DO LEARNED PARTICLE DYNAMICS TRANSFER ACROSS BATTERY OPERATING CONDITIONS?

Anonymous authors
Paper under double-blind review

ABSTRACT

Machine-learned simulators of physical systems are usually trained and tested under similar conditions, yet for batteries what matters is how electrode particles behave when the charging rate, temperature or cell age changes, and such measurements are scarce. We introduce a benchmark of operando optical microscopy movies, recorded while the batteries run, in which the same particles are filmed under different conditions: 22 graphite particles at slow (C/20) and faster (C/2) charging over 6–45 °C (about 300 hours of recording), 24 particles of a nickel manganese cobalt oxide (NMC) cathode at three rates, and four further cells. Forecasters predict a particle's future image from its recent frames and the known current, and are compared with baselines that replay the same particle's earlier behaviour. First, with one training rate, a network that steps forward in charge passed rather than in time transfers far better from slow to fast charging (graphite: skill gains of 0.42 and 0.51 over matched time steps at short and long horizons), because charge builds in the known speed-up. Second, same-particle baselines largely catch up with a step-by-step network at long horizons, whereas a U-Net that predicts each horizon directly stays ahead on six graphite rate and temperature splits but does not clearly beat NMC's best baselines. Third, particles first convert during lithiation in a reproducible order at one rate and temperature but not detectably across rates, and forecasts from slow to fast charging miss when and where particles convert, relative to a repeat experiment. For the foundation model Poseidon, performance depends strongly on how the output is scaled; Poseidon beats the direct U-Net at short horizons when data are scarce, but under the tested recipes neither Poseidon nor Walrus beats it at long horizons.

1 INTRODUCTION

Batteries are central to electrifying transport and to storing renewable electricity. How fast a battery can charge and how long it lasts depend on how the particles in its electrodes take up and release lithium. These particles do not fill uniformly: they switch between phases at different moments, and the pattern changes with how hard the cell is driven (Li et al., 2014; Lim et al., 2016). Operando microscopy can now film individual particles while a cell charges and discharges (Merryweather et al., 2021; Xu et al., 2022; Zhao et al., 2023). Such movies are hard to obtain: a cell must stay in focus under a microscope while it is cycled for hours to days (one slow half-cycle takes about a day), and each particle must be followed across frames and across experiments; the main graphite field of view here took about 300 hours of recording. The underlying physics is harder still to simulate faithfully, because it couples electrochemistry, phase separation and mechanics across many interacting particles whose parameters are unknown. Learning directly from measured movies is therefore attractive, and data of this kind are rare.

Machine-learned simulators, from neural operators (networks that map one field to another) to foundation models pretrained on simulated solutions of partial differential equations (PDEs) (Li et al., 2021; Herde et al., 2024; McCabe et al., 2026), are often trained and tested under similar conditions. For batteries the useful question is different: does a forecaster trained where data are plentiful, for example at slow charging, still work at a new charging rate, temperature or age? Benchmarks with real measurements vary the physical system or a control parameter (Hu et al., 2026; Mukhopadhyay

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

et al., 2026), and in many distribution-shift benchmarks a new condition also brings new objects (Koh et al., 2021). A loss of accuracy then does not tell us which change matters. We keep the objects fixed by filming the same particles under different conditions.

Our benchmark consists of interferometric scattering (iSCAT) movies of the same particles under several conditions: 22 graphite particles and 24 particles of an NMC cathode, plus four further cells (Sec. 3). A forecaster sees a particle's last four frames and the known current and predicts the particle's image after a given amount of further charge; particle outlines, image alignment and the capacity used to scale charge are supplied. We compare neural forecasters and pretrained foundation models with persistence, statistical extrapolation, replay of the same particle's earlier behaviour and a predictor of the particle's mean image value (Secs. 4 and 5).

Three findings emerge (Sec. 6). First, when training covers one charging rate, a network that steps forward in charge passed rather than in seconds transfers far better to a faster rate: at constant current, the charge clock builds in the known speed-up, whereas a time-stepped network caught up only when trained at several rates and given the rate (on NMC). Second, simple baselines that replay the same particle's history largely catch up with a step-by-step network at long horizons, while a U-Net that predicts each horizon directly stays ahead on six graphite rate and temperature splits, partly by capturing patterns inside each particle, but does not clearly beat NMC's best baselines. Third, particles first convert during lithiation in a reproducible order at one rate and temperature but not detectably across rates, and models trained at the slow rate forecast when and where they convert at the fast rate less accurately than a repeat experiment. For the foundation model Poseidon, performance depends strongly on how the output is scaled; it leads the direct U-Net at short horizons when data are scarce, but not at long ones.

Contributions. (i) A benchmark of real operando movies that re-observe the same particles across charging rates, temperatures and cycles, in two materials and six cells (supporting different tests), with common charge-based targets, strong same-particle baselines, and data and code for review. (ii) Controlled comparisons of stepping in time or in charge, read against repeated experiments. (iii) An evaluation of pretrained PDE foundation models with recipes chosen on validation data and controls trained from scratch, showing how output scaling changes the benefit of pretraining. Our contribution is the data and the evaluation, not a new architecture.

2 Related work

Learning from battery imaging. Operando iSCAT microscopy follows individual particles while a cell is cycled (Merryweather et al., 2021; Xu et al., 2022), and X-ray microscopy shows that reaction pathways inside particles depend on the rate (Lim et al., 2016). Zhao et al. (2023) recover reaction kinetics and their variation across particles from X-ray movies by fitting a physical model; other work estimates concentrations from optical images (Kerner et al., 2024) or forecasts operando infrared spectra (Ye et al., 2026). We instead forecast future optical images and test how the forecasts transfer when the operating condition changes, following the same particles across experiments.

Learned simulators and their transfer. Video prediction with recurrent and direct models is well established (Shi et al., 2015; Tan et al., 2023); predicting each horizon directly rather than iterating one-step forecasts is a classical trade-off (Marcellino et al., 2006), and step-by-step neural PDE solvers drift over long rollouts (Brandstetter et al., 2022; Lippe et al., 2023). We also evaluate neural operators (Li et al., 2021) and PDE foundation models (Herde et al., 2024; McCabe et al., 2026) pretrained on simulated data such as the Well (Ohana et al., 2024). Benchmarks with real measurements are recent: RealPDEBench pairs measurements and simulations for five physical systems (Hu et al., 2026), and Walrus has been transferred to laboratory turbulence (Mukhopadhyay et al., 2026). Generalisation across environments and coefficients has mainly been studied on simulations (Yin et al., 2021a; Kirchmeyer et al., 2022; Subramanian et al., 2023). Our splits follow distribution-shift and structured cross-validation practice (Koh et al., 2021; Roberts et al., 2017) and guard against leakage and weak baselines (Kapoor & Narayanan, 2023; McGreivy & Hakim, 2024). Measuring progress in charge is standard in battery science, from Coulomb counting (integrating the current; Ng et al., 2009) and incremental-capacity analysis (Dubarry et al., 2006) to ageing models driven by the charge passed (Wang et al., 2011; Schmalstieg et al., 2014), lifetime prediction from voltage curves (Severson et al., 2019) and images compared at matched state of charge (Li et al., 2014; Lu

et al., 2023b). Changing the time variable of a learned model has been studied for recurrent networks and for neural differential equations driven by an input signal (Tallec & Ollivier, 2018; Kidger et al., 2020), and rate-dependent reaction pathways (Lim et al., 2016; Merryweather et al., 2022) suggest why a charge axis alone may not be enough.

3 SAME-PARTICLE MOVIES ACROSS OPERATING CONDITIONS

Graphite. Graphite is the material of most lithium-ion battery anodes (negative electrodes). As lithium enters it (lithiation), a graphite particle passes through a sequence of ordered phases ("stages"), and each change of phase changes how strongly the particle reflects light. Interferometric scattering (iSCAT) microscopy measures this reflected light very sensitively, so a movie shows where and when each particle transforms. Our primary dataset is previously unpublished iSCAT microscopy of a graphite composite electrode, recorded with a setup similar to that of Han et al. (2026) (none of their movies are reused), together with the cell's voltage and current while it is charged and discharged at constant current. Charging rates are given as nominal C-rates: at C/20 a full charge would take about 20 hours, at C/2 about 2 hours and at 2C about half an hour; actual durations follow from the measured current and capacity. The same field of view was imaged in 18 recording runs (acquisitions): two formation cycles (the first cycles of a new cell, during which a surface layer forms and some lithium is lost), four at C/20, eight at C/2 and four at 2C, the last with a magnified view. Their currents of 12.8, 130 and 270 µA span a factor of 21, at 15–45 °C (C/20 and 2C) and 6–45 °C (C/2; Appendix A).

Particles. After alignment and quality filtering, 22 particle identities yield 246 movies across 13 widefield acquisitions (the first formation cycle is excluded because its field of view rotates). The magnified 2C acquisitions retain three of them: the central flake and two neighbours. Each movie is a 64×64 crop at 0.26 µm per pixel in one common orientation, so that pixels of a particle correspond across acquisitions. The 13 widefield acquisitions hold 11,670 stored frames (Table 3); the C/20→C/2 split alone scores 7,585 forecast windows at $\Delta = 0.02$ over the 22 particles. Particle outlines are drawn from whole acquisitions, so the geometry is given to the forecaster rather than predicted.

Charge. We integrate the measured current over time to obtain the charge $q(t)$ passed into the electrode, interpolated to each frame. Charge is measured in units of a fixed reference capacity, $Q_{\text{ref}} = 280\,\mu\text{Ah}$ (a typical C/20 half-cycle), with two coordinates: the charge state $s = q/Q_{\text{ref}}$, counted from the start of the acquisition, and the charge u passed since the start of the current half-cycle (one lithiation, or one delithiation, in which lithium leaves the electrode). Every acquisition except the second formation cycle starts from a delithiated state, so s counts charge relative to that state; it is not a measured lithium content of each particle. Faster and colder half-cycles pass less charge before the voltage limit (264–305 µAh at C/20, about 90 µAh at C/2 and 6 °C).

NMC. As a second material we use the iSCAT movies of Xu et al. (2022): single-crystal particles of $\text{LiNi}_x\text{Mn}_y\text{Co}_z\text{O}_2$ (NMC), a layered nickel manganese cobalt oxide widely used in cathodes (positive electrodes), recorded at C/3, 1C and 2C. Unlike graphite, NMC takes up lithium largely as a solid solution: the composition of one crystal phase changes gradually instead of distinct phases forming, so its optical signal changes smoothly rather than through moving phase boundaries. We identify 24 particles at all three rates (48×48 crops at 69 nm per pixel; $Q_{\text{ref}} = 561$ µAh, the measured C/3 capacity, also when C/3 is held out, so this calibration is retrospective; Appendix A).

Further electrodes. Four more cells, recorded in separate runs, extend the evaluation (Appendix A.1). Two graphite cells each contain one large flake, which we follow: cell B at C/20 in its formation cycle and at C/2, and cell C in two C/20 formation cycles. These flakes are used only to test models trained on the main electrode; their charge is scaled by a reference capacity measured on each cell (126 and 165 µAh), and every test is repeated with a capacity matched afterwards to the test acquisition and with a larger assumed pixel size. A second NMC electrode gives five particles at 0.5C, 1C, 2C and 3C, and a cell of NMC811 (nickel, manganese and cobalt in the ratio 8:1:1) gives 60 particles in 14 cycles between cycles 3 and 158, during which the upper voltage limit is raised from 4.2 to 4.6 V. With one object per graphite cell and five NMC particles, results on the first three of these cells are descriptive.

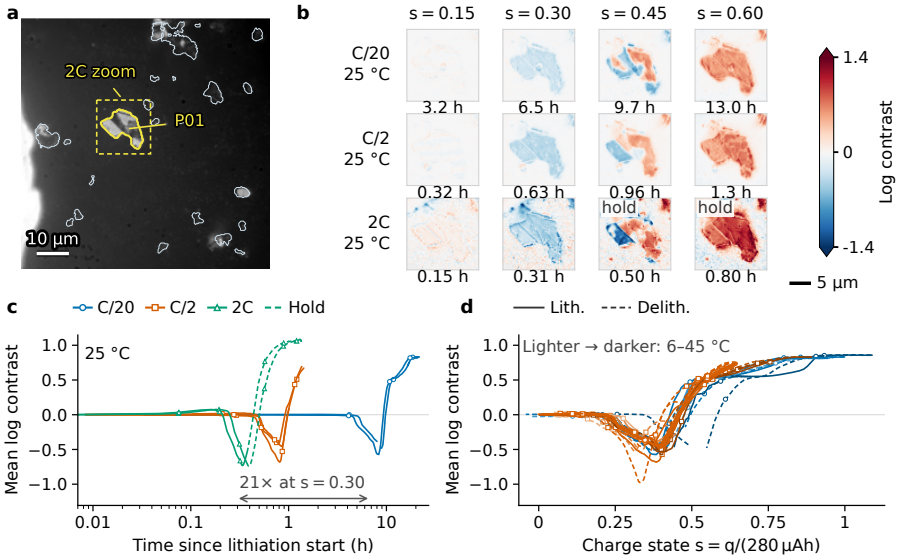

Figure 1: **The same particles, re-observed across operating conditions.** (a) The graphite field of view (C/20, 45 °C, mid-lithiation) with the 23 regions of the particle atlas, the reference map of particle identities (22 are retained); P01 is the central flake, also imaged in a magnified view at 2C. (b) P01 at matched Coulomb-counted charge state s during lithiation at C/20, C/2 and 2C (25 °C); numbers give the time since the start of lithiation (background-corrected log contrast; images during constant-voltage holds are illustrative and excluded from forecasting). (c) The P01 signal against physical time spreads by more than an order of magnitude between rates. (d) Against the charge state, the lithiation (solid) and delithiation (dashed) trajectories of P01 in all C/20 and C/2 acquisitions at 6–45 °C align much more closely.

4 Forecasting task, clocks and transfer splits

Task. Each stored movie follows one particle through an acquisition, and forecasting examples are drawn from its constant-current half-cycles. From the $k = 4$ most recent frames and the conditions known in advance, a forecaster predicts the particle's image after a further charge Δ has passed, $\hat{x}(u_0 + \Delta)$, where u_0 is the charge coordinate at the start of the forecast and the horizon $\Delta \in \{0.01, 0.02, 0.04, 0.08, 0.16, 0.32\}$ (in units of Q_{ref}) says how far ahead it reaches. The known conditions are the applied current, which sets the nominal rate and whether the electrode is being lithiated or delithiated, the temperature, the cycle index and the charge coordinates u and s. The cell voltage is a response of the cell, so forecasters may use only its past. Forecasts start from observed frames and use only frames recorded up to that point. At $\Delta = 0.16$ the forecast reaches 209 min ahead at C/20, 21 min at C/2 and 10 min at 2C.

Clocks. A step-by-step (autoregressive) forecaster repeatedly predicts the next state and feeds its own prediction back in, so it needs a step size, which we call its clock. The *frame clock* steps from one recorded frame to the next, the *time clock* by a fixed number of seconds (by default 60 s for graphite and 30 s for NMC), and the *charge clock* by a fixed amount of charge ($0.005\,Q_{\text{ref}}$ by default; Appendix J tests other step sizes). The *matched time clock* uses the duration of one charge step at the training rate (392.5 s at C/20). At constant current, one charge step lasts ten times less at C/2 than at C/20, so the charge clock builds in the known speed-up. Our splits test whether this built-in rescaling helps, not whether a model trained at one rate can learn how the dynamics depend on the rate. Whatever its clock, every forecaster is scored on the same observed target frames.

Transfer splits. Because the same particles are re-observed, we can hold out one condition at a time, although the order of acquisitions, voltage limits and cell history are not varied independently. *New rate*: train at C/20 and test at C/2, the reverse, and train at two rates to test at a third. *New temperature*: train at 25 °C and test at 6–15 °C or 35–45 °C (C/2), or at 15 and 45 °C (C/20). *New*

particle: five folds over particles, with a control that removes held-out particles' pixels from neighbouring training crops (Appendix B). *Later cycles* and *formation*: train on early or regular cycles and test on later or formation cycles. *Repeats*: train and test at the same condition, to measure in-distribution skill.

Scoring. Squared pixel errors are summed inside each particle's outline (SSE) and reported as skill, $1 - \sum \mathrm{SSE}_{\mathrm{model}} / \sum \mathrm{SSE}_{\mathrm{persistence}}$, where persistence predicts that the image does not change: skill 1 is perfect, 0 is no better than persistence, and negative skill is worse. We report means over training runs with different random seeds and 95% confidence intervals (CIs) of paired differences, which resample whole particles (or blocks of forecast origins for single flakes) and describe the recorded electrodes. A difference is *resolved* when its CI excludes zero; CIs for five or fewer particles or for single flakes are descriptive. The error splits exactly into an error of the particle's mean image value and an error of the pattern inside the particle; we report skill for each part, relative to the corresponding part of the persistence error, and a forecaster that only changes the mean scores zero on the pattern (Appendix H). For forecasts over a whole half-cycle we also compare, pixel by pixel, the charge at which each pixel changes fastest (a *switch map*), and the charge at which each particle's staging transition occurs. The same quantities computed between two real movies of one condition give *repeat agreement*: a reference for how reproducible the observed behaviour is under repeated nominal conditions, including imaging noise and differing histories. It is not an upper bound, because a forecaster that sees the particle's recent history can beat a noisy repeat. Most analyses are prospective, using only information available when a forecast starts; analyses selected with the observed future, such as forecasts started just before an observed transition, are retrospective diagnostics. A compact protocol for evaluating new methods is given in the reproducibility statement.

5 Forecasters

Baselines. *Persistence* repeats the last image. The *trend* baseline extrapolates each pixel's recent slope; a *damped trend*, whose extrapolated slope decays, and a per-pixel linear *AR(2)* extrapolation are fitted on training data (AR(2) is a two-lag linear statistical model, not a neural network). *Replay* adds to the last image the change that the same particle underwent over the same charge interval in a training movie. *History-aligned replay* first shifts that reference in charge to match the particle's recent history, and *population replay* uses the average change of other particles. The *particle-mean ridge* baseline predicts only the change of the particle's mean image value, by linear regression with a penalty on large coefficients (ridge regression), and leaves the pattern inside the particle unchanged. All baselines use only information available when the forecast is made (Appendix C).

Learned models. Our standard networks predict the change from the last frame, given $k = 4$ past frames, the particle outline and the conditions. The main model is a U-Net (Ronneberger et al., 2015), a convolutional network that encodes and reconstructs the image at several scales, whose features are scaled and shifted according to the conditions (FiLM); it is trained to predict one clock step ahead and rolled out step by step: the *autoregressive U-Net*. We also train a convolutional recurrent network (ConvLSTM; Shi et al., 2015), a Fourier neural operator acting on spatial Fourier modes (FNO; Li et al., 2021), a patch-based video transformer (Dosovitskiy et al., 2021) and a U-Net that also receives the same particle's earlier movies. As a physics-shaped comparator, a reaction–diffusion model of Allen–Cahn type (Allen & Cahn, 1979) evolves the optical contrast with a learned local reaction term and a diffusion-like coupling of neighbouring pixels; it forecasts optical contrast and is not a calibrated battery model. The *direct U-Net* predicts each horizon in a single pass, with the horizon as an input. By default, models are told the direction of the current and the charge state, written (d, s); (d, s, r) adds the log rate, log step time and log step charge. Model selection uses held-out blocks of the training acquisitions only (Appendix C).

Foundation models. We fine-tune the public pretrained weights of two foundation models for physical fields, Poseidon (Herde et al., 2024) at its two smaller sizes (T and B) and Walrus (McCabe et al., 2026), which were pretrained on simulated PDE solutions. Unlike our standard networks, their adapters do not receive the particle outline; all models use the same outlines in the loss and in scoring. We choose the adaptation recipe on validation data and compare the models with copies trained from scratch and with the U-Net (Appendix D).

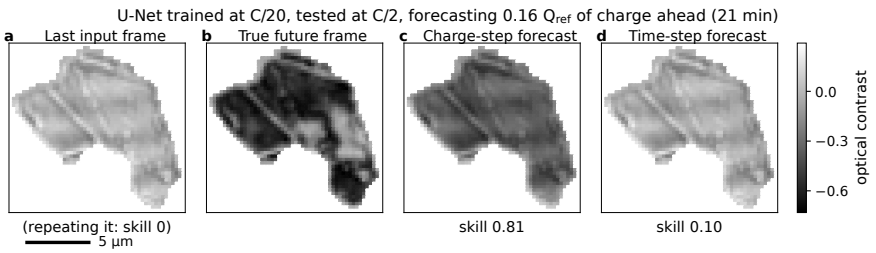

Figure 2: **One training rate, two clocks.** An autoregressive U-Net trained at C/20 forecasts P01 at C/2, $\Delta = 0.16$ ahead, using the charge clock (c) or matched time clock (d); window skills shown, window chosen by a declared median rule (Appendix E; aggregate results in Table 1).

Table 1: **Charge clock versus time clocks** (autoregressive U-Net, three seeds). Entries are $\Delta = 0.02$ / 0.16; skill is 1 for a perfect forecast and 0 for repeating the last image. Differences are paired on identical windows and seeds, with the 95% interval at $\Delta = 0.16$ below. Matched time: time steps equal to one charge step at the training rate. (d, s, r): the model is also given the rate and step sizes. Trained at the faster rate and tested at the slower one (C/2 → C/20, 2C → C/3), time-clock forecasts fall far below repeating the last image (skill -2 to -41), hence the large differences. †Three test particles (descriptive). Full table: Table 45.

Training → test rate	Charge-clock skill	Charge − time	Charge − matched time	Charge − time, (d, s, r)
Graphite				
C/20 → C/2	.51 / .67	+.23 / +.45 [.35, .49]	+.42 / +.51 [.37, .56]	+.16 / +.30 [.14, .35]
C/2 → C/20	.38 / .57	+6.50 / +2.57 [2.27, 3.11]	+6.73 / +5.06 [4.54, 6.10]	+5.54 / +3.10 [2.54, 3.91]
C/20 + C/2 → 2C†	.49 / .05	+.28 / −.19 [−.53, −.17]	n/r	+.26 / +.15 [.03, .16]
C/20 + 2C → C/2	.49 / .59	+.22 / +.38 [.25, .43]	n/r	+.24 / +.45 [.28, .52]
NMC				
C/3 → 1C, 2C	.79 / .89	+.61 / +.52 [.46, .55]	n/r	+.66 / +.51 [.43, .55]
C/3 + 1C → 2C	.81 / .89	+.35 / +.38 [.32, .42]	n/r	−.01 / +.01 [−.02, .04]
2C → C/3	.61 / .73	+19.6 / +33.2 [29.7, 43.6]	n/r	+30.1 / +42.1 [36.5, 58.1]
C/3 + 2C → 1C	.83 / .91	+.31 / +.15 [.10, .21]	n/r	+.07 / +.05 [.03, .07]

6 RESULTS

All comparisons use identical forecast windows within a split. The charge-clock lead holds in every test acquisition, and the direct U-Net's long-horizon gain has a positive point estimate in seven of eight (six resolved; Tables 41 and 44); both persist when forecasts start earlier (Table 55).

6.1 FINDING 1: STEPPING IN CHARGE HELPS TRANSFER FROM A SINGLE TRAINING RATE

With only one training rate, a model cannot learn from the data how the behaviour changes with the rate, but the charge clock builds in the known speed-up (Sec. 4). Trained at C/20 and tested at C/2, the autoregressive U-Net with the charge clock reaches skill 0.51 at $\Delta = 0.02$ and 0.67 at $\Delta = 0.16$. It beats the same network stepped in matched time by 0.42 [0.38, 0.44] and 0.51 [0.37, 0.56], and stepped in the default 60-s time steps by 0.23 and 0.45 (Table 1). Figure 2 shows why: the time-clock model predicts far too little change at C/2, as expected if it ignores the tenfold faster charging. A fit of the predicted against the observed change of the particle's mean image value over $\Delta = 0.16$ has slope 0.14 for the matched time clock, close to the 0.10 ratio of step durations, and 0.65 for the charge clock, where 1 would match the observed change (Table 53). No tested time

Table 2: **Skill of the U-Nets and of the best baseline** at $\Delta = 0.02$ / 0.16 (seed means). The best baseline is the strongest of eleven, including replay of the same particle and a particle-mean predictor. Last two columns: gain over the best baseline at $\Delta = 0.16$ with its 95% interval. New particles: five-fold averages, except †fold 0 only (five particles, descriptive); on fold 0 the best baseline scores .31 / .65 and the step-by-step U-Net .53 / .71. n/r: not run. Full table with all forecasters and splits: Table 31.

Test split	Best baseline	Step-by-step U-Net	Direct U-Net	Step-by-step − best	Direct − best
Graphite					
C/20 → C/2	.34 / .68	.51 / .65	.46 / .79	−.03 [−.13, .01]	+.11 [.05, .13]
C/2 → C/20	.20 / .46	.38 / .54	.30 / .63	+.08 [.04, .15]	+.16 [.13, .24]
Repeat at C/2	.32 / .66	.54 / .71	n/r	+.04 [.01, .14]	n/r
New particles	.28 / .57	.44 / .50	.45 / .75†	−.07 [−.10, .02]	+.11† [.09, .23]
Later cycles	.48 / .88	.67 / .86	.61 / .91	−.02 [−.08, .16]	+.04 [−.03, .22]
NMC					
C/3 → 1C, 2C	.75 / .88	.81 / .90	.79 / .91	+.02 [−.01, .05]	+.03 [−.002, .06]
2C → C/3	.58 / .82	.54 / .71	.42 / .83	−.11 [−.20, −.06]	+.01 [−.06, .04]

step closes the gap (at most 0.22 at $\Delta = 0.16$, against 0.52 to 0.77 for every tested charge step; Appendix J), and NMC shows the same ordering ($+0.52$ at 0.16). Nor is the advantage an artefact of which frames each clock sees: given the same last four frames, a direct U-Net told the horizon in charge rather than in time is better at $\Delta = 0.16$ by 0.61 and 0.64 in the two graphite directions and by 0.48 on NMC (three seeds; Table 54).

Training at several rates can partly replace the charge clock. Trained at C/3 and 1C and told the rate and step sizes ((d, s, r); Sec. 5), a time-clock model extrapolating to 2C has no resolved difference from the charge clock on NMC ($+0.01$ at 0.16). On graphite, extrapolation from C/20 and C/2 to 2C can only be tested on the three particles filmed at 2C: with these inputs the charge clock keeps a lead at $\Delta = 0.02$ and 0.16 (descriptive intervals), and without them it falls behind the time clock at long horizons (Table 45).

6.2 Finding 2: strong baselines catch the step-by-step network; direct prediction stays ahead

Replaying the same particle's history and predicting its mean image value are strong baselines (Table 2). At short horizons ($\Delta = 0.02$) the autoregressive U-Net beats the best baseline by 0.14 to 0.24 on six graphite rate and temperature splits with 17 to 22 test particles (Table 31). At long horizons ($\Delta = 0.16$) this lead mostly disappears: the U-Net is ahead on two of these splits, not resolved from the best baseline on three and behind on one. The direct U-Net, in contrast, stays ahead of the best baseline at $\Delta = 0.16$ on all six, by 0.10 to 0.21 (for example $+0.11$ [0.05, 0.13] from C/20 to C/2 and $+0.16$ in reverse; five seeds), and descriptively on the three-particle 2C split and new-particle fold 0. Direct prediction also helps ConvLSTM and, less consistently, the FNO (Table 56), and beats step-by-step prediction with a validation-tuned step size (Appendix J.4.1). About three quarters of the direct U-Net's gain comes from the pattern inside each particle, which mean-only baselines cannot predict by construction: within-particle structure contributes 0.085 of its 0.11 lead over the particle-mean ridge from C/20 to C/2 (Table 108).

Two failure cases are informative. On new particles, the autoregressive U-Net beats the particle-mean ridge only at short horizons (Table 29). On the second formation cycle, the learned models that use the charge state, and replay aligned by charge, all fall below persistence at $\Delta = 0.16$ (the autoregressive U-Net scores -1.41, and -0.86 even without the charge inputs), while AR(2), which only extrapolates recent history, reaches 0.25. This cycle does not start from a delithiated state, so

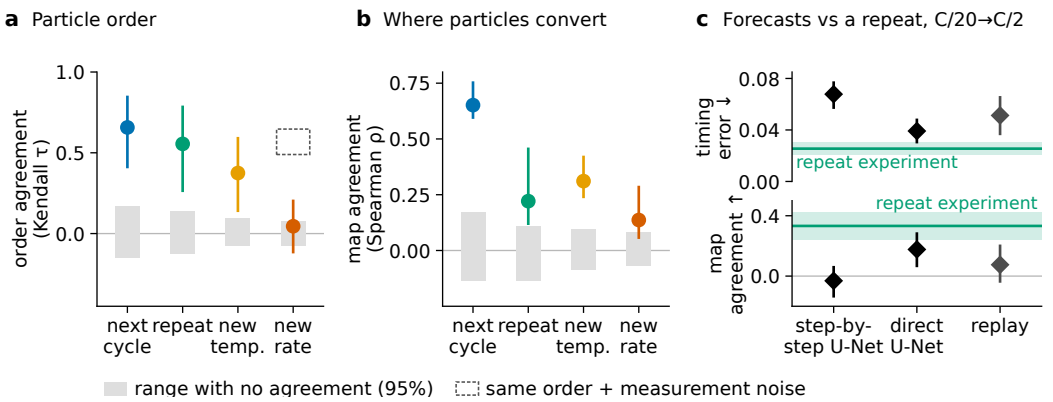

Figure 3: **Particles repeat their conversion order at one rate; forecasts from slow to fast charging miss when and where they convert.** First transition during lithiation, widefield recordings. (a) Agreement of the order in which particles convert, and (b) agreement of switch maps (when each pixel changes fastest), between two recordings: the next cycle of the same run, a repeat at the same rate and temperature, a different temperature and a different rate. Grey: range with no agreement; dashed box: agreement expected if the order were unchanged, given noise in the measured transitions. (c) Forecasts from C/20 to C/2 started just before 51 observed transitions (30 with maps; seeds 0–1), against a repeat experiment (green). Bars: 95% CIs. Full version with delithiation, magnified 2C views and second transitions: Fig. 4.

its charge count is offset. In a diagnostic re-score, shifting only the U-Net's charge-state input by an offset estimated from the voltage lifts the same trained model to 0.09 and removes about 62% of its squared error (Table 60); a change in the dynamics is not excluded.

6.3 Finding 3: Conversion order repeats at one rate; slow-to-fast forecasts miss when and where particles convert

Repeated movies at the same condition show how reproducible the observed behaviour is. For the first staging transition in lithiation, the order in which particles convert agrees between repeats at the same rate and temperature (Kendall rank correlation $\tau = 0.56$ [0.27, 0.78], where 1 is the same order and 0 no agreement). Across rates, which in this cell were recorded in separate blocks of the acquisition sequence, we detect no agreement (widefield $\tau = 0.05$ [−0.11, 0.20]; Fig. 3a), far below the 0.57 that noise in the measured transitions alone would leave (Appendix G). Switch maps, which record when each pixel changes fastest, agree best within an acquisition and least across rates; between repeats they agree only moderately, with an interval that overlaps the new-temperature value (Fig. 3b), so where conversion starts may be partly random and the repeat reference itself carries uncertainty; a mismatch alone does not show a different growth law (Appendix H.3).

Models trained at the slow rate reproduce these features less well than repeats do (Fig. 3c). In a retrospective diagnostic, forecasts start one charge step before each of 51 observed first lithiation transitions from C/20 to C/2 (19 particles, four acquisitions; seeds 0–1). The autoregressive U-Net's mean absolute error in the charge at which the transition occurs is 0.068, against 0.025 between repeats, and its switch maps have no resolved correlation with the observed ones ($\rho = -0.03$, against 0.33 between repeats; 30 targets with defined maps). The direct U-Net reduces the timing error to 0.039 and raises the correlation to 0.18, but still falls short of repeat agreement (Tables 109 and 110). Forecasts that do not use the observed future show the same gap: full rollouts of the autoregressive U-Net have timing error 0.105 against 0.030 between repeats and no map agreement ($\rho = 0.01$ against 0.40; Table 105), and the direct U-Net's fixed-start maps, mostly beyond its trained horizons, also show no agreement ($\rho = 0.01$ against 0.34 between repeats; Table 109). The exceptions concern timing only: in reverse-rate lithiation both models time transitions better than repeats but their maps do not agree ($\rho = -0.08$ and -0.02 against 0.26), and in delithiation full rollouts time them about as well as repeats (0.026 against 0.022). Training at the target condition, which also changes the training data, helps at C/2 (by 0.04 to 0.29 at $\Delta = 0.16$) but not at C/20 (-0.21; Tables 70 and 71).

6.4 FOUNDATION MODELS

Our one-step Poseidon adapter diverges on long C/2→C/20 rollouts, so validation selected direct prediction of each horizon. The direct U-Net then beats Poseidon-T by 0.17 to 0.21 at $\Delta = 0.16$ on three of four splits (common seeds 0–1), but most of this gap comes from how our adapter scaled its output. Among four adapter settings compared afterwards, validation selects scaling the predicted change by the spread of the input frames rather than by how much they changed; this alone lifts Poseidon-T by 0.10 to 0.18 at $\Delta = 0.16$ on three splits (Poseidon's own per-layer learning rates add almost nothing), leaving the U-Net ahead by 0.03 to 0.07 (Tables 24 and 25). Pretraining gains at $\Delta = 0.16$ change from 0.04–0.19 to 0.02–0.07, decreasing on three splits but increasing on fold 0. With the original adapter and one C/20 acquisition, Poseidon-T beats the direct U-Net at $\Delta = 0.02$ by 0.19 [0.16, 0.27] (the data-size test is mixed); its lead over the step-by-step U-Net is resolved on only one of the two single acquisitions tested ($+0.09$ [0.0005, 0.12]; Tables 20 and 21). Walrus, tested only with step-by-step training (a long-horizon result not matched to the direct U-Net's), is close to the direct U-Net at $\Delta = 0.02$ but falls below persistence at 0.16 (Table 17).

6.5 A SOLID-SOLUTION CONTRAST AND FURTHER ELECTRODES

In NMC, whose optical signal changes smoothly, simple extrapolation is a much stronger baseline than in graphite. From C/3 to 1C and 2C the autoregressive U-Net reaches 0.90 at $\Delta = 0.16$, but the particle-mean ridge (0.88) is not resolved from it; from 2C to C/3, AR(2) beats a single autoregressive U-Net but not, in seed means, separate U-Nets per current direction (Tables 2, 30 and 66). Direct prediction also helps here at long horizons: at $\Delta = 0.16$ it improves on step-by-step prediction by 0.01 and 0.12 (both resolved) without clearly beating the best baselines; it leads them at $\Delta = 0.32$ ($+0.04$ and $+0.06$), although from 2C to C/3 it trails AR(2) at $\Delta = 0.02$ (Table 68).

The single-rate charge-clock ordering recurs, descriptively, on independent electrodes (one flake per graphite cell; Table 92): on cell C's second formation cycle, main-electrode U-Nets trained at C/2 score 0.50 at $\Delta = 0.16$ with the charge clock and -0.72 with the time clock (ahead in all three seeds); on cell B at C/2, C/20-trained models show the same ordering (0.47 against 0.13) only with a capacity matched afterwards to the test acquisition (Tables 96 and 97). Yet the transferred models are weak there: all fall below persistence on both first formation cycles and seldom beat baselines carried over from the main electrode. In the NMC811 cell, an autoregressive U-Net trained on early cycles scores 0.87 at $\Delta = 0.16$ on cycles 99–117, but AR(2) leads on the last two of four later, higher-voltage cycles (pooled 0.52 against 0.42; Appendix F).

7 DISCUSSION AND LIMITATIONS

Following the same particles across conditions separates two questions: whether a model has learned the dynamics, and whether the dynamics stay the same under the new condition. Three lessons follow: measure progress in charge when training covers one rate; compare against same-object replay and a particle-mean predictor, not only persistence; and read transfer errors against the agreement between repeated experiments. The charge inputs describe electrode-level progress and direction (without them, skill at $\Delta = 0.16$ falls from 0.65 to 0.47 forward and 0.57 to 0.05 in reverse), while the images show each particle's starting state and where it changes. Direct prediction helps at long horizons on both main electrodes but not on every cell (it trails AR(2) on the second NMC electrode, 0.41 against 0.87), and poor transfer alongside lower repeat agreement does not establish a cause.

Limitations. The main graphite movies come from one field of view in one cell, recorded in sequence, so condition and cell history are confounded. The same particles can be re-observed only within one cell, and this series took about 300 hours of imaging, so repeating it on independent fields, though necessary, is slow. The further graphite cells contain one flake each, and the transition-order and map conclusions remain confined to the main electrode. Optical regions may contain several touching grains, s counts charge for the whole electrode, and the optical contrast is not a calibrated lithium fraction. Outlines, alignment and charge scale are fixed afterwards; the charge clock needs a known current schedule, reference capacity and charge origin, and is scored only on constant-current segments, where it exactly rescales time; rests, constant-voltage holds and variable currents are not tested. Foundation-model results depend on the adapter (its output scale alone is worth up to 0.18; Appendix D), and better adaptations may close the remaining gap.

AI USE STATEMENT

In this work, we used generative AI tools to help design and implement the experiments and analyses, process the data, analyse and interpret results, produce figures and tables, survey the literature, and draft and edit the text. Generative AI tools were not used to generate data or to prove mathematical claims. We have reviewed all AI-assisted work: the authors set the research questions and scope, checked the code, results, figures and text, and verified every reference against its published record. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Section 4 and Appendix C specify the forecasting task, the clocks, the transfer splits and the metrics; Appendix A describes how the datasets were built from the raw movies, including registration, particle identity, masks, the optical signal and the charge axis. All forecasters, baselines, training settings and the exact commands of every sweep are part of the anonymised code in the supplementary material (also browsable at `https://anonymous.4open.science/r/paper-code-7F3C`), together with the split definitions and the scripts that regenerate every table and figure from the saved evaluation records. Results are reported with the seed spread and with particle-level bootstrap intervals. For new methods we recommend a compact protocol: the splits C/20→C/2, C/2→C/20, the five new-particle folds and NMC C/3→1C, 2C; skill and pattern skill (Section 4) at $\Delta = 0.02$ and 0.16 on the released test windows, against persistence and the best released baseline, with paired particle-cluster intervals; and, for C/20→C/2, transition-timing error and switch-map correlation against repeat agreement. The released split files and code implement all of these metrics. The processed datasets of the five previously unpublished cells (per-particle crops with electrochemistry, masks, split files and checksums; 4.2 GB) are available to reviewers at an anonymous view-only link, `https://osf.io/4xryw/?view_only=982d53d950324a5bb1bc1b6b227f2aae` and the processed third-party NMC dataset (1.4 GB, same format) at a second one, `https://osf.io/phj7f/?view_only=fad5ccaee1ab4d5089b9f02ac7d52731`; both will be released under CC BY 4.0 with the final version. The supplementary package contains a processed-data sample that reproduces stored results on CPU (Appendix B.4), and the code can also rebuild the NMC dataset from the original source data.

REFERENCES

Samuel M. Allen and John W. Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica*, 27(6):1085–1095, 1979. doi: 10.1016/0001-6160(79)90196-2.

Peter M. Attia, Aditya Grover, Norman Jin, Kristen A. Severson, Todor M. Markov, Yang-Hung Liao, Michael H. Chen, Bryan Cheong, Nicholas Perkins, Zi Yang, Patrick K. Herring, Muratahan Aykol, Stephen J. Harris, Richard D. Braatz, Stefano Ermon, and William C. Chueh. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578(7795): 397–402, 2020. doi: 10.1038/s41586-020-1994-5.

Peng Bai, Daniel A. Cogswell, and Martin Z. Bazant. Suppression of phase separation in $LiFePO_4$ nanoparticles during battery discharge. *Nano Letters*, 11(11):4890–4896, 2011. doi: 10.1021/nl202764f.

Martin Z. Bazant. Theory of chemical kinetics and charge transfer based on nonequilibrium thermodynamics. *Accounts of Chemical Research*, 46(5):1144–1160, 2013. doi: 10.1021/ar300145c.

Johannes Brandstetter, Daniel E. Worrall, and Max Welling. Message passing neural PDE solvers. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=vSix3HPYKSU`.

J. R. Dahn. Phase diagram of Li_xC_6. *Physical Review B*, 44(17):9170–9177, 1991. doi: 10.1103/physrevb.44.9170.

Haitao D. Deng, Hongbo Zhao, Norman Jin, Lauren Hughes, Benjamin H. Savitzky, Colin Ophus, Dimitrios Fraggedakis, András Borbély, Young-Sang Yu, Eder G. Lomeli, Rui Yan, Jueyi Liu, David A. Shapiro, Wei Cai, Martin Z. Bazant, Andrew M. Minor, and William C. Chueh. Correlative image learning of chemo-mechanics in phase-transforming solids. *Nature Materials*, 21(5): 547–554, 2022. doi: 10.1038/s41563-021-01191-0.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Matthieu Dubarry, Vojtech Svoboda, Ruey Hwu, and Bor Yann Liaw. Incremental capacity analysis and close-to-equilibrium OCV measurements to quantify capacity fade in commercial rechargeable lithium batteries. *Electrochemical and Solid-State Letters*, 9(10):A454, 2006. doi: 10.1149/1.2221767.

Todd R. Ferguson and Martin Z. Bazant. Phase transformation dynamics in porous battery electrodes. *Electrochimica Acta*, 146:89–97, 2014. doi: 10.1016/j.electacta.2014.08.083.

Dimitrios Fraggedakis, Neel Nadkarni, Tao Gao, Tingtao Zhou, Yirui Zhang, Yu Han, Ryan M. Stephens, Yang Shao-Horn, and Martin Z. Bazant. A scaling law to determine phase morphologies during ion intercalation. *Energy & Environmental Science*, 13(7):2142–2152, 2020. doi: 10.1039/d0ee00653j.

Yinsheng Guo, Raymond B. Smith, Zhonghua Yu, Dmitri K. Efetov, Junpu Wang, Philip Kim, Martin Z. Bazant, and Louis E. Brus. Li intercalation into graphite: Direct optical imaging and Cahn–Hilliard reaction dynamics. *The Journal of Physical Chemistry Letters*, 7(11):2151–2156, 2016. doi: 10.1021/acs.jpclett.6b00625.

Jayesh K Gupta and Johannes Brandstetter. Towards multi-spatiotemporal-scale generalized PDE modeling. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=dPSTDbGtBY.

Jiho Han, George S. Phillips, Alice J. Merryweather, Juhwan Lim, Christoph Schnedermann, Robert L. Jack, Clare P. Grey, and Akshay Rao. Avalanche-like intercalation and intraparticle correlations in graphite. *Nature*, 656(8127):364–369, 2026. doi: 10.1038/s41586-026-10862-4.

Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for PDEs. In *Advances in Neural Information Processing Systems*, volume 37, pp. 72525–72624, 2024.

Peiyan Hu, Haodong Feng, Hongyuan Liu, Tongtong Yan, Wenhao Deng, Tianrun Gao, Rong Zheng, Haoren Zheng, Chenglei Yu, Chuanrui Wang, Kaiwen Li, Zhi-Ming Ma, Dezhi Zhou, Xingcai Lu, Dixia Fan, and Tailin Wu. RealPDEBench: A benchmark for complex physical systems with real-world data. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=y3oHMcoItR.

Stuart H. Hurlbert. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54(2):187–211, 1984. doi: 10.2307/1942661.

Sayash Kapoor and Arvind Narayanan. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9):100804, 2023. doi: 10.1016/j.patter.2023.100804.

Nadine Kerner, Yug Joshi, Monica Mead, Sebastian Eich, Roham Talei, and Guido Schmitz. Machine learning-assisted measurement of lithium transport using operando optical microscopy. *Acta Materialia*, 270:119867, 2024. doi: 10.1016/j.actamat.2024.119867.

Patrick Kidger, James Morrill, James Foster, and Terry Lyons. Neural controlled differential equations for irregular time series. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6696–6707, 2020.

Matthieu Kirchmeyer, Yuan Yin, Jérémie Dona, Nicolas Baskiotis, Alain Rakotomamonjy, and Patrick Gallinari. Generalizing to new physical systems via context-informed dynamics model. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 11283–11301. PMLR, 2022.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton Earnshaw, Imran Haque, Sara M Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. WILDS: A benchmark of in-the-wild distribution shifts. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5637–5664. PMLR, 2021.

Vincent Le Guen and Nicolas Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11471–11481, 2020. doi: 10.1109/cvpr42600.2020.01149.

Jizhou Li, Nikhil Sharma, Zhisen Jiang, Yang Yang, Federico Monaco, Zhengrui Xu, Dong Hou, Daniel Ratner, Piero Pianetta, Peter Cloetens, Feng Lin, Kejie Zhao, and Yijin Liu. Dynamics of particle network in composite battery cathodes. *Science*, 376(6592):517–521, 2022. doi: 10.1126/science.abm8962.

Yiyang Li, Farid El Gabaly, Todd R. Ferguson, Raymond B. Smith, Norman C. Bartelt, Joshua D. Sugar, Kyle R. Fenton, Daniel A. Cogswell, A. L. David Kilcoyne, Tolek Tyliszczak, Martin Z. Bazant, and William C. Chueh. Current-induced transition from particle-by-particle to concurrent intercalation in phase-separating battery electrodes. *Nature Materials*, 13(12):1149–1156, 2014. doi: 10.1038/nmat4084.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Jongwoo Lim, Yiyang Li, Daan Hein Alsem, Hongyun So, Sang Chul Lee, Peng Bai, Daniel A. Cogswell, Xuzhao Liu, Norman Jin, Young-sang Yu, Norman J. Salmon, David A. Shapiro, Martin Z. Bazant, Tolek Tyliszczak, and William C. Chueh. Origin and hysteresis of lithium compositional spatiodynamics within battery primary particles. *Science*, 353(6299):566–571, 2016. doi: 10.1126/science.aaf4914.

Phillip Lippe, Bas Veeling, Paris Perdikaris, Richard Turner, and Johannes Brandstetter. PDE-Refiner: Achieving accurate long rollouts with neural PDE solvers. In *Advances in Neural Information Processing Systems*, volume 36, pp. 67398–67433, 2023.

Edward N. Lorenz. Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric Sciences*, 26(4):636–646, 1969. doi: 10.1175/1520-0469(1969)26⟨636:aparbn⟩2.0.co;2.

Jiahuan Lu, Rui Xiong, Jinpeng Tian, Chenxu Wang, and Fengchun Sun. Deep learning to estimate lithium-ion battery state of health without additional degradation experiments. *Nature Communications*, 14(1):2760, 2023a. doi: 10.1038/s41467-023-38458-w.

Xuekun Lu, Marco Lagnoni, Antonio Bertei, Supratim Das, Rhodri E. Owen, Qi Li, Kieran O'Regan, Aaron Wade, Donal P. Finegan, Emma Kendrick, Martin Z. Bazant, Dan J. L. Brett, and Paul R. Shearing. Multiscale dynamics of charging and plating in graphite electrodes coupling operando microscopy and phase-field modelling. *Nature Communications*, 14(1):5127, 2023b. doi: 10.1038/s41467-023-40574-6.

Zhengyan Lun, Alice J. Merryweather, Amoghavarsha Mahadevegowda, Shrinidhi S. Pandurangi, Chao Xu, Simon Fairclough, Vikram S. Deshpande, Norman A. Fleck, Caterina Ducati, Christoph Schnedermann, Akshay Rao, and Clare P. Grey. Operando single-particle imaging reveals that asymmetric ion flux contributes to capacity degradation in aged Ni-rich layered cathodes. *Energy & Environmental Science*, 18(9):4097–4107, 2025. doi: 10.1039/d5ee00267b.

Massimiliano Marcellino, James H. Stock, and Mark W. Watson. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, 135 (1–2):499–526, 2006. doi: 10.1016/j.jeconom.2005.07.020.

Michael McCabe, Payel Mukhopadhyay, Tanya Marwah, Bruno Régaldo-Saint Blancard, François Rozet, Cristiana Diaconu, Lucas Thibaut Meyer, Kaze W. K. Wong, Mohammad-Hadi Sotoudeh, Alberto Bietti, Irina Espejo Morales, Rio Alexa Fear, Siavash Golkar, Tom Hehir, Keiya Hirashima, Geraud Krawezik, Francois Lanusse, Rudy Morel, Ruben Ohana, Liam Holden Parker, Mariel Pettee, Jeff Shen, Kyunghyun Cho, Miles Cranmer, and Shirley Ho. Walrus: A cross-domain foundation model for continuum dynamics. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, 2026. URL https://openreview.net/forum?id=kcyOjXoUZu. arXiv:2511.15684.

Nick McGreivy and Ammar Hakim. Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nature Machine Intelligence*, 6 (10):1256–1269, 2024. doi: 10.1038/s42256-024-00897-5.

Alice J. Merryweather, Christoph Schnedermann, Quentin Jacquet, Clare P. Grey, and Akshay Rao. Operando optical tracking of single-particle ion dynamics in batteries. *Nature*, 594(7864):522–528, 2021. doi: 10.1038/s41586-021-03584-2.

Alice J. Merryweather, Quentin Jacquet, Steffen P. Emge, Christoph Schnedermann, Akshay Rao, and Clare P. Grey. Operando monitoring of single-particle kinetic state-of-charge heterogeneities and cracking in high-rate Li-ion anodes. *Nature Materials*, 21(11):1306–1313, 2022. doi: 10.1038/s41563-022-01324-z.

Payel Mukhopadhyay, Stefan S. Nixon, Romain Watteaux, Michael McCabe, Alberto Bietti, Kyunghyun Cho, Cristiana Diaconu, Irina Espejo Morales, David Fouhey, Siavash Golkar, Tom Hehir, Shirley Ho, Jake Kovalic, Geraud Krawezik, Francois Lanusse, Tanya Marwah, Rudy Morel, Mariel Pettee, Helen Qu, Jeff Shen, Hadi Sotoudeh, Stuart B. Dalziel, and Miles Cranmer. Emergent transfer of a physics foundation model from simulation to laboratory turbulence. *arXiv preprint arXiv:2606.01470*, 2026.

Allan H. Murphy. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, 116(12):2417–2424, 1988. doi: 10.1175/1520-0493(1988)116⟨2417:ssbotm⟩2.0.co;2.

Karla Negrete, Marco-Tulio Fonseca Rodrigues, Daniel P. Abraham, and Maureen H. Tang. Forecasting battery electrode performance via electrochemical fluorescence microscopy and machine-learning. *ACS Applied Materials & Interfaces*, 17(50):67906–67913, 2025. doi: 10.1021/acsami.5c17708.

Kong Soon Ng, Chin-Sien Moo, Yi-Ping Chen, and Yao-Ching Hsieh. Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries. *Applied Energy*, 86(9):1506–1511, 2009. doi: 10.1016/j.apenergy.2008.11.021.

Ruben Ohana, Michael McCabe, Lucas Meyer, Rudy Morel, Fruzsina J. Agocs, Miguel Beneitez, Marsha Berger, Blakesley Burkhart, Stuart B. Dalziel, Drummond B. Fielding, Daniel Fortunato, Jared A. Goldberg, Keiya Hirashima, Yan-Fei Jiang, Rich R. Kerswell, Suryanarayana Maddu, Jonah Miller, Payel Mukhopadhyay, Stefan S. Nixon, Jeff Shen, Romain Watteaux, Bruno Régaldo-Saint Blancard, François Rozet, Liam Parker, Miles Cranmer, and Shirley Ho. The Well: A large-scale collection of diverse physics simulations for machine learning. In *Advances in Neural Information Processing Systems (Datasets and Benchmarks Track)*, volume 37, pp. 44989–45037, 2024.

Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

David R. Roberts, Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J. Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig, and Carsten F. Dormann. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8): 913–929, 2017. doi: 10.1111/ecog.02881.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351 of *Lecture Notes in Computer Science*, pp. 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28.

Johannes Schmalstieg, Stefan Käbitz, Madeleine Ecker, and Dirk Uwe Sauer. A holistic aging model for Li(NiMnCo)O_2 based 18650 lithium-ion batteries. *Journal of Power Sources*, 257:325–334, 2014. doi: 10.1016/j.jpowsour.2014.02.012.

Veronika Šedajová, Gabriela Horwitz, Jiho Han, Alice J. Merryweather, George S. Phillips, Vikram S. Deshpande, Norman A. Fleck, Akshay Rao, and Clare P. Grey. Interparticle communication and lithium dynamics in faceted nickel-rich NMC cathodes. *Journal of the American Chemical Society*, 148(4):4097–4109, 2026. doi: 10.1021/jacs.5c15171.

Kristen A. Severson, Peter M. Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H. Chen, Muratahan Aykol, Patrick K. Herring, Dimitrios Fraggedakis, Martin Z. Bazant, Stephen J. Harris, William C. Chueh, and Richard D. Braatz. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5):383–391, 2019. doi: 10.1038/s41560-019-0356-8.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Gogi K. Singh, Gerbrand Ceder, and Martin Z. Bazant. Intercalation dynamics in rechargeable battery materials: General theory and phase-transformation waves in $LiFePO_4$. *Electrochimica Acta*, 53(26):7599–7613, 2008. doi: 10.1016/j.electacta.2008.03.083.

Raymond B. Smith and Martin Z. Bazant. Multiphase porous electrode theory. *Journal of The Electrochemical Society*, 164(11):E3291–E3310, 2017. doi: 10.1149/2.0171711jes.

Raymond B. Smith, Edwin Khoo, and Martin Z. Bazant. Intercalation kinetics in multiphase-layered materials. *The Journal of Physical Chemistry C*, 121(23):12505–12523, 2017. doi: 10.1021/acs.jpcc.7b00185.

Shashank Subramanian, Peter Harrington, Kurt Keutzer, Wahid Bhimji, Dmitriy Morozov, Michael Mahoney, and Amir Gholami. Towards foundation models for scientific machine learning: Characterizing scaling and transfer behavior. In *Advances in Neural Information Processing Systems*, volume 36, pp. 71242–71262, 2023.

Valentin Sulzer, Peyman Mohtat, Antti Aitio, Suhak Lee, Yen T. Yeh, Frank Steinbacher, Muhammad Umer Khan, Jang Woo Lee, Jason B. Siegel, Anna G. Stefanopoulou, and David A. Howey. The challenge and opportunity of battery lifetime prediction from field data. *Joule*, 5(8):1934–1955, 2021. doi: 10.1016/j.joule.2021.06.005.

Corentin Tallec and Yann Ollivier. Can recurrent neural networks warp time? In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=SJcKhk-Ab.

Cheng Tan, Siyuan Li, Zhangyang Gao, Wenfei Guan, Zedong Wang, Zicheng Liu, Lirong Wu, and Stan Z. Li. OpenSTL: A comprehensive benchmark of spatio-temporal predictive learning. In *Advances in Neural Information Processing Systems (Datasets and Benchmarks Track)*, volume 36, pp. 69819–69831, 2023.

John Wang, Ping Liu, Jocelyn Hicks-Garner, Elena Sherman, Souren Soukiazian, Mark Verbrugge, Harshad Tataria, James Musser, and Peter Finamore. Cycle-life model for graphite-$LiFePO_4$ cells. *Journal of Power Sources*, 196(8):3942–3948, 2011. doi: 10.1016/j.jpowsour.2010.11.134.

Chao Xu, Alice J. Merryweather, Shrinidhi S. Pandurangi, Zhengyan Lun, David S. Hall, Vikram S. Deshpande, Norman A. Fleck, Christoph Schnedermann, Akshay Rao, and Clare P. Grey. Operando visualization of kinetically induced lithium heterogeneities in single-particle layered Ni-rich cathodes. *Joule*, 6(11):2535–2546, 2022. doi: 10.1016/j.joule.2022.09.008.

Kaiqi Yang, Yifan Cao, Youtian Zhang, Shaoxun Fan, Ming Tang, Daniel Aberg, Babak Sadigh, and Fei Zhou. Self-supervised learning and prediction of microstructure evolution with convolutional recurrent neural networks. *Patterns*, 2(5):100243, 2021. doi: 10.1016/j.patter.2021.100243.

Shuquan Ye, Ben Fei, Hongbin Xu, Jiaying Lin, and Wanli Ouyang. From static spectra to operando infrared dynamics: Physics informed flow modeling and a benchmark. *arXiv preprint arXiv:2602.18551*, 2026.

Yuan Yin, Ibrahim Ayed, Emmanuel de Bézenac, Nicolas Baskiotis, and Patrick Gallinari. LEADS: Learning dynamical systems that generalize across environments. In *Advances in Neural Information Processing Systems*, volume 34, pp. 7561–7573, 2021a.

Yuan Yin, Vincent Le Guen, Jérémie Dona, Emmanuel de Bézenac, Ibrahim Ayed, Nicolas Thome, and Patrick Gallinari. Augmenting physical models with deep networks for complex dynamics forecasting. In *International Conference on Learning Representations*, 2021b. URL https://openreview.net/forum?id=kmG8vRXTFv.

Quan Zeng, Shahed Rezaei, Luis Carrillo, Rachel Davidson, Bai-Xiang Xu, Sarbajit Banerjee, and Yu Ding. Chemomechanical damage prediction from phase-field simulation video sequences using a deep-learning-based methodology. *iScience*, 27(9):110822, 2024. doi: 10.1016/j.isci.2024.110822.

Han Zhang, Xiaofan Gui, Shun Zheng, Ziheng Lu, Yuqi Li, and Jiang Bian. BatteryML: An open-source platform for machine learning on battery degradation. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=sxGugrYhP9.

Hongbo Zhao, Brian D. Storey, Richard D. Braatz, and Martin Z. Bazant. Learning the physics of pattern formation from images. *Physical Review Letters*, 124(6):060201, 2020. doi: 10.1103/physrevlett.124.060201.

Hongbo Zhao, Haitao Dean Deng, Alexander E. Cohen, Jongwoo Lim, Yiyang Li, Dimitrios Fraggedakis, Benben Jiang, Brian D. Storey, William C. Chueh, Richard D. Braatz, and Martin Z. Bazant. Learning heterogeneous reaction kinetics from X-ray videos pixel by pixel. *Nature*, 621(7978):289–294, 2023. doi: 10.1038/s41586-023-06393-x.

Peichen Zhong, Bowen Deng, Tanjin He, Zhengyan Lun, and Gerbrand Ceder. Deep learning of experimental electrochemistry for battery cathodes across diverse compositions. *Joule*, 8(6):1837–1854, 2024. doi: 10.1016/j.joule.2024.03.010.

GUIDE TO THE APPENDICES

Use this table to go directly from a main-text finding to its evidence and controls. All results retain the support, uncertainty and scope stated at the destination; the guide introduces no additional comparisons.

Empty table cells. Many appendix tables are laid out as full grids of forecasters, splits and horizons, but the experiments were targeted: each forecaster and control was run only on the splits where it tests a stated question, since training every model on every split was not needed and the foundation models are costly to train. Cells marked n/r were not planned, n/a are not applicable (for example, same-particle replay has no earlier recording of a new particle), n/e mark a planned cell that was not evaluated, and – mark an undefined value (no common windows or zero persistence error).

Main-text claim or comparison	Appendix evidence
The benchmark follows matched particles across conditions (Sec. 3).	Appendix A; dataset Tables 3, 4 and 6.
Causal forecasts use retrospective geometry and calibration (Secs. 4–5).	Input audit: Appendix B.3, Table 12; implementation: Appendix C.
The charge clock aids single-rate transfer; rate inputs alter the comparison (Sec. 6.1).	Appendix E; clock Table 45 and progress Table 53.
Charge lead encoding wins at long horizons with identical histories; earlier origins retain headline contrasts (Sec. 6.1).	Appendix E; Tables 54 and 55.
Direct gains depend on horizon and architecture (Sec. 6.2).	Appendix E; Tables 31 and 56.
Some direct U-Net gains come from within-particle changes (Sec. 6.2).	Appendix H.1; Table 108; example in Fig. 7.
Pixel blanking leaves pooled new-particle skill changes unresolved; formation forecasts depend on the input offset (Sec. 6.2).	Appendix B.2, Table 10; Appendix E, Table 60.
First-transition lithiation order agrees at the same rate and temperature; across-rate agreement is unresolved (Sec. 6.3).	Appendix G; Tables 101 and 104; concurrency in Fig. 5.
Slow-to-fast forecasts fall short of repeat agreement; reverse timing is an exception (Sec. 6.3).	Appendix H.2; matched-support Tables 109 and 110.
Poseidon gains depend on recipe, data and budget (Sec. 6.4).	Appendix D; Tables 16, 20 and 22.
Local extrapolation is strong on solid-solution NMC (Sec. 6.5).	Appendix E; Tables 30 and 66.
Further graphite transfer depends on capacity calibration (Sec. 6.5).	Appendix F; Tables 92 and 96.
Further NMC electrodes test rate, ageing and voltage-window transfer (Sec. 6.5).	Appendix F; Tables 94 and 95; NMC811 in Appendix F.1.
The calibrated simulator fails all four claim gates (Sec. 6.3).	Appendix I.6; gate definitions in Table 116.
The step grid has no universal optimum; transition specialists give no consistent gain (Secs. 6.1, 6.3).	Appendix J.4, Fig. 9; Appendix K, Table 132.
Data and evaluation are the contribution (Sec. 1).	Scientific and forecasting context in Appendix L.

A DATASET CONSTRUCTION

This appendix documents sequence construction, calibration and particle matching.

Each HDF5 sequence contains one particle throughout an acquisition, including rests/holds. Graphite's 23-region atlas yields 246 retained sequences (02–14; P01–P22) from one graphite cell. Regions may contain touching flakes; operating conditions and cycling history are confounded.

Table 3: Graphite acquisitions. Rates are nominal; $I_{99.5}$ is the 99.5th percentile of absolute current, $V_{\min}$ the lower cut-off, and Δt the stored median cadence. Charges are integrated-charge ranges at stored half-cycle frames, paired lithiation/delithiation in cycle order. Dashes denote missing metadata; 02 is formation. †From the frame-interval-averaged current stored with the crops (no high-rate current index for 15–16). ‡The delithiation record of 15 contains a second constant-current bout and is not tabulated.

ID	Rate	$I_{99.5}$ (μA)	T (°C)	$V_{\min}$ (mV)	Δt (s)	Frames	Half-cycle charges (μAh)
02	C/20	7.34	25	—	120.20	2101	209.5/282.5
03	C/20	12.84	25	5	120.23	1380	279.7/276.8
04	C/20	12.84	45	5	120.22	1422	304.5/271.7
05	C/20	12.84	15	5	120.21	1269	265.9/272.2
06	C/20	12.84	25	5	120.20	1334	264.1/276.2
07	C/2	130.27	15	—	30.11	290	124.1/120.5
08	C/2	130.27	15	1	30.14	290	122.1/122.3
09	C/2	130.27	6	1	30.13	230	90.4/89.3
10	C/2	130.27	25	1	30.17	399	186.4/192.6
11	C/2	130.27	25	5	30.16	592	133.5/135.1; 152.0/153.4
12	C/2	130.27	35	5	30.17	457	213.6/214.4
13	C/2	130.27	45	5	30.18	521	249.3/247.9
14	C/2	130.27	25	5	30.23	1385	178.1/174.4; 176.7/176.7; 177.4/177.6; 178.2/178.0
15	2C	270.22†	25	5	4	4255	216.8/—‡
16	2C	270.33†	25	5	4	5000	213.9/224.0; 119.4/106.5; 106.4/105.8; 105.2/104.9
17	2C	270.29	45	5	4	6132	256.2/263.2; 169.1/155.3; 155.1/154.4; 154.5/153.5
18	2C	270.40	15	5	4	4862	184.8/194.2; 80.5/65.4; 66.8/66.6; 66.7/66.4

Graphite registration correlates high-passed log images on boundary collars (dilate 3, exclude erode 2 pixels), avoiding changing interiors and stationary background. Every frame is matched independently to fixed references with subpixel refinement. Activity masks mapped to file-04 START define the atlas by an eight-of-14 vote and watershed splitting. Placement uses START-anchored transforms and ± 6-pixel local refinement. Validity requires acceptable placement/background flags, correlation ≥ 0.30, refinement ≤ 5.5 pixels, area 0.60–1.60 times the particle's median, and $\geq 50\%$ usable fault-free frames. Acquisition 01 is excluded for rotation/rescaling uncorrected by translation-only extraction.

Population crops are 64×64 at 0.26 μm/pixel, bilinearly sampled from 4×4 mean-binned intensities into each acquisition's START frame, retaining every fourth source frame at C/20 and sixth at C/2. Stored frames are $\log(\mathrm{raw}/\mathrm{ref})$; reference images are medians of the first three good frames. The default signal subtracts the frame's median log ratio over background outside all masks dilated three pixels, reverting to the log ratio if fewer than 50 background pixels or no finite estimate exist. Full masks and one-pixel-eroded cores are retained. Quality 0 means good, 1 located artefacts (retained), and 2 unobservable (excluded from inputs/targets without bridging gaps). Missing registration or nonfinite mask pixels trigger quality 2, including 102 frames in 10 and the last 326 in 11.

Zoom files 15–18 use 5×5 mean binning, Gaussian smoothing and every fourth raw frame. Similarity transforms anchored at the start of each acquisition ($2.48\times$ magnification, $-114.7°$ rotation) and per-frame displacements map P01/P09/P15 onto file-04 crop grids. The transforms give each particle a common orientation across acquisitions. The acquisition records list 258 sequences, but

the binned indices supplied cover only 17–18; for 15–16, Table 3 gives frame counts, current and charges computed from the stored crops.

Trapezoidal integration of high-rate current gives $q(t) = -\int_0^t I(\tau)\,d\tau$ after unit conversion: negative current lithiates both materials. Charge and voltage are interpolated to image times; current is interval-averaged. From each segment's first stored CC frame t_0, we use $u = d[q(t) - q(t_0)]/Q_{\mathrm{ref}}$ and $s = q(t)/Q_{\mathrm{ref}}$, with $d = +1/-1$ for lithiation/delithiation. Fixed Q_{ref} is $280\,\mu$Ah (typical slow-rate graphite capacity) or $560.93\,\mu$Ah (measured NMC C/3 charge, including when C/3 is held out). Graphite starts at OCV ≥ 0.77 V except 02 (0.18 V), supporting a delithiated origin for "absolute SOC" elsewhere; 02 retains an unknown offset. NMC's zero is its acquisition origin. These electrode charge coordinates do not calibrate particle stoichiometry. A fraction of a half-cycle's own capacity requires its future endpoint and is not causal. Rests/holds are excluded from forecasting segments.

NMC movies from Xu et al. (2022) retain every source-stabilised frame (Table 4), 2×2-binned into 48×48 crops at $0.0694\,\mu$m/pixel. Background-weighted Lucas–Kanade shifts, median-filtered over nine frames, are applied by cubic interpolation; maximum residual drift is 0.216 binned pixels. Inter-acquisition matching uses translations. References use the initial 12/19/28 frames, at least 10 s before current onset; zero-filled borders are avoided and cores eroded two pixels. Signal/quality conventions match graphite; 7/8/7 blurred frames have quality 1. Timing uses the documented $+15$ s offset, not independently identifiable from optical kinks (estimates 0.3–37 s). Temperature is unmeasured; stored 25°C is nominal. Cycle offsets 4/9 for 1C/2C remain unverified.

Table 4: NMC acquisitions: OCV, CC charge to nominal 4.3 V, CC discharge to 3.0 V, 3.0 V hold, then OCV. I is charge current; capacities are positive high-rate-record integrals, with $Q_{\mathrm{d,all}}$ including the hold.

ID	Rate	I (μA)	Δt (s)	Frames	Q_{c}	$Q_{\mathrm{d,CC}}$ (μAh)	$Q_{\mathrm{d,all}}$
04	C/3	172.02	10.000	3060	560.93	471.78	512.26
05	1C	515.00	5.989	3000	489.76	403.65	461.76
10	2C	1029.52	3.996	3150	392.67	317.03	385.15

NMC segmentation uses 98th–2nd percentile activity over ten-frame medians, watershed seeds at log amplitude 0.10, and linear-activity half-maximum boundaries above local background. Acceptance requires median log amplitude ≥ 0.15, positive dominant contrast, area ≥ 25 pixels, solidity ≥ 0.75, and a two-pixel border margin. Every acquisition must pass, with all three pairwise links satisfying mutual best matching, IoU ≥ 0.5, centroid distance ≤ 2 pixels and runner-up IoU < 0.2. Accepted masks use a two-of-three vote. A retrospective gate requiring median raw-frame mask IoU ≥ 0.5 over seven QA frames per acquisition demoted six identities: 24 accepted particles (72 sequences) remain, with 30 provisional identities (54 sequences) excluded from the main cross-rate test set. Provisional identities enter training and the new-particle folds only. Multi-lobed regions remain flagged. Both datasets use complete acquisitions to curate static masks; neither supplies online segmentation or calibrated particle lithium content (App. B.3).

A.1 Further electrodes

We document the construction and scope of the additional-electrode datasets.

Four further cells were recorded in separate experimental runs with the same kind of setup (Table 6). Each dataset has its own pre-registration (object definition, gates, Q_{ref}, splits, sweeps and reporting rule), written before any forecasting result on it; later changes were appended as dated amendments with their reasons, and all are released with the code.

Graphite cells B and C. The raw movies are registered within each acquisition (phase correlation, then a similarity transform; rotations 0.21° and 0.27°), and the main flake of each cell is segmented by its optical activity; its identity across the cell's two acquisitions passes the pre-declared rule (mask IoU 0.97 and 0.94, centroid residuals 0.48 and 0.90 common pixels). The pre-registered gate asked for at least eight objects per cell. It was not met: apart from the flake, the fields contain only small

Table 5: Candidate electrode staging peaks (V versus Li) from incremental-capacity analysis; first cycles unless specified. Dashes mean unresolved/outside support, not phase absence; † indicates ambiguous assignment. Exploratory detection uses 0.5 mV bins, 4 mV Gaussian smoothing, 3.5% prominence and 16 mV separation. The note contains no 2C peaks.

Acquisition	Lithiation: dilute / 2L $\rightarrow$ 2 / 2 $\rightarrow$ 1	Delithiation: 1 $\rightarrow$ 2 / 2 $\rightarrow$ 2L / higher
02 (formation)	— / 0.116 / 0.078	0.101 / 0.141 / 0.225
03	0.200 / 0.111 / 0.074	0.105 / 0.146 / 0.228
04	0.202 / 0.110 / 0.078	0.095 / 0.130 / 0.220
05	0.187 / 0.100 / 0.061	0.122 / 0.159 / 0.238
06	0.199 / 0.111 / 0.074	0.106 / 0.146 / 0.229
07	— / — / —	— / 0.251 / 0.346
08	— / — / —	— / $0.252^{\dagger}$ / 0.348
09	— / — / —	— / — / —
10	0.138 / — / $0.026^{\dagger}$	0.130 / 0.148 / 0.263
11 (cycle 1)	0.124 / 0.058 / —	— / 0.228 / 0.323
11 (cycle 2)	0.112 / 0.026 / —	— / 0.211 / —
12	0.151 / 0.060 / 0.019	0.151 / 0.193 / 0.269
13	0.172 / 0.081 / 0.048	0.125 / 0.163 / 0.248
14 (cycles 1–4)	0.123–0.140 / 0.033–0.039 / —	— / 0.207–0.208 / —

Table 6: Further electrodes. Frames are the stored frames per acquisition and Δt the stored cadence; objects are those in the test sets. Graphite cells B and C are used for evaluation only (no training on them).

Cell	Material	Acquisitions	Frames	Δt (s)	Objects	Q_{ref} (µAh)
B	graphite	C/20 formation; C/2	1,801; 701	120; 30	1 flake	126.2 (162.4)
C	graphite	C/20 formation 1; 2	755; 559	300; 300	1 flake	164.6 (167.1)
NMC-2	NMC	0.5C, 1C, 2C, 3C	3,333–1,500	3.0–1.0	5 particles	466.2
NMC811	NMC811	14 C/2 cycles, 3–158	per cycle	10	60 particles	665.6

optical regions, which the experimenters judged too unreliable to identify and quantify as particles, so each cell contributes its flake alone (a 64×64 crop centred on it), no model is trained on these cells, and all statistics are descriptive moving-block intervals (blocks of ten consecutive forecast origins within a half-cycle). This scope change was made after the history-only baseline scores on the earlier, multi-region version had been inspected but before any neural model was evaluated on these cells. The current sign and electrochemistry rows were verified (negative current lithiates, as in the main dataset); charge is the trapezoidal integral of the measured current, interpolated to the frames, and constant-current segments exclude rests and voltage-hold tails. The echem–frame offset is assumed zero: optical responses at current reversals lie within two stored frames of the current switch except in cell C's second formation cycle (three frames, 15 min at a 5 min cadence, about 0.01 in s). The native pixel size is assumed equal to the main electrode's (0.065 µm); edge-width fits suggest 0.076–0.086 µm, so every evaluation is repeated with a frozen 0.084 µm anchor. Both native-calibration exports retain $0.26\,\mu$m/pixel output crops; the assumed native scale changes the sampled footprint. Q_{ref} is the first complete C/20 constant-current delithiation capacity; cell B's comes from its formation cycle, whose lithiation stopped early, so its C/2 delithiation passes $1.29\,Q_{\text{ref}}$, beyond the main electrode's range, and every evaluation is repeated with the capacity-matched values in parentheses (the largest complete constant-current delithiation). An independent audit recomputed the processing from the raw files with separate code and passed all seven of its checks. The capacity-matched scale is retrospective: it uses the evaluated C/2 delithiation in cell B and second formation delithiation in cell C. For cell B, $\Delta = 0.16$ therefore represents 20.19 µAh under the primary scale and 25.98 µAh under the sensitivity; histories, origins and eligibility also change. These are different physical forecast targets, not an isolated test of correcting capacity. The amendment preceded inspection of cross-electrode neural scores, after history-only baselines had already been inspected.

Second NMC electrode (NMC-2). One field of a single-crystal NMC composite electrode was imaged in four one-cycle acquisitions (constant-current charge to 4.4 V at 357.5, 714.5, 1428.6 and 2142.6 µA). Segmentation, identity and quality rules are those of the first NMC dataset, except that the activity threshold is scaled by each rate's constant-current charge swing (the 3C swing is 40% of the 0.5C swing, so a fixed threshold is biased against fast rates); this amendment was made before any forecasting result. Six particles are accepted at all four rates, five of which fit the 48×48 crop and form the test sets, below the pre-registered gate of eight, so the dataset is a descriptive replication; provisional particles are used for training only. The documented $+15$ s echem offset was verified (optical turning points within 6 s of the current reversal at every rate), and Q_{ref} is the measured 0.5C charge capacity. Splits: 0.5C to 1C, 2C and 3C; 3C to 0.5C; and 0.5C with 3C to 1C and 2C.

Multi-cycle NMC811. One NMC811 coin cell was imaged in sessions over 160 cycles at C/2 (500.9 µA), with the upper cut-off raised from 4.2 V to 4.4 V and later to 4.6 V. We use two consecutive cycles from each of seven life stages (cycles 3–4, 28–29, 56–57, 99–100 and 116–117 at 4.2 V; 127–128 at 4.4 V; 157–158 at 4.6 V), chosen for the fewest autofocus steps; a session whose activity maps are dominated by interference rings is excluded. Particles are tracked from master masks of the first session into every session (shape IoU after centring ≥ 0.5, displacement residual ≤ 3 pixels); the best-matched 60 of 72 accepted particles are stored as 40×40 crops at 95 nm per pixel, and frames with autofocus steps or blur are flagged. The charge capacity falls from 669 to 631 µAh between cycles 3 and 117 and rises to 837 and 871 µAh with the higher cut-offs; Q_{ref} is that of the earliest clean C/2 cycle. Splits: cycles 3–4 and 28–29 to cycles 99–100 and 116–117 at 4.2 V (ageing), the same training cycles to the 4.4 and 4.6 V cycles (voltage window, confounded with cycle number), and new particles (five folds).

B Data audits: crop overlap, pixel blanking, input exposure and reproduction

This appendix audits information exposure and code-package reproduction.

This appendix audits what information reaches a forecaster: whether pixels of a held-out particle occur in training crops (crop overlap), a retraining control that removes them (pixel blanking), which inputs and preprocessing steps are causal and which use completed acquisitions (input exposure),

and whether a table entry can be regenerated from the anonymous code package alone (reproduction).

B.1 Crop overlap in the new-particle folds

We measure how held-out particles' pixels enter neighbouring training crops. The new-particle folds hold out particle *identities* inside one shared field of view. A held-out particle can therefore appear, unlabelled, in the crops of its training neighbours. For a held-out sequence i at stored frame t, let C_i be its crop lattice, M_i its target mask and U_t the union of the crop rectangles of all training sequences of the same acquisition that are available at t. We report $|C_i \cap U_t|/|C_i|$ (crop exposure) and $|M_i \cap U_t|/|M_i|$ (mask exposure), summing numerators and denominators over the observable constant-current frames (quality $\neq 2$) of each sequence, particle, fold or material. Overlapping training crops are unioned, so every pixel is counted once, and the crop denominator is the nominal 64×64 (graphite) or 48×48 (NMC) raster. The count is geometric: it uses the stored crop origins and the per-frame registration shifts, which are identical for all crops of an acquisition (maximum discrepancy 0); crop-specific references and background subtraction mean that the same location need not carry identical intensities in two crops, and no overlap across acquisitions is counted. Two training pools are audited: the split's full training pool before validation is carved out, and the block-validation training sets of seeds 0–2 that the learned models actually use. The audit measures an information route, not how much skill that route contributes.

Table 7 gives the per-fold exposure. Before validation exclusions, neighbouring training crops cover 33.4% of the held-out graphite mask pixel-frames and 48.2% of the accepted NMC ones; 12 of 22 graphite and 22 of 24 NMC held-out particles have some mask exposure. After each seed's validation exclusions the mask exposure is 27.6–29.5% (graphite) and 42.2–44.1% (NMC). In every audited sequence the intersection of a held-out target mask with the *training target masks* is empty: the exposure comes from the surroundings of training particles, not from overlapping labels. The configured graphite fold 2 lists a 23rd atlas region that has no retained sequence; it is missing rather than unexposed. The NMC folds were configured over 30 identities, six of which later failed the identity gate (App. A); including them changes the pooled NMC mask exposure to 49.0%.

Table 7: Spatial exposure of held-out particles in the new-particle folds, pooled over observable constant-current frames, before validation exclusions. *Crop* and *mask*: fraction of the held-out crop and mask pixel-frames covered by same-acquisition training crops. *Equal-particle*: mask fraction averaged with equal weight per particle. *Exposed*: held-out particles with any covered mask pixel. Per-particle and per-sequence counts are in the supplementary audit tables.

Material	Fold	Particles	Sequences	Crop (%)	Mask (%)	Equal-particle (%)	Exposed
Graphite	0	5	53	44.2	50.0	31.3	2
Graphite	1	5	45	37.7	13.9	34.3	2
Graphite	2	4	46	27.5	10.9	25.2	2
Graphite	3	4	42	58.9	32.7	66.8	4
Graphite	4	4	41	53.0	63.2	49.9	2
Graphite	all	22	227	43.8	33.4	40.7	12
NMC	0	5	15	58.8	49.1	64.9	5
NMC	1	6	18	64.3	29.1	55.4	5
NMC	2	4	12	73.2	55.9	73.2	4
NMC	3	5	15	79.6	76.4	67.2	4
NMC	4	4	12	62.8	40.4	54.8	4
NMC	all	24	72	67.6	48.2	62.7	22

An earlier, incomplete control. One loss-masking control existed before this audit: on graphite fold 4 (four held-out particles) a U-Net told (d, u, s) was retrained with zero loss weight outside the target mask. With the same two seeds and exactly matched windows, horizons and particles, skill changes from 0.483 to 0.482 at $\Delta = 0.02$ and from 0.574 to 0.151 at $\Delta = 0.16$; the per-seed changes at $\Delta = 0.16$ are -0.471 $[-0.880, -0.204]$ and -0.374 $[-0.488, -0.247]$ (paired particle bootstrap, 1,000 resamples). This control cannot separate information from optimisation: it keeps the neighbouring pixels in the *inputs*, removes direct supervision on them, and changes the objective

of every training crop; it also covers one fold only. It motivated the blanking control below, which removes held-out pixels from inputs and targets while leaving the objective unchanged.

Skill against exposure in the existing models (exploratory). For the existing (d, s) U-Nets of all ten folds (seeds 0 and 1), Table 8 compares held-out particles with and without any mask exposure, using equal particle weights (seed skills are averaged per particle), 5,000 particle-bootstrap resamples within each group and a Spearman correlation between particle skill and exposure fraction. Exposed graphite particles have higher long-horizon skill and exposed NMC particles lower skill. With 22 graphite and 24 NMC particles, and only two unexposed NMC particles, this association is weak evidence either way: particle size, intrinsic predictability, rate and direction coverage and fold-specific models can all covary with exposure. It neither certifies absence of an exposure effect nor replaces the retraining control.

Table 8: Existing U-Net (d, s) skill of held-out particles with and without spatial exposure (equal particle weights; 95% particle-bootstrap intervals, 5,000 resamples within each group). ρ: Spearman correlation of particle skill with mask exposure fraction (exposure measured on the full training pool). Exploratory.

Material (exp. / unexp.)	Δ	Exposed skill	Unexposed skill	Difference	ρ
Graphite (12 / 10)	0.02	0.469 [0.453, 0.486]	0.394 [0.280, 0.496]	0.075 [−0.029, 0.193]	0.172 [−0.284, 0.596]
	0.16	0.623 [0.569, 0.674]	0.457 [0.295, 0.593]	0.166 [0.021, 0.329]	0.274 [−0.180, 0.636]
NMC (22 / 2)	0.02	0.676 [0.617, 0.733]	0.820 [0.748, 0.891]	−0.143 [−0.256, −0.033]	−0.389 [−0.666, −0.012]
	0.16	0.817 [0.735, 0.878]	0.927 [0.887, 0.967]	−0.111 [−0.209, −0.024]	−0.126 [−0.493, 0.250]

B.2 Pixel-blanking control for new particles

We test whether removing held-out particles' pixels changes forecast skill. The blanking control retrains the new-particle models with every pixel within a Euclidean distance D of any held-out particle's mask set to the background fill value in *all* training and validation frames, before windows are created. Distances are measured on the registered pixel-centre lattice of each acquisition, the dilation extends beyond crop edges, and the union over held-out particles is blanked in every training and validation crop of that acquisition. The fill value is 0, the value of the background-subtracted log contrast used by all models (the grid's missing-pixel fill is also 0). Masks, condition channels and their standardisation statistics, window validity, loss weights (4 inside, 1 outside the target mask), the optimiser, the step budget (6,000 steps, $k = 4$, four unrolled steps, batch 32, augmentation) and the test data are unchanged; without the flag, training is bit-identical to the default trainer. Verification checks, on the full training grids of fold 0 for both materials, that every covered training or validation pixel equals the fill value and that every uncovered pixel and every test pixel is unchanged, and one before/after image per acquisition was inspected. The primary distance is $D = 2$ pixels, which blanks 0.3–2.1% of the graphite and 0.8–1.5% of the NMC training and validation crop pixel-frames, depending on the fold; $D = 5$ (3.0% and 2.8% on fold 0) is a sensitivity control, because in the inspected images some optical signal extends beyond the mask plus two pixels, especially for NMC. $D = 2$ is therefore described by its exact spatial exclusion, not as removal of every optical halo.

The design comprises 24 retrained U-Nets on the charge clock (Table 9); each is paired with an existing default run of identical split, seed, conditioning and hyperparameters, and evaluated with causal origins on the unchanged test sets. The comparison of interest is the paired, seed-matched difference in skill between the blanked and the default model on exactly matched test windows, by material and horizon, with particle-cluster bootstrap intervals. The state-conditioned fold-4 arm connects the new control to the earlier loss-masking control on the same seeds.

Retraining all five folds with these pixels removed changes pooled new-particle skill at $\Delta s = 0.16$ by $+0.006$ [-0.026, 0.017] for graphite and $+0.002$ [-0.009, 0.005] for NMC, although one graphite fold drops by 0.086 [0.069, 0.114] (Table 10); new-particle results are therefore reported as identity-held-out forecasts with supplied geometry whose pooled skill is not driven by this exposure.

B.3 Input exposure: causal and retrospective inputs

We classify the information available to each forecaster. The benchmark is offline: images are registered, segmented, quality-flagged and placed on the charge axis using completed acquisitions.

Table 9: Pixel-blanking design (U-Net, charge clock, grouped validation as in the default runs, causal evaluation).

Arm	Folds	Seeds	Conditioning, D	Runs
Main, graphite	0–4	0, 1	(d, s), 2 px	10
Main, NMC	0–4	0, 1	(d, s), 2 px	10
Link to loss-masking control, graphite	4	0, 1	(d, u, s), 2 px	2
Distance sensitivity, graphite and NMC	0	0	(d, s), 5 px	2

Table 10: Pixel-blanking new-particle control. Training and validation pixels within D pixels of any held-out particle mask are set to zero; the comparator is the identical unblanked configuration and seed, scored on identical windows. Difference = blanked minus comparator skill, 95% paired particle-cluster interval; the five-fold row weights folds equally and needs all five folds. Pending rows have no finished records yet.

