Studies in Knowledge and Innovation
Vol. 2 No. 11 (2026)

The Best Baseline for a Battery Charge Label Changes With the Observation Window

Claire Warren · Matthew Kelley · Luke W Johnston

Published: November 26, 2026 | Article 1190

Abstract

The reported state of charge stayed at the same integer across 2,947 of 3,960 consecutive minute intervals in a public photovoltaic battery archive. Current integrated over each interval was positive in every case. At one minute, predicting no change in the reported state had a mean absolute error of 0.270 percentage points, compared with 0.376 for a reconstruction using current observed during the interval. At five minutes the order reversed. The comparison used 22 charge traces and reserved complete recording dates for evaluation. The result identifies a feature of the public reference series and of the selected loss. It gives a concrete reason to specify the evaluation interval when comparing charge tracking methods.

The Best Baseline for a Battery Charge Label Changes With the Observation Window

Abstract

The reported state of charge stayed at the same integer across 2,947 of 3,960 consecutive minute intervals in a public photovoltaic battery archive. Current integrated over each interval was positive in every case. At one minute, predicting no change in the reported state had a mean absolute error of 0.270 percentage points, compared with 0.376 for a reconstruction using current observed during the interval. At five minutes the order reversed. The comparison used 22 charge traces and reserved complete recording dates for evaluation. The result identifies a feature of the public reference series and of the selected loss. It gives a concrete reason to specify the evaluation interval when comparing charge tracking methods.

When an unchanged label receives the best score

A battery charge signal can move on a different timescale from the label used to judge a tracking algorithm. A current meter reports an electrical quantity at each observation. The state of charge supplied for validation may be available to the analyst only as a whole percentage. For two successive observations with the same reported state, a constant prediction is scored as exact. A positive estimated change can receive an error even if there is continuing current. This tension is unimportant if the target of the analysis is the behavior of that particular reported channel. It becomes consequential when evidence about it is used to judge how well a method traces electrical change.

Joint estimation of charge state, health, and power has motivated designs that use a shared measurement model or enforce coherence among independently estimated variables (Xu et al., 2026) (Zhang et al., 2021). Separate state-of-power research makes the timing used to characterize electrical response part of the model specification (Lai et al., 2024). Classical state estimators use circuit and observation models with explicit distinctions between latent states and measured outputs (Plett, 2004a) (Plett, 2004b) (Chen and Rincon-Mora, 2006) (Hu et al., 2012). A check against a stored state label need not reproduce the information requirements of the full estimation problem. Earlier experiments on measurement timescale in battery characterization demonstrate how different intervals can describe different properties even using closely related electrical signals (Barai et al., 2018).

The test here uses one publicly archived photovoltaic charging experiment. It asks how often the reported charge-state label has the same value at the ends of windows of different lengths on the observed charging traces. It then asks which of three simple comparison rules produces the smallest deviation from the observed label changes on recording dates left out of calibration. This is a reconstruction with the recorded current for the complete interval available to the analyst. The exercise isolates a question about the time scale of reference measurement and its evaluation. A different design would be needed for forecasting a future state without having first observed the electrical input to that interval.

Source and paired comparison

The URI-PBEST archive contains measured photovoltaic charging traces and other battery operating records (Rahman et al., 2026). Its associated estimator study illustrates the interest of sampled photovoltaic experiments for tracking a battery state in the presence of changing current (Rahman and Adhikari, 2025). We analyzed its 22 charging worksheets. They contain 181 equally spaced readings per sheet and cover three hours apiece, with an observation at each endpoint. There are 3,982 rows including sheet endpoints, 3,960 adjacent one-minute intervals, and 21 distinct recording dates. A single date supplied two three-hour traces. A real date cell in the first row was used where it could be parsed as an Excel date. Two separate malformed date cells were normalized to 5 and 14 August 2024 by matching their visible calendar digits and the worksheet names. No time or measurement values from either sheet were changed. Date normalization matters because both traces recorded on a shared date are held out together.

