

A Unified Continuous Time Framework for Adaptive Smartphone Battery Drain Prediction with Finite Horizon Error Analysis

Chuhan Wang

2024110886@stu.hit.edu.cn

Harbin Institute of Technology

Xinxin Li

Harbin Institute of Technology

Yiran Sun

Harbin Institute of Technology

Shunxiang Liu

Harbin Institute of Technology

Haifeng Zhang

Harbin Institute of Technology

Article

Keywords: smartphone battery, continuous-time modeling, adaptive calibration, error propagation, capacity aging, time to empty

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11289998/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Smartphone battery prediction requires a consistent connection between hardware demand, electrical dynamics and capacity aging. We develop a unified continuous-time framework that integrates a second-order equivalent circuit, a temperature-dependent resistance model, empirical aging and adaptive load calibration. A finite-horizon error analysis connects current bias and capacity uncertainty to state-of-charge drift, while exact interval updates provide a reference for numerical verification. Evaluation combines new Android telemetry, 1000 public smartphone workload profiles and five laboratory cell trajectories. On a later ZTE session, static scale calibration reduces power mean absolute error from 1.582 to 0.232 W; a predict-before-update recursive comparator achieves 0.117 W. Across 17 forward temporal blocks, the recursive method yields a pooled error of 0.155 W, with condition-dependent failures. On public Pixel 8 profiles, linear hardware features explain 89.4% of variation in a nominal-voltage power proxy. Dual-exponential capacity extrapolation achieves 0.00651 Ah root-mean-square error on the UCL cell, but offers no general advantage over simpler models on four NASA cells. The results establish a reproducible route from load calibration to charge-balance error assessment and identify the conditions governing model deployment. Complete-discharge validation remains necessary to quantify end-to-end runtime accuracy.

1 Introduction

Smartphone endurance emerges from the interaction of changing hardware demand with the electrical and aging state of the battery. Processor activity, display content and communication alter power demand within seconds, whereas temperature, polarization and capacity loss determine how that demand translates into voltage decline and charge depletion. A useful prediction framework must connect these time scales through a common physical interface and remain adaptable when operating conditions change.

Equivalent circuit models and joint state-parameter estimation provide an efficient representation of battery dynamics [1, 2], while uncertainty in initial state, capacity and measurements remains central to SOC estimation [3]. On the demand side, AndroWatts models device and component consumption from system metrics [4, 5]. EDATA uses current and voltage telemetry for energy debugging [6], and direct power-rail investigations show the importance of distinguishing observed component consumption from indirect estimates [7]. Display and wireless studies further motivate workload-dependent features [8, 9]. These developments create an opportunity to connect software-visible demand with physically consistent battery-state evolution.

Recent smartphone studies establish several complementary routes to battery prediction. Wu et al. [10] formulate continuous state and lifetime prediction from sensor information; Long et al. [11] combine a second-order circuit, modular power demand and temperature dependence for PinePhone scenario analysis. Zheng et al. [12] connect multiple physical effects and capacity aging in SOC estimation, while Xu et al. [13] distinguish instantaneous and delayed drain in device logs. These studies motivate a more

specific question at the interface between workload estimation and battery dynamics: how can a load model be revised as observations arrive, and how can errors in its resulting current input and in available capacity be carried into an interpretable charge-balance assessment? Supplementary Table S2 compares the aims and evidence supporting these related approaches.

We develop an adaptive continuous-time architecture that connects this observation interface to the battery-state equations. Hardware demand is mapped to battery-side power and current, which drive a second-order RC model and charge balance with temperature-dependent resistance and cycle-dependent available capacity. Within this architecture, a finite-horizon error relation separates the contributions of initial-state error, accumulated current error, capacity uncertainty and numerical error. Exact interval updates supply a common reference for checking the state implementation. The contribution is the explicit connection between calibration decisions and their consequences for state prediction: it makes clear which errors online load correction can address, which require capacity or state identification, and which originate in numerical integration.

Our newly collected ZTE phone sessions provide the primary empirical evidence for this architecture's adaptive load interface. We compare fixed calibration with recursive adaptation on a later session and on forward temporal blocks, scoring each recursive prediction before its target is used for updating. Reporting signed bias, time-weighted errors and local failures alongside pooled performance reveals both the benefit and the operating dependence of telemetry feedback. Independent public smartphone profiles and five laboratory cell trajectories provide complementary tests of load representation and capacity extrapolation. A prescribed numerical experiment verifies the state updates and finite-horizon error relation. Together, these elements contribute a reproducible framework and an evaluation design that connects real device observations, module selection and analytical state-error assessment. The empirical tests establish performance at the power and capacity interfaces; complete-discharge SOC and TTE accuracy remains a separate validation endpoint.

2 Principles and Methods

2.1 Battery states and the power interface

Let z denote state of charge as a fraction, I the battery discharge current in amperes, and V_t the terminal voltage in volts. Positive I denotes discharge. Time is expressed in seconds and available capacity Q_{av} in ampere-hours. For an individual prediction interval, Q_{av} is treated as fixed after initialization for the battery condition. The second-order RC representation consists of an ohmic resistance and two polarization branches [1, 2]:

$$V_t = U_{oc}(z, T) - IR_0 - v_1 - v_2$$

$$\frac{dv_j}{dt} = -\frac{v_j}{R_j C_j} + \frac{I}{C_j}, \quad j = 1, 2$$

2

$$\frac{dz}{dt} = -\frac{I}{3600 Q_{av}}$$

3

The factor 3600 converts Ah to ampere-seconds. Equations (1)–(3) use a unit discharge charge-balance factor and a fixed positive available capacity within each prediction interval. If capacity varies continuously and SOC is defined relative to that varying capacity, its derivative must also enter the SOC equation.

The load model supplies battery-side power P_{bat} . If component powers are measured at internal supply rails, their sum and the unmodeled base load must first be converted using an appropriate DC/DC efficiency η_{dc} . If the powers are already on a battery-side basis, no second efficiency correction is applied:

$$P_{bat} = \frac{P_{base} + \sum P_{component}}{\eta_{dc}}, \quad I = \frac{P_{bat}}{V_t}$$

4

Equation (4) closes the interface between the load and battery modules. Empirical evaluation uses observed current to assess this interface; deployment requires a forecast or prescribed scenario for future hardware demand.