Material	Fold	Inputs	D (px)	Comparison	Δ	Blanked	Comparator	Difference [95% CI]	Seeds
NMC	NMC particle fold 0	(d,s)	2	blanked minus unblanked	0.020	0.744	0.752	-0.008 [-0.045, 0.001]	2
NMC	NMC particle fold 1	(d,s)	2	blanked minus unblanked	0.020	0.811	0.800	0.010 [-0.001, 0.045]	2
NMC	NMC particle fold 2	(d,s)	2	blanked minus unblanked	0.020	0.805	0.806	-0.000 [-0.009, 0.006]	2
NMC	NMC particle fold 3	(d,s)	2	blanked minus unblanked	0.020	0.771	0.757	0.014 [0.012, 0.020]	2
NMC	NMC particle fold 4	(d,s)	2	blanked minus unblanked	0.020	0.703	0.710	-0.007 [-0.022, 0.005]	2
NMC	NMC particle fold 0	(d,s)	5	blanked minus unblanked	0.020	0.758	0.765	-0.007 [-0.017, 0.001]	1
graphite	New particle, fold 0	(d,s)	2	blanked minus unblanked	0.020	0.496	0.525	-0.029 [-0.031, -0.014]	2
graphite	graphite particle fold 1	(d,s)	2	blanked minus unblanked	0.020	0.220	0.189	0.030 [-0.005, 0.044]	2
graphite	graphite particle fold 2	(d,s)	2	blanked minus unblanked	0.020	0.526	0.526	0.000 [-0.002, 0.009]	2
graphite	graphite particle fold 3	(d,s)	2	blanked minus unblanked	0.020	0.464	0.468	-0.004 [-0.015, -0.001]	2
graphite	graphite particle fold 4	(d,s)	2	blanked minus unblanked	0.020	0.470	0.468	0.002 [-0.008, 0.014]	2
graphite	New particle, fold 0	(d,s)	5	blanked minus unblanked	0.020	0.506	0.515	-0.010 [-0.028, -0.002]	1
graphite	graphite particle fold 4	(d,u,s)	2	blanked minus unblanked	0.020	0.480	0.483	-0.003 [-0.006, 0.000]	2
graphite	graphite particle fold 4	(d,u,s)	2	blanked minus outside-weight-zero	0.020	0.480	0.482	-0.002 [-0.005, 0.003]	2
NMC	5-fold mean	(d,s)	2	blanked minus unblanked	0.020	0.767	0.765	0.002 [-0.008, 0.009]	2
graphite	5-fold mean	(d,s)	2	blanked minus unblanked	0.020	0.435	0.435	-0.000 [-0.007, 0.005]	2
NMC	NMC particle fold 0	(d,s)	2	blanked minus unblanked	0.160	0.873	0.864	0.008 [-0.026, 0.014]	2
NMC	NMC particle fold 1	(d,s)	2	blanked minus unblanked	0.160	0.916	0.910	0.005 [-0.017, 0.008]	2
NMC	NMC particle fold 2	(d,s)	2	blanked minus unblanked	0.160	0.941	0.944	-0.003 [-0.010, 0.005]	2
NMC	NMC particle fold 3	(d,s)	2	blanked minus unblanked	0.160	0.883	0.892	-0.008 [-0.034, 0.000]	2
NMC	NMC particle fold 4	(d,s)	2	blanked minus unblanked	0.160	0.876	0.869	0.007 [-0.003, 0.018]	2
NMC	NMC particle fold 0	(d,s)	5	blanked minus unblanked	0.160	0.889	0.874	0.015 [0.003, 0.038]	1
graphite	New particle, fold 0	(d,s)	2	blanked minus unblanked	0.160	0.661	0.747	-0.086 [-0.114, -0.069]	2
graphite	graphite particle fold 1	(d,s)	2	blanked minus unblanked	0.160	0.188	0.082	0.106 [-0.025, 0.146]	2
graphite	graphite particle fold 2	(d,s)	2	blanked minus unblanked	0.160	0.624	0.633	-0.009 [-0.032, -0.006]	2
graphite	graphite particle fold 3	(d,s)	2	blanked minus unblanked	0.160	0.695	0.674	0.021 [-0.004, 0.023]	2

Table 12 classifies each input or preprocessing operation as *causal* (C, available at the forecast origin), *causal given a stated assumption* (C*: a calibration, a completed reference library or a known protocol) or *retrospective* (R, constructed from later frames, completed half-cycles or the whole collection). A field counts as an exposure even when it is not a network channel, for instance when it selects windows or defines the loss. Future labels of *training* acquisitions are ordinary supervision (R relative to a training origin, C* relative to deployment). The causal-origin rule used for every reported number keeps future images out of the history; it does not make the geometry or the calibration causal. Training grids are built offline by two-sided interpolation of completed acquisitions, whereas evaluation snaps each origin back to a stored frame and interpolates histories only from the observed prefix.

Condition sets are defined in App. C; temperature, rate, cycle and voltage are not inputs of the headline (d, s) networks merely because they are stored with the data. A supplementary table enumerates all 98 saved training configurations (family, condition vector, history length, clock, validation grouping, loss weights and reference-pool settings); a saved configuration is not evidence of a completed result. Replay without a usable reference falls back to persistence rather than to any held-out target, and history-aligned replay and population residual fit their shift or gain on the observed target history only.

Disclosure. We evaluate forecasting on retrospectively curated, registered optical crops with supplied particle geometry and a fixed charge calibration. Causal origins restrict image histories to observations available at the forecast origin. Segmentation, registration, quality curation and part of the calibration use completed acquisitions; cross-clock mapping and step conditioning use the recorded electrochemical trajectory, so a prospective deployment would need a known protocol and preprocessing fixed on prior data, and reference-based methods additionally assume a completed

Table 11: Pixel-blanking new-particle control. (continued).

Material	Fold	Inputs	D (px)	Comparison	Δ	Blanked	Comparator	Difference [95% CI]	Seeds
graphite	graphite particle fold 4	(d,s)	2	blanked minus unblanked	0.160	0.491	0.491	-0.000 [-0.034, 0.036]	2
graphite	New particle, fold 0	(d,s)	5	blanked minus unblanked	0.160	0.652	0.769	-0.117 [-0.133, 0.025]	1
graphite	graphite particle fold 4	(d,u,s)	2	blanked minus unblanked	0.160	0.616	0.574	0.042 [-0.012, 0.078]	2
graphite	graphite particle fold 4	(d,u,s)	2	blanked minus outside-weight-zero	0.160	0.616	0.151	0.465 [0.205, 0.725]	2
NMC	5-fold mean	(d,s)	2	blanked minus unblanked	0.160	0.898	0.896	0.002 [-0.009, 0.005]	2
graphite	5-fold mean	(d,s)	2	blanked minus unblanked	0.160	0.532	0.526	0.006 [-0.026, 0.017]	2

Table 12: Input exposure ledger (condensed). C: causal; C*: causal under the stated assumption; R: retrospective. The full ledger, with the inputs of every forecaster family and code anchors, is in the supplementary material.

Input or operation	Label	Construction and what a prospective use would need
Observed image history	C / R	Frame values are observed; registration, background region and quality flags applied to them use completed acquisitions (below).
Masks, crop placement, identity	R	Graphite atlas and placement from whole-acquisition activity and votes across acquisitions; NMC masks from whole-movie activity with a vote over three acquisitions. Needs geometry fixed before forecasting or online segmentation.
Background correction	C / R	Per-frame median is causal; the particle-free region excludes whole-movie particle regions.
Intensity reference	C*	First good frames (graphite) or pre-onset frames (NMC); frame selection uses quality labels.
Registration	R	Graphite: retrospective atlas and drift cache. NMC: residual shifts median-filtered over nine frames (up to four future frames) after upstream stabilisation.
Segments, half-cycle labels	C / R	Timestamps and current sign are causal; current thresholds and segment boundaries are estimated from whole records.
Charge u, s and Q_{ref}	C*	Coulomb counting from a known initial anchor with fixed Q_{ref} (graphite 280 µAh; NMC the completed C/3 charge, so the reverse-rate NMC test uses a held-out-acquisition calibration). Neither coordinate is a calibrated particle composition.
Future step conditions, cross-clock mapping	R	For the time and frame clocks and step-conditioned models they come from the recorded future current and time; prospective use needs the imposed protocol.
Direction, rate, cycle	C	Commanded current sign and nominal rate; cycle is the local half-cycle index.
Temperature	C*	Stated set point (graphite); nominal and unmeasured for NMC. Not an input of the headline (d, s) models.
Voltage	C / R	Voltage held at the origin is causal; the future-voltage oracle is retrospective and appears only in a separately labelled oracle variant.
Quality and accepted-particle flags	R	Curation from multiple frames and acquisitions (NMC blur flags use a centred nine-frame filter).
Activity strata	R	Computed from the observed future; evaluation labels only (App. D, K).
Other acquisitions (replay, reference models)	C*	Valid only if those acquisitions were completed and registered before the forecast origin; chronological precedence is not checked by the splits.

library of earlier acquisitions. Charge coordinates and optical contrast are measurement proxies, not calibrated compositions.

B.4 Reproduction from the code package

We check whether the anonymous code package regenerates reported scores. To check that the released code alone regenerates a table entry, the anonymous code package was built, extracted into an empty directory and run on CPU against a single processed graphite file, without checkpoints, saved predictions or any file from the development checkout. Persistence and history-aligned replay on C/20→C/2 (causal origins, origin lattice every 0.02 from $u = 0.14$, acquisition reference pool,

Table 13: Per-model architecture and training limits. All optimisers are AdamW with peak learning rate 10^{-3} and weight decay 10^{-4}. Patience counts validation checks, spaced 500 steps apart; steps are maximum optimiser updates. The two T0 smoke runs use 300 steps.

Model	Architecture / widths	Optimiser	Steps	Batch	Unroll	Patience
U-Net	Four levels: 32, 64, 128, 256; seven residual blocks	AdamW	6,000	32	4	8
ConvLSTM	Stride-2 encoder; two width-64 cells; decoder 128, 64	AdamW	6,000	32	4	8
FNO	Four layers; width 64, rank 32, modes 16	AdamW	6,000	32	4	8
ViT	Patch 8; width 192; six layers and six heads	AdamW	6,000	32	4	8
`ac_rd`	MLP $c \to 16 \to 16 \to 4$; global D	AdamW	6,000	32	4	8
`fick`	MLP $c \to 16 \to 16 \to 1$; global D, L, b_0	AdamW	6,000	32	4	8
`unet_ref`	U-Net; eight reference images and availability flag	AdamW	6,000	32	4	8

$Q_{\text{ref}} = 280\,\mu\text{Ah}$) reproduce the four stored headline cells to floating-point precision: replay skill 0.248 and 0.655 at $\Delta = 0.02$ and 0.16 on 7,585 and 5,005 windows of 22 particles, with maximum absolute differences of 2.2×10^{-16} in skill and 5.6×10^{-17} in masked RMSE, and identical window counts. The run took 31 s on one core and scored 416 test segments against 144 training reference segments. The causal- origin flag is required (it is not the evaluator default). The package contains the code, split and identity manifests, sweep commands, the audit tables cited in this appendix and a processed-data sample on which this check runs, but no model weights. The full processed datasets of the five previously unpublished cells and of the third-party NMC electrode are released separately; the code can also rebuild the NMC dataset from the original source data. Foundation-model adapters require the public pretrained weights, which are not redistributed. The final anonymous archive passed 295 CPU clean-room tests with no failures or errors (one CUDA test and four optional sample-grid tests were skipped), reproduced the bundled sample and stored baselines from two processed-only released graphite files, regenerated the headline, Walrus and forecast-progress tables from saved numerical records, reran the particle-change and growth statistics from bundled inputs, verified the bundled dependence-sensitivity (A8) and NMC811 per-cycle summaries, reran the full forecast-progress diagnostic with external saved records, and had no unresolved identifier or scrub-replacement scan hits.

C Implementation details

This appendix specifies the models, baselines and evaluation protocol.

Architectures and hyperparameters. Tables 13 and 14 describe the default constructors and the combinations in the stored sweep commands. All use four history frames. The U-Net concatenates history and mask, uses two 3×3 convolutions per residual block with GroupNorm and SiLU, average-pool downsampling, bilinear upsampling and skip concatenation. Each FiLM map has a 64-unit hidden layer and predicts channel scales and shifts. ConvLSTM encodes each image with the mask at half resolution, processes the four frames with two recurrent layers, and decodes through FiLM-conditioned residual blocks; recurrent states reset at each forecast step. FNO lifts history, mask and two spatial coordinates, combining four spectral convolutions with local 1×1 paths and FiLM. The two Fourier corners have independent complex weights, with 16 modes per axis and a rank-32 channel bottleneck. ViT embeds 8×8 image patches and the mask, adds learned temporal and interpolated spatial positions, and appends a condition token; its six pre-normalised transformer layers have six heads, dimension 192, feed-forward width 768, GELU and zero dropout. Final residual heads are initialised to zero.

Grey-box discretisation. Both models use one explicit update per clock step, with effective step length one: dt and du are optional MLP inputs, not multipliers of the update. Diffusion is a global

Table 14: Exact real parameter counts at 64×64, $k = 4$, for conditioning dimensions used in the sweeps. A dash denotes an unswept combination. The $c = 7$ sets are S1's legacy *known* and NMC's *known* without temperature; $c = 8$ is full *known*. Reference models additionally condition on three reference features. Counts also apply at 48×48; FNO counts real and imaginary components separately.

Model	None ($c = 0$)	State ($c = 3$)	Stateplus ($c = 5$)	$c = 7$	Known ($c = 8$)
U-Net	2,058,657	2,245,281	2,247,073	2,248,865	2,249,761
ConvLSTM	—	—	—	1,067,201	—
FNO	—	—	—	4,280,449	—
ViT	—	—	—	2,795,392	—
`ac_rd`	5	405	437	469	485
`fick`	4	356	—	420	—
`unet_ref`	—	2,250,849	—	—	2,255,329

softplus-transformed scalar initialised to $D = 0.01$. Fluxes vanish across mask boundaries, holes and crop edges. For `fick`, the increment inside the mask is $D\nabla_m^2 x + b(\mathbf{c})\exp(-r/L) + b_0$, where r is Euclidean distance in pixels, zero at the first interior rim, L is a positive learned scalar initialised to 3, and b_0 is a learned uniform source. Condition MLPs have two SiLU hidden layers of width 16 (the implementation permits widths 1–16); without conditions their outputs are learned constants. Both models clamp the entire predicted crop to $[-3, 3]$ after every update. The reaction/source output layers start at zero except for the cubic coefficient, whose raw value starts at 0.1 ($-a_3^2 = -0.01$): at zero its derivative would vanish and the cubic term would never train. An earlier version had exactly this defect (it trained a quadratic reaction); all `ac_rd` results reported here use the corrected initialisation. Clipping bounds predictions but does not validate the reaction as a material law or establish long-horizon accuracy.

Sampling, loss and model selection. Windows contain four history and four target frames, all observable, consecutive and inside one constant-current segment. Training samples windows uniformly with replacement. Each sample receives one of eight rotations/reflections, shared by history, targets, mask and any reference frames. Predictions replace observations in the rolling history without teacher forcing or gradient detachment. The loss is

$$\mathcal{L} = \frac{1}{4BHW} \sum_{b=1}^{B} \sum_{j=1}^{4} \sum_{p} (1 + 3m_{b,p})(\hat{x}_{b,j,p} - x_{b,j,p})^2. \tag{1}$$

Thus background pixels retain weight one and the denominator is the full sample count, not the sum of weights. AdamW uses its default momentum coefficients $(0.9, 0.999)$ and numerical constant 10^{-8}, 200 linear warmup steps, cosine decay to zero over 6,000 steps, and gradient norm clipping at 1. Every validation window is evaluated at step zero, every 500 steps and at the final step; the minimum-loss checkpoint, including step zero, is retained. Training stops after eight checks without improvement. Validation holds out about 10% of the training segments in whole groups, never test conditions: seeded acquisition–half-cycle blocks when every current direction keeps at least three blocks in training, and whole particles otherwise (so that validation never removes a direction from training). Selected condition features are standardised using their occurrences as targets in training windows, including overlap multiplicity; validation and inference reuse those statistics.

Condition sets and sweep provenance. The full state vector is (d, u, s) with $s = q_{\mathrm{uAh}}/Q_{\mathrm{ref}}$ and $s = s_{\mathrm{start}} + d\,u$ within a half-cycle. Here s is an acquisition-origin Coulomb-counted coordinate, not an independent calibration of particle lithium content. In feature names, `dt_step_s` and `du_step` denote time and charge increments. *Stateplus* adds both to *state*; *known* also adds $\log_{10}$ C-rate, T_C and cycle. The grid computes step features with centred gradients (one-sided at endpoints), whereas rollout conditions use forward differences of the interpolated schedule. Cycle is the integer half-cycle index divided by two; stored acquisition `cycle_offset` metadata are not added.

S1 contains 60 charge-clock jobs for U-Net, ConvLSTM, FNO, ViT and `ac_rd`, with seven *known* features omitting s; S1b contains eight unconditioned U-Net jobs. S2 compares all three clocks and the four condition sets, with 168 U-Net jobs (seeds 0 and 1) and 84 `ac_rd` jobs (seed 0). S3 contains

45 NMC U-Net and 90 NMC grey-box jobs, using *none*, *state* and *known* without temperature, all with seed 0. NMC temperature is unmeasured; the builder's nominal 25 °C is not an experimental covariate. S4 contains 24 charge-clock `unet_ref` jobs with *state/known*, s alignment and seeds 0 and 1. These counts describe commands, not a claim that every job completed. All sweep commands exclude `E_V` and leave the evaluator's voltage-oracle flag disabled.

Replay baselines. With alignment coordinate $z = u$ or s, replay predicts $\hat{x}(z_0 + \delta) = x(z_0) + \langle x_r(z_0+\delta) - x_r(z_0)\rangle_r$, averaging over training segments r of the same particle and direction in other acquisitions, with $\delta = \Delta$ for u and $\delta = d\Delta$ for s ($d = \pm 1$ for lithiation and delithiation); references without observed support at $z_0 + \delta$ are dropped, and a forecast without any reference reverts to persistence. Pixel replay uses the full change image; mask-mean replay adds its mean inside the target mask uniformly. For history-aligned replay (Sec. 5), the charge-axis reference shift is at most 0.1 and best matches the particle's last 0.04 of mask-mean history. Population replay averages the changes of the other training particles at the same z; scaled population replay multiplies that change by a least-squares gain, clipped to $[0, 3]$, fitted on the particle's own recent history.

Reference retrieval. The memory U-Net has 14 image channels: four target history frames, their mask, eight reference frames (four history and up to four future positions), and an availability flag. Reference rate, temperature and cycle append three FiLM inputs, standardised over training reference segments with equal segment weights. Retrieval requires the same particle and direction, a different acquisition, and observable interpolation brackets throughout the requested s span. All test acquisitions are excluded from the reference pool. Training draws one eligible reference uniformly; validation and S4 evaluation average complete independent rollouts over eligible references. An unavailable reference invokes the learned no-reference path; lookahead beyond the supplied tape is zero-padded. This differs from replay's persistence fallback. The stored S4 commands use the training reference pool, not the optional training-plus-validation pool. They omit the evaluation seed flag, so evaluation reconstructs membership with seed 0 even for seed-1 models; that pool can differ from the checkpoint's fitting pool. Excluding all test acquisitions also removes earlier-cycle references from an acquisition used for later-cycle testing.

Common-target evaluation. Forecast origins start at $u_0 = 0.14$, every 0.02 in u. Horizons are $\Delta = 0.01, 0.02, 0.04, 0.08, 0.16$ and 0.32. The default `available` policy retains each observable horizon that ends within the half-cycle; it does not require the largest horizon to fit. Histories and targets are linearly sampled from stored original-sequence frames. Histories follow each model's clock, ending at the common origin. Autoregressive predictions are interpolated between the two clock steps bracketing each target; the observed origin supplies step zero. A final forecast step may extend beyond the segment to bracket an in-segment target, continuing only schedule coordinates. Targets are never extrapolated. Histories are intersected across models within an evaluation call; the sweep commands evaluate one checkpoint per call, so separate outputs require matching window IDs for a common-window comparison. Evaluation batches are at most 64 with a 256 MiB rollout staging budget. Reported horizon minutes are $60\Delta Q_{\mathrm{ref,uAh}}/|I_{\mathrm{uA}}|$, using the segment's median absolute current; observed elapsed minutes are also stored.

Errors and uncertainty. The main skill is $1 - \sum \mathrm{SSE}_{\mathrm{model}} / \sum \mathrm{SSE}_{\mathrm{persistence}}$ inside the particle mask. Crop errors use pixels where both target and origin are finite, identically for model and persistence. For window metrics, finite forecasts are clipped to $[-10, 10]$; nonfinite forecasts are replaced by $+10$ before subtracting the target and are counted, rather than dropped. We also retain RMSE, absolute mask-mean error and replay fallback rates. The sweep commands request 200 particle-cluster bootstrap resamples (the evaluator default is 1,000), with seed 0 and 95% percentile intervals. Each resample keeps a particle's acquisitions, cycles and windows together and recomputes the pooled SSE ratio. Groups are model/split/horizon, and additionally rate, temperature, direction, acquisition or particle. Zero persistence SSE gives undefined skill; a one-particle group produces a degenerate bootstrap interval. These intervals do not represent independent-cell replication or preprocessing uncertainty.

Long-horizon transitions. With `--long`, forecasts run from $u = 0.14$ to the segment endpoint on a common uniform charge axis with spacing at most 0.005. The complete target trace must be observable; this evaluates the remaining half-cycle, excluding its initial 0.14. A quadratic Savitzky–

Golay filter uses nominal width 0.03 in u, rounded to an odd sample count. The primary marker is the largest absolute derivative of the mask-mean signal. Additional peaks must reach 25% of that trace's maximum derivative and be separated by at least 0.03; endpoint peaks are allowed. Predicted and observed peaks are matched by minimum total absolute charge error, without a maximum matching distance, and unmatched counts are retained. Each pixel's switch charge is its maximum absolute smoothed derivative. We report switch-map Spearman correlation, median absolute charge error and identifiable-pixel coverage. Flat traces/pixels have undefined markers; nonfinite trajectories are rejected. Transition summaries average available segment-level metrics and bootstrap particle clusters.

Preprocessing and clocks. Graphite crops apply per-frame drift corrections and placed atlas masks to binned intensity images, giving 64×64 crops at $0.26\,\mu\mathrm{m}$ per pixel. NMC uses the source-stabilised movies, 2×2 intensity binning and residual subpixel drift correction, giving 48×48 crops at $0.0694\,\mu\mathrm{m}$ per pixel. Static masks and cross-acquisition identities are curated from complete acquisitions before forecasting, so mask geometry is supplied information (App. B.3). The stored signal is $\log(\mathrm{raw}/\mathrm{ref})$, with a per-pixel median reference from the first three good graphite frames or the initial pre-current NMC frames. The loader subtracts each frame's median log contrast over particle-free crop background when available, otherwise retaining the uncorrected log ratio. Training and scoring use the full particle mask, not the separately stored eroded core. Quality code 2 makes a frame unobservable; code-1 artefact frames remain eligible if finite.

Signed charge is obtained by trapezoidal integration of the high-rate current and interpolated to image times, with positive lithiation. We fix $Q_{\mathrm{ref}} = 280\,\mu\mathrm{Ah}$ for graphite and $560.93\,\mu\mathrm{Ah}$ for NMC (approximately 560.9). The latter is the measured C/3 charge capacity, used unchanged in every NMC split, including tests on C/3; it is not re-estimated using each split's training data. Constant-current segments alone are gridded, excluding rests and voltage holds. Charge spacing is 0.005; time spacing is 60 s for graphite and 30 s for NMC. The frame clock advances one stored sequence frame. In the graphite population builder these are every fourth raw frame at C/20 and every sixth at C/2, whereas NMC retains every source frame. Thus "frame" does not always mean native camera cadence. Grid images are stored as float16 and sampled batches converted to float32; invalid interpolation brackets invalidate a grid frame and windows never cross such breaks.

Compute. The recorded cluster runs use NVIDIA A100-SXM4 GPUs with 80 GB memory. The local benchmark log records 68.2 ms per U-Net optimiser step at batch 32, 64×64, history four and unroll four, with seven conditions, float32 and nondeterministic kernels (156.9 ms with deterministic kernels). This synthetic single-process timing excludes data loading and validation. Sweep commands omit mixed precision and request GPU-resident grids and two CPU threads per run. Four runs share each GPU for the learned sweeps; the grey-box job groups use seven runs for S2 and eight for S3, so the single-process timing is not per-run throughput under sharing.

D FOUNDATION-MODEL PROTOCOL AND RECIPES

This appendix documents adaptation recipes, controls and model selection.

A fixed training recipe can be unfair to a pretrained model, so the foundation-model comparison fixes the data, inputs, targets, validation and scoring, and compares *recipes* chosen on validation data only. The protocol below was registered before any new-recipe result existed and amended twice before any such result was inspected; earlier one-step runs are inherited, exploratory references. Conclusions are scoped to the evaluated checkpoints, recipes, horizons and splits.

D.1 MODELS AND ADAPTERS

We describe the pretrained models and the adapters used for microscopy. *Poseidon* (Herde et al., 2024) (scOT, a lead-time-conditioned Swin-v2 U-Net operator on 128×128 grids with patch size 4) is used at the T and B sizes from the public checkpoints. As in its fine-tuning code, the pretrained body is kept and the patch embedding and recovery layers are re-initialised for our channel layout: $k = 4$ context frames plus one constant map per condition in, one normalised change out. Each sample is standardised by the mean and standard deviation of its context; the crop is bilinearly resized to 128×128 and the output back to the native crop; the forecast is the last frame plus the

predicted change times the RMS of successive context differences. *Walrus* (McCabe et al., 2026) (about 1.3 billion parameters, 40 space–time blocks, causal in time) receives the 2-D crops as 3-D fields of unit depth resized to 128×128, with per-sample RMS normalisation of the context and its differences; conditions enter as constant fields that bypass the normalisation, and only bf16 or fp32 arithmetic is used. Neither adapter feeds the particle mask to the network; the mask enters the loss and the scoring exactly as for the other models. Fine-tuning modes are *full*, *LoRA* (rank 16 on the attention and MLP projections, plus the re-initialised input and output layers), *head* (input and output layers only) and *last-N* decoder layers.

Inherited one-step recipe (exploratory). The first fine-tuning runs trained one-step autoregression with a constant lead-time input (0.1) and were rolled out step by step, which ignores that scOT is lead-time conditioned by design. They used (d, u, s) conditioning (later (d, s)), 6,000 steps at batch 32, learning rates 10^{-4} (full), 3×10^{-4} (last four layers) and 5×10^{-4} (LoRA), grouped block validation and a checkpoint retention cap; Walrus used head fine-tuning for 1,500 steps at batch 16 and learning rate 10^{-4}, with a reduced evaluation (origins every 0.08, horizons 0.01–0.08, a fixed subset of test sequences). They are reported as inherited references and are not candidates in the recipe selection.

D.2 Matched recipes

We specify the adaptation recipes and controls compared on common targets. The original matched-recipe runs use four splits (Table 15: C/20→C/2, C/2→C/20, new-particle fold 0 and later cycles), seeds 0 and 1, conditioning (d, s), $k = 4$ history frames on the charge grid ($\Delta u = 0.005$), at most 6,000 optimiser steps with an effective batch of 32 (four accumulated micro-batches of 8), validation every 500 steps with patience 8, grouped validation (acquisition–half-cycle blocks, or whole particles where blocks would remove a direction), the common training loss (weight 4 inside the mask and 1 outside), the same augmentation and learning-rate schedule, and learning rate 10^{-4} for Poseidon-T (full fine-tuning) and 10^{-3} for the U-Net. The recipes differ only as follows. *One-step*: trained on the next grid step (unroll 1) and rolled out autoregressively. *Direct lead time*: one pass predicts $\hat{x}(u_0 + \Delta)$ from the origin history and the scheduled target conditions, with the lead-time input set to $\Delta/0.32 \in (0, 1]$; training samples (origin, horizon) pairs uniformly among the available pairs at the seven evaluation horizons $\Delta \in \{0.005, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32\}$. The scaling is a fixed dimensionless choice made before fitting, not a learned physical conversion, and there is no autoregression. *Unrolled*: training backpropagates through four autoregressive steps without teacher forcing. *Activity-weighted*: training windows are drawn with probability $0.2/N + 0.8\, a_i / \sum_j a_j$, where a_i is the masked RMS of successive *observed* history differences of window i (uniform if all vanish); validation and test windows are enumerated uniformly, and no future image enters the sampling. Poseidon uses it with one-step training; the U-Net uses it with four unrolled steps and has a matched uniform-sampling control. Two controls separate the recipe from the model. The *scratch* control trains the identical scOT-T architecture and configuration from random initialisation with the direct recipe and budget, so that a pretrained-minus-scratch difference isolates pretraining. The *direct U-Net* control gives the benchmark U-Net one additional conditioning input (through the same FiLM layers as the other conditions), the scaled horizon $\Delta/0.32$, and trains it on the same (origin, horizon) pairs with the same validation, so that a gain of direct prediction is not attributed to the foundation model. After the validation-only recipe selection below, three follow-ups use the selected recipe: Poseidon-B; the (d, u, s) conditioning ablation; and *per-direction* models, in which Poseidon-T and a matched U-Net are trained, validated and tested on lithiation-only or delithiation-only windows. Per-direction results are reported for each direction and for the recombined whole, obtained by summing SSE over directions before computing skill (direction skills are never averaged); the mixed-direction model remains a separate control. All forecast windows lie inside one constant-current segment: an audit of the 614 constant-current grid segments over all split, seed, partition and horizon cells (168) found no window crossing a current reversal, rest, voltage hold or positional gap. Walrus is gated: a Walrus run with the best recipe is prepared only if a new Poseidon recipe shows a positive paired difference against its matched one-step control on at least two splits at one headline horizon; otherwise the gate is recorded as not triggered.

Table 15: Graphite splits used by the recipe (App. D), transition-window (App. K) and step-size (App. J) studies, with the identifiers used in the supplementary code. Acquisitions as in Table 3; the first formation acquisition is always excluded.

Split	Identifier	Training	Test
C/20→C/2	`g_rate_C20_to_C2`	03–06	07–14
C/2→C/20	`g_rate_C2_to_C20`	07–14	03–06
C/20, C/2→2C	`g_rate_to_2C`	03–14	15–18
New particle, fold 0	`g_particle_fold0`	03–14, other particles	03–14, five held-out particles
Later cycles	`g_cycle_14`	14, first cycle	14, cycles 2–4
Same rate, C/20	`g_same_C20_25`	03	06
Same rate, C/2	`g_same_C2_25`	10, 11	14, first cycle
C/2, 25 °C→cold	`g_temp_C2_25_to_cold`	10, 11, 14	07–09
C/2, 25 °C→hot	`g_temp_C2_25_to_hot`	10, 11, 14	12, 13
C/20, 25 °C→15, 45 °C	`g_temp_C20_25_to_15_45`	03, 06	04, 05

D.3 Selection on validation only

We state how validation selects a recipe without using test scores. Checkpoints of every new recipe are selected on the same criterion: the mean, over $\Delta = 0.005$, 0.02 and 0.16, of the masked validation skill against persistence on at most 256 evenly spaced validation origins that are valid through $\Delta = 0.16$. For a given split and seed these origins, targets and masks are identical for direct and autoregressive recipes. Recipes are ranked only once the direct, unrolled and activity-weighted Poseidon runs have complete coverage (four splits by two seeds) by this frozen three-horizon validation mean; candidates within 0.05 of the best are resolved by lower mean validation inference time, then by a fixed recipe order. The selector refuses incomplete or mismatched validation matrices, reads no test output, and writes an immutable selection record that the follow-up runs read. No provisional winner is chosen from an incomplete matrix; a comparison that lacks complete coverage is reported as not evaluable rather than as passed or failed.

D.4 Evaluation, contrasts and reporting

We define the matched supports and contrasts used to evaluate adaptation. Test evaluation uses causal stored-frame origins every 0.02 in u from $u = 0.14$, the available-horizon policy, common masks, $Q_{\text{ref}} = 280\,\mu\text{Ah}$ and the seven horizons above; $\Delta = 0.005$ is one charge-grid step, not a native camera interval. The primary contrasts, at $\Delta = 0.005$, 0.02 and 0.16, are pretrained direct minus scratch direct, pretrained direct minus direct U-Net, and pretrained direct minus the matched one-step Poseidon, each also against the (d, s) autoregressive U-Net and history-aligned replay. Paired differences use exactly matched split, sequence, half-cycle, origin and horizon keys (with persistence SSE and masks checked), are averaged over seeds within each resample, and carry 95% percentile intervals from 2,000 physical-particle bootstrap resamples (seed 280923) that keep each particle's acquisitions, windows, models and seeds together; the intervals are descriptive and not adjusted for multiplicity. Practical equivalence requires the whole interval inside ± 0.05 skill; an interval that merely contains zero is "no resolved difference". A recipe is eligible for the main table only with both seeds on all four splits, and if it is best or practically equivalent to the best eligible competitor on at least two of four splits at one common headline horizon. Every table reports population skill, model and persistence masked RMSE, pixel and window coverage, the numbers of particles and acquisitions, the seed spread and the divergence rate: a window diverges if any masked prediction is non-finite or its raw masked SSE exceeds ten times the persistence SSE (a positive model error with zero persistence error also diverges); the benchmark's clipping to ± 10 log units is applied to the reported SSE and skill but not to the divergence test.

Strata. Two kinds of strata are reported and never mixed. *Prospective* origin-activity bins use the RMS over the four observations at charge offsets -0.015, -0.010, -0.005 and 0 before the origin, with tertile thresholds fitted once per split on training origins (seed 0) and frozen for every model, seed and horizon (ties: quiet $\le q_1 <$ moderate $\le q_2 <$ active); the active-bin subset scores every masked pixel, including pixels that never switch, on the same origins for all models and baselines.

Retrospective strata use the observed future and serve only as diagnostics: activity tertiles of the future persistence RMS, and pixel classes from the observed switch map (App. G), *front-adjacent change* (a switch within the horizon no farther than a radius r from a pixel that had already switched), *distant change* (farther than r, including no earlier switch) and *settled* (no switch within the horizon), with $r = 1, 2$ (primary) and 4 native pixels, i.e. 0.26, 0.52 and 1.04 µm. These optical-marker classes never select inputs, crops, origins, training pairs or models, and a skill difference between them does not identify deterministic growth or stochastic nucleation. For activity-weighted training we report the sampling probabilities, effective and distinct windows, and exposure by train-frozen activity bin.

D.5 Runs and results

We report completed runs, recipe failures and controlled comparisons. The original matched matrix comprises eight runs per recipe arm (four splits by two seeds): direct, scratch-direct, direct U-Net, unrolled, activity-weighted Poseidon, and activity-weighted and uniform U-Nets; the validation-gated follow-ups add eight best-recipe scratch, 16 per-direction Poseidon-T, 16 matched per-direction U-Net, eight Poseidon-B and eight (d, u, s) runs; existing checkpoints are re-evaluated with the strata without altering their stored evaluations. Poseidon runs share each GPU in pairs, U-Nets in fours and Walrus runs one per GPU. The later direct U-Net extension uses seeds 0–4 on seven graphite rate/temperature splits and new-particle fold 0; later cycles retain seeds 0–1. The registered FM3 pretraining and architecture contrasts retain their original common seeds 0–1, as do the budget controls below. Table 16 gives every recipe's skill by split and horizon with the paired contrasts quoted in Sec. 6.4: the pretraining gain (direct against scratch-direct), the direct U-Net with and without the mask input, Poseidon-B and Poseidon-T (d, u, s) with the selected recipe, and Walrus with the inherited one-step recipe (one seed, evaluated to $\Delta = 0.08$ only). The selector chose direct prediction. The unrolled recipe did not finish within the compute budget and two of eight activity-weighted runs diverged; a second amendment, written before any test score was read, made recipes that diverged or stopped on the budget ineligible, so neither could be selected. Relative to the autoregressive U-Net, Poseidon-B with the selected recipe is -0.06 [-0.12, -0.04] and -0.04 [-0.07, -0.00] at $\Delta = 0.02$ and 0.16 from C/20 to C/2, and Poseidon-T told (d, u, s) with the selected recipe is -0.18 [-0.20, -0.17] and -0.27 [-0.30, -0.14]. Per-direction models (Tables 18 and 19) are scored on their own direction's windows and recombined by summing errors. Four direct per-direction runs of each family failed because filtering by direction left no validation windows, so the recombined rows on both rate splits are marked failed; on new-particle fold 0 the recombined per-direction Poseidon-T gains $+0.14$ [0.03, 0.15] at 0.16 over the mixed-direction model and the U-Net does not change (-0.02 [-0.02, 0.01]).

Data efficiency, budget and learning rate. Three pre-registered controls test whether the comparison depends on the amount of training data or on the fine-tuning budget (released notes). Table 20 trains the selected direct recipe of Poseidon-T, its scratch control and the direct U-Net on one, two or four C/20 acquisitions and tests all on the same C/2 windows; the pre-registered hypothesis that the pretraining gain is larger with one acquisition than with four is mixed: supported at $\Delta = 0.02$ ($+0.014$ [0.002, 0.031]) and refuted at 0.16 (-0.060 [-0.095, -0.038]). The one- and two-acquisition arms use seeds 0–3, while four acquisitions retain seeds 0–1; shared particle draws include all these seeds in the difference of gains. With one acquisition, Poseidon minus direct U-Net is $+0.191$ [0.156, 0.269] at 0.02 and -0.263 [-0.283, -0.212] at 0.16; pretrained minus scratch is $+0.151$ [0.134, 0.177] and $+0.112$ [0.064, 0.191], respectively. With two acquisitions, Poseidon minus direct U-Net is $+0.069$ [0.044, 0.100] and -0.130 [-0.196, -0.104]. The original one-acquisition subset is acquisition 03 at 25 °C. The one- and two-acquisition subsets validate by particle, whereas the four-acquisition set uses acquisition–half-cycle blocks and adds temperatures. This is a subset-specific result, not a general data-efficiency law.

The acquisition-06 replication (seeds 0–1) gives Poseidon-minus-direct-U-Net gains of $+0.294$ [0.104, 0.388] at 0.02 and -0.112 [-0.194, -0.091] at 0.16, with pretrained-minus-scratch gains of $+0.104$ [0.080, 0.146] and $+0.107$ [0.035, 0.167]. Both seeds separately resolve the same signs. These are arithmetic means of seed-specific pooled-SSE contrasts on matched physical windows, using shared particle resamples across seeds and arms. All six arms match the original 416 test segments in acquisitions 07–14. Acquisition 06 has 15 indexed particles versus 21 in 03, so this descriptive replication does not isolate training sample size.

