of Charge Estimation for Lithium-Ion Batteries in Electric UAVs Based on Battery Model and Data-Driven Fusion. *Drones* **2025**, *9*, 247. [CrossRef]

30. Shibl, M.M.; Ismail, L.S.; Massoud, A.M. A Machine Learning-Based Battery Management System for State-of-Charge Prediction and State-of-Health Estimation for Unmanned Aerial Vehicles. *J. Energy Storage* **2023**, *66*, 107380. [CrossRef]
31. Monirul, I.M.; Qiu, L.; Ruby, R. Accurate State of Charge Estimation for UAV-Centric Lithium-Ion Batteries Using Customized Unscented Kalman Filter. *J. Energy Storage* **2025**, *107*, 114955. [CrossRef]
32. He, Z.; Martín Gómez, D.; de la Escalera Hueso, A.; Flores Peña, P.; Lu, X.; Armingol Moreno, J.M. Battery-SOC Estimation for Hybrid-Power UAVs Using Fast-OCV Curve with Unscented Kalman Filters. *Sensors* **2023**, *23*, 6429. [CrossRef] [PubMed]
33. Hua, M.; Xu, M.; Li, J.; Zhao, X.; Wang, Q.; Ni, X.; Wu, Y. Temporal-Aware Transformer Networks for State of Charge Multi-Output Prediction in Unmanned Aerial Vehicle Lithium-Ion Batteries. *Sci. Rep.* **2025**, *16*, 2543. [CrossRef] [PubMed]
34. Zhang, H.; Zhou, M.; Lan, X. State of Charge Estimation Algorithm for Unmanned Aerial Vehicle Power-Type Lithium Battery Packs Based on the Extended Kalman Filter. *Energies* **2019**, *12*, 3960. [CrossRef]
35. Yoo, M.Y.; Lee, J.H.; Choi, J.H.; Huh, J.S.; Sung, W. State-of-Charge Estimation of Batteries for Hybrid Urban Air Mobility. *Aerospace* **2023**, *10*, 550. [CrossRef]
36. Shen, L.; Li, J.; Meng, L.; Zhu, L.; Shen, H.T. Transfer Learning-Based State of Charge and State of Health Estimation for Li-Ion Batteries: A Review. *IEEE Trans. Transp. Electrif.* **2024**, *10*, 1465–1481. [CrossRef]
37. Qin, P.; Zhao, L. A Novel Transfer Learning-Based Cell SOC Online Estimation Method for a Battery Pack in Complex Application Conditions. *IEEE Trans. Ind. Electron.* **2024**, *71*, 1606–1615. [CrossRef]
38. Bian, C.; Yang, S.; Miao, Q. Cross-Domain State-of-Charge Estimation of Li-Ion Batteries Based on Deep Transfer Neural Network with Multiscale Distribution Adaptation. *IEEE Trans. Transp. Electrif.* **2021**, *7*, 1260–1270. [CrossRef]
39. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
40. Iwana, B.K.; Uchida, S. An Empirical Survey of Data Augmentation for Time Series Classification with Neural Networks. *PLoS ONE* **2021**, *16*, e0254841. [CrossRef] [PubMed]
41. Hu, C.; Cheng, F.; Zhao, Y.; Guo, S.; Ma, L. State of Charge Estimation for Lithium-Ion Batteries Based on Data Augmentation with Generative Adversarial Network. *J. Energy Storage* **2024**, *80*, 110004. [CrossRef]
42. Qiu, X.; Wang, S.; Chen, K. A Conditional Generative Adversarial Network-Based Synthetic Data Augmentation Technique for Battery State-of-Charge Estimation. *Appl. Soft Comput.* **2023**, *142*, 110281. [CrossRef]
43. Gu, X.; See, K.W.; Liu, Y.; Arshad, B.; Zhao, L.; Wang, Y. A Time-Series Wasserstein GAN Method for State-of-Charge Estimation of Lithium-Ion Batteries. *J. Power Sources* **2023**, *581*, 233472. [CrossRef]
44. Sifat, M.S.I.; Sawda, U.; Kabir, M.A.; Ahmed, M.U.; Lipu, M.S.H.; Islam, M.M.M. Generative Data Augmentation for Improving State Estimation and Prognostics in Lithium-Ion Batteries: Advances, Challenges, and Future Directions. *J. Energy Storage* **2026**, *153*, 121006. [CrossRef]
45. Sun, J.; Gu, A.; Kainz, J. A Solution Framework for the Experimental Data Shortage Problem of Lithium-Ion Batteries: Generative Adversarial Network-Based Data Augmentation for Battery State Estimation. *J. Energy Chem.* **2025**, *103*, 476–497. [CrossRef]