2.2 Temperature and capacity aging

A possible resistance correction is written in absolute temperature as:

$$R_j(T) = R_{j,ref} \exp \left[\beta_j \left(\frac{1}{T} - \frac{1}{T_{ref}} \right) \right], \beta_j > 0$$

5

T_{ref} is the calibration temperature and β_j has units of kelvin. This sign yields greater resistance at lower temperature. The correction must be identified over the intended battery and temperature range; the observational temperature strata reported below are not a controlled estimate of β_j .

For capacity aging, we use an empirical dual-exponential relation:

$$Q_{max}(N) = Q_{\infty} + A_f \exp(-a_f N) + A_s \exp(-a_s N)$$

N is cycle number; Q_{∞} , A_f and A_s are in Ah; a_f and a_s are in inverse cycles. Nonnegative amplitudes and positive rates enforce monotonic capacity loss within this model. Capacity retention $Q_{\max}(N)/Q_{\max}(1)$ is distinct from instantaneous SOC. The exponential terms represent different empirical time scales, but their fitted values do not independently identify solid-electrolyte-interphase growth or active-material loss. Battery aging involves multiple interacting mechanisms [14], and those mechanisms require evidence beyond a capacity fit.

2.3 Hardware load modeling and online updates

A component-based load layer may include processor, display, wireless and residual power. The processor contribution can be motivated by dynamic switching power proportional to switching activity, effective capacitance, squared supply voltage and frequency, together with leakage and idle power. An empirical model of frequency alone is an approximation to this relation and must be calibrated over the intended operating range. Display power may depend on brightness and image content [8], while wireless power depends on traffic and radio state [9]. These dependencies motivate feature selection; they do not establish universal coefficients or a causal ranking of drain factors.

The new session archive contains estimates of CPU, GPU, display, Wi-Fi, cellular and GPS power. Their sum is treated as a software estimator output. We assess low-dimensional aggregate calibration and use the richer public benchmark to evaluate hardware features; individual physical component coefficients are not identified from the ZTE archive.

For an identifiable linear regression with output y , feature vector φ and parameter vector θ , forgetting-factor recursive least squares can update parameters using current measurements. With covariance-like matrix P , gain K and forgetting factor $0 < \lambda \leq 1$, the proposed update is:

$$K_k = \frac{P_{k-1} \varphi_k}{\lambda + \varphi_k^T P_{k-1} \varphi_k}$$

7

$$\theta_k = \theta_{k-1} + K_k (y_k - \varphi_k^T \theta_{k-1})$$

8

$$P_k = \frac{1}{\lambda} [P_{k-1} - K_k \varphi_k^T P_{k-1}]$$

9

The dense update requires quadratic time and storage in the parameter count. We evaluate an affine load comparator with $\varphi = [1, x]$, initialized only from earlier observations, and a fixed forgetting factor of

0.995 per sample. Each prediction is recorded before the corresponding telemetry target updates the model. This implementation evaluates adaptive load calibration; it does not identify the RC or aging parameters.

2.4 Time to empty and finite horizon error

Time to empty is a first-passage quantity conditional on a future load trajectory. Starting at t_0 , it is the first time at which a specified SOC or terminal-voltage cutoff is reached:

$$\tau = \inf\{u \geq 0 : z(t_0 + u) \leq z_{cut} \text{ or } V_t(t_0 + u) \leq V_{cut}\}$$

10

An equivalent energy approximation equates the integral of battery-side power over the predicted interval to usable energy. Under constant power, $\tau = E_{usable}/P_{bat}$, giving seconds for joules and watts or hours for Wh and watts. A depletion endpoint, prediction origin and future-load assumption must be specified before TTE can be evaluated.

Continuous-time equations can be advanced by numerical integration or, for constant inputs and parameters over an interval, exact interval updates. For the latter case, each RC state decays exponentially towards R_{jl} , and the SOC decrement equals $I\Delta t/(3600Q_{av})$. Exact interval integration removes truncation error under those assumptions; it does not remove current bias, capacity error, unobserved load changes or model mismatch.

For the charge balance, let e denote estimated minus true SOC, $a = 1/(3600Q_{av})$, and $\hat{a} = 1/(3600\hat{Q}_{av})$. Hats denote estimated quantities; both true and estimated capacities are positive and fixed over the interval. Let $E_{num}(t)$ be a nonnegative bound on the absolute SOC error introduced by numerical solution. The triangle inequality gives:

$$|e(t)| \leq |e(0)| + a \int_0^t |\hat{I}(s) - I(s)| ds + |\hat{a} - a| \int_0^t |\hat{I}(s)| ds + E_{num}(t)$$

11

The integrals extend from 0 to t . The bound is finite for finite horizons and integrable errors. A constant current bias δI causes SOC drift $-\delta I t/(3600Q_{av})$, even with exact integration. Thus, Eq. (11) separates numerical error from the cumulative consequences of uncertain physical inputs and parameters. Section 2.7 verifies these effects against an analytic reference.

2.5 Experimental and public data

The new archive comprises 43,278 unique rows from 29 nonempty acquisitions on one ZTE 7552N phone with a T760 processor and Android 13, recorded during October 3–6, 2026. The 117 CSV files include automatic file rotations, which are joined within their parent acquisition. Raw JSONL logs are aligned one-to-one to CSV rows using nearest timestamps within 100 ms: 43,174 rows match, with a

maximum offset of 12 ms; one malformed raw line is excluded. A duplicated nested export is omitted. The raw BatteryManager current source is used consistently because CSV fallback sources contain unit and source inconsistencies. Android defines current in μA , positive for charging and negative for discharge [15].