Table 16: Foundation-model recipes. Seed mean$_{\pm \mathrm{SD}}$ skill on the U-Net core windows. One-step fine-tuning: next-step training with (d, s) inputs and block validation, or (d, u, s) inputs with validation-loss selection per seed. The remaining rows are matched 6,000-step recipes with automatic validation grouping (direct lead-time, unroll-4, activity-weighted, a scratch transformer and U-Net controls); their recipe choice uses validation results only (selected: direct lead-time recipe with diverged or budget-stopped recipes ineligible); recipes with a diverged run or a run stopped by the compute budget were ineligible. Marks: d one seed diverged (mean over the remaining seed); v validation grouping differs from the U-Net row; x mixed protocols; s fewer windows; stopped: training stopped by the 6-hour budget before 6,000 steps (no test evaluation). Walrus runs (one seed) were evaluated on a sparse origin lattice up to $\Delta = 0.08$ only, so their $\Delta = 0.16$ cells are n/r. Paired recipe contrasts: identical windows and common seeds with 95% particle-cluster bootstrap intervals. Walrus panel: skill at $\Delta = 0.02$ and 0.08 and Walrus minus the autoregressive U-Net (d, s) on identical windows and the common seed, same intervals. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Graphite	C/20→C/2		C/2→C/20		(C/20+C/2) →2C		C/2: 25 →6–15°C		New particle fold 0		Later cycles acq. 14	
Δ	.02	.16	.02	.16	.02	.16	.02	.16	.02	.16	.02	.16
U-Net, autoregressive (d, s)	$.51_{\pm.01}$	$.65_{\pm.05}$	$.38_{\pm.01}$	$.54_{\pm.02}$	$.50_{\pm.01}$	$.53_{\pm.06}$	$.46_{\pm.02}$	$.60_{\pm.09}$	$.53_{\pm.01}$	$.71_{\pm.06}$	$.67_{\pm.03}$	$.86_{\pm.04}$
Poseidon-T, one-step fine-tuning (full) (d, s)	$.45_{\pm.06}$	$.13_{\pm.02}$	$.21_{\pm.05}$	$-3.32_{\pm2.93}$	.41	−.26	$.10_{\pm.32}$	$.12_{\pm.08}$	$.26_{\pm.14}$	$-7.43_{\pm3.74}$	$.33^{\mathrm{v}}_{\pm.16}$	$-.82^{\mathrm{v}}_{\pm1.54}$
Poseidon-T, one-step fine-tuning (LoRA) (d, s)	.46	.12	n/r	n/r	n/r	n/r	−.03	−13.9	.35	−6.82	$.45^{\mathrm{v}}$	$.30^{\mathrm{v}}$
Poseidon-B, one-step fine-tuning (LoRA) (d, s)	.46	.11	.16	−.26	.19	−1.70	.31	.14	.30	−.69	$.44^{\mathrm{v}}$	$.28^{\mathrm{v}}$
Poseidon-B, one-step fine-tuning (last-4 layers) (d, s)	.38	.08	n/r	n/r	n/r	n/r	.28	.14	.39	−4.27	$.43^{\mathrm{v}}$	$.13^{\mathrm{v}}$
Walrus, one-step fine-tuning (head) (d, s)	$.12^{\mathrm{s}}$	n/r	n/r	n/r	n/r	n/r	$.17^{\mathrm{s}}$	n/r	$.20^{\mathrm{s}}$	n/r	$.11^{\mathrm{vs}}$	n/r
Walrus, one-step fine-tuning (last-4 layers) (d, s)	$.35^{\mathrm{s}}$	n/r	n/r	n/r	n/r	n/r	$.16^{\mathrm{s}}$	n/r	$.33^{\mathrm{s}}$	n/r	$.21^{\mathrm{vs}}$	n/r
Poseidon, one-step fine-tuning (d, u, s), val.-selected	.42	.09	n/r	n/r	n/r	n/r	.21	.06	n/e	n/e	n/r	n/r
Walrus, one-step fine-tuning (d, u, s), val.-selected	$.32^{\mathrm{s}}$	n/r	n/r	n/r	n/r	n/r	$-.55^{\mathrm{s}}$	n/r	n/r	n/r	n/r	n/r
Poseidon-T, direct lead-time (d, s), selected on validation	$.46_{\pm.02}$	$.63_{\pm.01}$	$.31_{\pm.00}$	$.46_{\pm.03}$	n/r	n/r	n/r	n/r	$.44_{\pm.01}$	$.56_{\pm.02}$	$.63_{\pm.01}$	$.87_{\pm.01}$
Poseidon-T, scratch direct (d, s)	$.32_{\pm.03}$	$.45_{\pm.00}$	$.23_{\pm.01}$	$.36_{\pm.02}$	n/r	n/r	n/r	n/r	$.35_{\pm.04}$	$.52_{\pm.07}$	$.43_{\pm.06}$	$.68_{\pm.01}$
Poseidon-T, unroll-4 (d, s), 8/8 stopped by budget	stopped	stopped	stopped	stopped	n/r	n/r	n/r	n/r	stopped	stopped	stopped	stopped
Poseidon-T, activity-weighted (d, s), 2/8 runs diverged	$.42_{\pm.09}$	$.19_{\pm.04}$	$.27^{\mathrm{d}}$	$.04^{\mathrm{d}}$	n/r	n/r	n/r	n/r	$.24^{\mathrm{d}}$	-2.15^{d}	$.65_{\pm.00}$	$.40_{\pm.01}$
U-Net, direct lead-time (d, s)	$.46_{\pm.01}$	$.79_{\pm.02}$	$.30_{\pm.01}$	$.63_{\pm.03}$	$.34_{\pm.05}$	$.65_{\pm.01}$	$.40_{\pm.10}$	$.73_{\pm.01}$	$.45_{\pm.02}$	$.75_{\pm.01}$	$.61_{\pm.06}$	$.91_{\pm.01}$
U-Net, activity-weighted (d, s)	$.46_{\pm.05}$	$.62_{\pm.01}$	$.32_{\pm.01}$	$.53_{\pm.02}$	n/r	n/r	n/r	n/r	$.52_{\pm.00}$	$.69_{\pm.02}$	$.64_{\pm.02}$	$.86_{\pm.03}$
U-Net, direct lead-time, no mask input (d, s)	$.46_{\pm.04}$	$.78_{\pm.01}$	$.33_{\pm.01}$	$.63_{\pm.00}$	n/r	n/r	n/r	n/r	$.43_{\pm.06}$	$.75_{\pm.01}$	$.62_{\pm.05}$	$.90_{\pm.03}$
U-Net, uniform sampling (d, s)	$.49_{\pm.03}$	$.73_{\pm.00}$	$.39_{\pm.00}$	$.54_{\pm.01}$	n/r	n/r	n/r	n/r	$.52_{\pm.00}$	$.71_{\pm.12}$	$.65_{\pm.02}$	$.85_{\pm.02}$
Poseidon-B, validation-selected recipe (d, s)	$.46_{\pm.00}$	$.63_{\pm.01}$	$.29_{\pm.01}$	$.45_{\pm.02}$	n/r	n/r	n/r	n/r	$.44_{\pm.02}$	$.61_{\pm.04}$	$.61_{\pm.02}$	$.85_{\pm.00}$
Poseidon-T, validation-selected recipe (d, u, s)	$.34_{\pm.06}$	$.40_{\pm.01}$	$.31_{\pm.03}$	$.44_{\pm.04}$	n/r	n/r	n/r	n/r	$.47_{\pm.01}$	$.64_{\pm.05}$	$.59_{\pm.05}$	$.84_{\pm.02}$

NMC	C/3→1C,2C		(C/3+1C) →2C		2C→C/3	
Δ	.02	.16	.02	.16	.02	.16
U-Net, autoregressive (d, s)	$.81_{\pm.01}$	$.90_{\pm.00}$	$.80_{\pm.01}$	$.89_{\pm.00}$	$.54_{\pm.08}$	$.71_{\pm.08}$
Poseidon-T, one-step fine-tuning (full) (d, s)	$.32_{\pm.16}$	$-9.30_{\pm13.6}$	n/r	n/r	$.55_{\pm.01}$	$.34_{\pm.01}$

Paired recipe contrasts	C/20→C/2		C/2→C/20		New particle fold 0		Later cycles acq. 14	
Δ	.02	.16	.02	.16	.02	.16	.02	.16
Pretraining: Poseidon-T direct − scratch	+.14 [.12,.16]	+.17 [.14,.24]	+.06 [.04,.08]	+.13 [.00,.21]	+.09 [.08,.16]	+.04 [.02,.21]	+.20 [.16,.24]	+.19 [.16,.27]
U-Net direct − Poseidon-T direct	+.02 [.01,.04]	+.17 [.16,.21]	−.01 [-.08,.02]	+.20 [.15,.29]	+.01 [.01,.06]	+.21 [.14,.21]	−.02 [-.03,.02]	+.04 [.03,.07]
U-Net direct, no mask − Poseidon-T direct	+.00 [-.02,.02]	+.15 [.13,.20]	+.02 [-.01,.04]	+.18 [.13,.29]	.00 [-.04,.03]	+.21 [.10,.21]	−.01 [-.02,.02]	+.03 [.02,.05]
U-Net activity-weighted − uniform	−.03 [-.09,-.02]	−.11 [-.16,-.10]	−.09 [-.21,-.04]	−.01 [-.07,.02]	+.00 [.00,.04]	−.02 [-.03,.08]	−.01 [-.02,.01]	+.01 [-.01,.02]

Walrus, $\Delta \leq 0.08$	C/20→C/2		C/2: 25 →6–15°C		New particle fold 0		Later cycles acq. 14	
Δ	.02	.08	.02	.08	.02	.08	.02	.08
Walrus, one-step fine-tuning (head)	.12	−.07	.17	.12	.21	−.38	.11	.12
− U-Net, autoregressive	−.35 [-.39,-.35]	−.72 [-.75,-.39]	−.30 [-.43,-.13]	−.45 [-.60,-.18]	−.29 [-.51,-.26]	−1.09 [-6.80,-.69]	−.54 [-.60,-.45]	−.74 [-.80,-.70]
Walrus, one-step fine-tuning (last-4 layers)	.35	.24	.16	−.22	.34	.17	.21	−.01
− U-Net, autoregressive	−.13 [-.33,-.11]	−.40 [-.50,-.21]	−.31 [-.39,-.25]	−.80 [-1.09,-.32]	−.16 [-.17,-.08]	−.54 [-.57,-.28]	−.44 [-.50,-.35]	−.88 [-1.00,-.60]
Walrus, one-step fine-tuning (d, u, s)	.32	.22	−.54	−9.54	n/r	n/r	n/r	n/r
− U-Net, autoregressive	−.15 [-.31,-.14]	−.43 [-.50,-.27]	−1.01 [-2.28,-.19]	−10.1 [-36.4,-.28]	n/r	n/r	n/r	n/r

Table 22 quadruples the budget to 24,000 updates with matched effective batches, not matched compute expenditure, and Table 23 gives Poseidon-T at two further learning rates, selected on validation only. At $\Delta = 0.16$, the completed forward-rate budget contrast narrows Poseidon's deficit by 0.0379 [0.0185, 0.0533], leaving a direct U-Net lead of 0.1366 [0.1227, 0.1814]. In reverse, the deficit narrows by 0.0807 [0.0504, 0.1166], leaving a direct U-Net lead of 0.1153 [0.0749, 0.1838]. The direct U-Net's reverse-rate change at $\Delta = 0.02$ is -0.0421 [-0.1722, 0.0135], so the "at most 0.02" change applies only to $\Delta = 0.16$. At reverse-rate $\Delta = 0.02$, the 24,000-update Poseidon lead over direct U-Net is resolved ($+0.091$ [0.006, 0.297]); the long-horizon direct U-Net lead survives on both rate splits. Table 17 gives the recipe-fair Walrus evaluation (last four blocks, one-step training

Table 17: Recipe-fair Walrus. Walrus last-four-block adaptation, one-step training at charge step 0.02, 6,000 requested steps, seeds 0–1; direct and autoregressive U-Net seeds 0–4; validation step-selected Poseidon-T seed 0 (step 0.02 forward, 0.04 reverse). All eligible particles and dense causal origins, through horizon 0.16. Walrus and U-Net use the prediction-independent frozen common windows; the Poseidon contrast further intersects its observable history. Skills are means with seed SD; contrasts average every named arm's seeds with 95% paired particle intervals (2,000 resamples). Each contrast uses its exact common support; displayed arm skills need not subtract to a contrast when support differs. Best baseline: strongest of eleven on their common windows, held fixed in the bootstrap. Inherited last-4: one seed, 1,500 requested steps, charge step 0.005, eight particles and sparse origins through 0.08, forward split only; fair minus inherited uses their sparse intersection. n/r: not run. At 0.16, frozen support keeps 5,004 forward windows (one U-Net window excluded) and 4,220 reverse windows (299 excluded). Per-contrast support and all six horizons remain in the released results. The preceding weaker Walrus recipes remain reported separately.

Forecaster or contrast	$\Delta = 0.02$	$\Delta = 0.08$	$\Delta = 0.16$
C/20→C/2			
Walrus, recipe-fair	0.447 ± 0.035	0.565 ± 0.024	-0.023 ± 0.216
seed 0	0.471	0.582	0.130
seed 1	0.422	0.547	-0.175
Direct U-Net	0.464 ± 0.014	0.678 ± 0.007	0.792 ± 0.018
Autoregressive U-Net	0.514 ± 0.012	0.623 ± 0.014	0.655 ± 0.048
Poseidon-T, validation step	0.472	0.605	0.529
Best baseline	0.344	0.508	0.681
Walrus − Direct U-Net	-0.017 [-0.031, 0.017]	-0.113 [-0.195, -0.078]	-0.814 [-0.989, -0.419]
Walrus − AR U-Net	-0.067 [-0.098, -0.055]	-0.058 [-0.098, -0.035]	-0.677 [-0.869, -0.253]
Walrus − Poseidon-T	-0.026 [-0.052, -0.012]	-0.040 [-0.106, -0.013]	-0.552 [-0.706, -0.200]
Walrus − best baseline	+0.103 [0.068, 0.118]	+0.056 [-0.076, 0.110]	-0.704 [-0.863, -0.320]
Inherited last-4, sparse windows	0.346	0.244	n/r
Fair − inherited, sparse windows	+0.051 [0.028, 0.065]	+0.392 [0.036, 0.435]	n/r
C/2→C/20			
Walrus, recipe-fair	0.369 ± 0.003	0.340 ± 0.010	-0.049 ± 0.012
seed 0	0.367	0.347	-0.058
seed 1	0.371	0.333	-0.041
Direct U-Net	0.315 ± 0.015	0.512 ± 0.025	0.614 ± 0.028
Autoregressive U-Net	0.374 ± 0.007	0.450 ± 0.012	0.525 ± 0.018
Poseidon-T, validation step	0.340	0.471	0.546
Best baseline	0.204	0.392	0.451
Walrus − Direct U-Net	+0.054 [-0.007, 0.095]	-0.172 [-0.239, -0.139]	-0.663 [-0.743, -0.604]
Walrus − AR U-Net	-0.005 [-0.040, 0.022]	-0.110 [-0.173, -0.077]	-0.575 [-0.670, -0.493]
Walrus − Poseidon-T	+0.033 [-0.012, 0.061]	-0.161 [-0.234, -0.130]	-0.660 [-0.750, -0.582]
Walrus − best baseline	+0.165 [0.087, 0.227]	-0.052 [-0.117, -0.008]	-0.500 [-0.572, -0.386]

Table 18: Per-direction Poseidon-T, one-step fine-tuning. Each direction model is scored on its own direction's windows; the whole is the SSE-recombined pair (never an average of skills) and needs both direction models of one seed. Comparator (Mixed): the mixed-direction model on identical windows and common seeds. Difference = per-direction minus mixed-direction skill, 95% paired particle-cluster interval. Particle fold 0 is a single identity fold. failed: training stopped because the direction filter left no validation windows for that seed.

Split	Scored on	Δ	Per-direction	Mixed	Difference [95% CI]	Seeds	Particles	Failed runs
C/20→C/2	lithiation	0.020	0.417	0.343	0.075 [0.050, 0.125]	1	22	–
C/20→C/2	delithiation	0.020	0.588	0.507	0.081 [0.070, 0.118]	1	22	–
C/20→C/2	whole	0.020	0.488	0.411	0.077 [0.061, 0.117]	1	22	–
C/20→C/2	lithiation	0.160	0.082	0.079	0.003 [-0.006, 0.018]	1	22	–
C/20→C/2	delithiation	0.160	0.206	0.192	0.014 [0.005, 0.058]	1	22	–
C/20→C/2	whole	0.160	0.125	0.119	0.007 [-0.001, 0.025]	1	22	–
C/2: 25 →6–15°C	lithiation	0.020	0.382	0.369	0.013 [-0.064, 0.083]	1	21	–
C/2: 25 →6–15°C	delithiation	0.020	failed	failed	failed	0	0	seed 0
C/2: 25 →6–15°C	whole	0.020	failed	failed	failed	0	0	delithiation seed 0
C/2: 25 →6–15°C	lithiation	0.160	0.077	0.049	0.028 [-0.039, 0.110]	1	21	–
C/2: 25 →6–15°C	delithiation	0.160	failed	failed	failed	0	0	seed 0
C/2: 25 →6–15°C	whole	0.160	failed	failed	failed	0	0	delithiation seed 0
New particle, fold 0	lithiation	0.020	-0.157	0.295	-0.452 [-0.537, 0.123]	1	5	–
New particle, fold 0	delithiation	0.020	0.592	0.402	0.190 [0.128, 0.196]	1	5	–
New particle, fold 0	whole	0.020	0.160	0.340	-0.180 [-0.221, 0.133]	1	5	–
New particle, fold 0	lithiation	0.160	-4.183	0.218	-4.400 [-4.928, 0.044]	1	5	–
New particle, fold 0	delithiation	0.160	0.199	-12.464	12.663 [0.020, 14.089]	1	5	–
New particle, fold 0	whole	0.160	-2.514	-4.610	2.095 [0.015, 2.338]	1	5	–
Later cycles acq. 14	lithiation	0.020	0.525	0.217	0.308 [0.251, 0.376]	1	15	–
Later cycles acq. 14	delithiation	0.020	0.782	0.689	0.093 [0.061, 0.110]	1	14	–
Later cycles acq. 14	whole	0.020	0.649	0.444	0.204 [0.172, 0.256]	1	15	–
Later cycles acq. 14	lithiation	0.160	0.172	0.126	0.046 [0.003, 0.057]	1	15	–
Later cycles acq. 14	delithiation	0.160	0.631	0.646	-0.015 [-0.132, 0.037]	1	14	–
Later cycles acq. 14	whole	0.160	0.294	0.265	0.029 [-0.020, 0.050]	1	15	–

at charge step 0.02, seeds 0–1): it is close to the direct U-Net at $\Delta = 0.02$ (0.447 against 0.464 from C/20 to C/2) and below persistence at 0.16 (skills -0.023 and -0.049 in the two directions).

Table 19: Per-direction Poseidon-T, direct lead-time (validation-selected recipe). Each direction model is scored on its own direction's windows; the whole is the SSE-recombined pair (never an average of skills) and needs both direction models of one seed. Comparator (Mixed): the mixed-direction model on identical windows and common seeds. Difference = per-direction minus mixed-direction skill, 95% paired particle-cluster interval. Particle fold 0 is a single identity fold. failed: training stopped because the direction filter left no validation windows for that seed.

Split	Scored on	Δ	Per-direction	Mixed	Difference [95% CI]	Seeds	Particles	Failed runs
C/20→C/2	lithiation	0.020	0.435	0.347	0.088 [0.041, 0.109]	1	22	seed 1
C/20→C/2	delithiation	0.020	0.563	0.553	0.009 [-0.004, 0.021]	1	22	seed 0
C/20→C/2	whole	0.020	failed	failed	failed	0	0	delithiation seed 0 lithiation seed 1
C/20→C/2	lithiation	0.160	0.639	0.571	0.068 [0.049, 0.102]	1	22	seed 1
C/20→C/2	delithiation	0.160	0.714	0.677	0.037 [-0.013, 0.063]	1	22	seed 0
C/20→C/2	whole	0.160	failed	failed	failed	0	0	delithiation seed 0 lithiation seed 1
C/2→C/20	lithiation	0.020	0.171	0.236	-0.065 [-0.111, -0.022]	1	22	seed 1
C/2→C/20	delithiation	0.020	0.429	0.400	0.029 [0.014, 0.057]	1	22	seed 0
C/2→C/20	whole	0.020	failed	failed	failed	0	0	delithiation seed 0 lithiation seed 1
C/2→C/20	lithiation	0.160	0.157	0.396	-0.239 [-0.374, -0.030]	1	22	seed 1
C/2→C/20	delithiation	0.160	0.549	0.517	0.032 [-0.006, 0.046]	1	22	seed 0
C/2→C/20	whole	0.160	failed	failed	failed	0	0	delithiation seed 0 lithiation seed 1
New particle, fold 0	lithiation	0.020	0.414	0.355	0.059 [0.032, 0.064]	2	5	–
New particle, fold 0	delithiation	0.020	0.562	0.527	0.035 [0.029, 0.086]	2	5	–
New particle, fold 0	whole	0.020	0.476	0.427	0.049 [0.048, 0.056]	2	5	–
New particle, fold 0	lithiation	0.160	0.604	0.513	0.090 [0.020, 0.098]	2	5	–
New particle, fold 0	delithiation	0.160	0.800	0.582	0.218 [0.044, 0.235]	2	5	–
New particle, fold 0	whole	0.160	0.678	0.539	0.139 [0.028, 0.150]	2	5	–
Later cycles acq. 14	lithiation	0.020	0.534	0.552	-0.018 [-0.034, 0.032]	2	15	–
Later cycles acq. 14	delithiation	0.020	0.728	0.705	0.022 [0.014, 0.031]	2	14	–
Later cycles acq. 14	whole	0.020	0.627	0.626	0.001 [-0.005, 0.031]	2	15	–
Later cycles acq. 14	lithiation	0.160	0.863	0.853	0.010 [0.003, 0.025]	2	15	–
Later cycles acq. 14	delithiation	0.160	0.936	0.926	0.009 [0.007, 0.018]	2	14	–
Later cycles acq. 14	whole	0.160	0.882	0.873	0.009 [0.005, 0.022]	2	15	–

Step-by-step U-Net with scarce data. Table 21 adds the step-by-step U-Net (the recipe of the main table, seeds 0–3) to the data-efficiency test. With one C/20 acquisition, Poseidon-T's short-horizon lead over it ($+0.124$ [-0.004, 0.165] and $+0.090$ [0.0005, 0.116] on acquisitions 03 and 06) is smaller than its lead over the direct U-Net and resolved on one of the two; with two acquisitions it is not resolved ($+0.056$ [-0.025, 0.086]). At $\Delta = 0.16$ both U-Nets lead Poseidon-T on both single acquisitions.

Reference fine-tuning settings and output scale. Our adapter used one learning rate for all Poseidon layers and scaled the predicted change by the RMS of successive context differences. Poseidon's own fine-tuning code gives the re-initialised input and output layers and the lead-time conditioned normalisation parameters a ten times larger learning rate (5×10^{-4}) than the pretrained body (5×10^{-5}, weight decay 10^{-6}), and an output scale tied to recent motion is small in quiet windows, for instance just before a transition. Table 24 re-runs the selected direct recipe with these per-layer rates, alone and with the change scaled by the standard deviation of the context instead (seeds 0–1, all four splits; not selected on validation, so a sensitivity analysis). The per-layer rates alone change skill at $\Delta = 0.16$ by -0.003 to $+0.079$. Adding the standard-deviation scale gains $+0.102$, $+0.148$, $+0.180$ and $+0.028$ over the original adapter (C/20→C/2, C/2→C/20, new-particle fold 0, later cycles; all resolved, fold 0 descriptive) and narrows the direct U-Net's lead to 0.065, 0.036, 0.034 and 0.014 (the last unresolved). At $\Delta = 0.02$ this Poseidon-T has no resolved difference from the direct U-Net from C/20 to C/2 and on fold 0, beats it from C/2 to C/20 ($+0.033$ [0.018, 0.049]) and trails it slightly on later cycles (-0.008 [-0.024, -0.003]). Table 25 adds the matching scratch controls and the one-rate arm with the standard-deviation scale.

Scratch controls and validation selection of the output scale. Table 25 adds Poseidon-T trained from scratch with the standard-deviation scale (one rate, and the reference per-layer rates) and pretrained Poseidon-T with one rate and this scale (seeds 0–1, all four splits). The scale accounts for most of the gain on three splits: with one rate it adds $+0.101$, $+0.139$, $+0.179$ and $+0.011$ at $\Delta = 0.16$ (C/20→C/2, C/2→C/20, new-particle fold 0, later cycles; the last unresolved), and the reference rates add at most 0.017 on top. The pretraining gain shrinks on three splits but grows on fold 0: pretrained minus scratch at $\Delta = 0.16$ is $+0.068$, $+0.024$, $+0.064$ and $+0.051$ with one

Table 20: Data efficiency of pretraining. C/20→C/2 with one, two or four C/20 training acquisitions (03; 03 and 06, both 25°C; 03–06, adding 15 and 45°C), identical test windows. Direct lead-time recipe, 6,000 steps; seeds 0–3 at one and two acquisitions, 0–1 at four. Each budget averages all its seeds; the difference of gains uses the same particle resamples without discarding the added seeds (skill: seed mean ± SD). Differences with 95% paired particle-cluster bootstrap intervals (2,000 resamples). Pre-registered hypothesis: the pretraining gain (pretrained − scratch) is larger with one training acquisition than with four; supported if the difference of the two gains has an interval above zero at either horizon, refuted if below, otherwise inconclusive. Stated confounds: validation groups particles with one or two acquisitions and blocks with four, and the step from two to four acquisitions adds temperatures as well as data. The separate acquisition-06 replication uses seeds 0–1 and the registered replication estimands: arithmetic mean of paired seed contrasts, shared particle resamples (seed 20260924); all per-seed and secondary-horizon results are retained. Acquisition 06 has 15 indexed particles versus 21 in acquisition 03, so this is not a fixed-sample-size acquisition comparison. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Training acquisitions	Forecaster or contrast	$\Delta = 0.02$	$\Delta = 0.16$
1 (acq. 03)	Poseidon-T, pretrained	0.424 ± 0.018	0.433 ± 0.025
	Poseidon-T, from scratch	0.274 ± 0.061	0.321 ± 0.117
	U-Net	0.234 ± 0.087	0.697 ± 0.020
	Pretrained − scratch	+0.151 [0.134, 0.177]	+0.112 [0.064, 0.191]
	Poseidon-T − U-Net	+0.191 [0.156, 0.269]	-0.263 [-0.283, -0.212]
2 (acq. 03, 06)	Poseidon-T, pretrained	0.444 ± 0.013	0.662 ± 0.010
	Poseidon-T, from scratch	0.291 ± 0.033	0.462 ± 0.053
	U-Net	0.375 ± 0.036	0.792 ± 0.005
	Pretrained − scratch	+0.153 [0.127, 0.171]	+0.200 [0.133, 0.269]
	Poseidon-T − U-Net	+0.069 [0.044, 0.100]	-0.130 [-0.196, -0.104]
4 (acq. 03–06)	Poseidon-T, pretrained	0.457 ± 0.017	0.625 ± 0.005
	Poseidon-T, from scratch	0.321 ± 0.026	0.453 ± 0.001
	U-Net	0.478 ± 0.005	0.800 ± 0.005
	Pretrained − scratch	+0.137 [0.122, 0.158]	+0.172 [0.137, 0.235]
	Poseidon-T − U-Net	-0.020 [-0.045, -0.008]	-0.174 [-0.207, -0.160]
Pretraining gain, 1 − 4 acquisitions		+0.014 [0.002, 0.031]	-0.060 [-0.095, -0.038]
Pre-registered verdict: mixed: supported at one horizon, refuted at the other ($\Delta = 0.02$: gain larger with one acquisition; $\Delta = 0.16$: gain smaller with one acquisition); at one acquisition Poseidon-T overtakes the U-Net at $\Delta = 0.02$ and trails the U-Net at 0.16.			
1 (acquisition 06)	Poseidon-T, pretrained	0.404 ± 0.030	0.667 ± 0.009
	Poseidon-T, from scratch	0.300 ± 0.020	0.560 ± 0.023
	U-Net	0.109 ± 0.100	0.780 ± 0.013
	Pretrained − scratch	+0.104 [0.080, 0.146]	+0.107 [0.035, 0.167]
	Poseidon-T − U-Net	+0.294 [0.104, 0.388]	-0.112 [-0.194, -0.091]

Table 21: Scarce training data, C/20 → C/2: forecasters trained on one C/20 acquisition (03 or 06) or on two, tested on the same C/2 windows. Skill: seed means (step-by-step U-Net seeds 0–3; Poseidon-T, its scratch control and the direct U-Net seeds 0–3, or 0–1 with acquisition 06). Differences at each Δ: 95% paired particle-cluster bootstrap intervals over all seeds of both forecasters (2,000 resamples); bold: resolved. The baselines were not refitted to these reduced training sets.

Training data	Δ	Poseidon-T	Scratch	Direct U-Net	Step-by-step	Poseidon − step-by-step	Poseidon − direct	Direct − step-by-step
One acquisition (03)	0.02	0.424	0.274	0.234	0.300	+0.124 [−0.004, 0.165]	**+0.191 [0.156, 0.269]**	**−0.066 [−0.200, −0.015]**
	0.16	0.433	0.321	0.697	0.597	**−0.164 [−0.197, −0.055]**	**−0.263 [−0.283, −0.212]**	**+0.100 [0.070, 0.178]**
One acquisition (06)	0.02	0.404	0.300	0.109	0.314	**+0.090 [0.0005, 0.116]**	**+0.294 [0.107, 0.384]**	**−0.204 [−0.299, −0.066]**
	0.16	0.667	0.560	0.780	0.756	**−0.089 [−0.139, −0.068]**	**−0.112 [−0.193, −0.092]**	**+0.023 [0.011, 0.077]**
Two acquisitions	0.02	0.444	0.291	0.375	0.388	+0.056 [−0.025, 0.086]	**+0.069 [0.044, 0.100]**	−0.013 [−0.125, 0.020]
	0.16	0.662	0.462	0.792	0.693	−0.031 [−0.054, 0.074]	**−0.130 [−0.196, −0.104]**	**+0.099 [0.055, 0.257]**

rate (all resolved), against $+0.172$, $+0.132$, $+0.039$ and $+0.189$ with the original scale; with the reference rates it is $+0.029$, $+0.014$, -0.015 and $+0.041$ (resolved on the first and last). With the standard-deviation scale the direct U-Net still leads at $\Delta = 0.16$ by 0.066, 0.044, 0.034 and 0.031 (all resolved; fold 0 descriptive). Applying the original selection rule (equal-weight validation skill at $\Delta = 0.005$, 0.02 and 0.16, averaged over the four splits; the fastest setting within 0.05 of the best) to the four settings selects one rate with the standard-deviation scale, which also has the highest validation skill (0.519, against 0.513 with the reference rates and 0.464 and 0.462 without the scale); its test mean skill is the highest on three of the four splits. The settings were compared after the original results were known, so this is a validated follow-up rather than a pre-registered selection.

Table 22: Fine-tuning budget. Direct lead-time recipe with the default 6,000 steps and with 24,000 steps (evaluation every 1,000 steps, the same cosine schedule stretched to the budget), seeds 0 and 1, identical test windows; skill: seed mean ± SD; differences with 95% paired particle-cluster bootstrap intervals (2,000 resamples). Pre-registered hypothesis: the Poseidon-T − U-Net difference at $\Delta = 0.16$ shrinks with the fourfold budget; supported if its change (24,000 − 6,000 steps; positive means the foundation model gains on the U-Net) has an interval above zero. Both models get the same budget; nothing else changes. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Split	Forecaster or contrast	$\Delta = 0.02$	$\Delta = 0.16$
C/20→C/2	U-Net, 6,000 steps	0.478 ± 0.005	0.800 ± 0.005
	U-Net, 24,000 steps	0.500 ± 0.032	0.793 ± 0.002
	U-Net, 24,000 − 6,000	+0.023 [0.012, 0.053]	-0.007 [-0.013, 0.012]
	Poseidon-T, 6,000 steps	0.457 ± 0.017	0.625 ± 0.005
	Poseidon-T, 24,000 steps	0.461 ± 0.016	0.656 ± 0.009
	Poseidon-T, 24,000 − 6,000	+0.004 [-0.003, 0.011]	+0.031 [0.018, 0.055]
	Poseidon-T − U-Net, 6,000	-0.020 [-0.045, -0.008]	-0.174 [-0.207, -0.160]
	Poseidon-T − U-Net, 24,000	-0.040 [-0.073, -0.027]	-0.137 [-0.181, -0.123]
	Change of Poseidon-T − U-Net	-0.019 [-0.052, -0.006]	+0.038 [0.019, 0.053]
Pre-registered verdict, C/20→C/2: supported.			
C/2→C/20	U-Net, 6,000 steps	0.285 ± 0.004	0.663 ± 0.003
	U-Net, 24,000 steps	0.243 ± 0.027	0.651 ± 0.002
	U-Net, 24,000 − 6,000	-0.042 [-0.172, 0.013]	-0.012 [-0.022, 0.006]
	Poseidon-T, 6,000 steps	0.298 ± 0.004	0.467 ± 0.025
	Poseidon-T, 24,000 steps	0.333 ± 0.011	0.536 ± 0.010
	Poseidon-T, 24,000 − 6,000	+0.035 [0.024, 0.057]	+0.069 [0.043, 0.104]
	Poseidon-T − U-Net, 6,000	+0.013 [-0.020, 0.077]	-0.196 [-0.293, -0.148]
	Poseidon-T − U-Net, 24,000	+0.091 [0.006, 0.297]	-0.115 [-0.184, -0.075]
	Change of Poseidon-T − U-Net	+0.078 [0.016, 0.223]	+0.081 [0.050, 0.117]
Pre-registered verdict, C/2→C/20: supported.			

Table 23: Poseidon-T learning rate. Poseidon-T direct lead-time recipe at three learning rates (seed 0, 6,000 steps; the default run is the one of the foundation-model table). Validation: mean masked SSE over persistence SSE at the three common validation horizons (lower is better); the learning rate is selected on validation only. Test skill on identical windows; last column: Poseidon-T − direct U-Net (seeds 0–4) at $\Delta = 0.16$ with a descriptive 95% paired particle-cluster interval (one Poseidon-T seed, no interval claim). Pre-registered rule: the U-Net lead survives learning-rate tuning only if the validation-selected run trails the direct U-Net at $\Delta = 0.16$ on both rate splits (point estimates). n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Split	Recipe	Validation	$\Delta = 0.02$	$\Delta = 0.16$	− U-Net, $\Delta = 0.16$
C/20→C/2	Poseidon-T, 3×10^{-5} (selected)	0.720	0.485	0.587	-0.204 [-0.264, -0.175]
	Poseidon-T, 10^{-4} (default)	0.727	0.445	0.629	-0.163 [-0.214, -0.137]
	Poseidon-T, 3×10^{-4}	0.756	0.426	0.550	-0.241 [-0.290, -0.224]
	Direct U-Net, seeds 0, 1, 2, 3, 4	–	0.463 ± 0.014	0.792 ± 0.018	–
C/2→C/20	Poseidon-T, 3×10^{-5}	0.483	0.307	0.427	-0.224 [-0.270, -0.192]
	Poseidon-T, 10^{-4} (default) (selected)	0.464	0.301	0.449	-0.202 [-0.274, -0.159]
	Poseidon-T, 3×10^{-4}	0.541	0.265	0.346	-0.304 [-0.358, -0.198]
	Direct U-Net, seeds 0, 1, 2, 3, 4	–	0.284 ± 0.007	0.651 ± 0.025	–
Pre-registered verdict: the U-Net lead survives Poseidon-T learning-rate tuning (selected run trails on both rate splits).					

E ADDITIONAL RESULTS

This appendix holds the full result tables behind Sec. 6. Every cell uses the causal windows and the paired particle-cluster intervals of Table 2; the supplementary results files list the runs behind each cell.

Full transfer tables. Table 29 is Table 2 with the gains split by current direction. Table 31 extends Table 2 to every graphite split, forecaster and horizon with seed standard deviations, and Table 38 does the same for NMC. Table 41 splits the U-Net's gain over the strongest baseline by test acquisition, and Table 43 repeats the comparison on the forecast origins shared by all horizons, and Table 44 gives the headline contrasts at the acquisition level: per-acquisition signs, leave-one-acquisition-out ranges and a bootstrap that resamples particles and acquisitions independently (crossed), which is indicative only with four to eight acquisitions. The acquisition diagnostic and all eight graphite rate/temperature and fold-0 headline contrasts use five direct U-Net seeds (0–4); later cycles retain two. At $\Delta = 0.16$ from C/20 to C/2, the direct U-Net minus particle-mean ridge effect is 0.111 pooled and 0.082 with equal acquisition weights; the pooled effect has crossed interval [0.036, 0.143] and leave-one-acquisition-out estimates 0.098–0.119. Seven of eight acquisitions have positive point estimates: acquisition 09 (6 °C) gives $+0.003$, while 11 (25 °C) gives -0.008; both are

Table 24: Poseidon reference fine-tuning recipe. Poseidon-T with the fine-tuning settings of its reference code. All rows use the selected direct lead-time recipe, conditioning (d, s), 6,000 steps, seeds 0–1 and identical test windows within a split. One learning rate: the runs of Table 16 (10^{-4} for all layers, weight decay 10^{-4}). Reference per-layer rates: 5×10^{-5} for the pretrained body and 5×10^{-4} for the re-initialised input and output layers and the lead-time conditioned normalisation parameters, weight decay 10^{-6}. SD-scaled change: the predicted change is scaled by the standard deviation of the context instead of the RMS of its successive differences. Skill: seed mean ± SD; differences with 95% paired particle-cluster bootstrap intervals (2,000 resamples) at $\Delta = 0.16$; direct U-Net: seeds 0–4 (0–1 on later cycles); fold 0 has five test particles (descriptive). n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Split	Forecaster	$\Delta = 0.02$	$\Delta = 0.16$	− direct U-Net	− one rate
C/20→C/2	Poseidon-T, one learning rate (10^{-4})	0.457 ± 0.017	0.625 ± 0.005	-0.167 [-0.199, -0.152]	–
	Poseidon-T, reference per-layer rates	0.407 ± 0.028	0.622 ± 0.004	-0.169 [-0.205, -0.159]	-0.003 [-0.018, 0.017]
	Poseidon-T, reference rates, SD-scaled change	0.465 ± 0.031	0.727 ± 0.012	-0.065 [-0.088, -0.026]	+0.102 [0.084, 0.156]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.463 ± 0.014	0.792 ± 0.018	–	–
C/2→C/20	Poseidon-T, one learning rate (10^{-4})	0.298 ± 0.004	0.467 ± 0.025	-0.184 [-0.273, -0.139]	–
	Poseidon-T, reference per-layer rates	0.312 ± 0.005	0.486 ± 0.016	-0.165 [-0.263, -0.120]	+0.019 [0.005, 0.027]
	Poseidon-T, reference rates, SD-scaled change	0.317 ± 0.030	0.615 ± 0.014	-0.036 [-0.075, -0.009]	+0.148 [0.115, 0.201]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.284 ± 0.007	0.651 ± 0.025	–	–
New particle, fold 0	Poseidon-T, one learning rate (10^{-4})	0.427 ± 0.007	0.539 ± 0.050	-0.214 [-0.218, -0.137]	–
	Poseidon-T, reference per-layer rates	0.470 ± 0.002	0.618 ± 0.036	-0.135 [-0.142, -0.082]	+0.079 [0.055, 0.080]
	Poseidon-T, reference rates, SD-scaled change	0.490 ± 0.018	0.719 ± 0.013	-0.034 [-0.035, -0.019]	+0.180 [0.116, 0.194]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.444 ± 0.017	0.753 ± 0.012	–	–
Later cycles	Poseidon-T, one learning rate (10^{-4})	0.626 ± 0.005	0.873 ± 0.006	-0.042 [-0.074, -0.031]	–
	Poseidon-T, reference per-layer rates	0.620 ± 0.003	0.876 ± 0.002	-0.038 [-0.061, -0.032]	+0.004 [-0.001, 0.016]
	Poseidon-T, reference rates, SD-scaled change	0.601 ± 0.025	0.901 ± 0.018	-0.014 [-0.034, 0.002]	+0.028 [0.022, 0.068]
	Direct U-Net, seeds 0, 1	0.609 ± 0.063	0.915 ± 0.012	–	–

Table 25: Poseidon-T pretraining and adapter controls. Poseidon-T direct prediction, seeds 0–1, 6,000 steps. Matched pretrained-minus-scratch contrasts measure the pretraining gain under each adapter setting. One rate: 10^{-4}, weight decay 10^{-4}. Reference settings: body 5×10^{-5}, new/conditional parameters 5×10^{-4}, weight decay 10^{-6}; its contrast includes the weight-decay change. SD scales predicted changes by context SD instead of successive-difference RMS. Direct U-Net uses all five seeds on rate/fold splits and both seeds on later cycles. Fold 0 is descriptive (five particles). Skill: seed mean ± sample SD. Differences: 95% paired particle-cluster bootstrap intervals (2,000 resamples). Identical test windows within each split and horizon across all available arms and their seeds. An arm is pending until all its planned seeds are evaluated. Resolved means the interval excludes zero; five or fewer test particles are descriptive. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Setting or contrast	Quantity	$\Delta = 0.02$	$\Delta = 0.16$	− U-Net, 0.02	− U-Net, 0.16
C/20→C/2					
Pretrained, one rate	skill	0.457 ± 0.017	0.625 ± 0.005	-0.006 [-0.029, 0.016]	-0.167 [-0.199, -0.152]
Pretrained, reference rates + SD	skill	0.465 ± 0.031	0.727 ± 0.012	+0.001 [-0.031, 0.019]	-0.065 [-0.088, -0.026]
Pretrained, one rate + SD	skill	0.489 ± 0.019	0.726 ± 0.023	+0.025 [-0.005, 0.043]	-0.066 [-0.079, -0.026]
Scratch, one rate	skill	0.321 ± 0.026	0.453 ± 0.001	-0.143 [-0.168, -0.129]	-0.338 [-0.420, -0.300]
Scratch, reference rates + SD	skill	0.442 ± 0.006	0.697 ± 0.004	-0.022 [-0.040, -0.015]	-0.094 [-0.129, -0.081]
Scratch, one rate + SD	skill	0.401 ± 0.009	0.658 ± 0.000	-0.062 [-0.083, -0.054]	-0.134 [-0.191, -0.116]
Direct U-Net	skill	0.463 ± 0.014	0.792 ± 0.018	–	–
Pretraining gain, one rate	difference	+0.137 [0.122, 0.158]	+0.172 [0.137, 0.235]	–	–
Pretraining gain, reference rates + SD	difference	+0.023 [-0.002, 0.042]	+0.029 [0.019, 0.068]	–	–
Pretraining gain, one rate + SD	difference	+0.087 [0.065, 0.111]	+0.068 [0.043, 0.159]	–	–
SD-scale effect, pretrained	difference	+0.031 [0.005, 0.044]	+0.101 [0.082, 0.159]	–	–
Reference-rate effect with SD, pretrained	difference	-0.024 [-0.029, -0.020]	+0.001 [-0.013, 0.005]	–	–
SD-scale effect, scratch	difference	+0.081 [0.051, 0.100]	+0.204 [0.173, 0.259]	–	–
Reference-rate effect with SD, scratch	difference	+0.040 [0.037, 0.046]	+0.040 [0.022, 0.095]	–	–

unresolved. The support changes with horizon, and pooled SSE gives larger or more active particles greater weight. These intervals remain conditional on acquisitions of one cell.

Table 26: Poseidon-T pretraining and adapter controls. (continued) Conventions as in Table 25.

Setting or contrast	Quantity	$\Delta = 0.02$	$\Delta = 0.16$	− U-Net, 0.02	− U-Net, 0.16
C/2→C/20					
Pretrained, one rate	skill	0.298 ± 0.004	0.467 ± 0.025	+0.014 [-0.022, 0.080]	-0.184 [-0.273, -0.139]
Pretrained, reference rates + SD	skill	0.317 ± 0.030	0.615 ± 0.014	+0.033 [0.018, 0.049]	-0.036 [-0.075, -0.009]
Pretrained, one rate + SD	skill	0.328 ± 0.010	0.606 ± 0.020	+0.044 [0.029, 0.068]	-0.044 [-0.091, -0.015]
Scratch, one rate	skill	0.234 ± 0.000	0.335 ± 0.053	-0.050 [-0.086, 0.021]	-0.316 [-0.364, -0.246]
Scratch, reference rates + SD	skill	0.320 ± 0.014	0.601 ± 0.023	+0.036 [0.003, 0.115]	-0.050 [-0.073, -0.026]
Scratch, one rate + SD	skill	0.303 ± 0.011	0.583 ± 0.004	+0.019 [-0.009, 0.085]	-0.068 [-0.109, -0.040]
Direct U-Net	skill	0.284 ± 0.007	0.651 ± 0.025	–	–
Pretraining gain, one rate	difference	+0.064 [0.043, 0.084]	+0.132 [-0.0004, 0.211]	–	–
Pretraining gain, reference rates + SD	difference	-0.003 [-0.073, 0.026]	+0.014 [-0.004, 0.026]	–	–
Pretraining gain, one rate + SD	difference	+0.024 [-0.021, 0.045]	+0.024 [0.007, 0.044]	–	–
SD-scale effect, pretrained	difference	+0.030 [-0.013, 0.053]	+0.139 [0.108, 0.185]	–	–
Reference-rate effect with SD, pretrained	difference	-0.011 [-0.024, -0.004]	+0.009 [0.004, 0.017]	–	–
SD-scale effect, scratch	difference	+0.070 [0.047, 0.087]	+0.248 [0.166, 0.303]	–	–
Reference-rate effect with SD, scratch	difference	+0.017 [-0.001, 0.045]	+0.018 [0.006, 0.041]	–	–

Table 27: Poseidon-T pretraining and adapter controls. (continued) Conventions as in Table 25.