The retained channels are clock time, the reported current, voltage, and state label. The voltage is checked for numeric completeness. Analysis uses current and the stored label. The store provides values as whole percentages. The source designates the field as a reference state of charge. Its whole-percentage values are available directly, but the exact upstream procedure by which these label values were generated cannot be reconstructed from the worksheet cells. Contemporary work on measurements and reference preparation documents different error sources for integrating current, sensor and timing uncertainty, and sampled outputs (Movassagh et al., 2021) (Mohammadi, 2022) (Ramírez-Rasgado et al., 2026). The present design can observe equality in two stored state labels and compare it with an integrated electrical measurement. It cannot identify the operator's exact rounding, display updating, or upstream estimation process.

For each worksheet we formed complete non-overlapping 1-, 5-, 15-, and 30-minute windows, separately by width. In every comparison the response is the end label minus the beginning label. We calculate charge flow in ampere-hours by trapezoidal integration of the current over the recorded timestamps. Adjacent windows touch at a boundary and successive measurements on a worksheet remain dependent. We report the exact positive-current and unchanged-label counts without treating those intervals as independent samples. Slope and time differences are never calculated across separate sheets. We also repeat the grouping with later starting offsets of two, seven, and fifteen minutes for the three wider intervals. This second pass discards the incomplete leading and trailing pieces from the shifted windows.

Three predictors supply candidate label changes. Persistence predicts zero, equivalent to reusing the starting label if the target were the ending label. The training median predicts the median change observed in the other dates at that same interval length. Medians are defined in terms of the reported label changes using a consistent computation (Hyndman and Fan, 1996). The current-based predictor multiplies the integrated current of the test interval by a single coefficient from a least-squares regression through the origin. Its fit uses label differences and integrated current from dates other than the test date. That coefficient is a calibration against the archived labels. It does not estimate usable capacity because the labels and measurement boundary alone cannot establish actual state or usable charge. Enhanced counting methods and unified state methods include additional information that our transparent comparator omits (Ng et al., 2009) (Tang et al., 2024).

The analysis scores mean absolute error in percentage points. The median comparator minimizes that loss among constant predictions on the training sample, consistent with the distinction between forecasts aimed at a central quantile and those aimed at a conditional mean (Gneiting, 2011). The same error definition applies to all three rules on every held-out date. Date-level evaluation eliminates within-date sharing between fit and test. It does not validate performance on new cells or, because the other dates can

precede or follow a test date, guarantee a prospectively deployable model. Research on structured validation and serial dependence recommends matching the split to the unit that must transfer (Roberts et al., 2017) (Bergmeir et al., 2018) (Cerqueira et al., 2020). Here that unit is a recording date.

Each width has the same number of complete windows per sheet in the main partition. Pooled mean absolute error consequently gives identical weight to each of the 22 traces at a given width. It gives two trace weights to the date measured twice and one to other dates, although the two are always tested jointly. We also retain a separate per-trace error for each predictor so that a result averaged across intervals can be connected to one worksheet. The signed label changes are used directly throughout. Rounding model outputs back to integers before scoring would define a classifier over the recorded digits and might make the comparison with persistence different.

The comparison switches between one and five minutes

All current integrals were positive at every selected width. At one minute the start and end labels were identical in 2,947 of 3,960 windows, or 74.4%. The median integrated current in those flat-label windows was 0.0196 Ah. In the 215 flat-label five-minute windows the median integral was 0.0395 Ah, and in the 13 flat-label thirty-minute windows it was 0.1661 Ah. An exact equality of labels can therefore coexist with a measured flow across both short and extended periods. At each width the number of windows and the values in them change together. The observed drop in the fraction of flat windows is descriptive and does not say that all flat one-minute intervals turn into moving labels when appended to their neighbors.