Eligible power observations are unplugged, report discharge status, and have negative raw current and positive voltage. The reference is minus raw current divided by 10^6 , multiplied by raw battery-intent voltage in mV divided by 1000. It is an internal telemetry reference in watts. The estimator input x is the logged component-power sum in mW divided by 1000. The early acquisition 20261005_014926 supplies 8755 calibration rows and the later 20261005_143740 supplies 17,635 evaluation rows. Observed adjacent intervals of at most 5 s cover 13.22 h across the archive; acquisition spans total 37.88 h. Gaps are retained, and no shutdown endpoint is inferred from the minimum observed battery level of 2%.

The capacity dataset comprises the 400-cycle UCL INR18650 MJ1 record [16] and all four cells in NASA's BatteryAgingARC-FY08Q4 release: B0005, B0006, B0007 and B0018 [17]. The UCL experiment uses a 24°C chamber, 1.5 A charging to 4.2 V and 4 A discharge to 2.5 V. NASA records contain 636 discharge capacities: 168 each for B0005–B0007 and 132 for B0018, with recorded ambient temperature 24°C . We retain the publisher's Capacity field unchanged; the README defines it for discharge to 2.7 V despite different operating cutoffs. These laboratory cells provide capacity trajectories separately from the phone power measurements. Source conditions and extraction details appear in Supplementary Methods S1.

The AndroWatts public dataset provides 1000 controlled workload profiles on a Pixel 8 running Android 15 [4, 5]. Its source script converts the reported battery discharge-rate column using a fixed 4.4 V. We retain that definition and explicitly call the resulting target a nominal-voltage power proxy. Eleven non-power inputs describe RGB content, brightness, three CPU frequencies, two GPU frequencies, GPU memory and Wi-Fi bytes. Power rails, battery readings and identifiers are excluded from predictors. Source-coded missing values comprise 154 CPU-mid entries and 119 Wi-Fi entries; imputation uses training data only.

2.6 Empirical evaluation design

The original ZTE session split was retained for direct comparison with the initial analysis. Additional candidates and robustness checks were specified before their execution, after that evaluation session had already been inspected; they are exploratory secondary analyses. Six static candidates are compared: the logged sum, training mean, nonnegative scale kx , nonnegative affine $a + bx$, nonnegative CPU regression $a + b(u/100)$, and nonnegative CPU-plus-screen regression $a + b(u/100) + cs$. Here u is CPU utilization in percent and s is the screen-on indicator. All coefficients use training observations only. The scale estimate is:

$$k = \max(0, \frac{\sum x_i y_i}{\sum x_i^2}), \hat{y} = kx$$

The seventh candidate is affine FFRLS using Equations (7)–(9). Initialization uses training normal equations with a numerical ridge of 10^{-6} ; $\lambda = 0.995$ is fixed per observed sample. Predictions are clipped below at zero, while the update uses the ordinary linear innovation. Targets become available only after their current predictions are scored. Static models receive no within-session updates. FFRLS therefore evaluates online calibration with telemetry feedback, rather than an independent forecast of future demand. All seven candidates and their failures are retained.

We report equal-sample and observed-time-weighted errors. A valid time edge joins eligible rows that are adjacent in the original table, belong to the same acquisition, and are separated by at most 5 s. Half its duration weights each endpoint; gaps and excluded rows are not bridged. Minute aggregates require at least 20 rows and 30 s of support within a continuous-segment/minute bin. A forward check uses fixed 30 min blocks with at least 300 test rows and 600 training rows strictly earlier than block start minus 60 s. Each model is refitted or initialized from the past only. Screen-state and fixed temperature strata are descriptive. Full eligibility rules are in Supplementary Methods S2.

For each cell, the earliest floor($0.75n$) discharge observations calibrate the capacity model and the remainder form a chronological test. The five candidates are dual exponential, single exponential, global linear trend, last-value persistence and a recent linear trend using the final $\max(20, \text{ceiling}(0.2n_{\text{cal}}))$ calibration observations. Exponential amplitudes and asymptotes are bounded to $[0,4]$ Ah and rates to $[10^{-6}, 1]$ per cycle. Four single-exponential and 27 dual-exponential starts are compared by calibration error alone. Each cell is fitted separately; this evaluates repeated within-cell extrapolation across sources. The added NASA protocol was specified before evaluating those trajectories. NASA summaries weight the four cells equally and exclude the previously inspected UCL result.

For AndroWatts, five contiguous groups of 200 sorted profile IDs are held out in turn, training on the other 800. IDs have no verified time interpretation, so this is profile-block cross-validation. Fixed candidates are the training mean, CPU-only ordinary least squares (OLS), all-feature OLS and nonnegative regression with an intercept. Training means impute missing inputs and training standard deviations scale features; NNLS inputs are not centered. An additional first-700/final-300 ID split tests sensitivity to grouping. All rows are retained, and no identical full feature rows occur. Coefficients are learned separately from the ZTE experiment, with no claimed parameter transfer.

MAE, RMSE and signed bias use the physical target's units; R^2 uses the evaluation subset's mean as its reference. Capacity-normalized errors divide by the first reported capacity of each cell. Correlated timestamps, profile folds and temporal blocks are not counted as independent devices. We report descriptive dispersion and local failures rather than significance tests based on row-level independence.

2.7 Numerical verification of the state and error equations

We verify the coupled second-order RC and SOC state equations using a fully specified numerical experiment. Illustrative parameters are $Q = 4$ Ah, $z(0) = 0.95$, $R_0 = 0.040$ ohm, $R_1 = 0.012$ ohm, $C_1 = 2500$

F, $R_2 = 0.025\ \Omega$ and $C_2 = 9600\ \text{F}$, giving polarization time constants of 30 and 240 s. Initial polarization is zero, and the observation relation uses $U_{oc}(z) = 3.0 + 1.2z\ \text{V}$. Discharge current alternates between 1.5 and 0.3 A every 300 s for 7200 s. These explicitly illustrative choices define a numerical verification case; they are not fitted phone parameters or additional physical measurements.