46. Echabarri, S.; Do, P.; Vu, H.C.; Liegeois, P.Y. A Modified TimeGAN-Based Data Augmentation Approach for the State of Health Prediction of Lithium-Ion Batteries. *Reliab. Eng. Syst. Saf.* **2025**, *264*, 111297. [CrossRef]
47. Seol, S.; Lee, J.; Yoon, J.; Kim, B. Improving SOH Estimation for Lithium-Ion Batteries Using TimeGAN. *Mach. Learn. Sci. Technol.* **2023**, *4*, 045007. [CrossRef]
48. Mohanty, P.K.; Jena, P.; Padhy, N.P. TimeGAN-Based Diversified Synthetic Data Generation Following BERT-Based Model for EV Battery SOC Prediction: A State-of-The-Art Approach. *IEEE Trans. Ind. Appl.* **2025**, *61*, 4167–4185. [CrossRef]
49. Channegowda, J.; Maiya, V.; Lingaraj, C. A Deep Learning Approach Towards Generating High-Fidelity Diverse Synthetic Battery Datasets. *IEEE Trans. Ind. Appl.* **2023**, *59*, 4620–4627. [CrossRef]
50. Zhang, C.; Zhang, Y.; Li, Z.; Zhang, Z.; Nazir, M.S.; Peng, T. Enhancing State of Charge and State of Energy Estimation in Lithium-Ion Batteries Based on a TimesNet Model with Gaussian Data Augmentation and Error Correction. *Appl. Energy* **2024**, *359*, 122669. [CrossRef]
51. Yao, W.; Lai, R.; Tian, Y.; Li, X.; Tian, J. State of Health Estimation of Lithium-Ion Batteries Using Data Augmentation and Feature Mapping. *IEEE Trans. Transp. Electrif.* **2025**, *11*, 4895–4905. [CrossRef]
52. Naaz, F.; Herle, A.; Channegowda, J.; Raj, A.; Lakshminarayanan, M. A Generative Adversarial Network-Based Synthetic Data Augmentation Technique for Battery Condition Evaluation. *Int. J. Energy Res.* **2021**, *45*, 19120–19135. [CrossRef]
53. Selvaraj, V.; Vairavasundaram, I. A Comprehensive Review of State of Charge Estimation in Lithium-Ion Batteries Used in Electric Vehicles. *J. Energy Storage* **2023**, *72*, 108777. [CrossRef]
54. Lipu, M.S.H.; Hannan, M.A.; Hussain, A.; Hoque, M.M.; Ker, P.J.; Saad, M.H.M.; Ayob, A. A Review of State of Health and Remaining Useful Life Estimation Methods for Lithium-Ion Battery in Electric Vehicles: Challenges and Recommendations. *J. Clean. Prod.* **2018**, *205*, 115–133. [CrossRef]
55. Xu, Y.; Hu, M.; Fu, C.; Cao, K.; Su, Z.; Yang, Z. State of Charge Estimation for Lithium-Ion Batteries Based on Temperature-Dependent Second-Order RC Model. *Electronics* **2019**, *8*, 1012. [CrossRef]

56. Dong, G.; Gao, G.; Lou, Y.; Yu, J.; Chen, C.; Wei, J. Hybrid Physics and Data-Driven Electrochemical States Estimation for Lithium-Ion Batteries. *IEEE Trans. Energy Convers.* **2024**, *39*, 2689–2700. [CrossRef]
57. Sharma, S.; Panigrahi, B.K. Aging Responsive State of Charge Prediction of Lithium-Ion Battery Using Attention Mechanism Based Convolutional Neural Networks. *IEEE Trans. Ind. Appl.* **2024**, *60*, 7342–7355. [CrossRef]
58. Savargaonkar, M.; Chehade, A.; Hussein, A.A. A Novel Neural Network with Gaussian Process Feedback for Modeling the State-of-Charge of Battery Cells. *IEEE Trans. Ind. Appl.* **2022**, *58*, 5340–5352. [CrossRef]
59. Yoon, J.; Jarrett, D.; van der Schaar, M. Time-series Generative Adversarial Networks. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32, pp. 5508–5518.
60. Pohlmann, S.; Mashayekh, A.; Kuder, M.; Neve, A.; Weyh, T. Data Augmentation and Feature Selection for the Prediction of the State of Charge of Lithium-Ion Batteries From Natural Working Conditions. *Energies* **2023**, *16*, 6750. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.