Of the 1,013 one-minute intervals with different reported labels, 983 advanced by one point and 30 advanced by more than one. The largest jump was seven points. The median current integral in advancing windows was 0.0226 Ah, about 0.0030 Ah above the median for the flat group. The overall maximum one-minute integral was also observed in a flat interval. This rules out a strict current-integral cutoff that can reproduce all one-minute transitions in the observed data. At longer widths the associated median flow differs more clearly between moving and flat labels. At thirty minutes those medians are 0.6438 and 0.1661 Ah. This describes how the selected grouping changes the overlap and says nothing about a causal threshold for a move of one reported state point.

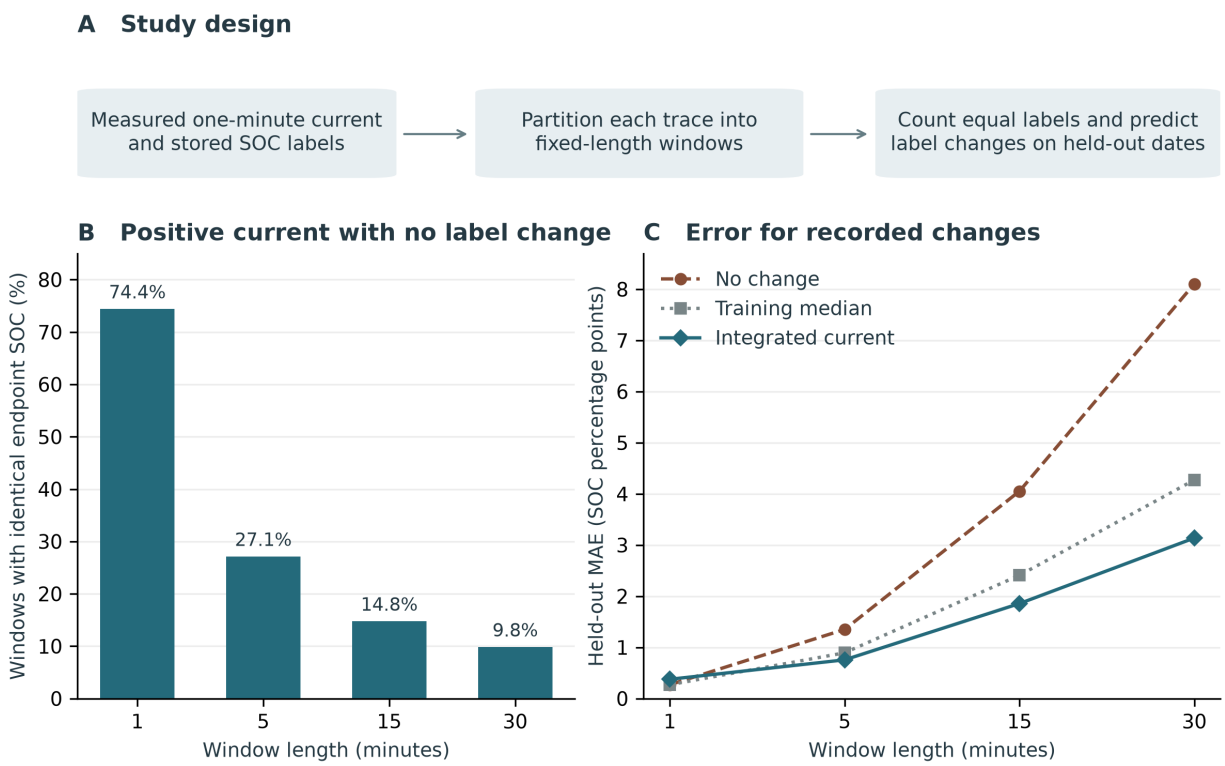


Figure 1. The measurement design uses the original charging traces. The two empirical panels track equal observed endpoint labels and held-date mean absolute errors at the four evaluated widths. Current integration uses measurements during each interval.

Table 1. Positive trapezoidal current integration was observed in all windows. Errors are in percentage points of the stored state-of-charge label, pooled over held-date predictions.