Setting or contrast	Quantity	$\Delta = 0.02$	$\Delta = 0.16$	− U-Net, 0.02	− U-Net, 0.16
New particle, fold 0					
Pretrained, one rate	skill	0.427 ± 0.007	0.539 ± 0.050	-0.016 [-0.046, -0.015]	-0.214 [-0.218, -0.137]
Pretrained, reference rates + SD	skill	0.490 ± 0.018	0.719 ± 0.013	+0.046 [-0.022, 0.056]	-0.034 [-0.035, -0.019]
Pretrained, one rate + SD	skill	0.491 ± 0.011	0.719 ± 0.006	+0.047 [-0.008, 0.054]	-0.034 [-0.037, -0.013]
Scratch, one rate	skill	0.338 ± 0.046	0.500 ± 0.091	-0.106 [-0.188, -0.094]	-0.253 [-0.419, -0.235]
Scratch, reference rates + SD	skill	0.444 ± 0.017	0.734 ± 0.014	+0.000 [-0.099, 0.013]	-0.019 [-0.090, -0.012]
Scratch, one rate + SD	skill	0.358 ± 0.096	0.654 ± 0.042	-0.086 [-0.202, -0.071]	-0.099 [-0.282, -0.079]
Direct U-Net	skill	0.444 ± 0.017	0.753 ± 0.012	–	–
Pretraining gain, one rate	difference	+0.089 [0.079, 0.161]	+0.039 [0.020, 0.207]	–	–
Pretraining gain, reference rates + SD	difference	+0.046 [0.043, 0.085]	-0.015 [-0.024, 0.070]	–	–
Pretraining gain, one rate + SD	difference	+0.133 [0.125, 0.195]	+0.064 [0.042, 0.267]	–	–
SD-scale effect, pretrained	difference	+0.063 [0.020, 0.069]	+0.179 [0.113, 0.202]	–	–
Reference-rate effect with SD, pretrained	difference	-0.001 [-0.022, 0.002]	+0.001 [-0.008, 0.003]	–	–
SD-scale effect, scratch	difference	+0.020 [-0.018, 0.023]	+0.154 [0.087, 0.157]	–	–
Reference-rate effect with SD, scratch	difference	+0.086 [0.084, 0.104]	+0.080 [0.067, 0.195]	–	–

Table 28: Poseidon-T pretraining and adapter controls. (continued) Conventions as in Table 25.

Setting or contrast	Quantity	$\Delta = 0.02$	$\Delta = 0.16$	− U-Net, 0.02	− U-Net, 0.16
Later cycles					
Pretrained, one rate	skill	0.626 ± 0.005	0.873 ± 0.006	+0.017 [-0.017, 0.026]	-0.042 [-0.074, -0.031]
Pretrained, reference rates + SD	skill	0.601 ± 0.025	0.901 ± 0.018	-0.008 [-0.024, -0.003]	-0.014 [-0.034, 0.002]
Pretrained, one rate + SD	skill	0.594 ± 0.058	0.884 ± 0.036	-0.015 [-0.032, -0.012]	-0.031 [-0.053, -0.021]
Scratch, one rate	skill	0.426 ± 0.060	0.684 ± 0.014	-0.182 [-0.252, -0.165]	-0.231 [-0.349, -0.216]
Scratch, reference rates + SD	skill	0.580 ± 0.016	0.860 ± 0.001	-0.029 [-0.061, -0.020]	-0.054 [-0.101, -0.048]
Scratch, one rate + SD	skill	0.565 ± 0.009	0.833 ± 0.010	-0.044 [-0.110, -0.029]	-0.082 [-0.182, -0.065]
Direct U-Net	skill	0.609 ± 0.063	0.915 ± 0.012	–	–
Pretraining gain, one rate	difference	+0.200 [0.163, 0.237]	+0.189 [0.163, 0.273]	–	–
Pretraining gain, reference rates + SD	difference	+0.021 [0.006, 0.048]	+0.041 [0.027, 0.097]	–	–
Pretraining gain, one rate + SD	difference	+0.029 [0.011, 0.083]	+0.051 [0.042, 0.129]	–	–
SD-scale effect, pretrained	difference	-0.032 [-0.040, -0.005]	+0.011 [-0.001, 0.045]	–	–
Reference-rate effect with SD, pretrained	difference	+0.007 [0.001, 0.017]	+0.017 [-0.012, 0.050]	–	–
SD-scale effect, scratch	difference	+0.139 [0.128, 0.148]	+0.149 [0.141, 0.170]	–	–
Reference-rate effect with SD, scratch	difference	+0.015 [0.004, 0.049]	+0.028 [0.008, 0.079]	–	–

Table 29: Learned forecasters against strong baselines. Skill $= 1 - \mathrm{SSE}/\mathrm{SSE}_{\mathrm{persistence}}$ on identical causal windows per split, fold and charge-axis horizon Δ; seed mean$_{\pm\mathrm{SD}}$. New-particle means need all five identity folds (superscript: folds available). Learned rows step in charge with inputs (d, s); the direct lead-time U-Net predicts each horizon in one step (matched-budget control of the foundation-model recipes). Direct seeds: 0–4 on rate/temperature splits and graphite fold 0; 0–1 on later cycles and NMC811. Bottom blocks: model minus the strongest available baseline on the same windows (P persistence, T trend, A AR(2), R SOC replay, H history-aligned replay, Q population replay, D damped trend, M history-mean persistence, X pixel replay, S scaled population replay, L particle-mean ridge), 95% paired particle-cluster bootstrap interval. Marks: i interim foundation-model recipe; v validation grouping differs from the U-Net row; x mixed protocols; s fewer windows. NMC811 new particle: †fold 0 only (only folds 0–1 of five are evaluated). n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows). Full results are in the appendix.

(a) Graphite	C/20→C/2		C/2→C/20		(C/20+C/2) →2C		C/2: 25 →6–15°C		C/2: 25 →35–45°C		C/20: 25 →15,45°C	
Δ	.02	.16	.02	.16	.02	.16	.02	.16	.02	.16	.02	.16
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	$-.06$	$-.76$	$-.36$	$-.95$	$-.02$	$-.58$	$-.21$	-1.81	$-.10$	$-.58$	$-.48$	-1.05
AR(2)	.18	.05	$-.27$	.00	.23	$-.04$	$-.22$	-1.19	$-.04$	-1.21	.09	.15
SOC replay	.16	.60	.13	.36	.11	.39	.17	.53	.18	.54	$-.01$	.17
History-aligned replay	.25	.65	.20	.46	.14	.48	.23	.60	.25	.55	.14	.38
Population replay	.19	.50	.14	.35	.09	.30	.16	.33	.14	.54	.09	.25
U-Net, autoregressive (d, s)	$.51_{\pm.01}$	$.65_{\pm.05}$	$.38_{\pm.01}$	$.54_{\pm.02}$	$.50_{\pm.01}$	$.53_{\pm.06}$	$.46_{\pm.02}$	$.60_{\pm.09}$	$.49_{\pm.03}$	$.68_{\pm.02}$	$.30_{\pm.04}$	$.35_{\pm.07}$
U-Net, direct lead-time (d, s)	$.46_{\pm.01}$	$.79_{\pm.02}$	$.30_{\pm.01}$	$.63_{\pm.03}$	$.34_{\pm.05}$	$.65_{\pm.01}$	$.40_{\pm.10}$	$.73_{\pm.01}$	$.37_{\pm.10}$	$.77_{\pm.02}$	$.19_{\pm.07}$	$.49_{\pm.02}$
FNO (d, s)	$.53_{\pm.00}$	$.70_{\pm.01}$	$.36_{\pm.01}$	$.49_{\pm.01}$	$.41_{\pm.04}$	$-1.25_{\pm1.79}$	$.43_{\pm.06}$	$.61_{\pm.03}$	$.46_{\pm.04}$	$.68_{\pm.06}$	$.27_{\pm.04}$	$.37_{\pm.05}$
ConvLSTM (d, s)	$.47_{\pm.01}$	$.72_{\pm.03}$	$.38_{\pm.00}$	$.59_{\pm.00}$	$.41_{\pm.01}$	$.60_{\pm.05}$	$.44_{\pm.00}$	$.68_{\pm.01}$	$.42_{\pm.00}$	$.73_{\pm.02}$	$.20_{\pm.06}$	$.40_{\pm.01}$
ViT (d, s)	$.12_{\pm.12}$	$.33_{\pm.25}$	$.20_{\pm.00}$	$.55_{\pm.01}$	$.14_{\pm.08}$	$.50_{\pm.17}$	$.22_{\pm.12}$	$.39_{\pm.31}$	$.14_{\pm.02}$	$.72_{\pm.01}$	$.07_{\pm.10}$	$.21_{\pm.30}$
Grey box AC-RD (d, s)	$.20_{\pm.01}$	$.58_{\pm.05}$	$.06_{\pm.04}$	$.19_{\pm.07}$	$.10_{\pm.05}$	$.56_{\pm.01}$	$.01_{\pm.00}$	$-.39_{\pm.21}$	$-.06_{\pm.07}$	$.14_{\pm.20}$	$.12_{\pm.05}$	$.35_{\pm.09}$
Replay-augmented U-Net (d, s)	$.50_{\pm.02}$	$.65_{\pm.05}$	$.33_{\pm.03}$	$.42_{\pm.07}$	$.48_{\pm.01}$	$.63_{\pm.01}$	$.31_{\pm.13}$	$.55_{\pm.17}$	$.43_{\pm.02}$	$.70_{\pm.02}$	$.27_{\pm.00}$	$.35_{\pm.05}$
Foundation model (val.-selected)	$.46_{\pm.02}$	$.63_{\pm.01}$	$.31_{\pm.00}$	$.46_{\pm.03}$	$.19^{\mathrm{i}}$	-1.70^{i}	$.31^{\mathrm{i}}$	$.14^{\mathrm{i}}$	n/r	n/r	n/r	n/r
U-Net − best baseline	$+.17^{\mathrm{L}}$ [.13,.20]	$-.03^{\mathrm{L}}$ [-.13,.01]	$+.18^{\mathrm{H}}$ [.11,.23]	$+.08^{\mathrm{H}}$ [.04,.15]	$+.23^{\mathrm{L}}$ [.06,.24]	$-.03^{\mathrm{L}}$ [-.17,-.02]	$+.14^{\mathrm{L}}$ [.02,.21]	$-.01^{\mathrm{H}}$ [-.11,.09]	$+.24^{\mathrm{H}}$ [.16,.29]	$+.12^{\mathrm{X}}$ [.09,.23]	$+.16^{\mathrm{H}}$ [.08,.22]	$-.03^{\mathrm{H}}$ [-.10,.03]
lithiation	$+.11^{\mathrm{L}}$ [.07,.17]	$-.11^{\mathrm{L}}$ [-.23,-.06]	$+.14^{\mathrm{H}}$ [.03,.21]	$.00^{\mathrm{H}}$ [-.05,.08]	$+.20^{\mathrm{L}}$ [.08,.21]	$-.02^{\mathrm{H}}$ [-.04,.20]	$+.16^{\mathrm{L}}$ [.05,.24]	$-.11^{\mathrm{H}}$ [-.24,-.03]	$+.14^{\mathrm{H}}$ [.07,.19]	$+.03^{\mathrm{H}}$ [-.08,.09]	$+.08^{\mathrm{H}}$ [-.04,.18]	$-.12^{\mathrm{H}}$ [-.18,-.07]
delithiation	$+.25^{\mathrm{L}}$ [.20,.28]	$+.13^{\mathrm{L}}$ [.06,.15]	$+.19^{\mathrm{L}}$ [.13,.24]	$+.08^{\mathrm{L}}$ [.00,.15]	$+.24^{\mathrm{A}}$ [.16,.24]	$-.04^{\mathrm{L}}$ [-.40,-.02]	$+.08^{\mathrm{L}}$ [-.07,.17]	$+.19^{\mathrm{L}}$ [.16,.25]	$+.23^{\mathrm{A}}$ [.01,.33]	$+.20^{\mathrm{Q}}$ [.16,.31]	$+.23^{\mathrm{H}}$ [.16,.30]	$+.04^{\mathrm{H}}$ [-.06,.18]
Direct U-Net − best baseline	$+.12^{\mathrm{L}}$ [.08,.13]	$+.11^{\mathrm{L}}$ [.05,.13]	$+.10^{\mathrm{H}}$ [.05,.14]	$+.16^{\mathrm{H}}$ [.13,.24]	$+.07^{\mathrm{L}}$ [-.03,.08]	$+.09^{\mathrm{L}}$ [.02,.10]	$+.08^{\mathrm{L}}$ [-.04,.16]	$+.13^{\mathrm{H}}$ [.06,.27]	$+.13^{\mathrm{H}}$ [-.02,.21]	$+.21^{\mathrm{X}}$ [.14,.36]	$+.04^{\mathrm{H}}$ [-.04,.12]	$+.11^{\mathrm{H}}$ [.06,.19]
lithiation	$+.09^{\mathrm{L}}$ [.02,.11]	$+.08^{\mathrm{L}}$ [.01,.10]	$+.09^{\mathrm{H}}$ [.01,.13]	$+.11^{\mathrm{H}}$ [.07,.20]	$+.08^{\mathrm{L}}$ [.02,.08]	$+.02^{\mathrm{H}}$ [.00,.19]	$+.09^{\mathrm{L}}$ [.00,.15]	$+.07^{\mathrm{H}}$ [.01,.18]	$+.02^{\mathrm{H}}$ [-.17,.12]	$+.13^{\mathrm{H}}$ [.10,.18]	$+.10^{\mathrm{H}}$ [.03,.20]	$+.03^{\mathrm{H}}$ [-.04,.08]
delithiation	$+.17^{\mathrm{L}}$ [.15,.21]	$+.17^{\mathrm{L}}$ [.11,.19]	$+.09^{\mathrm{L}}$ [.02,.14]	$+.14^{\mathrm{L}}$ [.10,.18]	$+.06^{\mathrm{A}}$ [.02,.07]	$+.14^{\mathrm{L}}$ [.03,.14]	$+.07^{\mathrm{L}}$ [-.16,.19]	$+.24^{\mathrm{L}}$ [.20,.31]	$+.13^{\mathrm{A}}$ [-.10,.24]	$+.26^{\mathrm{Q}}$ [.21,.37]	$-.01^{\mathrm{H}}$ [-.13,.08]	$+.18^{\mathrm{H}}$ [.11,.33]

(b) Graphite	New particle		New particle fold 0		Later cycles acq. 14		Later cycles 2C	
Δ	.02	.16	.02	.16	.02	.16	.02	.16
Persistence	.00	.00	.00	.00	.00	.00	.00	.00
Trend	$-.21$	$-.92$	$-.16$	$-.74$	.05	$-.89$	.01	$-.72$
AR(2)	.15	.08	.19	.07	.37	.03	.22	$-.08$
SOC replay	n/a	n/a	n/a	n/a	.42	.75	.30	.76
History-aligned replay	n/a	n/a	n/a	n/a	.40	.72	.22	.29
Population replay	.16	.42	.18	.52	.34	.69	.19	.67
U-Net, autoregressive (d, s)	$.44_{\pm.01}$	$.50_{\pm.02}$	$.53_{\pm.01}$	$.71_{\pm.06}$	$.67_{\pm.03}$	$.86_{\pm.04}$	$.67_{\pm.04}$	$.79_{\pm.10}$
U-Net, direct lead-time (d, s)	n/r$^{1/5}$	n/r$^{1/5}$	$.45_{\pm.02}$	$.75_{\pm.01}$	$.61_{\pm.06}$	$.91_{\pm.01}$	n/r	n/r
FNO (d, s)	n/r$^{1/5}$	n/r$^{1/5}$	$.55_{\pm.00}$	$.71_{\pm.04}$	$.60_{\pm.03}$	$.76_{\pm.16}$	n/r	n/r
ConvLSTM (d, s)	n/r$^{1/5}$	n/r$^{1/5}$	$.54_{\pm.01}$	$.75_{\pm.01}$	$.64_{\pm.05}$	$.83_{\pm.04}$	n/r	n/r
ViT (d, s)	n/r$^{1/5}$	n/r$^{1/5}$	$.33_{\pm.03}$	$.72_{\pm.03}$	$.43_{\pm.04}$	$.78_{\pm.04}$	n/r	n/r
Grey box AC-RD (d, s)	n/r$^{1/5}$	n/r$^{1/5}$	$.21_{\pm.02}$	$.61_{\pm.07}$	$.37_{\pm.01}$	$.79_{\pm.00}$	n/r	n/r
Replay-augmented U-Net (d, s)	n/r$^{0/5}$	n/r$^{0/5}$	n/a	n/a	n/a	n/a	n/r	n/r
Foundation model (val.-selected)	n/r	n/r	$.44_{\pm.01}$	$.56_{\pm.02}$	$.63_{\pm.01}$	$.87_{\pm.01}$	n/r	n/r
U-Net − best baseline	$+.16^{\mathrm{L}}$ [.13,.20]	$-.07^{\mathrm{L}}$ [-.10,.02]	$+.22^{\mathrm{L}}$ [.20,.29]	$+.06^{\mathrm{L}}$ [-.01,.07]	$+.19^{\mathrm{X}}$ [.04,.25]	$-.02^{\mathrm{X}}$ [-.08,.16]	$+.33^{\mathrm{L}}$ [.14,.34]	$-.02^{\mathrm{L}}$ [-.08,.01]
lithiation	$+.10^{\mathrm{L}}$ [.08,.17]	$-.13^{\mathrm{L}}$ [-.18,.01]	$+.18^{\mathrm{L}}$ [.15,.30]	$-.02^{\mathrm{L}}$ [-.07,.02]	$+.21^{\mathrm{R}}$ [.12,.28]	$-.03^{\mathrm{X}}$ [-.10,.21]	$+.24^{\mathrm{R}}$ [.10,.25]	$-.14^{\mathrm{X}}$ [-.14,-.04]
delithiation	$+.22^{\mathrm{L}}$ [.20,.28]	$+.02^{\mathrm{L}}$ [-.04,.09]	$+.27^{\mathrm{L}}$ [.21,.33]	$+.19^{\mathrm{L}}$ [.01,.19]	$+.02^{\mathrm{X}}$ [-.06,.04]	$+.00^{\mathrm{X}}$ [-.04,.02]	$+.37^{\mathrm{H}}$ [.28,.37]	$+.23^{\mathrm{L}}$ [.14,.24]
Direct U-Net − best baseline	n/r	n/r	$+.14^{\mathrm{L}}$ [.13,.20]	$+.11^{\mathrm{L}}$ [.09,.23]	$+.13^{\mathrm{X}}$ [-.02,.19]	$+.04^{\mathrm{X}}$ [-.03,.22]	n/r	n/r
lithiation	n/r	n/r	$+.10^{\mathrm{L}}$ [.08,.20]	$+.05^{\mathrm{L}}$ [.03,.23]	$+.16^{\mathrm{R}}$ [.07,.24]	$+.04^{\mathrm{X}}$ [-.04,.27]	n/r	n/r
delithiation	n/r	n/r	$+.19^{\mathrm{L}}$ [.14,.23]	$+.20^{\mathrm{L}}$ [.12,.23]	$-.06^{\mathrm{X}}$ [-.13,-.03]	$+.01^{\mathrm{X}}$ [-.03,.02]	n/r	n/r

Table 30: Learned forecasters against strong baselines (continued). Same matching, seed and interval rules as the preceding panel.

(c) NMC	C/3→1C,2C		(C/3+1C) →2C		2C→C/3		NMC: new particle	
Δ	.02	.16	.02	.16	.02	.16	.02	.16
Persistence	.00	.00	.00	.00	.00	.00	.00	.00
Trend	.62	.51	.40	.10	.03	.53	.31	.48
AR(2)	.69	.80	.68	.80	.58	.82	.61	.76
SOC replay	.54	.69	.51	.81	.13	.30	n/a	n/a
History-aligned replay	.57	.70	.68	.85	.11	.32	n/a	n/a
Population replay	.56	.67	.50	.66	.26	.42	.45	.71
U-Net, autoregressive (d, s)	$.81_{\pm.01}$	$.90_{\pm.00}$	$.80_{\pm.01}$	$.89_{\pm.00}$	$.54_{\pm.08}$	$.71_{\pm.08}$	$.76_{\pm.01}$	$.90_{\pm.00}$
Foundation model (val.-selected)	$.32^{\mathrm{i}}_{\pm.16}$	$-9.30^{\mathrm{i}}_{\pm13.6}$	n/r	n/r	$.55^{\mathrm{i}}_{\pm.01}$	$.34^{\mathrm{i}}_{\pm.01}$	n/r	n/r
U-Net − best baseline	$+.06^{\mathrm{S}}$ [.02,.09]	$+.02^{\mathrm{L}}$ [-.01,.05]	$+.12^{\mathrm{H}}$ [.05,.19]	$+.02^{\mathrm{L}}$ [-.06,.07]	$-.04^{\mathrm{A}}$ [-.10,-.01]	$-.11^{\mathrm{A}}$ [-.20,-.06]	$.00^{\mathrm{L}}$ [-.02,.01]	$+.01^{\mathrm{L}}$ [-.01,.03]
lithiation	$+.12^{\mathrm{S}}$ [.10,.16]	$+.04^{\mathrm{L}}$ [.02,.06]	$+.07^{\mathrm{H}}$ [-.02,.17]	$-.01^{\mathrm{H}}$ [-.04,.02]	$+.15^{\mathrm{A}}$ [.03,.21]	$+.06^{\mathrm{A}}$ [-.06,.11]	$-.02^{\mathrm{L}}$ [-.03,.00]	$-.01^{\mathrm{L}}$ [-.02,.01]
delithiation	$+.01^{\mathrm{S}}$ [-.06,.06]	$-.01^{\mathrm{X}}$ [-.07,.07]	$+.03^{\mathrm{S}}$ [-.10,.11]	$.00^{\mathrm{S}}$ [-.11,.05]	$-.42^{\mathrm{D}}$ [-.53,-.33]	$-.26^{\mathrm{A}}$ [-.33,-.21]	$+.01^{\mathrm{L}}$ [-.02,.03]	$+.02^{\mathrm{L}}$ [.00,.06]

(d) NMC811	NMC811 ageing		NMC811 voltage window		NMC811: new particle†	
Δ	.02	.16	.02	.16	.02	.16
Persistence	.00	.00	.00	.00	.00	.00
Trend	−2.31	−6.38	−1.41	−2.89	−1.58	−2.86
AR(2)	.41	.77	.30	.52	.39	.71
SOC replay	.17	.85	.11	.39	n/a	n/a
History-aligned replay	.19	.83	.11	.45	n/a	n/a
Population replay	.16	.78	.12	.37	.22	.64
U-Net, autoregressive (d, s)	$.61_{\pm.00}$	$.87_{\pm.02}$	$.31_{\pm.01}$	$.42_{\pm.05}$	$.62_{\pm.01}$	$.78_{\pm.01}$
U-Net, direct lead-time (d, s)	$.54_{\pm.06}$	$.89_{\pm.01}$	$-.02_{\pm.11}$	$.50_{\pm.01}$	$.55_{\pm.01}$	$.82_{\pm.00}$
U-Net − best baseline	$+.20^{\mathrm{A}}$ [.19,.22]	$+.02^{\mathrm{L}}$ [.01,.03]	$+.01^{\mathrm{A}}$ [-.04,.04]	$-.11^{\mathrm{A}}$ [-.21,-.04]	$+.23^{\mathrm{A}}$ [.21,.26]	$-.01^{\mathrm{L}}$ [-.05,.01]
lithiation	$+.20^{\mathrm{A}}$ [.19,.22]	$+.04^{\mathrm{L}}$ [.03,.05]	$+.08^{\mathrm{A}}$ [.06,.09]	$-.04^{\mathrm{A}}$ [-.11,.00]	$+.20^{\mathrm{A}}$ [.18,.22]	$+.01^{\mathrm{L}}$ [-.01,.03]
delithiation	$+.20^{\mathrm{A}}$ [.19,.21]	$-.01^{\mathrm{R}}$ [-.02,.01]	$-.06^{\mathrm{A}}$ [-.12,.00]	$-.16^{\mathrm{A}}$ [-.29,-.07]	$+.25^{\mathrm{L}}$ [.21,.29]	$-.03^{\mathrm{L}}$ [-.08,.00]
Direct U-Net − best baseline	$+.13^{\mathrm{A}}$ [.12,.14]	$+.04^{\mathrm{L}}$ [.03,.06]	$-.32^{\mathrm{A}}$ [-.48,-.20]	$-.02^{\mathrm{A}}$ [-.11,.05]	$+.15^{\mathrm{A}}$ [.14,.17]	$+.02^{\mathrm{L}}$ [-.01,.04]
lithiation	$+.14^{\mathrm{A}}$ [.12,.15]	$+.04^{\mathrm{L}}$ [.03,.07]	$+.02^{\mathrm{A}}$ [.00,.04]	$+.06^{\mathrm{A}}$ [.01,.10]	$+.14^{\mathrm{A}}$ [.13,.16]	$+.04^{\mathrm{L}}$ [.04,.06]
delithiation	$+.12^{\mathrm{A}}$ [.11,.13]	$+.02^{\mathrm{R}}$ [.02,.04]	$-.62^{\mathrm{A}}$ [-.87,-.41]	$-.08^{\mathrm{A}}$ [-.21,.02]	$+.15^{\mathrm{L}}$ [.10,.19]	$+.01^{\mathrm{L}}$ [-.04,.03]

Table 31: All forecasters, splits and horizons. Skill $= 1 - \mathrm{SSE}/\mathrm{SSE}_{\mathrm{persistence}}$ on identical causal windows within each split, fold and horizon (charge-axis horizon Δ); seed mean$_{\pm\mathrm{SD}}$ over available training seeds (SD describes seeds, not experimental replication). New-particle means need all five identity folds (superscript: folds available); fold 0 is also shown on its own. Learned rows step in charge with direction and charge-state inputs (d, s) and roll out autoregressively, except the direct lead-time U-Net, which predicts each horizon in one step (matched-budget control of the foundation-model recipes, same windows). Direct seeds: 0–4 on the seven graphite rate/temperature splits and graphite fold 0; 0–1 on later cycles, NMC-2 and NMC811. Bottom block: U-Net, and direct lead-time U-Net, minus the strongest available baseline (up to eleven; chosen on the same windows, identically for both) (superscript: P persistence, T trend, A AR(2), R SOC replay, H history-aligned replay, Q population replay, D damped trend, M history-mean persistence, X pixel replay, S scaled population replay, L particle-mean ridge) with 95% paired particle-cluster bootstrap intervals (2,000 resamples of particles, seed-averaged SSE; conditional on the observed acquisitions and electrode). The foundation-model row is the recipe selected on validation only where that selection exists; marks: i interim recipe (inherited one-step fine-tuning chosen by validation loss); v validation grouping differs from the U-Net row; x seeds with different protocols; s scored on fewer windows than the core rows. Rows without any planned run in a panel are omitted. NMC811 ageing: cycles 3–29 to 99–117 (4.2 V upper cut-off); NMC811 voltage window: the same training cycles to cycles at a 4.4 or 4.6 V cut-off (confounded with cycle number). NMC811 new particle: †fold 0 only; only folds 0–1 of five are evaluated, so its five-fold mean is withheld from paper tables and retained as missing in the CSV. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

	C/20→C/2						C/2→C/20					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	.12	−.06	−.22	−.40	−.76	−1.46	−.21	−.36	−.41	−.55	−.95	−.90
AR(2)	.28	.18	.10	.05	.05	.25	−.24	−.27	−.16	−.07	.00	.13
SOC replay	.10	.16	.24	.43	.60	.73	.09	.13	.22	.32	.36	.54
History-aligned replay	.19	.25	.34	.50	.65	.75	.13	.20	.31	.41	.46	.65
Population replay	.15	.19	.27	.36	.50	.70	.10	.14	.22	.29	.35	.59
Damped trend	.28	.19	.10	.04	.01	−.01	.07	.07	.07	.04	.01	.00
History-mean persistence	−.87	−.39	−.17	−.07	−.03	.00	−.72	−.30	−.15	−.07	−.03	−.01
SOC pixel replay	−.22	−.11	.05	.39	.61	.75	−.01	.04	.15	.27	.33	.53
Scaled population replay	.17	.15	.17	.15	.18	.35	.14	.17	.19	.06	−.47	−.07
Particle-mean ridge	.30	.34	.40	.51	.68	.78	.04	.07	.12	.22	.34	.63
U-Net, autoregressive (d, s)	$.49_{\pm.01}$	$.51_{\pm.01}$	$.54_{\pm.01}$	$.62_{\pm.01}$	$.65_{\pm.05}$	$.70_{\pm.05}$	$.34_{\pm.00}$	$.38_{\pm.01}$	$.43_{\pm.01}$	$.47_{\pm.01}$	$.54_{\pm.02}$	$.72_{\pm.02}$
U-Net, autoregressive, no inputs	$.48_{\pm.01}$	$.48_{\pm.02}$	$.49_{\pm.02}$	$.56_{\pm.05}$	$.47_{\pm.09}$	$.15_{\pm.22}$	$.36_{\pm.01}$	$.39_{\pm.00}$	$.44_{\pm.00}$	$.38_{\pm.01}$	$-.02_{\pm.09}$	$.13_{\pm.06}$
U-Net, autoregressive (d, u, s)	$.38_{\pm.02}$	$.37_{\pm.01}$	$.33_{\pm.02}$	$.33_{\pm.01}$	$.34_{\pm.07}$	$.45_{\pm.14}$	$.35_{\pm.00}$	$.39_{\pm.01}$	$.44_{\pm.01}$	$.46_{\pm.02}$	$.47_{\pm.07}$	$.71_{\pm.04}$
U-Net, direct lead-time (d, s)	$.41_{\pm.02}$	$.46_{\pm.01}$	$.55_{\pm.02}$	$.68_{\pm.01}$	$.79_{\pm.02}$	$.86_{\pm.00}$	$.22_{\pm.01}$	$.30_{\pm.01}$	$.42_{\pm.02}$	$.53_{\pm.02}$	$.63_{\pm.03}$	$.80_{\pm.01}$
FNO (d, s)	$.50_{\pm.01}$	$.53_{\pm.00}$	$.57_{\pm.00}$	$.63_{\pm.00}$	$.70_{\pm.01}$	$.76_{\pm.05}$	$.34_{\pm.01}$	$.36_{\pm.01}$	$.40_{\pm.01}$	$.44_{\pm.00}$	$.49_{\pm.01}$	$.70_{\pm.02}$
ConvLSTM (d, s)	$.45_{\pm.02}$	$.47_{\pm.01}$	$.52_{\pm.00}$	$.62_{\pm.03}$	$.72_{\pm.03}$	$.78_{\pm.04}$	$.35_{\pm.00}$	$.38_{\pm.00}$	$.45_{\pm.00}$	$.52_{\pm.01}$	$.59_{\pm.00}$	$.76_{\pm.01}$
ViT (d, s)	$.09_{\pm.09}$	$.12_{\pm.12}$	$.17_{\pm.16}$	$.25_{\pm.23}$	$.33_{\pm.25}$	$.49_{\pm.25}$	$.14_{\pm.00}$	$.20_{\pm.00}$	$.33_{\pm.01}$	$.45_{\pm.01}$	$.55_{\pm.01}$	$.75_{\pm.00}$
Grey box AC-RD (d, s)	$.14_{\pm.01}$	$.20_{\pm.01}$	$.29_{\pm.02}$	$.42_{\pm.03}$	$.58_{\pm.05}$	$.84_{\pm.01}$	$.00_{\pm.05}$	$.06_{\pm.04}$	$.14_{\pm.02}$	$.17_{\pm.00}$	$.19_{\pm.07}$	$.57_{\pm.11}$
Replay-augmented U-Net (d, s)	$.47_{\pm.02}$	$.50_{\pm.02}$	$.52_{\pm.01}$	$.56_{\pm.01}$	$.65_{\pm.05}$	$.79_{\pm.07}$	$.31_{\pm.01}$	$.33_{\pm.03}$	$.35_{\pm.04}$	$.35_{\pm.07}$	$.42_{\pm.07}$	$.65_{\pm.10}$
Foundation model (val.-selected)	$.44_{\pm.01}$	$.46_{\pm.02}$	$.52_{\pm.02}$	$.59_{\pm.00}$	$.63_{\pm.01}$	$.64_{\pm.01}$	$.25_{\pm.01}$	$.31_{\pm.00}$	$.38_{\pm.01}$	$.40_{\pm.01}$	$.46_{\pm.03}$	$.60_{\pm.04}$
U-Net − best baseline	$+.19^{\mathrm{L}}$ [.15,.21]	$+.17^{\mathrm{L}}$ [.13,.20]	$+.13^{\mathrm{L}}$ [.09,.16]	$+.11^{\mathrm{L}}$ [.00,.16]	$-.03^{\mathrm{L}}$ [-.13,.01]	$-.09^{\mathrm{L}}$ [-.11,-.06]	$+.20^{\mathrm{S}}$ [.11,.26]	$+.18^{\mathrm{H}}$ [.11,.23]	$+.12^{\mathrm{H}}$ [.05,.17]	$+.07^{\mathrm{H}}$ [.00,.11]	$+.08^{\mathrm{H}}$ [.04,.15]	$+.06^{\mathrm{H}}$ [.00,.17]
Direct U-Net − best baseline	$+.11^{\mathrm{L}}$ [.06,.12]	$+.12^{\mathrm{L}}$ [.08,.13]	$+.14^{\mathrm{L}}$ [.10,.16]	$+.17^{\mathrm{L}}$ [.10,.19]	$+.11^{\mathrm{L}}$ [.05,.13]	$+.08^{\mathrm{L}}$ [.07,.10]	$+.08^{\mathrm{S}}$ [-.08,.15]	$+.10^{\mathrm{H}}$ [.05,.14]	$+.11^{\mathrm{H}}$ [.04,.15]	$+.12^{\mathrm{H}}$ [.08,.17]	$+.16^{\mathrm{H}}$ [.13,.24]	$+.15^{\mathrm{H}}$ [.11,.24]

	(C/20+C/2) →2C						C/2: 25 →6–15°C					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	n/e
Trend	.08	−.02	−.11	−.25	−.58	−.98	−.04	−.21	−.39	−.80	−1.81	n/e
AR(2)	.35	.23	.11	.03	−.04	−.11	−.14	−.22	−.49	−.77	−1.19	n/e
SOC replay	.09	.11	.16	.26	.39	.36	.12	.17	.26	.42	.53	n/e
History-aligned replay	.10	.14	.18	.29	.48	.46	−.57	.23	.34	.49	.60	n/e
Population replay	.07	.09	.13	.21	.30	.26	.13	.16	.19	.23	.33	n/e
Damped trend	.26	.18	.08	.01	−.04	−.13	.19	.18	.10	.06	.03	n/e
History-mean persistence	−.87	−.36	−.15	−.06	−.01	.02	−.76	−.35	−.14	−.06	−.04	n/e
SOC pixel replay	.07	.09	.15	.26	.39	.35	−.06	.05	.19	.42	.54	n/e
Scaled population replay	.13	.15	.18	.23	.29	.11	.11	.03	−.25	−.53	−.11	n/e
Particle-mean ridge	.25	.27	.29	.36	.56	.60	.26	.32	.36	.43	.54	n/e
U-Net, autoregressive (d, s)	$.45_{\pm.01}$	$.50_{\pm.01}$	$.54_{\pm.02}$	$.58_{\pm.02}$	$.53_{\pm.06}$	$.19_{\pm.29}$	$.41_{\pm.01}$	$.46_{\pm.02}$	$.49_{\pm.05}$	$.52_{\pm.09}$	$.60_{\pm.09}$	n/e
U-Net, autoregressive, no inputs	$.49_{\pm.01}$	$.53_{\pm.02}$	$.55_{\pm.03}$	$.58_{\pm.02}$	$.57_{\pm.04}$	$.47_{\pm.06}$	$.38_{\pm.02}$	$.41_{\pm.02}$	$.39_{\pm.04}$	$.32_{\pm.02}$	$.30_{\pm.12}$	n/e
U-Net, autoregressive (d, u, s)	$.43_{\pm.01}$	$.49_{\pm.01}$	$.55_{\pm.02}$	$.64_{\pm.02}$	$.68_{\pm.03}$	$.60_{\pm.14}$	$.38_{\pm.01}$	$.42_{\pm.00}$	$.44_{\pm.03}$	$.43_{\pm.11}$	$.47_{\pm.31}$	n/e
U-Net, direct lead-time (d, s)	$.32_{\pm.04}$	$.34_{\pm.05}$	$.43_{\pm.05}$	$.55_{\pm.02}$	$.65_{\pm.01}$	$.74_{\pm.01}$	$.26_{\pm.16}$	$.40_{\pm.10}$	$.51_{\pm.08}$	$.63_{\pm.03}$	$.73_{\pm.01}$	n/e
FNO (d, s)	$.37_{\pm.05}$	$.41_{\pm.04}$	$.41_{\pm.02}$	$.15_{\pm.20}$	$-1.25_{\pm1.79}$	$-5.19_{\pm4.96}$	$.38_{\pm.05}$	$.43_{\pm.06}$	$.45_{\pm.06}$	$.51_{\pm.01}$	$.61_{\pm.03}$	n/e
ConvLSTM (d, s)	$.37_{\pm.01}$	$.41_{\pm.01}$	$.46_{\pm.01}$	$.53_{\pm.05}$	$.60_{\pm.05}$	$.72_{\pm.01}$	$.41_{\pm.01}$	$.44_{\pm.00}$	$.49_{\pm.02}$	$.57_{\pm.02}$	$.68_{\pm.01}$	n/e
ViT (d, s)	$.10_{\pm.05}$	$.14_{\pm.08}$	$.22_{\pm.13}$	$.36_{\pm.17}$	$.50_{\pm.17}$	$.62_{\pm.18}$	$.17_{\pm.10}$	$.22_{\pm.12}$	$.27_{\pm.15}$	$.34_{\pm.21}$	$.39_{\pm.31}$	n/e
Grey box AC-RD (d, s)	$.02_{\pm.07}$	$.10_{\pm.05}$	$.21_{\pm.02}$	$.38_{\pm.00}$	$.56_{\pm.01}$	$.72_{\pm.06}$	$-.06_{\pm.04}$	$.01_{\pm.00}$	$.09_{\pm.05}$	$.08_{\pm.15}$	$-.39_{\pm.21}$	n/e
Replay-augmented U-Net (d, s)	$.43_{\pm.01}$	$.48_{\pm.01}$	$.54_{\pm.02}$	$.61_{\pm.04}$	$.63_{\pm.01}$	$.49_{\pm.26}$	$.35_{\pm.04}$	$.31_{\pm.13}$	$.32_{\pm.15}$	$.48_{\pm.06}$	$.55_{\pm.17}$	n/e
Foundation model (val.-selected)	$.40^{\mathrm{i}}$	$.19^{\mathrm{i}}$	$-.61^{\mathrm{i}}$	-1.68^{i}	-1.70^{i}	-3.85^{i}	$.32^{\mathrm{i}}$	$.31^{\mathrm{i}}$	$.24^{\mathrm{i}}$	$.16^{\mathrm{i}}$	$.14^{\mathrm{i}}$	n/e
U-Net − best baseline	$+.10^{\mathrm{A}}$ [.10,.18]	$+.23^{\mathrm{L}}$ [.06,.24]	$+.25^{\mathrm{L}}$ [.08,.26]	$+.22^{\mathrm{L}}$ [.02,.23]	$-.03^{\mathrm{L}}$ [-.17,-.02]	$-.41^{\mathrm{L}}$ [-.73,-.39]	$+.15^{\mathrm{L}}$ [.03,.22]	$+.14^{\mathrm{L}}$ [.02,.21]	$+.13^{\mathrm{L}}$ [.01,.21]	$+.03^{\mathrm{H}}$ [-.03,.14]	$-.01^{\mathrm{H}}$ [-.11,.09]	n/e
Direct U-Net − best baseline	$-.03^{\mathrm{A}}$ [-.04,.05]	$+.07^{\mathrm{L}}$ [-.03,.08]	$+.14^{\mathrm{L}}$ [.08,.15]	$+.19^{\mathrm{L}}$ [.13,.19]	$+.09^{\mathrm{L}}$ [.02,.10]	$+.14^{\mathrm{L}}$ [.10,.15]	$+.00^{\mathrm{L}}$ [-.16,.10]	$+.08^{\mathrm{L}}$ [-.04,.16]	$+.15^{\mathrm{L}}$ [.04,.22]	$+.14^{\mathrm{H}}$ [.05,.28]	$+.13^{\mathrm{H}}$ [.06,.27]	n/e

Table 32: All forecasters, splits and horizons (continued).