The reference solution advances analytically at the known current-switching events and evaluates intermediate times in closed form. Forward Euler, classical RK4 and exact interval updates are tested at all prespecified steps of 20, 10, 5, 2, 1 and 0.5 s. Every step divides the 300-s input intervals, and all RK stages use the current from their own interval. The reported errors are maximum absolute differences at integration nodes. Apparent convergence orders are estimated using the three finest steps. No step or input case is selected according to its observed error.

A separate analytical test applies a constant + 20 mA current bias, an assumed capacity of 3.8 Ah instead of 4 Ah, and their combination. With $a = 1/(3600Q)$, the SOC error is compared with the finite-horizon bound a times the integrated absolute current error plus the absolute error in a times integrated absolute estimated current. The initial state is exact and numerical error is excluded from this propagation test.

2.8 AI assisted preparation and analysis

OpenAI Codex (OpenAI; used on 6–7 October 2026) assisted with manuscript drafting and revision, literature checking, development of research-design and validation options, and the writing, debugging and execution of data-processing, statistical-analysis, numerical-verification and plotting code. The workflow used the supplied phone records and identified public datasets; empirical observations were not generated by an AI model. Numerical scenarios were defined explicitly and are reported separately from observed data. Figures were rendered programmatically from the analytical data or numerical experiments. Separate computational implementations were used to check reported metrics and selected numerical results. The analysis scripts, data-source identifiers and verification records are included in the reproducibility materials. AI assistance does not replace author judgement or responsibility for the scientific content.

3 Results and Discussion

3.1 Adaptive power calibration on the recorded phone

The logged aggregate exhibits a substantial scale mismatch against battery telemetry. Its later-session MAE is 1.5816 W and signed bias is + 1.5815 W. A scale fitted to the earlier session, $k = 0.444327$, reduces MAE to 0.2324 W. FFRLS further reduces contemporaneous MAE to 0.1165 W with a bias of + 0.0129 W. Table 1 compares every specified candidate, and Fig. 1 shows how the evaluation time scale changes the reported errors.

Table 1

Table 1

ZTE later-session and forward-block power errors. Session: 17,635 rows. Forward: 24,636 scored rows in 17 correlated blocks. FFRLS receives each telemetry target only after prediction; static models remain frozen within each evaluated session or block.

Model	Session MAE W	Session RMSE W	Time MAE W	Forward MAE W
Logged sum	1.5816	1.8148	1.6710	1.6867
Training mean	0.8306	0.8874	0.8321	0.4822
Scale NNLS	0.2324	0.3265	0.2573	0.1884
Affine NNLS	0.5276	0.5665	0.5259	0.2113
CPU NNLS	0.3981	0.4772	0.4151	0.2107
CPU + screen NNLS	0.3981	0.4772	0.4151	0.1609
Affine FFRLS	0.1165	0.1948	0.1215	0.1546

Performance depends on both calibration flexibility and operating conditions. A nonnegative intercept increases the static affine MAE to 0.5276 W, while the fitted CPU-plus-screen model assigns zero to the additional screen coefficient and matches CPU-only predictions. This coefficient does not imply zero physical display power. Time weighting raises scale-model MAE to 0.2573 W, compared with 0.1215 W for FFRLS. The available evaluation support is 18,679 s, so these values characterize observed intervals rather than energy use across unobserved gaps.

The forward check yields pooled MAEs of 0.1884 W for scale, 0.1609 W for CPU-plus-screen and 0.1546 W for FFRLS. FFRLS nevertheless has higher MAE than scale in 3 of 17 blocks and higher MAE than CPU-plus-screen in 8; four FFRLS blocks have negative R^2 . Equal-block mean MAEs are 0.2223 W for scale and 0.1876 W for FFRLS, showing the effect of uneven block sizes. The full block comparison is provided in Supplementary Figure S2.

Residual structure also matters. Neighboring scale and FFRLS residuals have correlations of 0.636 and 0.526; after minute aggregation the corresponding values remain 0.607 and 0.559. The $\geq 40^\circ\text{C}$ stratum has 629 observations and scale MAE 0.8172 W, whereas the $< 30^\circ\text{C}$ stratum contains only 36 observations. These strata combine changes in workload, time and temperature. They identify conditions for further investigation without isolating a causal temperature effect.

3.2 Capacity extrapolation across independent cell records

The UCL reference favors the dual-exponential representation: its held-out RMSE is 0.00651 Ah, compared with 0.04100 Ah for a single exponential and 0.01437 Ah for the recent linear trend. Capacity declines from 3.37 to 2.88 Ah across the record. The same fixed evaluation procedure on the four added NASA cells reveals a more varied pattern (Table 2).

Table 2

Table 2

Capacity RMSE in Ah for the fixed chronological split. NASA means weight cells equally. UCL is the previously inspected reference. Each model is calibrated on that cell's early history; these are not unchanged-parameter transfer tests.

Cell and train/test	Dual exp.	Single exp.	Global linear	Persistence	Recent linear
UCL EIL-015 300/100	0.00651	0.04100	0.06025	0.04336	0.01437
B0005 126/42	0.01893	0.01893	0.03449	0.06336	0.02891
B0006 126/42	0.02099	0.02099	0.09652	0.10713	0.03456
B0007 126/42	0.01912	0.01912	0.04027	0.05620	0.01089
B0018 99/33	0.06494	0.06496	0.08828	0.02954	0.04913
NASA mean	0.03099	0.03100	0.06489	0.06406	0.03087

The dual-exponential NASA mean RMSE is 0.03099 Ah, close to 0.03100 Ah for the single exponential and 0.03087 Ah for the recent linear trend. Recent linear extrapolation performs best on B0007, while persistence performs best on B0018. On B0018, the dual-exponential error reaches 0.06494 Ah and R^2 is -3.884 . Thus, increasing aging-model complexity is useful for the UCL trajectory but does not ensure better extrapolation across all observed degradation patterns.

Parameter behavior explains why empirical model selection is necessary. One dual-exponential amplitude approaches zero on B0005–B0007, effectively reducing the fit to a single exponential. The UCL asymptote reaches its lower bound, and its slow time scale of about 4227 cycles extends far beyond the observation window. These fits describe capacity trajectories without separately identifying aging mechanisms or a long-term capacity floor. Supplementary Figure S3 presents every cell trajectory and the overlapping forecasts.