Window length (min)	Complete windows	Identical labels	Median integrated Ah among identical labels	Persistence MAE	Training median MAE	Integrated-current MAE
1	3,960	2,947	0.0196	0.270	0.270	0.376
5	792	215	0.0395	1.351	0.894	0.756
15	264	39	0.0945	4.053	2.409	1.857
30	132	13	0.1661	8.106	4.273	3.137

At one minute the training median is zero on every date holdout and agrees with persistence. The latter scores better than the integrated-current comparator on every sheet. At five minutes the current reconstruction has a pooled error of 0.756 percentage points, 0.137 points below the training median and 0.595 points below no change. It improves on the training median in 19 out of 22 traces. The improvement becomes 0.552 points pooled at fifteen minutes and 1.136 points at thirty minutes. At the thirty-minute width the current model still has a greater raw error than at five minutes. The gain within width and the larger absolute error across widths answer different questions because both the response magnitude and the number of observations change.

The fitted conversion slopes across date holdouts were 15.99–16.51 label points per Ah for one-minute windows and 15.69–16.23 for thirty-minute windows. Those overlapping ranges put the contrast in perspective. A large change in the average fitted conversion was unnecessary for the ranking of the same two model families to switch. The tolerance implicit in absolute error behaves differently once the response distribution has more changes greater than one point. At five minutes 304 windows advanced by

more than one point. At thirty minutes 116 of 132 did. Comparing methods at several widths can therefore test whether an apparently superior baseline is supported only where the archived label is mostly constant.

The aligned windows put a divider at the start of every sheet. With the shifted dividers the fraction of equal endpoint labels was 27.7%, 14.0%, and 9.1% for five, fifteen, and thirty minutes. In each shifted case the current reconstruction again has the smallest pooled error. The largest observed thirty-minute integral among flat labels is 0.259 Ah in the original partition and 0.353 Ah in the shifted partition. These maxima summarize particular windows and are sensitive to their placement. Agreement in the model ordering under both partitions does not require their extrema or the sheet-level improvement counts to coincide. The inference about longer windows rests on the comparison run separately in each partition.

Implications for benchmark design

A stronger one-minute result for persistence follows from the way the loss treats a held constant value. If the criterion is absolute deviation from the next stored integer, reducing error over the frequently unchanged labels can dominate the total. A reviewer interested in the rarer transition intervals can also ask how well their frequency, magnitudes, or timing are recovered. A characterization based on accumulated current over a longer interval has more space for many small positive readings to affect the comparison against the label. The shift in ordering here is evidence on the scope of the reference and scoring protocol. It provides no inference that labels with more decimals, or another upstream reference system, would produce the same switch.

There is also an operational difference between requiring an algorithm to report meaningful changes and allowing it to mirror a reference readout. The first goal needs a reference that can resolve movement at the period being studied. The second can be supported by measuring agreement with the displayed value and retaining the observed prevalence of repeats. If both goals must be supported by one archive, the readout series could be analyzed alongside the original electrical measurements with separate error criteria for stable labels and transitions. A single pooled score otherwise blends the two tasks and can reward a design that matches the behavior of one channel particularly well.

Work on data for cell ageing and stochastic characterization illustrates why a measurement archive can be valuable for one purpose while a given derived field has a narrower interpretation (Luh and Blank, 2024) (Khan et al., 2025) (Kirkaldy et al., 2024). Battery degradation instrumentation and cross-cell variability analyses similarly define a comparison in terms of explicit conditions and protocols (Birkl et al., 2017) (Rüther et al., 2023). The present source supplies one photovoltaic arrangement, and the held-date results remain within that arrangement. Work on learning early indicators of cycle life illustrates a different target and group-transfer problem (Severson et al., 2019). Those domains motivate recording what identifier was left out and how the reference quantity was produced. The positive current integrals here cannot establish a change in long-term health.