	C/2: 25 →35–45°C						C/20: 25 →15,45°C					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	.07	−.10	−.23	−.37	−.58	−.95	−.38	−.48	−.49	−.60	−1.05	−.91
AR(2)	.17	−.04	−.33	−.66	−1.21	−3.38	.11	.09	.09	.10	.15	.39
SOC replay	.14	.18	.26	.40	.54	.28	−.04	−.01	.06	.13	.17	.43
History-aligned replay	.08	.25	.32	.43	.55	.36	.08	.14	.27	.37	.38	.60
Population replay	.11	.14	.19	.31	.54	.33	.05	.09	.17	.24	.25	.56
Damped trend	.26	.21	.13	.06	.02	.00	−.01	.02	.05	.04	.00	.00
History-mean persistence	−.86	−.38	−.17	−.07	−.03	−.01	−.77	−.31	−.15	−.07	−.04	−.02
SOC pixel replay	.04	.11	.22	.39	.56	.29	−.46	−.31	−.18	−.02	.04	.40
Scaled population replay	.15	.11	.00	−.18	−.05	.04	.10	.11	.13	−.08	−.94	−.27
Particle-mean ridge	−9.94	−11.7	−14.0	−13.3	−5.62	−3.56	−.76	−.82	−.88	−.56	−.17	.53
U-Net, autoregressive (d, s)	$.47_{\pm.03}$	$.49_{\pm.03}$	$.47_{\pm.04}$	$.53_{\pm.03}$	$.68_{\pm.02}$	$.79_{\pm.03}$	$.30_{\pm.02}$	$.30_{\pm.04}$	$.37_{\pm.04}$	$.42_{\pm.07}$	$.35_{\pm.07}$	$.64_{\pm.05}$
U-Net, autoregressive, no inputs	$.49_{\pm.01}$	$.49_{\pm.02}$	$.44_{\pm.04}$	$.25_{\pm.15}$	$-.26_{\pm.26}$	$.04_{\pm.26}$	$.33_{\pm.02}$	$.37_{\pm.03}$	$.41_{\pm.05}$	$.30_{\pm.06}$	$-.06_{\pm.09}$	$-.03_{\pm.19}$
U-Net, autoregressive (d, u, s)	$.46_{\pm.03}$	$.47_{\pm.02}$	$.49_{\pm.03}$	$.56_{\pm.04}$	$.68_{\pm.05}$	$.76_{\pm.07}$	$.29_{\pm.03}$	$.31_{\pm.04}$	$.43_{\pm.02}$	$.50_{\pm.05}$	$.44_{\pm.05}$	$.70_{\pm.03}$
U-Net, direct lead-time (d, s)	$.25_{\pm.18}$	$.37_{\pm.10}$	$.46_{\pm.05}$	$.60_{\pm.02}$	$.77_{\pm.02}$	$.87_{\pm.00}$	$-.01_{\pm.21}$	$.19_{\pm.07}$	$.33_{\pm.02}$	$.45_{\pm.01}$	$.49_{\pm.02}$	$.70_{\pm.00}$
FNO (d, s)	$.43_{\pm.04}$	$.46_{\pm.04}$	$.49_{\pm.04}$	$.56_{\pm.05}$	$.68_{\pm.06}$	$.72_{\pm.15}$	$.28_{\pm.01}$	$.27_{\pm.04}$	$.36_{\pm.01}$	$.41_{\pm.03}$	$.37_{\pm.05}$	$.59_{\pm.03}$
ConvLSTM (d, s)	$.44_{\pm.01}$	$.42_{\pm.00}$	$.41_{\pm.03}$	$.57_{\pm.03}$	$.73_{\pm.02}$	$.83_{\pm.00}$	$.18_{\pm.08}$	$.20_{\pm.06}$	$.29_{\pm.03}$	$.38_{\pm.02}$	$.40_{\pm.01}$	$.64_{\pm.00}$
ViT (d, s)	$.07_{\pm.02}$	$.14_{\pm.02}$	$.32_{\pm.00}$	$.53_{\pm.00}$	$.72_{\pm.01}$	$.84_{\pm.01}$	$.04_{\pm.06}$	$.07_{\pm.10}$	$.13_{\pm.18}$	$.19_{\pm.26}$	$.21_{\pm.30}$	$.30_{\pm.43}$
Grey box AC-RD (d, s)	$-.08_{\pm.08}$	$-.06_{\pm.07}$	$.01_{\pm.03}$	$.07_{\pm.05}$	$.14_{\pm.20}$	$.52_{\pm.22}$	$.06_{\pm.02}$	$.12_{\pm.05}$	$.20_{\pm.07}$	$.29_{\pm.09}$	$.35_{\pm.09}$	$.61_{\pm.11}$
Replay-augmented U-Net (d, s)	$.46_{\pm.01}$	$.43_{\pm.02}$	$.41_{\pm.08}$	$.56_{\pm.04}$	$.70_{\pm.02}$	$.77_{\pm.04}$	$.23_{\pm.04}$	$.27_{\pm.00}$	$.36_{\pm.01}$	$.42_{\pm.03}$	$.35_{\pm.05}$	$.61_{\pm.03}$
U-Net − best baseline	$+.22^{\mathrm{D}}$ [.05,.31]	$+.24^{\mathrm{H}}$ [.16,.29]	$+.15^{\mathrm{H}}$ [.00,.27]	$+.11^{\mathrm{H}}$ [-.05,.23]	$+.12^{\mathrm{X}}$ [.09,.23]	$+.42^{\mathrm{H}}$ [.38,.45]	$+.19^{\mathrm{A}}$ [.12,.23]	$+.16^{\mathrm{H}}$ [.08,.22]	$+.11^{\mathrm{H}}$ [.03,.17]	$+.05^{\mathrm{H}}$ [-.02,.14]	$-.03^{\mathrm{H}}$ [-.10,.03]	$+.04^{\mathrm{H}}$ [.00,.12]
Direct U-Net − best baseline	$.00^{\mathrm{D}}$ [-.38,.19]	$+.13^{\mathrm{H}}$ [-.02,.21]	$+.14^{\mathrm{H}}$ [-.02,.24]	$+.17^{\mathrm{H}}$ [.09,.28]	$+.21^{\mathrm{X}}$ [.14,.36]	$+.50^{\mathrm{H}}$ [.45,.57]	$-.12^{\mathrm{A}}$ [-.33,.04]	$+.04^{\mathrm{H}}$ [-.04,.12]	$+.06^{\mathrm{H}}$ [-.01,.12]	$+.09^{\mathrm{H}}$ [.04,.17]	$+.11^{\mathrm{H}}$ [.06,.19]	$+.10^{\mathrm{H}}$ [.06,.18]

	New particle						New particle fold 0					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	−.06	−.21	−.32	−.50	−.92	−1.15	.00	−.16	−.27	−.42	−.74	−1.13
AR(2)	.17	.15	.13	.09	.08	.26	.27	.19	.12	.07	.07	.27
SOC replay	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
History-aligned replay	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Population replay	.12	.16	.23	.31	.42	.61	.13	.18	.25	.36	.52	.69
Damped trend	.18	.15	.11	.05	.01	−.01	.21	.17	.11	.05	.01	−.02
History-mean persistence	−.79	−.34	−.16	−.07	−.03	−.01	−.82	−.36	−.16	−.07	−.03	.00
SOC pixel replay	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Scaled population replay	.16	.17	.13	.01	−.24	.04	.18	.20	.19	.12	.09	.25
Particle-mean ridge	.24	.28	.36	.46	.57	.73	.26	.31	.38	.48	.65	.82
U-Net, autoregressive (d, s)	$.41_{\pm.01}$	$.44_{\pm.01}$	$.47_{\pm.01}$	$.49_{\pm.02}$	$.50_{\pm.02}$	$.70_{\pm.04}$	$.48_{\pm.01}$	$.53_{\pm.01}$	$.61_{\pm.00}$	$.67_{\pm.02}$	$.71_{\pm.06}$	$.82_{\pm.04}$
U-Net, autoregressive, no inputs	$.40_{\pm.01}$	$.42_{\pm.01}$	$.44_{\pm.01}$	$.39_{\pm.02}$	$.10_{\pm.05}$	$.06_{\pm.04}$	$.47_{\pm.00}$	$.51_{\pm.01}$	$.55_{\pm.02}$	$.57_{\pm.04}$	$.51_{\pm.07}$	$.23_{\pm.04}$
U-Net, autoregressive (d, u, s)	$.41_{\pm.00}$	$.44_{\pm.01}$	$.47_{\pm.02}$	$.50_{\pm.00}$	$.55_{\pm.02}$	$.74_{\pm.02}$	$.47_{\pm.01}$	$.53_{\pm.01}$	$.60_{\pm.01}$	$.66_{\pm.01}$	$.69_{\pm.06}$	$.82_{\pm.06}$
U-Net, direct lead-time (d, s)	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$.41_{\pm.01}$	$.45_{\pm.02}$	$.54_{\pm.01}$	$.65_{\pm.01}$	$.75_{\pm.01}$	$.85_{\pm.01}$
FNO (d, s)	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$.50_{\pm.00}$	$.55_{\pm.00}$	$.60_{\pm.01}$	$.65_{\pm.03}$	$.71_{\pm.04}$	$.82_{\pm.03}$
ConvLSTM (d, s)	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$.50_{\pm.00}$	$.54_{\pm.01}$	$.59_{\pm.01}$	$.66_{\pm.01}$	$.75_{\pm.01}$	$.84_{\pm.02}$
ViT (d, s)	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$.25_{\pm.02}$	$.33_{\pm.03}$	$.45_{\pm.03}$	$.59_{\pm.04}$	$.72_{\pm.03}$	$.81_{\pm.02}$
Grey box AC-RD (d, s)	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$\mathrm{n/r}^{1/5}$	$.15_{\pm.01}$	$.21_{\pm.02}$	$.31_{\pm.03}$	$.45_{\pm.05}$	$.61_{\pm.07}$	$.85_{\pm.01}$
Replay-augmented U-Net (d, s)	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$	n/a	n/a	n/a	n/a	n/a	n/a
Foundation model (val.-selected)	n/r	n/r	n/r	n/r	n/r	n/r	$.42_{\pm.01}$	$.44_{\pm.01}$	$.48_{\pm.02}$	$.54_{\pm.03}$	$.56_{\pm.02}$	$.60_{\pm.03}$
U-Net − best baseline	$+.17^{\mathrm{L}}$ [.12,.20]	$+.16^{\mathrm{L}}$ [.13,.20]	$+.12^{\mathrm{L}}$ [.10,.20]	$+.04^{\mathrm{L}}$ [.00,.12]	$-.07^{\mathrm{L}}$ [-.10,.02]	$-.02^{\mathrm{L}}$ [-.04,.03]	$+.21^{\mathrm{A}}$ [.19,.36]	$+.22^{\mathrm{L}}$ [.20,.29]	$+.22^{\mathrm{L}}$ [.16,.30]	$+.19^{\mathrm{L}}$ [.10,.19]	$+.06^{\mathrm{L}}$ [-.01,.07]	$+.01^{\mathrm{L}}$ [-.01,.03]
Direct U-Net − best baseline	n/r	n/r	n/r	n/r	n/r	n/r	$+.14^{\mathrm{A}}$ [.12,.28]	$+.14^{\mathrm{L}}$ [.13,.20]	$+.16^{\mathrm{L}}$ [.14,.27]	$+.17^{\mathrm{L}}$ [.15,.30]	$+.11^{\mathrm{L}}$ [.09,.23]	$+.03^{\mathrm{L}}$ [.02,.12]

Table 33: All forecasters, splits and horizons (continued).

	Later cycles acq. 14						Later cycles 2C					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	.24	.05	$-.16$	$-.36$	$-.89$	-2.54	.10	.01	$-.05$	$-.21$	$-.72$	-1.40
AR(2)	.46	.37	.24	.13	.03	$-.15$	.30	.22	.13	.05	$-.08$	$-.44$
SOC replay	.35	.42	.51	.64	.75	.75	.25	.30	.40	.56	.76	.67
History-aligned replay	.34	.40	.50	.62	.72	.70	.20	.22	.30	.30	.29	.35
Population replay	.28	.34	.44	.57	.69	.75	.16	.19	.26	.43	.67	.59
Damped trend	.36	.29	.18	.08	.02	$-.07$	.26	.16	.08	.03	$-.02$	$-.11$
History-mean persistence	$-.92$	$-.41$	$-.18$	$-.08$	$-.03$	.02	$-.88$	$-.37$	$-.16$	$-.07$	$-.01$	.07
SOC pixel replay	.31	.48	.67	.82	.88	.90	.21	.30	.45	.63	.79	.71
Scaled population replay	.28	.30	.34	.40	.47	.58	.12	.14	.11	.03	.10	.30
Particle-mean ridge	.34	.40	.49	.62	.75	.77	.26	.34	.41	.58	.81	.78
U-Net, autoregressive (d, s)	$.61_{\pm.02}$	$.67_{\pm.03}$	$.74_{\pm.03}$	$.81_{\pm.04}$	$.86_{\pm.04}$	$.86_{\pm.04}$	$.60_{\pm.03}$	$.67_{\pm.04}$	$.74_{\pm.05}$	$.79_{\pm.06}$	$.79_{\pm.10}$	$.79_{\pm.11}$
U-Net, autoregressive, no inputs	$.60_{\pm.01}$	$.63_{\pm.01}$	$.67_{\pm.02}$	$.67_{\pm.04}$	$.51_{\pm.14}$	$.17_{\pm.35}$	$.61_{\pm.00}$	$.70_{\pm.01}$	$.78_{\pm.01}$	$.79_{\pm.02}$	$.70_{\pm.02}$	$.50_{\pm.19}$
U-Net, autoregressive (d, u, s)	$.61_{\pm.03}$	$.67_{\pm.03}$	$.73_{\pm.03}$	$.80_{\pm.03}$	$.86_{\pm.02}$	$.84_{\pm.03}$	$.62_{\pm.01}$	$.69_{\pm.02}$	$.76_{\pm.02}$	$.82_{\pm.04}$	$.83_{\pm.11}$	$.68_{\pm.41}$
U-Net, direct lead-time (d, s)	$.54_{\pm.06}$	$.61_{\pm.06}$	$.74_{\pm.03}$	$.86_{\pm.01}$	$.91_{\pm.01}$	$.93_{\pm.01}$	n/r	n/r	n/r	n/r	n/r	n/r
FNO (d, s)	$.57_{\pm.03}$	$.60_{\pm.03}$	$.65_{\pm.05}$	$.73_{\pm.09}$	$.76_{\pm.16}$	$.65_{\pm.31}$	n/r	n/r	n/r	n/r	n/r	n/r
ConvLSTM (d, s)	$.59_{\pm.04}$	$.64_{\pm.05}$	$.70_{\pm.05}$	$.77_{\pm.04}$	$.83_{\pm.04}$	$.76_{\pm.08}$	n/r	n/r	n/r	n/r	n/r	n/r
ViT (d, s)	$.35_{\pm.04}$	$.43_{\pm.04}$	$.55_{\pm.04}$	$.69_{\pm.03}$	$.78_{\pm.04}$	$.82_{\pm.01}$	n/r	n/r	n/r	n/r	n/r	n/r
Grey box AC-RD (d, s)	$.28_{\pm.01}$	$.37_{\pm.01}$	$.50_{\pm.00}$	$.67_{\pm.00}$	$.79_{\pm.00}$	$.86_{\pm.01}$	n/r	n/r	n/r	n/r	n/r	n/r
Replay-augmented U-Net (d, s)	n/a	n/a	n/a	n/a	n/a	n/a	n/r	n/r	n/r	n/r	n/r	n/r
Foundation model (val.-selected)	$.57_{\pm.01}$	$.63_{\pm.01}$	$.72_{\pm.01}$	$.82_{\pm.01}$	$.87_{\pm.01}$	$.90_{\pm.01}$	n/r	n/r	n/r	n/r	n/r	n/r
U-Net − best baseline	$+.15^{\mathrm{A}}$ [.11,.17]	$+.19^{\mathrm{X}}$ [.04,.25]	$+.06^{\mathrm{X}}$ [-.05,.12]	$-.01^{\mathrm{X}}$ [-.08,.10]	$-.02^{\mathrm{X}}$ [-.08,.16]	$-.03^{\mathrm{X}}$ [-.08,.19]	$+.29^{\mathrm{A}}$ [.27,.55]	$+.33^{\mathrm{L}}$ [.14,.34]	$+.29^{\mathrm{X}}$ [.20,.30]	$+.17^{\mathrm{X}}$ [.04,.17]	$-.02^{\mathrm{L}}$ [-.08,.01]	$+.01^{\mathrm{L}}$ [-.03,.08]
Direct U-Net − best baseline	$+.08^{\mathrm{A}}$ [-.01,.11]	$+.13^{\mathrm{X}}$ [-.02,.19]	$+.06^{\mathrm{X}}$ [-.04,.12]	$+.03^{\mathrm{X}}$ [-.05,.14]	$+.04^{\mathrm{X}}$ [-.03,.22]	$+.03^{\mathrm{X}}$ [-.03,.24]	n/r	n/r	n/r	n/r	n/r	n/r

	C/20 repeat 03→06						C/2 same condition					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	$-.07$	$-.23$	$-.33$	$-.44$	$-.62$	$-.74$	.19	.01	$-.19$	$-.42$	$-.97$	-2.64
AR(2)	.16	.10	.07	.08	.13	.35	.01	$-.73$	-3.83	-30.0	-220.5	-216.6
SOC replay	$-.12$	$-.07$	$-.02$	.07	.23	.46	$-.21$	$-.21$	$-.22$	$-.14$	.22	.32
History-aligned replay	.03	.06	.09	.34	.51	.57	$-.78$	.07	.22	.40	.59	.40
Population replay	.00	.02	.02	.05	.15	.42	.01	.03	.06	.11	.19	.22
Damped trend	.15	.11	.07	.04	.01	.00	.30	.20	.11	.05	.01	$-.03$
History-mean persistence	$-.77$	$-.33$	$-.15$	$-.06$	$-.03$	$-.01$	$-.90$	$-.39$	$-.18$	$-.08$	$-.03$	.02
SOC pixel replay	$-.73$	$-.51$	$-.33$	$-.09$	.16	.43	-1.05	$-.90$	$-.76$	$-.52$	.00	$-.02$
Scaled population replay	.03	.03	$-.07$	$-.24$	$-.61$	.10	.18	.14	$-.03$	$-.32$	$-.60$	$-.41$
Particle-mean ridge	.14	.18	.23	.33	.55	.80	.26	.32	.39	.51	.66	.66
U-Net, autoregressive (d, s)	$.39_{\pm.01}$	$.40_{\pm.03}$	$.41_{\pm.06}$	$.41_{\pm.05}$	$.56_{\pm.05}$	$.77_{\pm.06}$	$.51_{\pm.03}$	$.54_{\pm.04}$	$.58_{\pm.06}$	$.69_{\pm.07}$	$.71_{\pm.19}$	$.58_{\pm.33}$
U-Net, autoregressive, no inputs	$.34_{\pm.08}$	$.28_{\pm.05}$	$.35_{\pm.05}$	$.41_{\pm.04}$	$.22_{\pm.19}$	$.21_{\pm.12}$	$.52_{\pm.03}$	$.52_{\pm.04}$	$.48_{\pm.03}$	$.38_{\pm.10}$	$-.18_{\pm.70}$	$-2.72_{\pm3.41}$
U-Net, autoregressive (d, u, s)	$.35_{\pm.04}$	$.38_{\pm.03}$	$.41_{\pm.04}$	$.45_{\pm.02}$	$.59_{\pm.03}$	$.76_{\pm.04}$	$.48_{\pm.04}$	$.44_{\pm.12}$	$.37_{\pm.26}$	$.60_{\pm.07}$	$.76_{\pm.06}$	$.73_{\pm.07}$
U-Net − best baseline	$+.23^{\mathrm{A}}$ [.14,.27]	$+.22^{\mathrm{L}}$ [.14,.25]	$+.18^{\mathrm{L}}$ [.07,.21]	$+.07^{\mathrm{H}}$ [-.06,.26]	$+.01^{\mathrm{L}}$ [-.09,.05]	$-.03^{\mathrm{L}}$ [-.06,.00]	$+.20^{\mathrm{D}}$ [.15,.23]	$+.22^{\mathrm{L}}$ [.16,.24]	$+.19^{\mathrm{L}}$ [.11,.23]	$+.17^{\mathrm{L}}$ [.10,.20]	$+.04^{\mathrm{L}}$ [.01,.14]	$-.08^{\mathrm{L}}$ [-.14,.14]

Table 34: All forecasters, splits and horizons (continued).

	Formation						C/20→2C					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	−.25	−.52	−.71	−1.02	−1.30	−1.10	.08	−.02	−.11	−.25	−.58	−.98
AR(2)	.20	.16	.14	.16	.25	.37	.28	.17	.07	.01	−.03	−.06
SOC replay	−.33	−.51	−.82	−1.26	−1.48	−.67	.09	.12	.17	.25	.38	.37
History-aligned replay	−.20	−.27	−.32	−.46	−.57	−.05	.10	.10	.12	.27	.43	.43
Population replay	−.23	−.38	−.62	−1.04	−1.21	−.60	.06	.08	.11	.16	.25	.28
Damped trend	.12	.09	.05	.02	.02	.02	.25	.15	.06	.02	−.01	−.05
History-mean persistence	−.75	−.35	−.15	−.07	−.05	−.04	−.87	−.36	−.15	−.06	−.01	.02
SOC pixel replay	−.85	−1.05	−1.38	−1.74	−1.91	−.87	.03	.07	.13	.23	.37	.36
Scaled population replay	.00	.00	−.04	−.17	−.49	.16	.11	.13	.16	.17	.31	.09
Particle-mean ridge	−.02	−.02	−.14	−.57	−1.20	.01	.25	.26	.26	.32	.53	.60
U-Net, autoregressive (d, s)	$.17_{\pm.01}$	$.13_{\pm.02}$	$-.10_{\pm.08}$	$-.67_{\pm.41}$	$-1.41_{\pm1.16}$	$-.56_{\pm.55}$	$.32_{\pm.02}$	$.36_{\pm.02}$	$.41_{\pm.03}$	$.46_{\pm.06}$	$.40_{\pm.13}$	$-.30_{\pm.72}$
U-Net, autoregressive, no inputs	$.21_{\pm.02}$	$.22_{\pm.06}$	$.00_{\pm.24}$	$-.31_{\pm.17}$	$-.86_{\pm.51}$	$-.77_{\pm.77}$	$.35_{\pm.00}$	$.37_{\pm.00}$	$.38_{\pm.00}$	$.42_{\pm.02}$	$.37_{\pm.04}$	$.34_{\pm.11}$
U-Net, autoregressive (d, u, s)	$.20_{\pm.02}$	$.21_{\pm.01}$	$.13_{\pm.05}$	$-.09_{\pm.16}$	$-.44_{\pm.41}$	$-.33_{\pm.37}$	$.24_{\pm.01}$	$.24_{\pm.01}$	$.24_{\pm.02}$	$.25_{\pm.05}$	$.26_{\pm.15}$	$.15_{\pm.49}$
U-Net − best baseline	$-.03^{\mathrm{A}}$ [-.07,.01]	$-.04^{\mathrm{A}}$ [-.10,.01]	$-.24^{\mathrm{A}}$ [-.43,-.12]	$-.83^{\mathrm{A}}$ [-1.36,-.56]	-1.66^{A} [-2.12,-1.40]	$-.93^{\mathrm{A}}$ [-1.19,-.58]	$+.04^{\mathrm{A}}$ [.04,.10]	$+.10^{\mathrm{L}}$ [.04,.11]	$+.14^{\mathrm{L}}$ [.09,.15]	$+.13^{\mathrm{L}}$ [.02,.14]	$-.13^{\mathrm{L}}$ [-.30,-.12]	$-.90^{\mathrm{L}}$ [-1.86,-.83]

	Mixed to cold						C/20→C/2, new particle					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
Trend	−.04	−.21	−.39	−.80	−1.81	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
AR(2)	.12	.07	−.07	−.29	−.68	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
SOC replay	.11	.19	.29	.47	.58	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
History-aligned replay	.03	.26	.38	.54	.62	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
Population replay	.14	.16	.18	.30	.38	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
Damped trend	.19	.18	.10	.06	.03	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
History-mean persistence	−.76	−.35	−.14	−.06	−.04	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
SOC pixel replay	−.03	.12	.26	.50	.62	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
Scaled population replay	.19	.20	.05	.09	.09	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
Particle-mean ridge	.26	.29	.33	.43	.56	n/e	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$	$\mathrm{n/e}^{0/5}$
U-Net, autoregressive (d, s)	$.40_{\pm.02}$	$.46_{\pm.02}$	$.50_{\pm.03}$	$.49_{\pm.11}$	$.58_{\pm.11}$	n/e	$.44_{\pm.01}$	$.44_{\pm.01}$	$.44_{\pm.02}$	$.47_{\pm.03}$	$.51_{\pm.03}$	$.66_{\pm.06}$
U-Net, autoregressive, no inputs	$.38_{\pm.01}$	$.41_{\pm.00}$	$.40_{\pm.01}$	$.36_{\pm.05}$	$.39_{\pm.15}$	n/e	$.42_{\pm.00}$	$.41_{\pm.01}$	$.37_{\pm.02}$	$.38_{\pm.04}$	$.30_{\pm.08}$	$.09_{\pm.18}$
U-Net, autoregressive (d, u, s)	$.40_{\pm.01}$	$.46_{\pm.01}$	$.49_{\pm.02}$	$.44_{\pm.07}$	$.59_{\pm.03}$	n/e	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$	$\mathrm{n/r}^{0/5}$
U-Net − best baseline	$+.14^{\mathrm{L}}$ [.07,.19]	$+.17^{\mathrm{L}}$ [.10,.23]	$+.12^{\mathrm{H}}$ [.06,.19]	$-.05^{\mathrm{H}}$ [-.10,-.02]	$-.04^{\mathrm{X}}$ [-.26,.05]	n/e	n/e	n/e	n/e	n/e	n/e	n/e

Table 35: All forecasters, splits and horizons (continued).

	(C/20+2C) →C/2						C/3→1C,2C					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	.12	−.06	−.22	−.40	−.76	−1.46	.35	.62	.62	.59	.51	.65
AR(2)	.34	.25	.15	.08	.06	.20	.33	.69	.74	.77	.80	.83
SOC replay	−.02	.00	.05	.22	.46	.73	.21	.54	.60	.64	.69	.73
History-aligned replay	.01	.04	.16	.35	.65	.77	.25	.57	.62	.66	.70	.75
Population replay	.16	.20	.28	.40	.55	.72	.25	.56	.61	.65	.67	.68
Damped trend	.28	.19	.10	.04	.01	−.01	.35	.65	.67	.66	.61	.59
History-mean persistence	−.87	−.39	−.17	−.07	−.03	.00	−.59	−.56	−.25	−.12	−.06	−.03
SOC pixel replay	−.33	−.25	−.12	.13	.45	.74	.19	.52	.60	.65	.70	.75
Scaled population replay	.19	.21	.18	.19	.19	.35	.37	.75	.80	.82	.83	.84
Particle-mean ridge	.29	.33	.39	.50	.69	.77	.23	.69	.75	.85	.88	.89
U-Net, autoregressive (d, s)	$.48_{\pm.02}$	$.49_{\pm.04}$	$.51_{\pm.04}$	$.58_{\pm.04}$	$.59_{\pm.07}$	$.66_{\pm.07}$	$.41_{\pm.00}$	$.81_{\pm.01}$	$.84_{\pm.00}$	$.87_{\pm.00}$	$.90_{\pm.00}$	$.91_{\pm.01}$
U-Net, autoregressive, no inputs	n/r	n/r	n/r	n/r	n/r	n/r	$.21_{\pm.29}$	$.40_{\pm.57}$	$.42_{\pm.59}$	$.43_{\pm.60}$	$.43_{\pm.60}$	$.43_{\pm.61}$
U-Net, autoregressive (d, u, s)	n/r	n/r	n/r	n/r	n/r	n/r	$.40_{\pm.02}$	$.80_{\pm.01}$	$.84_{\pm.01}$	$.87_{\pm.00}$	$.90_{\pm.01}$	$.91_{\pm.01}$
U-Net, direct lead-time (d, s)	$.40_{\pm.02}$	$.46_{\pm.01}$	$.55_{\pm.02}$	$.68_{\pm.01}$	$.80_{\pm.01}$	$.86_{\pm.01}$	n/r	n/r	n/r	n/r	n/r	n/r
Foundation model (val.-selected)	n/r	n/r	n/r	n/r	n/r	n/r	$.25^{\mathrm{i}}_{\pm.00}$	$.32^{\mathrm{i}}_{\pm.16}$	$.46^{\mathrm{i}}_{\pm.14}$	$-.55^{\mathrm{i}}_{\pm1.37}$	$-9.30^{\mathrm{i}}_{\pm13.6}$	$-3.98^{\mathrm{i}}_{\pm5.96}$
U-Net − best baseline	$+.14^{\mathrm{A}}$ [.12,.18]	$+.16^{\mathrm{L}}$ [.13,.20]	$+.12^{\mathrm{L}}$ [.09,.16]	$+.08^{\mathrm{L}}$ [-.01,.11]	$-.10^{\mathrm{L}}$ [-.17,-.08]	$-.11^{\mathrm{L}}$ [-.13,-.07]	$+.04^{\mathrm{S}}$ [.02,.05]	$+.06^{\mathrm{S}}$ [.02,.09]	$+.05^{\mathrm{S}}$ [.01,.08]	$+.03^{\mathrm{L}}$ [.00,.04]	$+.02^{\mathrm{L}}$ [-.01,.05]	$+.02^{\mathrm{L}}$ [.01,.03]
Direct U-Net − best baseline	$+.06^{\mathrm{A}}$ [.04,.10]	$+.13^{\mathrm{L}}$ [.08,.14]	$+.16^{\mathrm{L}}$ [.12,.17]	$+.18^{\mathrm{L}}$ [.13,.20]	$+.10^{\mathrm{L}}$ [.08,.13]	$+.09^{\mathrm{L}}$ [.08,.13]	n/r	n/r	n/r	n/r	n/r	n/r

	(C/3+1C) →2C						2C→C/3					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	.22	.40	.39	.28	.10	.42	−.34	.03	.20	.28	.53	.47
AR(2)	.28	.68	.72	.76	.80	.84	.48	.58	.69	.77	.82	.85
SOC replay	.16	.51	.69	.76	.81	.82	−.18	.13	.24	.28	.30	.28
History-aligned replay	.29	.68	.76	.81	.85	.88	−.16	.11	.21	.28	.32	.31
Population replay	.15	.50	.59	.63	.66	.68	−.10	.26	.40	.44	.42	.36
Damped trend	.23	.45	.47	.44	.39	.50	−.30	.09	.32	.49	.68	.61
History-mean persistence	−.50	−.63	−.24	−.12	−.06	−.02	−.35	−.32	−.22	−.13	−.07	−.04
SOC pixel replay	.15	.52	.71	.78	.83	.84	−.23	.10	.22	.28	.30	.29
Scaled population replay	.28	.65	.73	.77	.79	.74	−.27	.14	.35	.40	.42	.37
Particle-mean ridge	.20	.66	.73	.82	.87	.89	−137216.7	−56551.3	−23644.8	−8602.9	−2624.8	−623.0
U-Net, autoregressive (d, s)	$.32_{\pm.00}$	$.80_{\pm.01}$	$.84_{\pm.01}$	$.87_{\pm.00}$	$.89_{\pm.00}$	$.91_{\pm.01}$	$.56_{\pm.11}$	$.54_{\pm.08}$	$.56_{\pm.10}$	$.61_{\pm.11}$	$.71_{\pm.08}$	$.82_{\pm.05}$
U-Net, autoregressive, no inputs	$.32_{\pm.00}$	$.77_{\pm.00}$	$.80_{\pm.01}$	$.81_{\pm.02}$	$.80_{\pm.03}$	$.82_{\pm.01}$	$.68_{\pm.04}$	$.69_{\pm.03}$	$.77_{\pm.02}$	$.78_{\pm.04}$	$.78_{\pm.06}$	$.76_{\pm.08}$
U-Net, autoregressive (d, u, s)	$.34_{\pm.00}$	$.82_{\pm.00}$	$.86_{\pm.00}$	$.88_{\pm.00}$	$.89_{\pm.01}$	$.90_{\pm.02}$	$.36_{\pm.27}$	$.35_{\pm.32}$	$.42_{\pm.26}$	$.54_{\pm.14}$	$.71_{\pm.03}$	$.80_{\pm.02}$
Foundation model (val.-selected)	n/r	n/r	n/r	n/r	n/r	n/r	$.55^{\mathrm{i}}_{\pm.01}$	$.55^{\mathrm{i}}_{\pm.01}$	$.53^{\mathrm{i}}_{\pm.01}$	$.43^{\mathrm{i}}_{\pm.01}$	$.34^{\mathrm{i}}_{\pm.01}$	$.23^{\mathrm{i}}_{\pm.04}$
U-Net − best baseline	$+.03^{\mathrm{H}}$ [.01,.06]	$+.12^{\mathrm{H}}$ [.05,.19]	$+.08^{\mathrm{H}}$ [.01,.15]	$+.04^{\mathrm{L}}$ [-.02,.08]	$+.02^{\mathrm{L}}$ [-.06,.07]	$+.01^{\mathrm{L}}$ [-.01,.03]	$+.08^{\mathrm{A}}$ [.01,.14]	$-.04^{\mathrm{A}}$ [-.10,-.01]	$-.14^{\mathrm{A}}$ [-.21,-.10]	$-.16^{\mathrm{A}}$ [-.26,-.12]	$-.11^{\mathrm{A}}$ [-.20,-.06]	$-.02^{\mathrm{A}}$ [-.06,.03]

Table 36: All forecasters, splits and horizons (continued).

	(C/3+2C) →1C						NMC: new particle					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	.55	.79	.80	.81	.77	.77	.12	.31	.41	.43	.48	.53
AR(2)	.45	.71	.76	.80	.82	.83	.36	.61	.68	.73	.76	.77
SOC replay	.19	.46	.73	.82	.86	.86	n/a	n/a	n/a	n/a	n/a	n/a
History-aligned replay	.39	.68	.81	.86	.88	.88	n/a	n/a	n/a	n/a	n/a	n/a
Population replay	.26	.51	.66	.71	.73	.72	.16	.45	.62	.69	.71	.71
Damped trend	.55	.81	.82	.82	.76	.63	.13	.35	.50	.57	.62	.59
History-mean persistence	−.73	−.51	−.25	−.12	−.06	−.03	−.50	−.44	−.23	−.12	−.06	−.03
SOC pixel replay	.16	.44	.73	.83	.88	.88	n/a	n/a	n/a	n/a	n/a	n/a
Scaled population replay	.46	.54	.78	.83	.86	.87	.24	.46	.66	.70	.72	.69
Particle-mean ridge	.46	.74	.82	.87	.89	.90	.41	.77	.82	.87	.89	.91
U-Net, autoregressive (d, s)	$.56_{\pm.01}$	$.83_{\pm.01}$	$.86_{\pm.01}$	$.88_{\pm.02}$	$.91_{\pm.02}$	$.91_{\pm.03}$	$.48_{\pm.02}$	$.76_{\pm.01}$	$.82_{\pm.01}$	$.86_{\pm.01}$	$.90_{\pm.00}$	$.92_{\pm.01}$
U-Net, autoregressive, no inputs	n/r	n/r	n/r	n/r	n/r	n/r	$.22_{\pm.11}$	$.36_{\pm.14}$	$.38_{\pm.14}$	$.38_{\pm.15}$	$.37_{\pm.16}$	$.39_{\pm.12}$
U-Net, autoregressive (d, u, s)	n/r	n/r	n/r	n/r	n/r	n/r	$.49_{\pm.02}$	$.77_{\pm.01}$	$.83_{\pm.00}$	$.87_{\pm.01}$	$.90_{\pm.01}$	$.93_{\pm.01}$
U-Net − best baseline	$+.01^{\mathrm{D}}$ [-.01,.03]	$+.02^{\mathrm{D}}$ [-.02,.08]	$+.03^{\mathrm{D}}$ [.00,.08]	$+.01^{\mathrm{L}}$ [-.02,.04]	$+.01^{\mathrm{L}}$ [-.02,.06]	$+.02^{\mathrm{L}}$ [.00,.04]	$+.08^{\mathrm{L}}$ [.06,.08]	$.00^{\mathrm{L}}$ [-.02,.01]	$+.01^{\mathrm{L}}$ [-.02,.02]	$-.01^{\mathrm{L}}$ [-.03,.00]	$+.01^{\mathrm{L}}$ [-.01,.03]	$+.01^{\mathrm{L}}$ [-.01,.02]

	NMC811 ageing						NMC811 voltage window					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	−1.02	−2.31	−3.68	−5.31	−6.38	−6.52	−.81	−1.41	−1.64	−2.17	−2.89	−3.66
AR(2)	.28	.41	.50	.67	.77	.82	.23	.30	.32	.41	.52	.64
SOC replay	.01	.17	.41	.70	.85	.91	.01	.11	.21	.30	.39	.44
History-aligned replay	.04	.19	.41	.69	.83	.89	.00	.11	.22	.33	.45	.52
Population replay	.03	.16	.38	.64	.78	.84	.02	.12	.20	.29	.37	.41
Damped trend	−.54	−.61	−.29	−.05	.03	.04	−.39	−.26	.01	.07	.07	.04
History-mean persistence	.35	.24	.07	−.02	−.03	−.02	.17	.02	−.07	−.07	−.04	−.03
SOC pixel replay	−.16	.02	.33	.67	.84	.91	−.13	.02	.17	.29	.39	.44
Scaled population replay	−.08	−.01	.19	.42	.54	.60	−.06	.03	.14	.23	.31	.33
Particle-mean ridge	.13	.28	.47	.73	.85	.91	.06	.08	.05	.14	.36	.67
U-Net, autoregressive (d, s)	$.53_{\pm.00}$	$.61_{\pm.00}$	$.65_{\pm.01}$	$.79_{\pm.02}$	$.87_{\pm.02}$	$.90_{\pm.03}$	$.34_{\pm.02}$	$.31_{\pm.01}$	$.25_{\pm.01}$	$.31_{\pm.03}$	$.42_{\pm.05}$	$.60_{\pm.03}$
U-Net, autoregressive, no inputs	$.44_{\pm.05}$	$.48_{\pm.06}$	$.40_{\pm.11}$	$.33_{\pm.22}$	$.27_{\pm.32}$	$.27_{\pm.47}$	$.28_{\pm.04}$	$.21_{\pm.08}$	$.10_{\pm.09}$	$.03_{\pm.01}$	$-.03_{\pm.09}$	$-.08_{\pm.19}$
U-Net, autoregressive (d, u, s)	$.53_{\pm.00}$	$.61_{\pm.00}$	$.65_{\pm.00}$	$.80_{\pm.01}$	$.87_{\pm.01}$	$.91_{\pm.02}$	$.30_{\pm.01}$	$.21_{\pm.08}$	$.13_{\pm.15}$	$.17_{\pm.19}$	$.30_{\pm.17}$	$.57_{\pm.06}$
U-Net, direct lead-time (d, s)	$.44_{\pm.07}$	$.54_{\pm.06}$	$.61_{\pm.03}$	$.79_{\pm.02}$	$.89_{\pm.01}$	$.93_{\pm.01}$	$-.13_{\pm.18}$	$-.02_{\pm.11}$	$.13_{\pm.02}$	$.32_{\pm.01}$	$.50_{\pm.01}$	$.70_{\pm.02}$
U-Net − best baseline	$+.18^{\mathrm{M}}$ [.15,.21]	$+.20^{\mathrm{A}}$ [.19,.22]	$+.15^{\mathrm{A}}$ [.13,.17]	$+.07^{\mathrm{L}}$ [.06,.09]	$+.02^{\mathrm{L}}$ [.01,.03]	$-.01^{\mathrm{L}}$ [-.02,.00]	$+.11^{\mathrm{A}}$ [.08,.13]	$+.01^{\mathrm{A}}$ [-.04,.04]	$-.07^{\mathrm{A}}$ [-.13,-.02]	$-.10^{\mathrm{A}}$ [-.18,-.04]	$-.11^{\mathrm{A}}$ [-.21,-.04]	$-.07^{\mathrm{L}}$ [-.14,-.03]
Direct U-Net − best baseline	$+.10^{\mathrm{M}}$ [.06,.12]	$+.13^{\mathrm{A}}$ [.12,.14]	$+.11^{\mathrm{A}}$ [.10,.14]	$+.06^{\mathrm{L}}$ [.05,.09]	$+.04^{\mathrm{L}}$ [.03,.06]	$+.02^{\mathrm{L}}$ [.02,.03]	$-.36^{\mathrm{A}}$ [-.53,-.23]	$-.32^{\mathrm{A}}$ [-.48,-.20]	$-.19^{\mathrm{A}}$ [-.32,-.10]	$-.09^{\mathrm{A}}$ [-.19,-.01]	$-.02^{\mathrm{A}}$ [-.11,.05]	$+.03^{\mathrm{L}}$ [.01,.05]

Table 37: All forecasters, splits and horizons (continued).

	NMC811: new particle†						NMC-2: 0.5C →1C,2C,3C					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	−.86	−1.58	−1.94	−2.41	−2.86	−3.14	−.39	−.85	−.82	−.76	−.36	.01
AR(2)	.27	.39	.47	.60	.71	.79	.23	.60	.74	.81	.87	.92
SOC replay	n/a	n/a	n/a	n/a	n/a	n/a	.11	.36	.52	.65	.73	.81
History-aligned replay	n/a	n/a	n/a	n/a	n/a	n/a	.06	.20	.39	.70	.80	.89
Population replay	.06	.22	.39	.54	.64	.71	.13	.41	.57	.65	.70	.78
Damped trend	−.43	−.33	−.03	.07	.07	.05	−.35	−.60	−.32	.05	.35	.32
History-mean persistence	.22	.07	−.05	−.06	−.04	−.03	−.01	−.17	−.12	−.09	−.05	−.03
SOC pixel replay	n/a	n/a	n/a	n/a	n/a	n/a	.03	.29	.50	.65	.74	.82
Scaled population replay	−.01	.14	.30	.54	.62	.67	.12	.48	.65	.69	.69	.62
Particle-mean ridge	.15	.37	.54	.71	.79	.87	.04	.17	.49	.69	.84	.93
U-Net, autoregressive (d, s)	$.51_{\pm.01}$	$.62_{\pm.01}$	$.64_{\pm.01}$	$.71_{\pm.01}$	$.78_{\pm.01}$	$.85_{\pm.01}$	$.26_{\pm.04}$	$.57_{\pm.06}$	$.63_{\pm.17}$	$.66_{\pm.16}$	$.74_{\pm.14}$	$.90_{\pm.03}$
U-Net, autoregressive, no inputs	$.45_{\pm.01}$	$.51_{\pm.02}$	$.46_{\pm.07}$	$.38_{\pm.18}$	$.30_{\pm.30}$	$.29_{\pm.41}$	n/r	n/r	n/r	n/r	n/r	n/r
U-Net, autoregressive (d, u, s)	$.51_{\pm.00}$	$.63_{\pm.00}$	$.65_{\pm.00}$	$.72_{\pm.00}$	$.78_{\pm.00}$	$.85_{\pm.00}$	n/r	n/r	n/r	n/r	n/r	n/r
U-Net, direct lead-time (d, s)	$.40_{\pm.01}$	$.55_{\pm.01}$	$.62_{\pm.00}$	$.73_{\pm.00}$	$.82_{\pm.00}$	$.89_{\pm.01}$	$.04_{\pm.05}$	$.20_{\pm.28}$	$.28_{\pm.40}$	$.34_{\pm.48}$	$.41_{\pm.57}$	$.44_{\pm.62}$
U-Net − best baseline	$+.24^{A}$ [.23,.25]	$+.23^{A}$ [.21,.26]	$+.10^{L}$ [.07,.14]	$+.00^{L}$ [-.05,.03]	$-.01^{L}$ [-.05,.01]	$-.02^{L}$ [-.07,.00]	$+.03^{A}$ [-.01,.07]	$-.03^{A}$ [-.19,.03]	$-.11^{A}$ [-.34,-.01]	$-.16^{A}$ [-.46,-.03]	$-.13^{A}$ [-.42,-.02]	$-.03^{L}$ [-.09,.01]
Direct U-Net − best baseline	$+.13^{A}$ [.11,.14]	$+.15^{A}$ [.14,.17]	$+.08^{L}$ [.05,.11]	$+.02^{L}$ [-.01,.04]	$+.02^{L}$ [-.01,.04]	$+.02^{L}$ [.00,.04]	$-.20^{A}$ [-.25,-.16]	$-.40^{A}$ [-.48,-.28]	$-.45^{A}$ [-.55,-.33]	$-.47^{A}$ [-.56,-.38]	$-.46^{A}$ [-.51,-.40]	$-.49^{L}$ [-.52,-.47]

	NMC-2: 3C→0.5C						NMC-2: (0.5C+3C) →1C,2C					
Δ	.01	.02	.04	.08	.16	.32	.01	.02	.04	.08	.16	.32
Persistence	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Trend	−.52	−.26	−.07	−.05	−.14	−.29	−.51	−.78	−.72	−.65	−.13	.01
AR(2)	.44	.58	.65	.74	.80	.89	.33	.60	.73	.80	.86	.91
SOC replay	.08	.12	.15	.15	.08	.00	.16	.43	.61	.70	.74	.66
History-aligned replay	−.12	.09	.17	.20	.15	.00	.14	.39	.55	.68	.73	.71
Population replay	.07	.12	.14	.14	.07	.00	.12	.46	.65	.74	.81	.76
Damped trend	−.16	.19	.26	.19	.11	.06	−.40	−.38	−.03	.25	.31	.19
History-mean persistence	−.18	−.25	−.20	−.12	−.06	−.04	.00	−.15	−.11	−.09	−.05	−.03
SOC pixel replay	−.09	.06	.13	.14	.08	.00	−.03	.29	.55	.69	.75	.66
Scaled population replay	.09	.16	.18	.15	.07	.00	.08	.49	.72	.79	.81	.63
Particle-mean ridge	−205122.3	−87475.9	−29770.8	−9867.0	−3030.9	−900.2	.23	.55	.75	.83	.87	.89
U-Net, autoregressive (d, s)	$-.77_{\pm.74}$	$-1.03_{\pm1.14}$	$-1.23_{\pm1.41}$	$-1.31_{\pm1.64}$	$-1.09_{\pm1.64}$	$-.82_{\pm1.37}$	$.55_{\pm.01}$	$.75_{\pm.02}$	$.81_{\pm.02}$	$.81_{\pm.02}$	$.85_{\pm.01}$	$.89_{\pm.04}$
U-Net, direct lead-time (d, s)	$-.80_{\pm.09}$	$-.18_{\pm.11}$	$.24_{\pm.03}$	$.48_{\pm.14}$	$.55_{\pm.28}$	$.64_{\pm.25}$	n/r	n/r	n/r	n/r	n/r	n/r
U-Net − best baseline	-1.21^{A} [-1.58,-1.06]	-1.61^{A} [-2.18,-1.25]	-1.88^{A} [-2.45,-1.40]	-2.05^{A} [-2.56,-1.62]	-1.89^{A} [-2.20,-1.68]	-1.71^{A} [-1.98,-1.52]	$+.21^{A}$ [.18,.24]	$+.15^{A}$ [.11,.24]	$+.06^{L}$ [.01,.10]	$-.01^{L}$ [-.10,.03]	$-.02^{L}$ [-.10,.05]	$-.03^{A}$ [-.12,.04]
Direct U-Net − best baseline	-1.24^{A} [-1.48,-.95]	$-.76^{A}$ [-.91,-.53]	$-.41^{A}$ [-.48,-.25]	$-.26^{A}$ [-.32,-.12]	$-.25^{A}$ [-.29,-.15]	$-.25^{A}$ [-.31,-.17]	n/r	n/r	n/r	n/r	n/r	n/r

Table 38: NMC forecasting. 0.02 / 0.16 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –. NMC rate tests use the 24 accepted identities; particle folds retain their prescribed populations, including provisional identities. NMC811 ageing: cycles 3–29 to 99–117 (4.2 V); NMC811 voltage window: cycles 3–29 at 4.2 V to cycles at a 4.4 or 4.6 V upper cut-off (cut-off and cycle number are confounded). The direct lead-time U-Net uses seeds 0–1 on NMC811; it predicts each horizon in one step. NMC811 new particle: † fold 0 only; only folds 0–1 of five are evaluated, so the five-fold mean is withheld. Full fold coverage remains in the CSV.