3.3 Public workload benchmark for the hardware layer

On the 1000 AndroWatts profiles, all-feature OLS achieves an out-of-fold MAE of 0.2959 W and R^2 of 0.8935 for the nominal-voltage power proxy. CPU-only OLS yields 0.3469 W MAE, showing an aggregate benefit from the additional hardware features. Nonnegative regression produces 0.4061 W MAE and R^2 of 0.8081 (Table 3 and Fig. 2). The added sign constraints therefore do not improve prediction in this benchmark.

Table 3

Table 3
AndroWatts out-of-fold power-proxy errors

Model	MAE W	RMSE W	Bias W	R ²
Training mean	0.9242	1.2168	0.0000	-0.0012
CPU OLS	0.3469	0.4512	0.0006	0.8623
All-feature OLS	0.2959	0.3968	0.0024	0.8935
Nonnegative hardware model	0.4061	0.5327	0.0664	0.8081

The first-700/final-300 ID sensitivity split gives MAEs of 0.3093 W for all-feature OLS and 0.4110 W for nonnegative regression, preserving this ordering. This public-data result complements the temporal ZTE experiment through a wider range of controlled hardware states. It establishes neither transfer of coefficients across the two phones nor physical identification of each component's contribution. Negative OLS coefficients may reflect correlated inputs and proxy limitations, while nonnegative constraints can reduce flexibility when features incompletely describe device behavior.

3.4 Numerical behavior and finite horizon error propagation

The numerical experiment resolves the expected distinction between state-equation consistency and integration accuracy. Maximum terminal-voltage error decreases from 3.368e-03 to 6.000e-05 V for Euler and from 1.825e-05 to 4.322e-12 V for RK4 as the step decreases from 20 to 0.5 s. The apparent orders are 1.01 and 4.03, respectively. Exact interval updates remain within 3.09e-13 V of the event-level analytical reference. All schemes integrate SOC exactly up to floating-point rounding for this piecewise-constant current, with maximum SOC discrepancy 2.57e-13.

The true charge throughput is 1.8 Ah, taking SOC from 0.95 to 0.50 over two hours. A constant + 20 mA current bias produces 1.00 percentage point of SOC error at two hours despite analytical integration. A 5% underestimated capacity produces 2.37 percentage points, and the combined case produces 3.42 percentage points. The computed errors satisfy the finite-horizon bound throughout the reporting grid. For the chosen same-direction perturbations the bound is tight; its increase with horizon is a feature of accumulated input and parameter uncertainty, not a numerical integration defect.

Together, the convergence and perturbation experiments verify the computational implementation of the unified dynamics and the explicit error-propagation relation. They support using exact interval updates where sampled loads are held constant, while retaining uncertainty terms when the input current or effective capacity is imperfectly known.

3.5 Implications for unified battery prediction

The main empirical contribution comes from the self-collected phone sessions: they show that a component-power estimate can require substantial calibration before it serves as a reliable battery-side

input, and that recursive correction can improve its contemporaneous agreement with telemetry while still failing in particular operating blocks. The predict-before-update design makes this result operationally interpretable because the observation used to correct the model is excluded from that observation's prediction. It evaluates the benefit of sequential telemetry feedback under an explicit information schedule. The public workload and cell records then test whether greater feature or aging-model complexity consistently helps; their mixed results support selecting each module by its own evaluation target and operating conditions.

The analytical contribution connects these calibration choices to charge-balance consequences. Current bias accumulates with the prediction horizon, while capacity error rescales the charge throughput; refining the numerical step cannot eliminate either effect. By separating those terms, the framework identifies distinct interventions: load calibration for input mismatch, capacity identification for charge-scale error, and an appropriate state update for truncation error. This use of established circuit dynamics, recursive estimation and integral inequalities supplies a coherent basis for diagnosing an adaptive predictor. The comparison with related smartphone models [10–13] therefore concerns the combination of an explicit observation-to-state interface, sequential evaluation on newly collected phone data and finite-horizon error accounting, rather than a claim that the individual circuit or estimation methods are new.

The finite-horizon analysis is especially useful because a small instantaneous error can still produce a material endurance error when it persists. Reducing mean power error and bias addresses this pathway, but a TTE prediction also depends on initial charge, usable capacity, cutoff conditions and the future workload. A complete deployment should calibrate the phone's open-circuit-voltage relation and RC parameters, initialize capacity for the device condition, and use Eq. (10) with a stated workload scenario or forecast. The component datasets should inform these modules without being joined as though they were one measured phone trajectory.

The principal empirical limitations are one self-recorded phone, previously inspected session data, internal telemetry as the reference, and separate calibration within each public cell or phone dataset. Brightness in the ZTE archive takes only two values, GPU utilization is unavailable, and several radio-state flags have no variation. The sample-based forgetting factor also depends on acquisition rate. These conditions restrict inference about component coefficients, controlled thermal response and device-population performance. Regression error is not a complete metrological uncertainty budget.

The next validation stage should synchronize external electrical measurements with phone telemetry, repeat controlled workloads on independent devices, and follow prospective predictions to observed discharge endpoints. Separate frozen and adaptive evaluations would determine how much telemetry feedback is required in deployment. Numerical examples establish properties of the specified equations, while complete-discharge experiments are needed to quantify end-to-end SOC and TTE accuracy. This distinction preserves a clear path from the present unified architecture to device-level validation.

4 Conclusion

We develop a unified continuous-time framework connecting smartphone hardware demand, RC battery dynamics, temperature-dependent resistance, capacity aging and online load calibration. Its finite-horizon error analysis makes the influence of current and capacity uncertainty explicit, and its analytic interval updates provide a transparent numerical reference. The architecture supports a coordinated evaluation of model structure, calibration and error propagation.