Evaluation protocols for battery-management algorithms explicitly distinguish functions and measurement checks (Campestrini et al., 2016). Data inventories and reproducibility guidance provide examples of retaining source context, raw inputs and analysis assumptions for later comparison (dos Reis et al., 2021) (Gebru et al., 2021) (Peng, 2011). For an experiment where it matters to distinguish small changes from repeated stored labels, a compact result could carry the label resolution, timestamp resolution, window width, and persistence error together. Additional validation with an independently documented reference stream would address whether good label reconstruction also means accurate state estimation. A dashboard that simply compares two errors in percentage points, without the time interval or the reference behavior, can otherwise draw contradictory conclusions from adjacent views of the same charging record.

Data and code availability

The raw archive is data/raw/uri_pbest_1190.zip, verified against the Zenodo-published checksum. Run python scripts/uri_pbest_windows_1190.py to reproduce results in analysis/uri_pbest_windows_1190.json. Run python scripts/plot_uri_pbest_windows_1190.py to render the PNG and SVG.

References

Barai, A., Uddin, K., Widanage, W. D., McGordon, A., & Jennings, P. (2018). A study of the influence of measurement timescale on internal resistance characterisation methodologies for lithium-ion cells. *Scientific Reports, 8(1), 21*. https://doi.org/10.1038/s41598-017-18424-5

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, *120*, 70-83. https://doi.org/10.1016/j.csda.2017.11.003

Birkl, C. R., Roberts, M. R., McTurk, E., Bruce, P. G., & Howey, D. A. (2017). Degradation diagnostics for lithium ion cells. *Journal of Power Sources, 341, 373–386*. https://doi.org/10.1016/j.jpowsour.2016.12.011

Campestrini, C., Horsche, M. F., Zilberman, I., Heil, T., Zimmermann, T., & Jossen, A. (2016). Validation and benchmark methods for battery management system functionalities: State of charge estimation algorithms. *Journal of Energy Storage*, *7*, 38-51. https://doi.org/10.1016/j.est.2016.05.007

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: an empirical study on performance estimation methods. *Machine Learning*, *109*(11), 1997-2028. https://doi.org/10.1007/s10994-020-05910-7

Chen, M., & Rincon-Mora, G. (2006). Accurate electrical battery model capable of predicting runtime and I-V performance. *IEEE Transactions on Energy Conversion, 21(2), 504–511*. https://doi.org/10.1109/tec.2006.874229

dos Reis, G., Strange, C., Yadav, M., & Li, S. (2021). Lithium-ion battery data and where to find it. *Energy and AI, 5, 100081*. https://doi.org/10.1016/j.egyai.2021.100081

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., III, H. D., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64(12), 86–92*. https://doi.org/10.1145/3458723

Gneiting, T. (2011). Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, *106*(494), 746-762. https://doi.org/10.1198/jasa.2011.r10138

Hu, X., Li, S., & Peng, H. (2012). A comparative study of equivalent circuit models for Li-ion batteries. *Journal of Power Sources, 198, 359–367*. https://doi.org/10.1016/j.jpowsour.2011.10.013

Hyndman, R. J., & Fan, Y. (1996). Sample Quantiles in Statistical Packages. *The American Statistician, 50(4), 361–365*. https://doi.org/10.1080/00031305.1996.10473566

Khan, M. A., Thatipamula, S., Tresca, L., Xu, L., Trewartha, A., & Onori, S. (2025). High-power lithium-ion battery characterization dataset for stochastic battery modeling. *Scientific Data, 12(1), 1506*. https://doi.org/10.1038/s41597-025-05725-y

Xu, D., Chen, H., & Gui, H. (2026). Unified Online Estimation Method for SOC, SOH, and Power Capacity Considering Safety Boundary Consistency in Battery Management Systems. *2026 International Conference on Electrical, Control and Information Technology (ECITech), 133–137*. https://doi.org/10.1109/ecitech69277.2026.11601281

Kirkaldy, N., Samieian, M. A., Offer, G. J., Marinescu, M., & Patel, Y. (2024). Lithium-ion battery degradation: Comprehensive cycle ageing data and analysis for commercial 21700 cells. *Journal of Power Sources, 603, 234185*. https://doi.org/10.1016/j.jpowsour.2024.234185