Forecaster	C/3→1C,2C	(C/3+1C) →2C	2C→C/3
Persistence	0.00 / 0.00 24/2/2688 ; 24/2/2016	0.00 / 0.00 24/1/1152 ; 24/1/816	0.00 / 0.00 24/1/1848 ; 24/1/1488
Trend	0.62 / 0.51 24/2/2688 ; 24/2/2016	0.40 / 0.10 24/1/1152 ; 24/1/816	0.03 / 0.53 24/1/1848 ; 24/1/1488
Damped trend	0.65 / 0.61 24/2/2688 ; 24/2/2016	0.45 / 0.39 24/1/1152 ; 24/1/816	0.09 / 0.68 24/1/1848 ; 24/1/1488
History-mean persistence	-0.56 / -0.06 24/2/2688 ; 24/2/2016	-0.63 / -0.06 24/1/1152 ; 24/1/816	-0.32 / -0.07 24/1/1848 ; 24/1/1488
AR(2)	0.69 / 0.80 24/2/2688 ; 24/2/2016	0.68 / 0.80 24/1/1152 ; 24/1/816	0.58 / 0.82 24/1/1848 ; 24/1/1488
SOC mean replay	0.54 / 0.69 24/2/2688 ; 24/2/2016	0.51 / 0.81 24/1/1152 ; 24/1/816	0.13 / 0.30 24/1/1848 ; 24/1/1488
SOC pixel replay	0.52 / 0.70 24/2/2688 ; 24/2/2016	0.52 / 0.83 24/1/1152 ; 24/1/816	0.10 / 0.30 24/1/1848 ; 24/1/1488
History-aligned SOC replay	0.57 / 0.70 24/2/2688 ; 24/2/2016	0.68 / 0.85 24/1/1152 ; 24/1/816	0.11 / 0.32 24/1/1848 ; 24/1/1488
Scaled population replay	0.75 / 0.83 24/2/2688 ; 24/2/2016	0.65 / 0.79 24/1/1152 ; 24/1/816	0.14 / 0.42 24/1/1848 ; 24/1/1488
Population SOC replay	0.56 / 0.67 24/2/2688 ; 24/2/2016	0.50 / 0.66 24/1/1152 ; 24/1/816	0.26 / 0.42 24/1/1848 ; 24/1/1488
Particle-mean ridge	0.69 / 0.88 24/2/2688 ; 24/2/2016	0.66 / 0.87 24/1/1152 ; 24/1/816	-56551.27 / -2624.79 24/1/1848 ; 24/1/1488
U-Net charge / d, s	0.81 ± 0.01 / 0.90 ± 0.00 24/2/2688 ; 24/2/2016	0.80 ± 0.01 / 0.89 ± 0.00 24/1/1152 ; 24/1/816	0.54 ± 0.08 / 0.71 ± 0.08 24/1/1848 ; 24/1/1488
U-Net charge / none	0.40 ± 0.57 / 0.43 ± 0.60 24/2/2688 ; 24/2/2016	0.77 ± 0.00 / 0.80 ± 0.03 24/1/1152 ; 24/1/816	0.69 ± 0.03 / 0.78 ± 0.06 24/1/1848 ; 24/1/1488
U-Net charge / d, u, s	0.80 ± 0.01 / 0.90 ± 0.01 24/2/2688 ; 24/2/2016	0.82 ± 0.00 / 0.89 ± 0.01 24/1/1152 ; 24/1/816	0.35 ± 0.32 / 0.71 ± 0.03 24/1/1848 ; 24/1/1488
U-Net direct / d, s	– / – 0/0/0 ; 0/0/0	– / – 0/0/0 ; 0/0/0	– / – 0/0/0 ; 0/0/0

Table 39: NMC forecasting. 0.02 / 0.16 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –. NMC rate tests use the 24 accepted identities; particle folds retain their prescribed populations, including provisional identities. NMC811 ageing: cycles 3–29 to 99–117 (4.2 V); NMC811 voltage window: cycles 3–29 at 4.2 V to cycles at a 4.4 or 4.6 V upper cut-off (cut-off and cycle number are confounded). The direct lead-time U-Net uses seeds 0–1 on NMC811; it predicts each horizon in one step. NMC811 new particle: † fold 0 only; only folds 0–1 of five are evaluated, so the five-fold mean is withheld. Full fold coverage remains in the CSV.

Forecaster	NMC: new particle	NMC811 ageing	NMC811 voltage window
Persistence	0.00 / 0.00 5/5 of 5 folds 30/3/5670 ; 30/3/4380	0.00 / 0.00 60/4/18960 ; 60/4/15600	0.00 / 0.00 60/4/26700 ; 60/4/23340
Trend	0.31 / 0.48 5/5 of 5 folds 30/3/5670 ; 30/3/4380	-2.31 / -6.38 60/4/18960 ; 60/4/15600	-1.41 / -2.89 60/4/26700 ; 60/4/23340
Damped trend	0.35 / 0.62 5/5 of 5 folds 30/3/5670 ; 30/3/4380	-0.61 / 0.03 60/4/18960 ; 60/4/15600	-0.26 / 0.07 60/4/26700 ; 60/4/23340
History-mean persistence	-0.44 / -0.06 5/5 of 5 folds 30/3/5670 ; 30/3/4380	0.24 / -0.03 60/4/18960 ; 60/4/15600	0.02 / -0.04 60/4/26700 ; 60/4/23340
AR(2)	0.61 / 0.76 5/5 of 5 folds 30/3/5670 ; 30/3/4380	0.41 / 0.77 60/4/18960 ; 60/4/15600	0.30 / 0.52 60/4/26700 ; 60/4/23340
SOC mean replay	– / – 0/0 of 5 folds 0/0/0 ; 0/0/0	0.17 / 0.85 60/4/18960 ; 60/4/15600	0.11 / 0.39 60/4/26700 ; 60/4/23340
SOC pixel replay	– / – 0/0 of 5 folds 0/0/0 ; 0/0/0	0.02 / 0.84 60/4/18960 ; 60/4/15600	0.02 / 0.39 60/4/26700 ; 60/4/23340
History-aligned SOC replay	– / – 0/0 of 5 folds 0/0/0 ; 0/0/0	0.19 / 0.83 60/4/18960 ; 60/4/15600	0.11 / 0.45 60/4/26700 ; 60/4/23340
Scaled population replay	0.46 / 0.72 5/5 of 5 folds 30/3/5670 ; 30/3/4380	-0.01 / 0.54 60/4/18960 ; 60/4/15600	0.03 / 0.31 60/4/26700 ; 60/4/23340
Population SOC replay	0.45 / 0.71 5/5 of 5 folds 30/3/5670 ; 30/3/4380	0.16 / 0.78 60/4/18960 ; 60/4/15600	0.12 / 0.37 60/4/26700 ; 60/4/23340
Particle-mean ridge	0.77 / 0.89 5/5 of 5 folds 30/3/5670 ; 30/3/4380	0.28 / 0.85 60/4/18960 ; 60/4/15600	0.08 / 0.36 60/4/26700 ; 60/4/23340
U-Net charge / d, s	0.76 ± 0.01 / 0.90 ± 0.00 5/5 of 5 folds 30/3/5670 ; 30/3/4380	0.61 ± 0.00 / 0.87 ± 0.02 60/4/18960 ; 60/4/15600	0.31 ± 0.01 / 0.42 ± 0.05 60/4/26700 ; 60/4/23340
U-Net charge / none	0.36 ± 0.14 / 0.37 ± 0.16 5/5 of 5 folds 30/3/5670 ; 30/3/4380	0.48 ± 0.06 / 0.27 ± 0.32 60/4/18960 ; 60/4/15600	0.21 ± 0.08 / -0.03 ± 0.09 60/4/26700 ; 60/4/23340
U-Net charge / d, u, s	0.77 ± 0.01 / 0.90 ± 0.01 5/5 of 5 folds 30/3/5670 ; 30/3/4380	0.61 ± 0.00 / 0.87 ± 0.01 60/4/18960 ; 60/4/15600	0.21 ± 0.08 / 0.30 ± 0.17 60/4/26700 ; 60/4/23340
U-Net direct / d, s	– / – 0/0 of 5 folds 0/0/0 ; 0/0/0	0.54 ± 0.06 / 0.89 ± 0.01 60/4/18960 ; 60/4/15600	-0.02 ± 0.11 / 0.50 ± 0.01 60/4/26700 ; 60/4/23340

Table 40: NMC forecasting. 0.02 / 0.16 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –. NMC rate tests use the 24 accepted identities; particle folds retain their prescribed populations, including provisional identities. NMC811 ageing: cycles 3–29 to 99–117 (4.2 V); NMC811 voltage window: cycles 3–29 at 4.2 V to cycles at a 4.4 or 4.6 V upper cut-off (cut-off and cycle number are confounded). The direct lead-time U-Net uses seeds 0–1 on NMC811; it predicts each horizon in one step. NMC811 new particle: †fold 0 only; only folds 0–1 of five are evaluated, so the five-fold mean is withheld. Full fold coverage remains in the CSV.

Forecaster	NMC811: new particle†
Persistence	0.00 / 0.00 12/14/15096 ; 12/14/12744
Trend	-1.58 / -2.86 12/14/15096 ; 12/14/12744
Damped trend	-0.33 / 0.07 12/14/15096 ; 12/14/12744
History-mean persistence	0.07 / -0.04 12/14/15096 ; 12/14/12744
AR(2)	0.39 / 0.71 12/14/15096 ; 12/14/12744
SOC mean replay	– / – 0/0/0 ; 0/0/0
SOC pixel replay	– / – 0/0/0 ; 0/0/0
History-aligned SOC replay	– / – 0/0/0 ; 0/0/0
Scaled population replay	0.14 / 0.62 12/14/15096 ; 12/14/12744
Population SOC replay	0.22 / 0.64 12/14/15096 ; 12/14/12744
Particle-mean ridge	0.37 / 0.79 12/14/15096 ; 12/14/12744
U-Net charge / d, s	0.62 ± 0.01 / 0.78 ± 0.01 12/14/15096 ; 12/14/12744
U-Net charge / none	0.51 ± 0.02 / 0.30 ± 0.30 12/14/15096 ; 12/14/12744
U-Net charge / d, u, s	0.63 ± 0.00 / 0.78 ± 0.00 12/14/15096 ; 12/14/12744
U-Net direct / d, s	0.55 ± 0.01 / 0.82 ± 0.00 12/14/15096 ; 12/14/12744

Table 41: Paired gains over the best baseline by test acquisition. U-Net (autoregressive or direct lead-time) minus the split's strongest baseline (the comparator of the main-table difference rows, fixed across acquisitions), on identical windows restricted to one test acquisition; 95% paired particle-cluster bootstrap within the acquisition (2,000 resamples). Rows 'all' reproduce the main-table difference. Seeds: autoregressive five, direct 0–4 on both rate splits (seed-averaged SSE).

Split	Test acquisition	Δ	U-Net	Best baseline	U-Net skill	Baseline skill	Difference [95% CI]	Particles	Windows
C/20→C/2	07 (C/2, 15°C)	0.020	autoregressive	Particle-mean ridge	0.419	0.306	0.113 [0.061, 0.161]	21	585
C/20→C/2	07 (C/2, 15°C)	0.020	direct	Particle-mean ridge	0.337	0.306	0.031 [-0.037, 0.083]	21	585
C/20→C/2	08 (C/2, 15°C)	0.020	autoregressive	Particle-mean ridge	0.421	0.242	0.180 [0.110, 0.277]	21	588
C/20→C/2	08 (C/2, 15°C)	0.020	direct	Particle-mean ridge	0.351	0.242	0.109 [0.033, 0.182]	21	588
C/20→C/2	09 (C/2, 6°C)	0.020	autoregressive	Particle-mean ridge	0.301	0.286	0.015 [-0.077, 0.085]	20	335
C/20→C/2	09 (C/2, 6°C)	0.020	direct	Particle-mean ridge	0.317	0.286	0.031 [-0.023, 0.070]	20	335
C/20→C/2	10 (C/2, 25°C)	0.020	autoregressive	Particle-mean ridge	0.494	0.312	0.181 [0.133, 0.220]	19	639
C/20→C/2	10 (C/2, 25°C)	0.020	direct	Particle-mean ridge	0.434	0.312	0.122 [0.052, 0.153]	19	639
C/20→C/2	11 (C/2, 25°C)	0.020	autoregressive	Particle-mean ridge	0.472	0.324	0.147 [0.093, 0.220]	16	512
C/20→C/2	11 (C/2, 25°C)	0.020	direct	Particle-mean ridge	0.375	0.324	0.051 [0.014, 0.064]	16	512
C/20→C/2	12 (C/2, 35°C)	0.020	autoregressive	Particle-mean ridge	0.558	0.340	0.218 [0.167, 0.260]	16	992
C/20→C/2	12 (C/2, 35°C)	0.020	direct	Particle-mean ridge	0.482	0.340	0.142 [0.104, 0.166]	16	992
C/20→C/2	13 (C/2, 45°C)	0.020	autoregressive	Particle-mean ridge	0.545	0.310	0.235 [0.171, 0.285]	17	1206
C/20→C/2	13 (C/2, 45°C)	0.020	direct	Particle-mean ridge	0.462	0.310	0.152 [0.108, 0.172]	17	1206
C/20→C/2	14 (C/2, 25°C)	0.020	autoregressive	Particle-mean ridge	0.520	0.386	0.134 [0.084, 0.156]	15	2728
C/20→C/2	14 (C/2, 25°C)	0.020	direct	Particle-mean ridge	0.513	0.386	0.126 [0.074, 0.140]	15	2728
C/20→C/2	all	0.020	autoregressive	Particle-mean ridge	0.513	0.344	0.170 [0.135, 0.200]	22	7585
C/20→C/2	all	0.020	direct	Particle-mean ridge	0.464	0.344	0.120 [0.076, 0.134]	22	7585
C/20→C/2	07 (C/2, 15°C)	0.160	autoregressive	Particle-mean ridge	0.534	0.529	0.005 [-0.233, 0.114]	20	300
C/20→C/2	07 (C/2, 15°C)	0.160	direct	Particle-mean ridge	0.618	0.529	0.089 [0.011, 0.127]	20	300
C/20→C/2	08 (C/2, 15°C)	0.160	autoregressive	Particle-mean ridge	0.556	0.542	0.014 [-0.156, 0.084]	21	294
C/20→C/2	08 (C/2, 15°C)	0.160	direct	Particle-mean ridge	0.633	0.542	0.091 [0.003, 0.129]	21	294
C/20→C/2	09 (C/2, 6°C)	0.160	autoregressive	Particle-mean ridge	0.656	0.714	-0.057 [-0.178, 0.028]	20	58
C/20→C/2	09 (C/2, 6°C)	0.160	direct	Particle-mean ridge	0.717	0.714	0.003 [-0.085, 0.057]	20	58
C/20→C/2	10 (C/2, 25°C)	0.160	autoregressive	Particle-mean ridge	0.560	0.661	-0.100 [-0.227, -0.051]	19	406
C/20→C/2	10 (C/2, 25°C)	0.160	direct	Particle-mean ridge	0.796	0.661	0.135 [0.093, 0.151]	19	406
C/20→C/2	11 (C/2, 25°C)	0.160	autoregressive	Particle-mean ridge	0.458	0.546	-0.088 [-0.368, -0.029]	16	288
C/20→C/2	11 (C/2, 25°C)	0.160	direct	Particle-mean ridge	0.538	0.546	-0.008 [-0.017, 0.035]	16	288
C/20→C/2	12 (C/2, 35°C)	0.160	autoregressive	Particle-mean ridge	0.677	0.716	-0.039 [-0.116, 0.004]	16	768
C/20→C/2	12 (C/2, 35°C)	0.160	direct	Particle-mean ridge	0.820	0.716	0.103 [0.003, 0.150]	16	768
C/20→C/2	13 (C/2, 45°C)	0.160	autoregressive	Particle-mean ridge	0.673	0.673	-0.001 [-0.082, 0.050]	17	975
C/20→C/2	13 (C/2, 45°C)	0.160	direct	Particle-mean ridge	0.783	0.673	0.110 [0.040, 0.147]	17	975
C/20→C/2	14 (C/2, 25°C)	0.160	autoregressive	Particle-mean ridge	0.703	0.711	-0.008 [-0.118, 0.024]	15	1916
C/20→C/2	14 (C/2, 25°C)	0.160	direct	Particle-mean ridge	0.843	0.711	0.131 [0.070, 0.150]	15	1916

Table 42: Paired gains over the best baseline by test acquisition. (continued).

Split	Test acquisition	Δ	U-Net	Best baseline	U-Net skill	Baseline skill	Difference [95% CI]	Particles	Windows
C/20→C/2	all	0.160	autoregressive	Particle-mean ridge	0.655	0.681	-0.026 [-0.127, 0.010]	22	5005
C/20→C/2	all	0.160	direct	Particle-mean ridge	0.792	0.681	0.111 [0.046, 0.133]	22	5005
C/2→C/20	03 (C/20, 25°C)	0.020	autoregressive	History-aligned replay	0.371	0.237	0.134 [0.065, 0.204]	21	1470
C/2→C/20	03 (C/20, 25°C)	0.020	direct	History-aligned replay	0.288	0.237	0.051 [-0.009, 0.115]	21	1470
C/2→C/20	04 (C/20, 45°C)	0.020	autoregressive	History-aligned replay	0.230	0.093	0.137 [0.096, 0.204]	22	1633
C/2→C/20	04 (C/20, 45°C)	0.020	direct	History-aligned replay	0.139	0.093	0.046 [0.011, 0.083]	22	1633
C/2→C/20	05 (C/20, 15°C)	0.020	autoregressive	History-aligned replay	0.533	0.266	0.267 [0.169, 0.319]	22	1471
C/2→C/20	05 (C/20, 15°C)	0.020	direct	History-aligned replay	0.469	0.266	0.203 [0.140, 0.248]	22	1471
C/2→C/20	06 (C/20, 25°C)	0.020	autoregressive	History-aligned replay	0.486	0.264	0.222 [0.033, 0.288]	15	1180
C/2→C/20	06 (C/20, 25°C)	0.020	direct	History-aligned replay	0.431	0.264	0.167 [0.056, 0.210]	15	1180
C/2→C/20	all	0.020	autoregressive	History-aligned replay	0.379	0.200	0.180 [0.106, 0.229]	22	5754
C/2→C/20	all	0.020	direct	History-aligned replay	0.303	0.200	0.104 [0.048, 0.141]	22	5754
C/2→C/20	03 (C/20, 25°C)	0.160	autoregressive	History-aligned replay	0.446	0.395	0.051 [-0.038, 0.156]	21	1071
C/2→C/20	03 (C/20, 25°C)	0.160	direct	History-aligned replay	0.548	0.395	0.154 [0.060, 0.270]	21	1071
C/2→C/20	04 (C/20, 45°C)	0.160	autoregressive	History-aligned replay	0.297	0.280	0.018 [-0.041, 0.151]	22	1246
C/2→C/20	04 (C/20, 45°C)	0.160	direct	History-aligned replay	0.388	0.280	0.109 [0.049, 0.250]	22	1246
C/2→C/20	05 (C/20, 15°C)	0.160	autoregressive	History-aligned replay	0.620	0.488	0.132 [0.050, 0.186]	22	1054
C/2→C/20	05 (C/20, 15°C)	0.160	direct	History-aligned replay	0.728	0.488	0.240 [0.185, 0.288]	22	1054
C/2→C/20	06 (C/20, 25°C)	0.160	autoregressive	History-aligned replay	0.781	0.664	0.116 [0.082, 0.214]	15	977
C/2→C/20	06 (C/20, 25°C)	0.160	direct	History-aligned replay	0.835	0.664	0.171 [0.149, 0.233]	15	977
C/2→C/20	all	0.160	autoregressive	History-aligned replay	0.540	0.463	0.077 [0.036, 0.152]	22	4348
C/2→C/20	all	0.160	direct	History-aligned replay	0.626	0.463	0.163 [0.132, 0.244]	22	4348

Table 43: Common-origin sensitivity, C/20→C/2. Every horizon is scored on the same forecast origins (those whose targets exist at every listed horizon), so skill changes with Δ are not changes of support. Best baseline (letter as in the main table) is re-selected on these windows; differences are U-Net (autoregressive, direct) minus it with 95% paired particle-cluster bootstrap intervals. Last column: autoregressive U-Net on all available origins (main table).

Origins	Δ	Origins	Acq.	Best	Best baseline	U-Net AR	U-Net direct	AR − best [95% CI]	Direct − best [95% CI]	AR, all origins
common origins, horizons ¡= 0.16	0.010	5003	8	L	0.320	0.509	0.425	0.189 [0.147, 0.208]	0.105 [0.060, 0.118]	0.493
common origins, horizons ¡= 0.16	0.020	5003	8	L	0.362	0.533	0.482	0.171 [0.137, 0.215]	0.120 [0.094, 0.128]	0.513
common origins, horizons ¡= 0.16	0.040	5003	8	L	0.425	0.549	0.570	0.125 [0.098, 0.183]	0.146 [0.113, 0.166]	0.536
common origins, horizons ¡= 0.16	0.080	5003	8	L	0.532	0.646	0.701	0.114 [0.017, 0.150]	0.169 [0.103, 0.190]	0.623
common origins, horizons ¡= 0.16	0.160	5003	8	L	0.681	0.654	0.792	-0.027 [-0.128, 0.010]	0.110 [0.045, 0.133]	0.655
common origins, horizons ¡= 0.32	0.010	2472	5	A	0.327	0.565	0.462	0.238 [0.193, 0.264]	0.134 [0.092, 0.162]	0.493
common origins, horizons ¡= 0.32	0.020	2472	5	L	0.360	0.602	0.519	0.242 [0.189, 0.261]	0.159 [0.126, 0.168]	0.513
common origins, horizons ¡= 0.32	0.040	2472	5	H	0.416	0.626	0.605	0.210 [0.191, 0.259]	0.190 [0.163, 0.254]	0.536
common origins, horizons ¡= 0.32	0.080	2472	5	H	0.528	0.674	0.718	0.147 [0.087, 0.181]	0.190 [0.159, 0.252]	0.623
common origins, horizons ¡= 0.32	0.160	2472	5	L	0.666	0.633	0.775	-0.033 [-0.138, 0.007]	0.109 [0.024, 0.140]	0.655
common origins, horizons ¡= 0.32	0.320	2472	5	L	0.784	0.697	0.864	-0.087 [-0.106, -0.055]	0.080 [0.065, 0.104]	0.697

Table 44: Acquisition-level evidence for the headline contrasts. Estimates are skill differences, model minus comparator, pooled over identical causal windows with seed-averaged SSE; they reproduce Table 1 and the difference rows of Table 2, whose strongest baseline is the comparator. Pooled: estimate and the paired particle-cluster 95% interval of those tables (test acquisitions held fixed). Sign: test acquisitions whose own estimate has the pooled sign. Per acquisition: smallest and largest estimate within one test acquisition. Leave one out: range of the pooled estimate after deleting each test acquisition in turn (a sensitivity range, not an interval; with two acquisitions it equals the per-acquisition values). Two-level: 95% percentile interval over 2,000 resamples of test acquisitions and then of the particles within each drawn acquisition. Crossed: particles and acquisitions resampled independently, each particle keeping its weight in every acquisition; because the same particles recur in every acquisition, the two-level interval can understate the particle component, and the crossed one is the conservative reading. Both average seeds without resampling them, need at least four test acquisitions, and are coarse with four to eight (four acquisitions allow only 35 distinct acquisition draws). Seeds: training seeds whose own pooled difference has the pooled sign. Repeat Direct U-Net uses seeds 0–4 on both rate splits; autoregressive baseline comparisons also use 0–4. Repeat rows: positive gaps and their range at $\Delta = 0.16$. Every cell rests on more than five particles (15–24 per acquisition), so no interval is withheld.

Test split	Δ	Pooled [95% CI]	Sign	Per acquisition	Leave one out	Two-level 95%	Crossed 95%	Seeds
U-Net (d, s): charge clock minus matched-time clock								
C/20→C/2	0.02	+0.42 [0.38, 0.44]	8/8	[0.22, 0.47]	[0.41, 0.43]	[0.39, 0.46]	[0.38, 0.46]	3/3
C/20→C/2	0.16	+0.51 [0.37, 0.56]	8/8	[0.32, 0.69]	[0.45, 0.53]	[0.39, 0.58]	[0.33, 0.59]	3/3
C/2→C/20	0.02	+6.73 [5.81, 8.39]	4/4	[6.06, 7.47]	[6.49, 7.09]	[5.80, 7.83]	[5.53, 8.58]	3/3
C/2→C/20	0.16	+5.06 [4.54, 6.10]	4/4	[4.20, 6.72]	[4.60, 5.38]	[4.22, 6.29]	[4.15, 6.46]	3/3
U-Net (d, s): charge clock minus time clock								
NMC C/3→1C,2C	0.02	+0.61 [0.54, 0.68]	2/2	[0.43, 0.84]	[0.43, 0.84]	–	–	3/3
NMC C/3→1C,2C	0.16	+0.52 [0.46, 0.55]	2/2	[0.44, 0.64]	[0.44, 0.64]	–	–	3/3
Direct U-Net minus the strongest baseline (C/20→C/2: particle-mean ridge; C/2→C/20: history-aligned replay)								
C/20→C/2	0.02	+0.12 [0.08, 0.13]	8/8	[0.03, 0.15]	[0.11, 0.13]	[0.08, 0.14]	[0.07, 0.14]	5/5
C/20→C/2	0.16	+0.11 [0.05, 0.13]	7/8	[−0.01, 0.13]	[0.10, 0.12]	[0.06, 0.14]	[0.04, 0.14]	5/5
C/2→C/20	0.02	+0.10 [0.05, 0.14]	4/4	[0.05, 0.20]	[0.08, 0.13]	[0.04, 0.19]	[0.03, 0.19]	5/5
C/2→C/20	0.16	+0.16 [0.13, 0.24]	4/4	[0.11, 0.24]	[0.14, 0.18]	[0.11, 0.22]	[0.10, 0.25]	5/5
Autoregressive U-Net minus the strongest baseline (C/20→C/2: particle-mean ridge; C/2→C/20: history-aligned replay)								
C/20→C/2	0.02	+0.17 [0.13, 0.20]	8/8	[0.01, 0.23]	[0.16, 0.19]	[0.13, 0.22]	[0.12, 0.22]	5/5
C/2→C/20	0.02	+0.18 [0.11, 0.23]	4/4	[0.13, 0.27]	[0.16, 0.20]	[0.12, 0.26]	[0.10, 0.26]	5/5
U-Net (d, s): same-condition repeat model minus rate-transfer model, one held-out acquisition each								
C/2	0.16	–	3/3 > 0	[0.04, 0.29]	–	–	–	–
C/20	0.16	–	0/2 > 0	[−0.21, −0.07]	–	–	–	–

Table 45: Controlled clock comparison, all rows (Table 1 shows a subset). Only the clock and the conditioning differ within a split (same protocol, windows and seeds); the last column of each panel interpolates between the training rates. Top line: charge-clock skill (seed mean$_{\pm \mathrm{SD}}$); other lines: charge minus the named clock in skill on identical windows, with a 95% paired particle-cluster interval (positive favours charge). (d, s): direction and charge state; (d, s, r): plus log rate, log step time and log step charge. Time steps: 60 s (graphite), 30 s (NMC); matched time: the training rate's duration of one 0.005 charge step. n/r: not run; n/e: not evaluated.

Graphite	C/20→C/2		C/2→C/20		(C/20+C/2) →2C		C/20→2C		(C/20+2C) →C/2	
Δ	.02	.16	.02	.16	.02	.16	.02	.16	.02	.16
Charge (d, s) skill	$.51_{\pm.02}$	$.67_{\pm.05}$	$.38_{\pm.00}$	$.57_{\pm.01}$	$.49_{\pm.04}$	$.05_{\pm.75}$	$.35_{\pm.01}$	$.45_{\pm.14}$	$.49_{\pm.04}$	$.59_{\pm.07}$
vs time	+.23 [.17,.26]	+.45 [.35,.49]	+6.50 [5.73,7.82]	+2.57 [2.27,3.11]	+.28 [.18,.29]	−.19 [-.53,-.17]	+.31 [.22,.31]	+.42 [.13,.44]	+.22 [.16,.25]	+.38 [.25,.43]
vs matched time	+.42 [.38,.44]	+.51 [.37,.56]	+6.73 [5.81,8.39]	+5.06 [4.54,6.10]	n/r	n/r	+.35 [.27,.36]	+.49 [.23,.50]	n/r	n/r
vs frame	+.04 [.02,.06]	+.25 [.15,.28]	+.61 [.40,1.02]	+.54 [.42,.69]	+.41 [.27,.41]	−.46 [-.73,-.45]	+.09 [.09,.21]	−.02 [-.13,.37]	+.04 [.02,.09]	+.23 [.11,.26]
Charge (d, s, r) skill	$.44_{\pm.02}$	$.55_{\pm.18}$	$.37_{\pm.01}$	$.52_{\pm.08}$	$.47_{\pm.01}$	$.57_{\pm.11}$	$.31_{\pm.00}$	$.39_{\pm.20}$	$.49_{\pm.03}$	$.61_{\pm.17}$
vs time	+.16 [.10,.19]	+.30 [.14,.35]	+5.54 [5.00,6.49]	+3.10 [2.54,3.91]	+.26 [.18,.27]	+.15 [.03,.16]	+.25 [.15,.26]	+.31 [-.26,.34]	+.24 [.16,.27]	+.45 [.28,.52]
vs matched time	+.35 [.32,.38]	+.41 [.22,.47]	+4.99 [4.34,6.26]	+3.93 [3.07,5.44]	n/r	n/r	+.31 [.23,.32]	+.46 [-.06,.49]	n/r	n/r
vs frame	−.05 [-.07,-.01]	+.14 [.02,.19]	+.48 [.30,.89]	+.31 [.22,.49]	+1.07 [.97,2.67]	+.16 [.13,.55]	+.03 [.03,.10]	−.06 [-.14,-.06]	+.06 [.05,.07]	+.29 [.17,.34]

NMC	C/3→1C,2C		(C/3+1C) →2C		2C→C/3		(C/3+2C) →1C	
Δ	.02	.16	.02	.16	.02	.16	.02	.16
Charge (d, s) skill	$.79_{\pm.02}$	$.89_{\pm.01}$	$.81_{\pm.01}$	$.89_{\pm.00}$	$.61_{\pm.11}$	$.73_{\pm.18}$	$.83_{\pm.01}$	$.91_{\pm.02}$
vs time	+.61 [.54,.68]	+.52 [.46,.55]	+.35 [.26,.40]	+.38 [.32,.42]	+19.6 [13.7,22.2]	+33.2 [29.7,43.6]	+.31 [.20,.41]	+.15 [.10,.21]
vs frame	+.25 [.22,.28]	+.28 [.25,.32]	+.20 [.17,.21]	+.18 [.13,.21]	+1.48 [.96,1.74]	+2.18 [1.70,2.65]	+.12 [.09,.16]	+.10 [.07,.16]
Charge (d, s, r) skill	$.81_{\pm.01}$	$.90_{\pm.00}$	$.76_{\pm.06}$	$.85_{\pm.02}$	$.49_{\pm.09}$	$.61_{\pm.12}$	$.83_{\pm.00}$	$.93_{\pm.00}$
vs time	+.66 [.52,.76]	+.51 [.43,.55]	−.01 [-.03,.02]	+.01 [-.02,.04]	+30.1 [25.1,33.0]	+42.1 [36.5,58.1]	+.07 [.03,.11]	+.05 [.03,.07]
vs frame	+.29 [.25,.32]	+.30 [.27,.34]	+.04 [.02,.10]	+.03 [-.01,.12]	+2.03 [1.77,2.37]	+3.06 [2.45,4.13]	+.09 [.08,.13]	+.09 [.07,.14]

Table 46: Controlled clock comparison. 0.02 / 0.16 / 0.32 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –. Clocks use matched training protocols and common seed IDs. Matched-time runs use the declared step controls.

Forecaster	C/20→C/2	C/2→C/20
Charge / d, s	0.51 ± 0.02 / 0.67 ± 0.05 / 0.75 ± 0.10 22/8/7194 ; 22/8/4632 ; 22/4/2243	0.38 ± 0.00 / 0.57 ± 0.01 / 0.73 ± 0.01 22/4/5926 ; 22/4/4519 ; 22/4/3454
Charge / d, s, r	0.44 ± 0.02 / 0.55 ± 0.18 / 0.58 ± 0.42 22/8/7194 ; 22/8/4632 ; 22/4/2243	0.37 ± 0.01 / 0.52 ± 0.08 / 0.67 ± 0.13 22/4/5926 ; 22/4/4519 ; 22/4/3454
Time / d, s	0.28 ± 0.03 / 0.22 ± 0.03 / 0.19 ± 0.05 22/8/7194 ; 22/8/4632 ; 22/4/2243	-6.12 ± 0.30 / -1.99 ± 0.95 / -0.78 ± 0.83 22/4/5926 ; 22/4/4519 ; 22/4/3454
Time / d, s, r	0.28 ± 0.04 / 0.25 ± 0.02 / 0.30 ± 0.18 22/8/7194 ; 22/8/4632 ; 22/4/2243	-5.17 ± 0.74 / -2.58 ± 1.41 / -1.25 ± 1.17 22/4/5926 ; 22/4/4519 ; 22/4/3454
Matched time / d, s	0.09 ± 0.01 / 0.16 ± 0.02 / 0.23 ± 0.04 22/8/7194 ; 22/8/4632 ; 22/4/2243	-6.35 ± 0.22 / -4.48 ± 4.18 / -6.90 ± 11.95 22/4/5926 ; 22/4/4519 ; 22/4/3454
Matched time / d, s, r	0.08 ± 0.02 / 0.14 ± 0.11 / 0.15 ± 0.22 22/8/7194 ; 22/8/4632 ; 22/4/2243	-4.62 ± 1.19 / -3.40 ± 2.37 / -1.02 ± 1.46 22/4/5926 ; 22/4/4519 ; 22/4/3454
Frame / d, s	0.48 ± 0.02 / 0.42 ± 0.06 / 0.40 ± 0.10 22/8/7194 ; 22/8/4632 ; 22/4/2243	-0.23 ± 0.02 / 0.03 ± 0.14 / 0.42 ± 0.15 22/4/5926 ; 22/4/4519 ; 22/4/3454
Frame / d, s, r	0.49 ± 0.01 / 0.40 ± 0.06 / 0.51 ± 0.14 22/8/7194 ; 22/8/4632 ; 22/4/2243	-0.10 ± 0.06 / 0.21 ± 0.13 / 0.54 ± 0.18 22/4/5926 ; 22/4/4519 ; 22/4/3454

Table 47: Controlled clock comparison. 0.02 / 0.16 / 0.32 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –. Clocks use matched training protocols and common seed IDs. Matched-time runs use the declared step controls.

Forecaster	(C/20+C/2) →2C	C/20→2C
Charge / d, s	0.49 ± 0.04 / 0.05 ± 0.75 / -2.14 ± 3.19 3/4/1296 ; 3/4/780 ; 3/4/387	0.35 ± 0.01 / 0.45 ± 0.14 / -1.01 ± 0.63 3/4/660 ; 3/4/351 ; 3/3/135
Charge / d, s, r	0.47 ± 0.01 / 0.57 ± 0.11 / 0.53 ± 0.24 3/4/1296 ; 3/4/780 ; 3/4/387	0.31 ± 0.00 / 0.39 ± 0.20 / -0.83 ± 2.08 3/4/660 ; 3/4/351 ; 3/3/135
Time / d, s	0.21 ± 0.02 / 0.24 ± 0.21 / -0.00 ± 0.83 3/4/1296 ; 3/4/780 ; 3/4/387	0.04 ± 0.01 / 0.03 ± 0.02 / -0.02 ± 0.02 3/4/660 ; 3/4/351 ; 3/3/135
Time / d, s, r	0.21 ± 0.03 / 0.42 ± 0.06 / 0.54 ± 0.16 3/4/1296 ; 3/4/780 ; 3/4/387	0.06 ± 0.01 / 0.09 ± 0.03 / 0.05 ± 0.11 3/4/660 ; 3/4/351 ; 3/3/135
Matched time / d, s	– / – / – 0/0/0 ; 0/0/0 ; 0/0/0	-0.00 ± 0.00 / -0.04 ± 0.01 / -0.23 ± 0.06 3/4/660 ; 3/4/351 ; 3/3/135
Matched time / d, s, r	– / – / – 0/0/0 ; 0/0/0 ; 0/0/0	-0.00 ± 0.00 / -0.06 ± 0.01 / -0.27 ± 0.15 3/4/660 ; 3/4/351 ; 3/3/135
Frame / d, s	0.08 ± 0.11 / 0.51 ± 0.16 / 0.14 ± 0.48 3/4/1296 ; 3/4/780 ; 3/4/387	0.26 ± 0.06 / 0.47 ± 0.18 / -0.24 ± 0.62 3/4/660 ; 3/4/351 ; 3/3/135
Frame / d, s, r	-0.59 ± 0.68 / 0.41 ± 0.07 / 0.41 ± 0.23 3/4/1296 ; 3/4/780 ; 3/4/387	0.28 ± 0.01 / 0.46 ± 0.23 / -2.06 ± 3.64 3/4/660 ; 3/4/351 ; 3/3/135

Conditioning. Table 46 gives the charge-clock skill for every conditioning set, Table 50 the paired contrasts between sets, and Table 52 the (d, u, s) against (d, s) contrast by direction. Conditioning is direction-specific: on C/2 delithiation, (d, s) exceeds (d, u) by 0.53 [0.37, 0.58] and (d, u, s) by 1.01 [0.84, 1.41] at 0.16, with neither contrast resolved on lithiation. The latter comparison removes a harmful feature; it does not by itself establish the benefit of s. For interpolation from C/3 and 2C to 1C, the (d, s, r) bundle (adding log rate, log step time and log step charge) reduces charge's lead over time to 0.07 [0.03, 0.11] and 0.05 [0.03, 0.07] at $\Delta = 0.02$ and 0.16, from 0.31 and 0.15 without the bundle. From C/20 and 2C to C/2, where the second training rate contains three particles, charge still leads by 0.45 [0.28, 0.52] at 0.16.

Forecast example. Figure 2 shows the lithiation window of the central particle P01 at C/2 whose charge-clock skill at $\Delta = 0.16$ is the lower median (rank 78) among the 156 windows scored by both clocks, a rule fixed before rendering and independent of skill with the matched time clock; saved errors are reproduced to 0.1% (script and selection record in the code package).

Forecast progress. Table 53 measures how far each clock's forecasts advance: the slope, through the origin, of the forecast change of the particle mean on the observed change, pooled over windows and seeds with particle-bootstrap intervals (declared before computing; optical progress, not lithium motion). Forecasts with the matched time clock advance at about the training rate's pace, under-

Table 48: Controlled clock comparison. 0.02 / 0.16 / 0.32 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –. Clocks use matched training protocols and common seed IDs. Matched-time runs use the declared step controls.