New phone telemetry demonstrates a reduction in later-session power MAE from 1.582 W to 0.232 W with static scale calibration and to 0.117 W with a predict-before-update recursive comparator. Forward blocks and condition-specific results reveal when this advantage varies. Public workload and multi-cell comparisons show that added features can improve prediction, whereas nonnegative constraints and additional exponential terms require empirical justification. These findings strengthen the framework's use as a reproducible basis for adaptive battery-drain prediction and guide the next stage of prospective device-level validation.

Declarations

Competing interests

The authors declare no competing interests.

Supplementary information

Supplementary Information, Supplementary Data 1 and Supplementary Code 1 accompany this manuscript.

Author Contribution

Chuhan Wang contributed to conceptualization, methodology, formal analysis, software development, and writing the original draft. Xinxin Li contributed to methodology, software development, validation, data curation, visualization, and writing the original draft. Yiran Sun contributed to software development, data curation, validation, visualization, resources, and manuscript review and editing. Shunxiang Liu contributed to conceptualization, methodology, investigation, project administration, and supervision. Haifeng Zhang contributed to supervision, resources, and manuscript review and editing. All authors reviewed and approved the final manuscript.

Acknowledgement

This work was supported by Heilongjiang Province's Key Research and Development Project (2023ZXJ03C02), the Longjiang Science and Technology Outstanding Talent (Innovative Talents) Program (RC2025CX027), and the National Natural Science Foundation of China (52075117).

Data Availability

The field-limited records from the authors' self-collected smartphone experiment supporting the reported results are provided in Supplementary Data 1, including all 43,278 aligned sample positions and 43,174 selected raw electrical records, a data dictionary and provenance summaries. Application and network identifiers and private diagnostic logs are excluded. Supplementary Data 1 also contains the source-derived capacity records and selected AndroWatts profile variables used in the supporting analyses, with source attribution and reuse information. The original public datasets are available from UCL (<https://doi.org/10.5522/04/12159462.v1>), the NASA Prognostics Data Repository (<https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>) and the AndroWatts deposit (<https://doi.org/10.5281/zenodo.14314943>). Designed numerical verification outputs are labeled separately from experimental observations.

Code availability

The analysis, plotting and independent verification code for the reported results is provided in Supplementary Code 1, with fixed protocols, dependency versions and reproduction instructions. Extracting Supplementary Data 1 and Supplementary Code 1 into the same directory reconstructs SR_Reproducibility/; running `python run_all.py` reproduces the analyses and figures offline from the included inputs. A CSV-based capacity-analysis entry preserves the original fitting functions, initializations and chronological splits. The original AndroWatts analysis script is included as read-only source documentation and is not executed by the reproduction workflow. The private acquisition archive and data-collection application source are not included.

References

1. Zhang, L., Peng, H., Ning, Z., Mu, Z. & Sun, C. Comparative research on RC equivalent circuit models for lithium-ion batteries of electric vehicles. *Appl. Sci.* **7**, 1002. <https://doi.org/10.3390/app7101002> (2017).
2. Plett, G. L. Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs. Part 3. State and parameter estimation. *J. Power Sources.* **134**, 277–292. <https://doi.org/10.1016/j.jpowsour.2004.02.033> (2004).
3. 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.* **78**, 834–854.
<https://doi.org/10.1016/j.rser.2017.05.001> (2017).
4. Guegain, E., Raes, R., Chachignot, N., Quinton, C. & Rouvoy, R. AndroWatts: unpacking the power consumption of mobile device's components. Zenodo (version v1) (2024).
<https://doi.org/10.5281/zenodo.14314943>
 5. Guégain, É., Raes, R., Chachignot, N., Quinton, C. & Rouvoy, R. AndroWatts: unpacking the power consumption of mobile device's components. In 2025 IEEE/ACM 12th International Conference on Mobile Software Engineering and Systems (MOBILESoft) 71–81 (IEEE, 2025).
<https://doi.org/10.1109/MOBILESoft66462.2025.00015>
 6. Blokland, E., Cruz, L. & van Deursen, A. EDATA: energy debugging and testing for Android. In 2025 IEEE/ACM 12th International Conference on Mobile Software Engineering and Systems (MOBILESoft) 94–104 (IEEE, 2025). <https://doi.org/10.1109/MOBILESoft66462.2025.00017>
 7. Gupta, A., Heidari, A., Kalipattapu, A., Jain, I. K. & Bharadia, D. 3 W's of smartphone power consumption: who, where and how much is draining my battery? In Proceedings of the 30th Annual International Conference on Mobile Computing and Networking 2248–2250 ACM, (2024).
<https://doi.org/10.1145/3636534.3695905>
 8. Abularach Arnez, J. et al. Current consumption analysis of LCD and AMOLED display technologies: an approach based on multi-text input modalities. *IEEE Lat Am. Trans.* **24**, 84–97.
<https://doi.org/10.1109/TLA.2026.11334047> (2026).
 9. Spachos, P. & Gregori, S. WiFi throughput and power consumption tradeoffs in smartphones. In 2017 24th International Conference on Telecommunications (ICT) 1–5 IEEE, (2017).
<https://doi.org/10.1109/ICT.2017.7998238>
 10. Wu, Z., Dong, W., Yang, Z. & Diao, M. A sensor-aware physics-based framework for continuous smartphone battery state and lifetime prediction. *Sensors* **26**, 4707.
<https://doi.org/10.3390/s26154707> (2026).
 11. Long, Y., Tian, S., Liu, X. & Peng, X. Predicting smartphone battery time-to-empty(TTE) on an open-source PinePhone platform: a modular electro-thermal continuous-time model. *PLOS ONE*. **21**, e0355305. <https://doi.org/10.1371/journal.pone.0355305> (2026).
 12. Zheng, Y., Li, Y., Song, L. & Dai, X. MTEC-SOC: a multi-physics aging-aware model for smartphone battery SOC estimation under diverse user behaviors. *Batteries* **12**, 130.
<https://doi.org/10.3390/batteries12040130> (2026).
 13. Xu, C., Wang, R. & Huang, C. Separating instantaneous and delayed battery-drain effects for smartphone remaining runtime prediction. *Electronics* **15**, 3257.
<https://doi.org/10.3390/electronics15153257> (2026).
 14. Vetter, J. et al. Ageing mechanisms in lithium-ion batteries. *J. Power Sources*. **147**, 269–281.
<https://doi.org/10.1016/j.jpowsour.2005.01.006> (2005).
 15. Android Developers. BatteryManager. Android API reference. October (2026).
<https://developer.android.com/reference/android/os/BatteryManager.html> (accessed 6.