Lai, X., Yuan, M., Tang, X., Zheng, Y., Zhu, J., Sun, Y., Zhou, Y., & Gao, F. (2024). State-of-power estimation for lithium-ion batteries based on a frequency-dependent integer-order model. *Journal of Power Sources, 594, 234000*. https://doi.org/10.1016/j.jpowsour.2023.234000

Luh, M., & Blank, T. (2024). Comprehensive battery aging dataset: capacity and impedance fade measurements of a lithium-ion NMC/C-SiO cell. *Scientific Data, 11(1), 1004*. https://doi.org/10.1038/s41597-024-03831-x

Mohammadi, F. (2022). Lithium-ion battery State-of-Charge estimation based on an improved Coulomb-Counting algorithm and uncertainty evaluation. *Journal of Energy Storage*, *48*, 104061. https://doi.org/10.1016/j.est.2022.104061

Movassagh, K., Raihan, A., Balasingam, B., & Pattipati, K. (2021). A Critical Look at Coulomb Counting Approach for State of Charge Estimation in Batteries. *Energies*, *14*(14), 4074. https://doi.org/10.3390/en14144074

Ng, K. S., Moo, C., Chen, Y., & Hsieh, Y. (2009). Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries. *Applied Energy*, *86*(9), 1506-1511. https://doi.org/10.1016/j.apenergy.2008.11.021

Peng, R. D. (2011). Reproducible Research in Computational Science. *Science, 334(6060), 1226–1227*. https://doi.org/10.1126/science.1213847

Plett, G. L. (2004a). Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1. Background. *Journal of Power Sources, 134(2), 252–261*. https://doi.org/10.1016/j.jpowsour.2004.02.031

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

Rahman, S., & Adhikari, K. (2025). Extended and Unscented Kalman Filter Variants for Tracking the State of Charge of PV Battery Systems. *2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 1-6. https://doi.org/10.1109/icecet63943.2025.11472367

Rahman, S., Adhikari, K., & Icer, K. (2026). University of Rhode Island PV Battery Energy Storage Dataset for SOC Tracking (URI-PBEST). *Zenodo, version 1*. https://doi.org/10.5281/zenodo.19487496

Ramírez-Rasgado, F., García-Godina, E., Hernández-González, O., Guerrero-Sánchez, M., Valencia-Palomo, G., & Garrido-Meléndez, J. (2026). State-of-Charge Estimation for Lithium-Ion Batteries Under Noisy and Sampled Output Measurements. *Processes*, *14*(19), 3048. https://doi.org/10.3390/pr14193048

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913-929. https://doi.org/10.1111/ecog.02881

Rüther, T., Schamel, M., Plank, C., Schomburg, F., Röder, F., & Danzer, M. A. (2023). Cell-to-cell variation beyond parameter analysis — Identification and correlation of processes in Lithium-Ion Batteries using a combined distribution of relaxation times analysis. *Journal of Power Sources, 587, 233677*. https://doi.org/10.1016/j.jpowsour.2023.233677

Severson, K. A., Attia, P. M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M. H., Aykol, M., Herring, P. K., Fraggedakis, D., Bazant, M. Z., Harris, S. J., Chueh, W. C., & Braatz, R. D. (2019). Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, *4*(5), 383-391. https://doi.org/10.1038/s41560-019-0356-8

Tang, A., Huang, Y., Xu, Y., Hu, Y., Yan, F., Tan, Y., Jin, X., & Yu, Q. (2024). Data-physics-driven estimation of battery state of charge and capacity. *Energy*, *294*, 130776. https://doi.org/10.1016/j.energy.2024.130776

Zhang, T., Guo, N., Sun, X., Fan, J., Yang, N., Song, J., & Zou, Y. (2021). A Systematic Framework for State of Charge, State of Health and State of Power Co-Estimation of Lithium-Ion Battery in Electric Vehicles. *Sustainability, 13(9), 5166*. https://doi.org/10.3390/su13095166