16. Heenan, T. et al. Lithium-ion battery INR18650 MJ1 data: 400 electrochemical cycles (EIL-015). University College London (2020). <https://doi.org/10.5522/04/12159462>

17. Saha, B. & Goebel, K. Battery Data Set. NASA Prognostics Data Repository, NASA Ames Research Center (2007). <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>

Figures

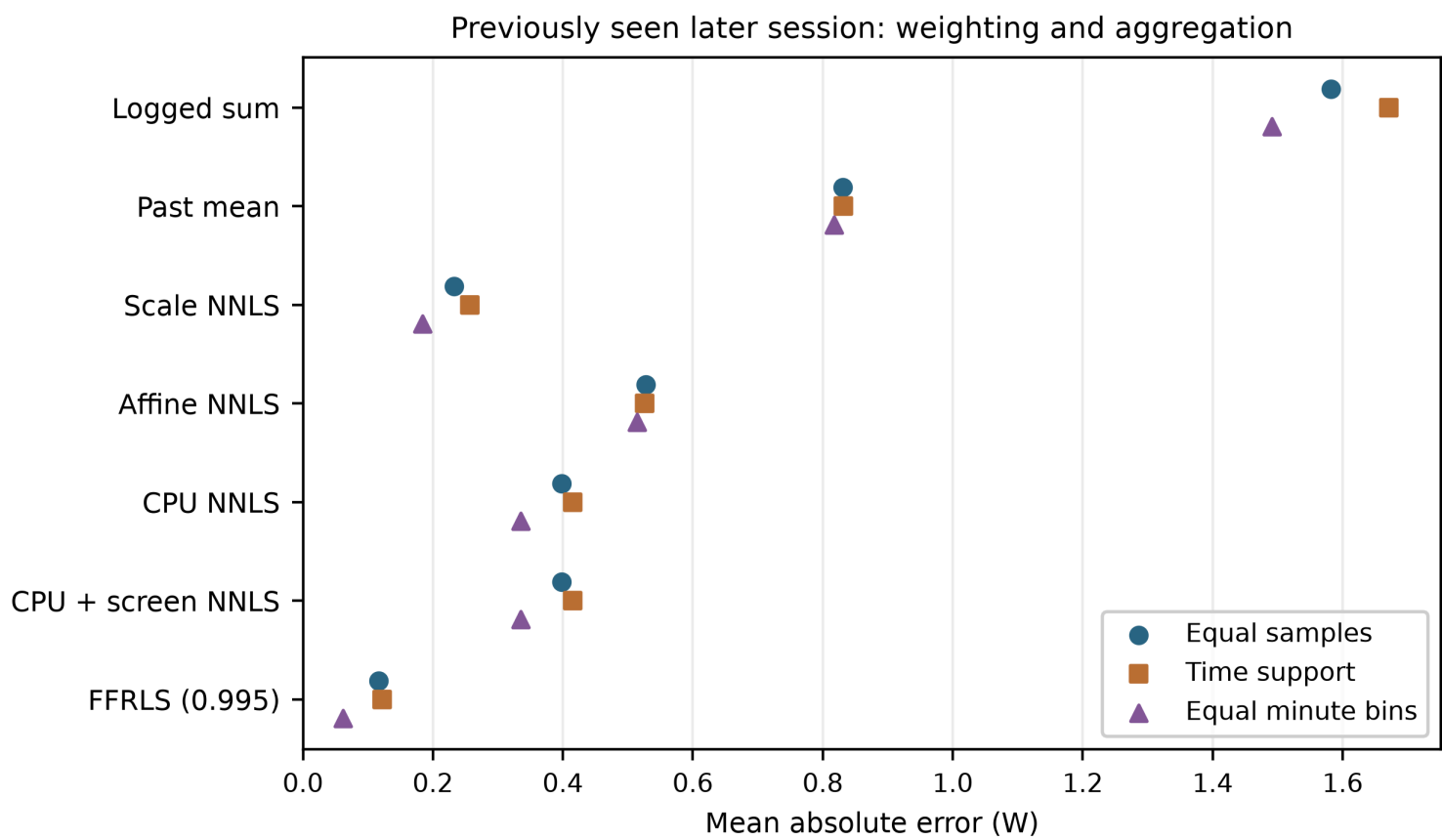


Figure 1

Power error under sample weighting, observed-time weighting and minute aggregation. All seven models are shown. The 297 qualifying minute bins use a coverage-qualified subset and a different temporal resolution; their lower errors do not represent improved instantaneous estimates.

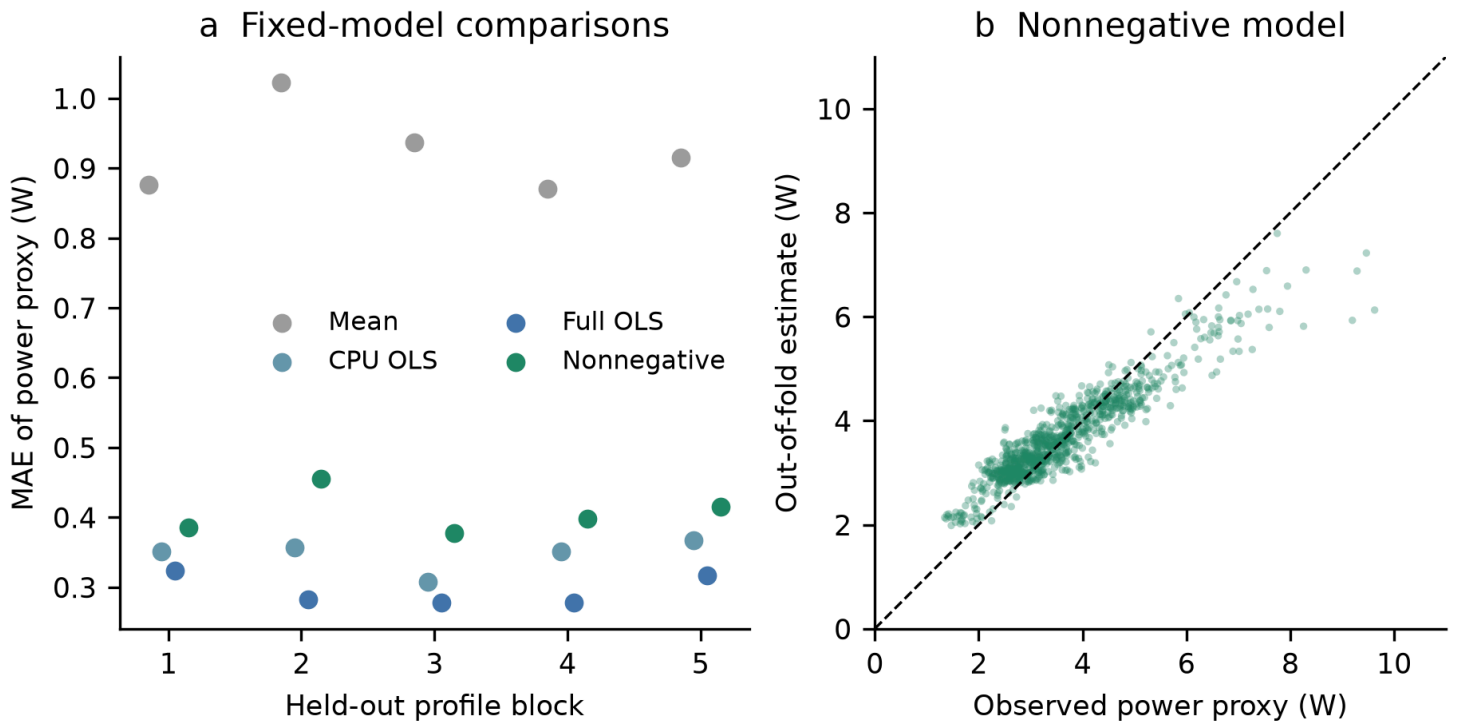


Figure 2

Independent reanalysis of AndroWatts. Panel a compares each held-out profile block; panel b shows the nonnegative model's out-of-fold estimates. The target uses the source's fixed 4.4 V conversion and is a power proxy. Blocks contain profiles from one Pixel 8, and their order is not treated as verified chronology.

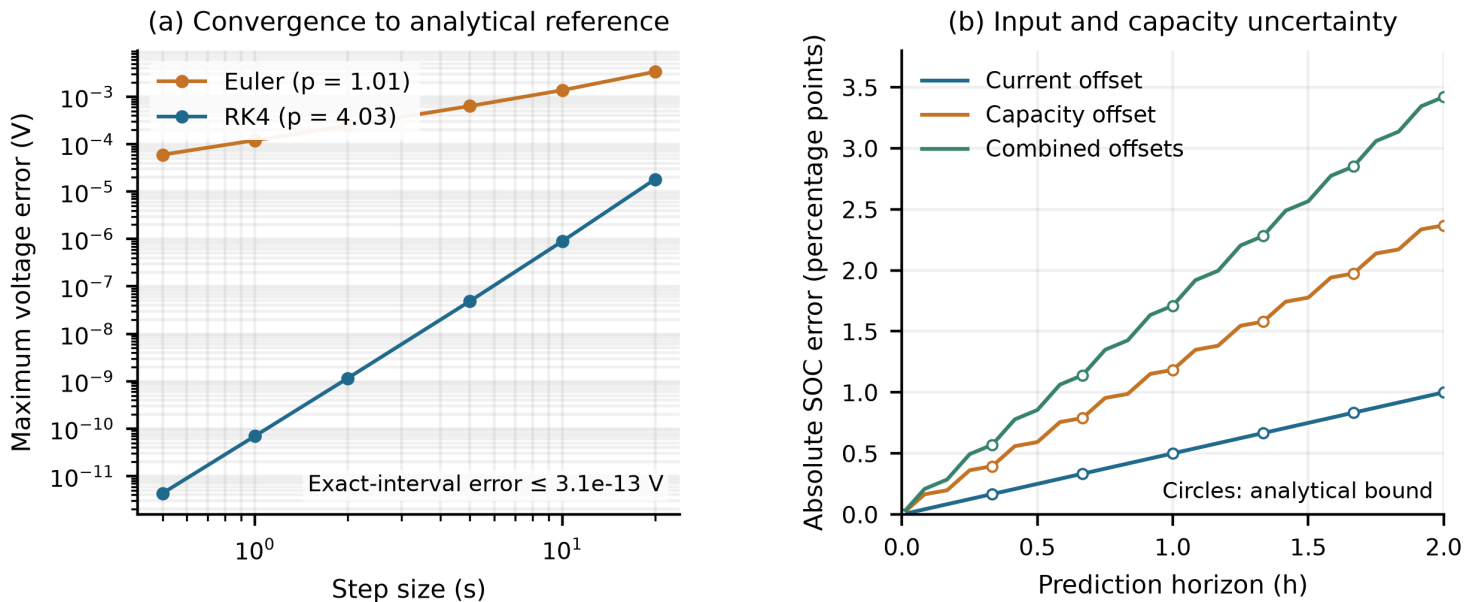


Figure 3

Numerical verification of the unified continuous-time equations with illustrative parameters. (a) Maximum terminal-voltage error relative to the event-level analytical solution for every prescribed step. The stated orders use the three finest steps; exact interval error is shown as a numerical value because it is at rounding level. (b) SOC errors for current bias, capacity bias and their combination; circles show the analytical finite-horizon bounds. The example verifies equations and numerical propagation, rather than physical-device or TTE accuracy.

Supplementary Files

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

- [SupplementaryCode1.zip](#)
- [SupplementaryData1.zip](#)
- [SupplementaryInformation.pdf](#)