Forecaster	C/3→1C,2C	(C/3+1C) →2C
Charge / d, s	0.79 ± 0.02 / 0.89 ± 0.01 / 0.90 ± 0.02 24/2/2688 ; 24/2/2016 ; 24/2/1248	0.81 ± 0.01 / 0.89 ± 0.00 / 0.91 ± 0.00 24/1/1152 ; 24/1/816 ; 24/1/432
Charge / d, s, r	0.81 ± 0.01 / 0.90 ± 0.00 / 0.91 ± 0.00 24/2/2688 ; 24/2/2016 ; 24/2/1248	0.76 ± 0.06 / 0.85 ± 0.02 / 0.86 ± 0.03 24/1/1152 ; 24/1/816 ; 24/1/432
Time / d, s	0.18 ± 0.14 / 0.37 ± 0.02 / 0.38 ± 0.01 24/2/2688 ; 24/2/2016 ; 24/2/1248	0.46 ± 0.01 / 0.51 ± 0.02 / 0.56 ± 0.02 24/1/1152 ; 24/1/816 ; 24/1/432
Time / d, s, r	0.15 ± 0.20 / 0.39 ± 0.04 / 0.41 ± 0.04 24/2/2688 ; 24/2/2016 ; 24/2/1248	0.77 ± 0.02 / 0.85 ± 0.01 / 0.87 ± 0.01 24/1/1152 ; 24/1/816 ; 24/1/432
Matched time / d, s	– / – / – 0/0/0 ; 0/0/0 ; 0/0/0	– / – / – 0/0/0 ; 0/0/0 ; 0/0/0
Matched time / d, s, r	– / – / – 0/0/0 ; 0/0/0 ; 0/0/0	– / – / – 0/0/0 ; 0/0/0 ; 0/0/0
Frame / d, s	0.54 ± 0.03 / 0.61 ± 0.03 / 0.61 ± 0.04 24/2/2688 ; 24/2/2016 ; 24/2/1248	0.61 ± 0.02 / 0.71 ± 0.01 / 0.72 ± 0.01 24/1/1152 ; 24/1/816 ; 24/1/432
Frame / d, s, r	0.53 ± 0.02 / 0.60 ± 0.04 / 0.61 ± 0.05 24/2/2688 ; 24/2/2016 ; 24/2/1248	0.72 ± 0.02 / 0.83 ± 0.02 / 0.86 ± 0.02 24/1/1152 ; 24/1/816 ; 24/1/432

Table 49: Controlled clock comparison. 0.02 / 0.16 / 0.32 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –. Clocks use matched training protocols and common seed IDs. Matched-time runs use the declared step controls.

Forecaster	2C→C/3
Charge / d, s	0.61 ± 0.11 / 0.73 ± 0.18 / 0.84 ± 0.10 24/1/1848 ; 24/1/1488 ; 24/1/1104
Charge / d, s, r	0.49 ± 0.09 / 0.61 ± 0.12 / 0.71 ± 0.05 24/1/1848 ; 24/1/1488 ; 24/1/1104
Time / d, s	-18.99 ± 5.82 / -32.51 ± 7.15 / -35.56 ± 9.98 24/1/1848 ; 24/1/1488 ; 24/1/1104
Time / d, s, r	-29.64 ± 4.07 / -41.48 ± 14.04 / -41.65 ± 12.85 24/1/1848 ; 24/1/1488 ; 24/1/1104
Matched time / d, s	– / – / – 0/0/0 ; 0/0/0 ; 0/0/0
Matched time / d, s, r	– / – / – 0/0/0 ; 0/0/0 ; 0/0/0
Frame / d, s	-0.87 ± 0.06 / -1.45 ± 0.26 / -1.52 ± 0.40 24/1/1848 ; 24/1/1488 ; 24/1/1104
Frame / d, s, r	-1.54 ± 0.50 / -2.45 ± 0.73 / -4.37 ± 3.95 24/1/1848 ; 24/1/1488 ; 24/1/1104

advancing from C/20 to C/2 and over-advancing in reverse at $\Delta = 0.02$ and 0.16 (at 0.32 the reverse slope is 0.62).

Identical-history clock control. The clock comparison of Table 1 lets each clock build its history on its own grid. Table 54 removes that difference: direct U-Nets all receive the same last four stored frames, and only the lead differs, given in charge or as the time the constant-current schedule needs to pass it (pre-registered in the released notes; three seeds). It also compares the charge lead with stored-frame history against the same lead with the charge-grid history. History-spacing effects are larger at short horizons and small at long horizons: in reverse at 0.16, the charge-grid gain is $+0.006$ [0.001, 0.016]. On NMC, identical stored-frame histories give a charge-lead advantage of $+0.475$ [0.401, 0.504] at 0.16 (three seeds), extending the graphite lead-encoding result. This identifies a direct-model lead-encoding contrast; it does not isolate numerical integration clocks in the autoregressive comparison.

Earlier forecast origins. The default requested origins are $u = 0.14 + 0.02k$, with history eligibility checked separately. Table 55 evaluates the frozen checkpoints from $u = 0.05 + 0.02k$ (three clock seeds, two direct seeds). This shifted lattice is not a strict extension of the default windows, and causal backward snapping can place actual origins below 0.05. On both rate splits, pooled gains of the charge clock over the time and matched time clocks, and of the direct U-Net over the best baseline, remain resolved at 0.02 and 0.16. The added-origin subset $0.05 \le u_0 < 0.14$ has no

Table 50: Charge-clock conditioning. 0.02 / 0.16 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –.

Forecaster	C/20→C/2	C/2→C/20	(C/20+C/2) →2C
d, s	0.51 ± 0.01 / 0.65 ± 0.05 22/8/7591 ; 22/8/5005	0.37 ± 0.01 / 0.57 ± 0.02 22/4/5926 ; 22/4/4519	0.50 ± 0.01 / 0.53 ± 0.06 3/4/1296 ; 3/4/780
d, u	0.36 ± 0.00 / 0.50 ± 0.04 22/8/7591 ; 22/8/5005	0.39 ± 0.00 / 0.47 ± 0.03 22/4/5926 ; 22/4/4519	– / – 0/0/0 ; 0/0/0
d, u, s	0.37 ± 0.01 / 0.34 ± 0.07 22/8/7591 ; 22/8/5005	0.38 ± 0.01 / 0.50 ± 0.07 22/4/5926 ; 22/4/4519	0.49 ± 0.01 / 0.68 ± 0.03 3/4/1296 ; 3/4/780
None	0.48 ± 0.02 / 0.47 ± 0.09 22/8/7591 ; 22/8/5005	0.38 ± 0.00 / 0.05 ± 0.08 22/4/5926 ; 22/4/4519	0.53 ± 0.02 / 0.57 ± 0.04 3/4/1296 ; 3/4/780
d	0.49 ± 0.04 / 0.39 ± 0.09 22/8/7591 ; 22/8/5005	0.38 ± 0.01 / 0.03 ± 0.20 22/4/5926 ; 22/4/4519	– / – 0/0/0 ; 0/0/0
Fixed-scale operating variables	0.12 ± 0.00 / -1.05 ± 1.57 22/8/7591 ; 22/8/5005	0.35 ± 0.01 / 0.49 ± 0.01 22/4/5926 ; 22/4/4519	0.44 ± 0.03 / 0.61 ± 0.10 3/4/1296 ; 3/4/780
Full state + fixed-scale variables	0.34 ± 0.04 / 0.38 ± 0.03 22/8/7591 ; 22/8/5005	0.33 ± 0.04 / 0.36 ± 0.21 22/4/5926 ; 22/4/4519	0.48 ± 0.04 / 0.58 ± 0.21 3/4/1296 ; 3/4/780
Origin voltage + state	0.38 / 0.54 22/8/7591 ; 22/8/5005	0.38 / 0.54 22/4/5926 ; 22/4/4519	– / – 0/0/0 ; 0/0/0

Table 51: Charge-clock conditioning. 0.02 / 0.16 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –.

Forecaster	C/2: 25 →6–15°C	C/2: 25 →35–45°C	C/20: 25 →15,45°C
d, s	0.46 ± 0.02 / 0.60 ± 0.09 22/3/1509 ; 21/3/652	0.49 ± 0.03 / 0.68 ± 0.02 17/2/2198 ; 17/2/1743	0.28 ± 0.03 / 0.42 ± 0.07 22/2/3234 ; 22/2/2429
d, u	0.32 ± 0.02 / 0.50 ± 0.09 22/3/1509 ; 21/3/652	0.34 ± 0.05 / 0.59 ± 0.03 17/2/2198 ; 17/2/1743	0.31 ± 0.05 / 0.54 ± 0.00 22/2/3234 ; 22/2/2429
d, u, s	0.42 ± 0.00 / 0.47 ± 0.31 22/3/1509 ; 21/3/652	0.47 ± 0.02 / 0.68 ± 0.05 17/2/2198 ; 17/2/1743	0.26 ± 0.07 / 0.50 ± 0.04 22/2/3234 ; 22/2/2429
None	0.41 ± 0.02 / 0.30 ± 0.12 22/3/1509 ; 21/3/652	0.49 ± 0.02 / -0.26 ± 0.26 17/2/2198 ; 17/2/1743	0.34 ± 0.02 / 0.06 ± 0.07 22/2/3234 ; 22/2/2429
d	0.46 ± 0.02 / 0.40 ± 0.09 22/3/1509 ; 21/3/652	0.50 ± 0.00 / 0.24 ± 0.40 17/2/2198 ; 17/2/1743	0.36 ± 0.02 / 0.27 ± 0.07 22/2/3234 ; 22/2/2429
Fixed-scale operating variables	0.42 ± 0.07 / 0.63 ± 0.03 22/3/1509 ; 21/3/652	0.48 ± 0.01 / 0.68 ± 0.07 17/2/2198 ; 17/2/1743	0.32 ± 0.02 / 0.47 ± 0.04 22/2/3234 ; 22/2/2429
Full state + fixed-scale variables	-0.30 ± 1.04 / 0.58 ± 0.13 22/3/1509 ; 21/3/652	0.47 ± 0.03 / 0.70 ± 0.02 17/2/2198 ; 17/2/1743	0.29 ± 0.06 / 0.51 ± 0.07 22/2/3234 ; 22/2/2429
Origin voltage + state	0.38 / 0.60 22/3/1509 ; 21/3/652	0.42 / 0.67 17/2/2198 ; 17/2/1743	0.29 ± 0.05 / 0.47 ± 0.04 22/2/3234 ; 22/2/2429

forward windows for the matched time clock: its inherited history spans about $0.152\,Q_{\text{ref}}$ and was not shortened. This is a separate sensitivity, not a paired comparison with the old lattice. The clock evaluations omitted the secondary horizon 0.005; those cells are unscored, not empty or zero.

Direct prediction with other architectures. Table 56 gives direct lead-time ConvLSTM and FNO against their autoregressive counterparts on seven comparable splits (two seeds). At 0.16, ConvLSTM resolves gains on six and FNO on five; the (C/20, 2C)→C/2 interpolation split has direct runs but no autoregressive counterpart. ConvLSTM gains $+0.063$ [0.044, 0.091] forward and $+0.048$ [0.018, 0.067] in reverse. FNO loses forward (-0.311 [-0.334, -0.217]) and on the C/20 temperature split (-0.176 [-0.238, -0.080]), but gains in reverse ($+0.123$ [0.092, 0.192]). Direct U-Net minus direct ConvLSTM is $+0.009$ [0.005, 0.014] forward and unresolved in reverse (-0.010 [-0.018, 0.009]); these comparisons use five U-Net seeds and two ConvLSTM seeds. At 0.02, direct ConvLSTM and FNO trail their autoregressive versions on both rate splits. Thus the original hypothesis requiring gains in all four rate-split architecture cells is not supported.

Table 52: Conditioning check by direction. C/20 to C/2. All four configurations use common seeds and windows within each direction. Adding u is isolated by d, s versus d, u, s; d, u versus d, s also replaces a feature. The expected direction asymmetry is an exploratory mechanism check, not causal identification.

Direction	Horizon	d, s minus	d, s	Comparator	Difference	95 percent low	95 percent high
all	0.020	U-Net (d, u)	0.516	0.359	0.156	0.107	0.177
all	0.020	U-Net (d, u, s)	0.516	0.376	0.140	0.099	0.154
all	0.020	U-Net none	0.516	0.477	0.038	0.025	0.068
all	0.160	U-Net (d, u)	0.666	0.497	0.168	0.105	0.188
all	0.160	U-Net (d, u, s)	0.666	0.299	0.366	0.306	0.493
all	0.160	U-Net none	0.666	0.522	0.143	0.107	0.266
lithiation	0.020	U-Net (d, u)	0.434	0.450	−0.016	−0.029	0.015
lithiation	0.020	U-Net (d, u, s)	0.434	0.424	0.010	0.004	0.023
lithiation	0.020	U-Net none	0.434	0.393	0.041	0.015	0.064
lithiation	0.160	U-Net (d, u)	0.603	0.630	−0.026	−0.042	0.003
lithiation	0.160	U-Net (d, u, s)	0.603	0.580	0.023	−0.003	0.070
lithiation	0.160	U-Net none	0.603	0.421	0.182	0.136	0.325
delithiation	0.020	U-Net (d, u)	0.629	0.233	0.397	0.319	0.423
delithiation	0.020	U-Net (d, u, s)	0.629	0.308	0.321	0.268	0.342
delithiation	0.020	U-Net none	0.629	0.595	0.034	0.011	0.099
delithiation	0.160	U-Net (d, u)	0.782	0.251	0.531	0.367	0.584
delithiation	0.160	U-Net (d, u, s)	0.782	−0.224	1.006	0.842	1.408
delithiation	0.160	U-Net none	0.782	0.710	0.072	0.030	0.203

Table 53: Forecast progress in the particle-mean optical signal. For each window, $x = \langle I_{\text{target}} - I_{\text{origin}} \rangle_M$ and $y = \langle I_{\text{forecast}} - I_{\text{origin}} \rangle_M$. Here I is the stored log-intensity signal and the forecast is the scored output (including the evaluator's existing $[-10, 10]$ clipping). The slope is $\hat{\beta} = \sum xy / \sum x^2$, with no intercept, pooled over the Table 1 common windows, both cycling directions and U-Net (d, s) seeds 0–2. Brackets give 95% particle-cluster percentile intervals (2,000 draws, seed 23), conditional on the fitted seeds and electrode (22 particles). The median is over individual window–seed ratios y/x with $|x| > 0.02$ in the stored log-intensity units; retained counts accompany the released numerical results. A slope of one matches the signed observed mean change; zero means no projected change. The matched-time step-duration ratio is $38.7/392.5 = 0.0986$ (C/2 over C/20), a naive constant-speed reference for forward transfer, not a mechanistic law.

		C/20→C/2		C/2→C/20	
Δ	Clock	Slope [95% CI]	Median	Slope [95% CI]	Median
0.02	Charge	0.639 [0.562, 0.722]	0.797	0.502 [0.444, 0.563]	0.736
0.02	Time (60 s)	0.296 [0.270, 0.324]	0.153	1.659 [1.501, 1.860]	3.857
0.02	Matched time	0.091 [0.074, 0.109]	0.114	1.697 [1.529, 1.911]	3.988
0.16	Charge	0.648 [0.588, 0.698]	0.839	0.772 [0.743, 0.801]	0.902
0.16	Time (60 s)	0.166 [0.141, 0.189]	0.070	1.644 [1.570, 1.730]	1.729
0.16	Matched time	0.140 [0.119, 0.161]	0.069	1.571 [1.490, 1.666]	1.743
0.32	Charge	0.838 [0.787, 0.886]	0.911	0.826 [0.798, 0.855]	0.906
0.32	Time (60 s)	0.169 [0.149, 0.189]	0.093	1.669 [1.582, 1.763]	1.320
0.32	Matched time	0.264 [0.214, 0.313]	0.110	0.616 [0.518, 0.713]	0.934

Table 54: Identical-history clock control. Direct U-Nets whose history is the same four stored frames up to the origin whatever the rate; one arm receives the lead in charge (Δs), the other in time (the time the half-cycle's constant-current schedule needs to pass Δs). Both arms draw identical training samples and validation origins and require the same charge-grid reference history. Seeds 0–2 in every stored-history arm; the descriptive graphite charge-grid comparison retains the registered common seeds 0–2. The NMC extension uses its accepted particle set and $Q_{\mathrm{ref}} = 560.93$ microampere-hours. skill: seed mean ± SD; differences with 95% paired particle-cluster bootstrap intervals (2,000 resamples). Pre-registered hypothesis: the charge lead beats the time lead at $\Delta = 0.16$ on both rate splits. The history comparison (charge grid − stored frames, both with a charge lead) is descriptive. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Split	Forecaster or contrast	$\Delta = 0.02$	$\Delta = 0.16$	$\Delta = 0.32$
C/20→C/2	Stored frames, lead in charge	0.320 ± 0.043	0.807 ± 0.003	0.872 ± 0.013
	Stored frames, lead in time	0.164 ± 0.045	0.200 ± 0.080	0.376 ± 0.156
	Charge-grid history, lead in charge (seeds 0, 1, 2)	0.472 ± 0.011	0.788 ± 0.021	0.864 ± 0.004
	Lead in charge − lead in time	+0.156 [0.123, 0.168]	+0.606 [0.501, 0.636]	+0.496 [0.448, 0.518]
	Charge-grid − stored-frame history	+0.151 [0.117, 0.163]	-0.019 [-0.037, 0.001]	-0.009 [-0.037, 0.004]
C/2→C/20	Stored frames, lead in charge	0.195 ± 0.046	0.655 ± 0.007	0.811 ± 0.005
	Stored frames, lead in time	-1.939 ± 0.075	0.016 ± 0.184	0.498 ± 0.090
	Charge-grid history, lead in charge (seeds 0, 1, 2)	0.281 ± 0.008	0.661 ± 0.003	0.811 ± 0.006
	Lead in charge − lead in time	+2.134 [1.659, 2.789]	+0.639 [0.511, 0.903]	+0.313 [0.247, 0.449]
	Charge-grid − stored-frame history	+0.086 [0.064, 0.124]	+0.006 [0.001, 0.016]	+0.000 [-0.002, 0.005]
NMC: C/3→1C,2C	Stored frames, lead in charge	0.784 ± 0.011	0.910 ± 0.006	0.932 ± 0.004
	Stored frames, lead in time	0.520 ± 0.019	0.435 ± 0.013	0.438 ± 0.013
	Lead in charge − lead in time	+0.264 [0.220, 0.293]	+0.475 [0.401, 0.504]	+0.493 [0.431, 0.515]
Pre-registered verdict: supported (charge lead better at $\Delta = 0.16$ on both rate splits).				

Table 55: Earlier-origin sensitivity. Existing checkpoints evaluated from requested charge origin 0.05 at spacing 0.02, with causal backward snapping. Subsets use actual origin u_0; all includes origins snapped below 0.05. This lattice is shifted relative to the original 0.14 lattice, not a strict extension. Mean paired skill differences across clock seeds 0–2 and direct seeds 0–1; 95% particle-cluster percentile intervals (2,000 shared resamples, bootstrap seed 20260924). Parentheses: particles/windows on exact common physical origins and targets. Best baseline: persistence, history-aligned replay, AR(2) or particle-mean ridge on their common support, reselected within each resample. Empty support is kept explicit; n/e means the replacement direct evaluations have not arrived. The original clock commands omitted horizon 0.005; it is marked unscored in the released results. Per-seed results, all seven horizon statuses, snap distances and dropped counts are in the released origin sensitivity results. These are conditional descriptive sensitivities.

Split	Origins	Contrast	$\Delta = 0.02$	$\Delta = 0.16$
C/20→C/2	$0.05 \le u_0 < 0.14$	Charge − time	+0.317 [0.262, 0.344] (22/1474)	+0.551 [0.477, 0.644] (22/1463)
C/20→C/2	$0.05 \le u_0 < 0.14$	Charge − matched time	empty support	empty support
C/20→C/2	$0.05 \le u_0 < 0.14$	Direct − best baseline	+0.209 [0.186, 0.254] (22/1471)	+0.171 [0.154, 0.197] (22/1460)
C/20→C/2	All returned	Charge − time	+0.244 [0.194, 0.269] (22/9256)	+0.475 [0.391, 0.498] (22/6674)
C/20→C/2	All returned	Charge − matched time	+0.414 [0.378, 0.435] (22/7023)	+0.505 [0.366, 0.553] (22/4476)
C/20→C/2	All returned	Direct − best baseline	+0.239 [0.219, 0.275] (22/9247)	+0.163 [0.122, 0.181] (22/6669)
C/20→C/2	$u_0 \ge 0.14$	Charge − time	+0.228 [0.174, 0.255] (22/7414)	+0.455 [0.357, 0.490] (22/4846)
C/20→C/2	$u_0 \ge 0.14$	Charge − matched time	+0.414 [0.378, 0.435] (22/7023)	+0.505 [0.366, 0.553] (22/4476)
C/20→C/2	$u_0 \ge 0.14$	Direct − best baseline	+0.130 [0.093, 0.144] (22/7410)	+0.119 [0.050, 0.143] (22/4846)
C/2→C/20	$0.05 \le u_0 < 0.14$	Charge − time	+11.272 [8.457, 15.900] (22/632)	+24.100 [18.967, 29.592] (22/632)
C/2→C/20	$0.05 \le u_0 < 0.14$	Charge − matched time	+14.702 [11.392, 20.271] (22/632)	+34.546 [22.503, 45.218] (22/632)
C/2→C/20	$0.05 \le u_0 < 0.14$	Direct − best baseline	+0.033 [0.013, 0.050] (22/632)	+0.238 [0.174, 0.347] (22/632)
C/2→C/20	All returned	Charge − time	+6.602 [5.768, 7.953] (22/6579)	+3.671 [3.136, 4.528] (22/5173)
C/2→C/20	All returned	Charge − matched time	+6.943 [5.962, 8.758] (22/6579)	+6.776 [5.916, 8.100] (22/5173)
C/2→C/20	All returned	Direct − best baseline	+0.099 [0.048, 0.128] (22/6429)	+0.169 [0.139, 0.247] (22/5024)
C/2→C/20	$u_0 \ge 0.14$	Charge − time	+6.433 [5.622, 7.792] (22/5789)	+2.493 [2.194, 2.981] (22/4383)
C/2→C/20	$u_0 \ge 0.14$	Charge − matched time	+6.634 [5.667, 8.405] (22/5789)	+5.098 [4.601, 6.017] (22/4383)
C/2→C/20	$u_0 \ge 0.14$	Direct − best baseline	+0.101 [0.048, 0.129] (22/5639)	+0.165 [0.136, 0.238] (22/4234)

Table 56: Direct prediction with other architectures. Direct lead-time ConvLSTM and FNO (the lead as one extra conditioning input, default widths, 6,000 steps) against their autoregressive counterparts (same history and inputs, four-step unrolled training), on identical windows and common seeds 0 and 1; direct U-Net with seeds 0–4 on all eight splits. Best baseline: the strongest declared baseline of the main table on its windows (letters as there), with the direct model restricted to those windows. Differences with 95% paired particle-cluster bootstrap intervals (2,000 resamples); on fold 0 (five test particles) the number of particles favouring the direct model is given in parentheses. Pre-registered rule: a cell (architecture, split) supports direct prediction only if its interval at $\Delta = 0.16$ lies above zero; the general statement needs all four rate-split cells. The recipes differ in validation: the direct runs select checkpoints on the common lead-time validation, the autoregressive ones on rollout validation. Kept its initial checkpoint because validation never improved on it (early stop; scores reported): direct FNO, C/20→C/2, seed 0; direct FNO, C/20: 25→15,45°C, seed 1; direct FNO, (C/20+2C)→C/2, seed 0. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Split	Forecaster or contrast	$\Delta = 0.02$	$\Delta = 0.16$
C/20→C/2	Direct ConvLSTM	0.323 ± 0.017	0.782 ± 0.007
	Autoregressive ConvLSTM	0.473 ± 0.006	0.719 ± 0.030
	Direct − autoregressive ConvLSTM	-0.150 [-0.194, -0.108]	+0.063 [0.044, 0.091]
	Direct ConvLSTM − best baseline	-0.021 [-0.061, -0.009]	+0.101 [0.038, 0.123]
	Direct FNO	0.169 ± 0.238	0.394 ± 0.557
	Autoregressive FNO	0.535 ± 0.004	0.705 ± 0.015
	Direct − autoregressive FNO	-0.366 [-0.394, -0.344]	-0.311 [-0.334, -0.217]
	Direct FNO − best baseline	-0.175 [-0.215, -0.158]	-0.287 [-0.311, -0.258]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.463 ± 0.014	0.792 ± 0.018
	Direct U-Net − direct ConvLSTM	+0.141 [0.120, 0.150]	+0.009 [0.005, 0.014]
	Direct U-Net − direct FNO	+0.295 [0.258, 0.313]	+0.398 [0.339, 0.414]
C/2→C/20	Direct ConvLSTM	0.215 ± 0.014	0.661 ± 0.003
	Autoregressive ConvLSTM	0.374 ± 0.000	0.613 ± 0.000
	Direct − autoregressive ConvLSTM	-0.158 [-0.279, -0.106]	+0.048 [0.018, 0.067]
	Direct ConvLSTM − best baseline	+0.051 [-0.036, 0.110]	+0.173 [0.146, 0.237]
	Direct FNO	0.190 ± 0.001	0.643 ± 0.001
	Autoregressive FNO	0.347 ± 0.013	0.519 ± 0.013
	Direct − autoregressive FNO	-0.157 [-0.240, -0.113]	+0.123 [0.092, 0.192]
	Direct FNO − best baseline	+0.027 [-0.058, 0.086]	+0.155 [0.123, 0.239]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.284 ± 0.007	0.651 ± 0.025
	Direct U-Net − direct ConvLSTM	+0.069 [0.033, 0.123]	-0.010 [-0.018, 0.009]
	Direct U-Net − direct FNO	+0.094 [0.065, 0.138]	+0.008 [-0.012, 0.019]

Table 57: Direct prediction with other architectures. (continued)

Split	Forecaster or contrast	$\Delta = 0.02$	$\Delta = 0.16$
New particle, fold 0	Direct ConvLSTM	0.334 ± 0.024	0.757 ± 0.012
	Autoregressive ConvLSTM	0.530 ± 0.005	0.755 ± 0.006
	Direct − autoregressive ConvLSTM	-0.196 [-0.253, -0.175] (0/5)	+0.002 [-0.019, 0.188] (3/5)
	Direct ConvLSTM − best baseline	+0.024 [0.018, 0.080]	+0.108 [0.037, 0.199]
	Direct FNO	0.317 ± 0.021	0.755 ± 0.013
	Autoregressive FNO	0.539 ± 0.002	0.716 ± 0.039
	Direct − autoregressive FNO	-0.221 [-0.289, -0.161] (0/5)	+0.039 [0.021, 0.197] (5/5)
	Direct FNO − best baseline	+0.006 [-0.007, 0.091]	+0.107 [0.074, 0.233]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.444 ± 0.017	0.753 ± 0.012
	Direct U-Net − direct ConvLSTM	+0.110 [0.103, 0.153]	-0.004 [-0.009, 0.056]
	Direct U-Net − direct FNO	+0.126 [0.107, 0.177]	-0.002 [-0.005, 0.019]
(C/20+C/2)→2C	Direct ConvLSTM	0.212 ± 0.044	0.675 ± 0.013
	Autoregressive ConvLSTM	0.405 ± 0.006	0.603 ± 0.050
	Direct − autoregressive ConvLSTM	-0.193 [-0.207, -0.173]	+0.072 [0.030, 0.074]
	Direct ConvLSTM − best baseline	-0.059 [-0.177, -0.051]	+0.116 [0.041, 0.121]
	Direct FNO	0.303 ± 0.012	0.644 ± 0.002
	Autoregressive FNO	0.415 ± 0.037	-1.250 ± 1.788
	Direct − autoregressive FNO	-0.111 [-0.122, 0.059]	+1.895 [1.855, 3.017]
	Direct FNO − best baseline	+0.032 [-0.091, 0.041]	+0.085 [0.004, 0.090]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.345 ± 0.049	0.650 ± 0.013
	Direct U-Net − direct ConvLSTM	+0.133 [0.129, 0.150]	-0.025 [-0.027, -0.024]
	Direct U-Net − direct FNO	+0.042 [0.040, 0.064]	+0.006 [-0.019, 0.014]

Table 58: Direct prediction with other architectures. (continued)

Split	Forecaster or contrast	$\Delta = 0.02$	$\Delta = 0.16$
C/2: 25→6–15°C	Direct ConvLSTM	0.399 ± 0.068	0.709 ± 0.008
	Autoregressive ConvLSTM	0.444 ± 0.004	0.679 ± 0.010
	Direct − autoregressive ConvLSTM	-0.045 [-0.136, 0.021]	+0.030 [0.009, 0.073]
	Direct ConvLSTM − best baseline	+0.086 [-0.078, 0.181]	+0.106 [0.042, 0.243]
	Direct FNO	0.261 ± 0.109	0.648 ± 0.063
	Autoregressive FNO	0.425 ± 0.063	0.611 ± 0.030
	Direct − autoregressive FNO	-0.164 [-0.273, -0.070]	+0.036 [0.002, 0.113]
	Direct FNO − best baseline	-0.054 [-0.242, 0.051]	+0.045 [-0.044, 0.229]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.397 ± 0.099	0.733 ± 0.011
	Direct U-Net − direct ConvLSTM	-0.002 [-0.052, 0.057]	+0.024 [0.005, 0.055]
	Direct U-Net − direct FNO	+0.136 [0.065, 0.213]	+0.085 [0.021, 0.115]
C/2: 25→35–45°C	Direct ConvLSTM	0.342 ± 0.044	0.749 ± 0.006
	Autoregressive ConvLSTM	0.420 ± 0.002	0.727 ± 0.021
	Direct − autoregressive ConvLSTM	-0.078 [-0.194, -0.021]	+0.022 [0.011, 0.034]
	Direct ConvLSTM − best baseline	+0.094 [-0.224, 0.252]	+0.190 [0.144, 0.324]
	Direct FNO	0.258 ± 0.007	0.760 ± 0.004
	Autoregressive FNO	0.456 ± 0.037	0.679 ± 0.058
	Direct − autoregressive FNO	-0.198 [-0.388, -0.108]	+0.081 [0.030, 0.120]
	Direct FNO − best baseline	+0.009 [-0.137, 0.102]	+0.201 [0.120, 0.382]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.375 ± 0.095	0.766 ± 0.016
	Direct U-Net − direct ConvLSTM	+0.033 [-0.049, 0.207]	+0.017 [-0.017, 0.081]
	Direct U-Net − direct FNO	+0.117 [0.082, 0.128]	+0.006 [-0.023, 0.020]

Table 59: Direct prediction with other architectures. (continued)

Split	Forecaster or contrast	$\Delta = 0.02$	$\Delta = 0.16$
C/20: 25→15,45°C	Direct ConvLSTM	0.138 ± 0.048	0.484 ± 0.084
	Autoregressive ConvLSTM	0.178 ± 0.031	0.459 ± 0.003
	Direct − autoregressive ConvLSTM	-0.040 [-0.056, -0.011]	+0.024 [0.008, 0.056]
	Direct ConvLSTM − best baseline	+0.017 [-0.060, 0.115]	+0.056 [-0.005, 0.145]
	Direct FNO	0.014 ± 0.020	0.263 ± 0.372
	Autoregressive FNO	0.198 ± 0.104	0.439 ± 0.032
	Direct − autoregressive FNO	-0.183 [-0.213, -0.119]	-0.176 [-0.238, -0.080]
	Direct FNO − best baseline	-0.088 [-0.156, 0.003]	-0.151 [-0.210, -0.064]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.167 ± 0.073	0.547 ± 0.015
	Direct U-Net − direct ConvLSTM	+0.029 [-0.019, 0.060]	+0.064 [0.045, 0.092]
	Direct U-Net − direct FNO	+0.153 [0.086, 0.196]	+0.284 [0.236, 0.317]
(C/20+2C)→C/2	Direct ConvLSTM	0.330 ± 0.064	0.763 ± 0.048
	Autoregressive ConvLSTM	n/r	n/r
	Direct − autoregressive ConvLSTM	n/r	n/r
	Direct ConvLSTM − best baseline	+0.000 [-0.094, 0.029]	+0.070 [0.046, 0.123]
	Direct FNO	0.169 ± 0.239	0.394 ± 0.557
	Autoregressive FNO	n/r	n/r
	Direct − autoregressive FNO	n/r	n/r
	Direct FNO − best baseline	-0.161 [-0.199, -0.145]	-0.300 [-0.318, -0.239]
	Direct U-Net, seeds 0, 1, 2, 3, 4	0.462 ± 0.014	0.798 ± 0.011
	Direct U-Net − direct ConvLSTM	+0.132 [0.113, 0.182]	+0.035 [0.005, 0.044]
	Direct U-Net − direct FNO	+0.293 [0.263, 0.310]	+0.404 [0.345, 0.419]

Pre-registered verdict: cells as they fall (not all four rate-split cells supported). C/20→C/2 ConvLSTM: supported; C/20→C/2 FNO: autoregressive better; C/2→C/20 ConvLSTM: supported; C/2→C/20 FNO: supported; New particle, fold 0 ConvLSTM: not resolved; New particle, fold 0 FNO: supported; (C/20+C/2)→2C ConvLSTM: supported; (C/20+C/2)→2C FNO: supported; C/2: 25→6–15°C ConvLSTM: supported; C/2: 25→6–15°C FNO: supported; C/2: 25→35–45°C ConvLSTM: supported; C/2: 25→35–45°C FNO: supported; C/20: 25→15,45°C ConvLSTM: supported; C/20: 25→15,45°C FNO: autoregressive better; (C/20+2C)→C/2 ConvLSTM: autoregressive arm not planned; (C/20+2C)→C/2 FNO: autoregressive arm not planned.

Architectures, new particles, later cycles and formation. The one-sentence summaries at the end of Sec. 6.2 rest on the following numbers, on the same causal windows and with the same intervals as Table 2. Among autoregressive models on the six evaluated rate and temperature splits, ConvLSTM has the highest long-horizon seed means (0.72 from C/20 to C/2); this ranking does not extend to later cycles. The FNO matches the U-Net except on the 2C extrapolation, where it is unstable, and the patch transformer is the weakest; the grey-box reaction model, an optical reaction–diffusion comparator rather than a calibrated battery model, is competitive only at long horizons on two splits. On new particles, where same-particle replay is unavailable, the U-Net reaches 0.44 and 0.50 against 0.16 and 0.42 for population replay. Neighbouring training crops can contain pixels of a held-out particle (Appendix B), so all five folds were retrained with every pixel within two of a held-out particle's mask removed from training inputs and targets: pooled new-particle skill changes by $+0.006$ [-0.026, 0.017] for graphite and $+0.002$ [-0.009, 0.005] for NMC at $\Delta s = 0.16$, although one graphite fold drops by 0.086 [0.069, 0.114] (Table 10). Later cycles give high skill to the U-Net (0.86) and pixel replay (0.88 at 0.16). On the formation cycle, without an offset the learned models and charge-aligned replay are worse than persistence at long horizons (U-Net -1.41, s-aligned mean replay -1.35 at 0.16). The formation cycle does not start from the charge state that s assumes, whereas AR(2), which extrapolates only the recent history, reaches 0.25; aligning replay on an offset estimated from the quasi-equilibrium voltage lifts it to $+0.42$, which implicates the starting charge state, although it does not rule out changed dynamics (Tables 61 and 62). Applying the same offset only to the U-Net's charge-state input gives Table 60. On the paper's common windows at $\Delta = 0.16$, it removes 62.2% of the original squared error; the corrected delithiation skill remains negative. This intervention identifies sensitivity to the input coordinate, not unchanged formation dynamics.

Table 60: Formation U-Net rescore on identical paper windows (acquisition 02, 19 particles, seeds 0–2). The offset $+0.282711$ changes only the charge-state input; weights, histories, targets and masks are fixed. Skills average seed SSE. Changes have paired particle-cluster 95% intervals (2,000 resamples), conditional on this acquisition. Removed SSE is relative to the unshifted model, not persistence.

Δ	Direction	Windows	Unshifted	Shifted	Skill change [95% CI]	SSE removed
0.02	Both	1102	0.126	0.289	$+0.162$ [0.122, 0.200]	18.6%
	Lithiation	532	0.043	0.132	$+0.089$ [0.020, 0.141]	9.3%
	Delithiation	570	0.205	0.436	$+0.231$ [0.199, 0.275]	29.0%
0.16	Both	703	-1.413	0.088	$+1.501$ [1.283, 1.786]	62.2%
	Lithiation	399	-0.605	0.251	$+0.856$ [0.552, 1.402]	53.3%
	Delithiation	304	-2.154	-0.062	$+2.092$ [1.812, 2.705]	66.3%

Table 61: Formation (first lithiation) forecasts. Acquisition 02, held out; models trained on the C/20 acquisitions 03–06. Main windows: every forecaster of the full main table on identical causal windows (seed mean ± SD). Alignment variants: replay on its own windows with the recorded charge state or with a charge state shifted by the voltage offset (0.283; a quasi-equilibrium reference, not a calibrated OCV), descriptive only.

Windows	Forecaster	$\Delta = 0.01$	$\Delta = 0.02$	$\Delta = 0.04$	$\Delta = 0.08$	$\Delta = 0.16$	$\Delta = 0.32$	Particles
main windows	Persistence	0.00	0.00	0.00	0.00	0.00	0.00	19
main windows	Trend	-0.25	-0.52	-0.71	-1.02	-1.30	-1.10	19
main windows	AR(2)	0.20	0.16	0.14	0.16	0.25	0.37	19
main windows	SOC replay	-0.33	-0.51	-0.82	-1.26	-1.48	-0.67	19
main windows	History-aligned replay	-0.20	-0.27	-0.32	-0.46	-0.57	-0.05	19
main windows	Population replay	-0.23	-0.38	-0.62	-1.04	-1.21	-0.60	19
main windows	Damped trend	0.12	0.09	0.05	0.02	0.02	0.02	19
main windows	History-mean persistence	-0.75	-0.35	-0.15	-0.07	-0.05	-0.04	19
main windows	SOC pixel replay	-0.85	-1.05	-1.38	-1.74	-1.91	-0.87	19
main windows	Scaled population replay	0.00	-0.00	-0.04	-0.17	-0.49	0.16	19
main windows	Particle-mean ridge	-0.02	-0.02	-0.14	-0.57	-1.20	0.01	19
main windows	U-Net, autoregressive (d, s)	0.17 ± 0.01	0.13 ± 0.02	-0.10 ± 0.08	-0.67 ± 0.41	-1.41 ± 1.16	-0.56 ± 0.55	19
main windows	U-Net, autoregressive, no inputs	0.21 ± 0.02	0.22 ± 0.06	0.00 ± 0.24	-0.31 ± 0.17	-0.86 ± 0.51	-0.77 ± 0.77	19
main windows	U-Net, autoregressive (d, u, s)	0.20 ± 0.02	0.21 ± 0.01	0.13 ± 0.05	-0.09 ± 0.16	-0.44 ± 0.41	-0.33 ± 0.37	19
alignment variants (own windows)	SOC replay, recorded charge state	-0.29	-0.41	-0.69	-1.09	-1.35	-0.59	19
alignment variants (own windows)	SOC replay, voltage-offset charge state	0.06	0.09	0.19	0.29	0.42	0.51	19
alignment variants (own windows)	SOC pixel replay, recorded charge state	-0.74	-0.85	-1.16	-1.50	-1.72	-0.78	19
alignment variants (own windows)	SOC pixel replay, voltage-offset charge state	-0.23	-0.17	-0.01	0.14	0.34	0.51	19
alignment variants (own windows)	Mean replay, half-cycle position	-0.10	-0.14	-0.26	-0.45	-0.75	-0.23	19
alignment variants (own windows)	Trend, half-cycle position	-0.31	-0.52	-0.66	-0.95	-1.23	-0.99	19

Table 62: Formation baseline alignment sensitivity. Starting voltage 0.179 V maps to s=0.28271 on the acquisition-03 C/20 lithiation voltage curve. The s-offset variant changes only the reference-query coordinate; targets, causal origins and u are unchanged. Training-only references and identical windows are used for u, unadjusted s and offset s. This is an exploratory voltage-alignment sensitivity, not a calibrated state estimate.

Baseline / alignment	Direction	Horizon	Skill	RMSE	Fallback	n particles
Persistence	all	0.020	0.000	0.103	0.000	19
Persistence	all	0.160	0.000	0.309	0.000	19
Mean replay / u	all	0.020	−0.140	0.110	0.026	19
Mean replay / u	all	0.160	−0.747	0.409	0.036	19
Mean replay / s offset	all	0.020	0.086	0.099	0.001	19
Mean replay / s offset	all	0.160	0.420	0.235	0.001	19
Mean replay / s original	all	0.020	−0.411	0.123	0.203	19
Mean replay / s original	all	0.160	−1.350	0.474	0.184	19
Pixel replay / u	all	0.020	−0.419	0.123	0.026	19
Pixel replay / u	all	0.160	−0.966	0.434	0.036	19
Pixel replay / s offset	all	0.020	−0.168	0.112	0.001	19
Pixel replay / s offset	all	0.160	0.340	0.251	0.001	19
Pixel replay / s original	all	0.020	−0.850	0.141	0.203	19
Pixel replay / s original	all	0.160	−1.720	0.510	0.184	19
Trend	all	0.020	−0.520	0.128	0.000	19
Trend	all	0.160	−1.234	0.462	0.000	19
Persistence	lithiation	0.020	0.000	0.113	0.000	19
Persistence	lithiation	0.160	0.000	0.304	0.000	19
Mean replay / u	lithiation	0.020	−0.482	0.138	0.016	19
Mean replay / u	lithiation	0.160	−1.864	0.514	0.014	19
Mean replay / s offset	lithiation	0.020	0.031	0.111	0.002	19
Mean replay / s offset	lithiation	0.160	0.408	0.234	0.002	19
Mean replay / s original	lithiation	0.020	−0.479	0.137	0.016	19
Mean replay / s original	lithiation	0.160	−1.874	0.515	0.014	19

Table 63: Formation baseline alignment sensitivity. (continued). Starting voltage 0.179 V maps to s=0.28271 on the acquisition-03 C/20 lithiation voltage curve. The s-offset variant changes only the reference-query coordinate; targets, causal origins and u are unchanged. Training-only references and identical windows are used for u, unadjusted s and offset s. This is an exploratory voltage-alignment sensitivity, not a calibrated state estimate.

Baseline / alignment	Direction	Horizon	Skill	RMSE	Fallback	n particles
Pixel replay / u	lithiation	0.020	−0.969	0.159	0.016	19
Pixel replay / u	lithiation	0.160	−2.275	0.550	0.014	19
Pixel replay / s offset	lithiation	0.020	−0.274	0.128	0.002	19
Pixel replay / s offset	lithiation	0.160	0.319	0.251	0.002	19
Pixel replay / s original	lithiation	0.020	−0.979	0.159	0.016	19
Pixel replay / s original	lithiation	0.160	−2.285	0.551	0.014	19
Trend	lithiation	0.020	−0.827	0.153	0.000	19
Trend	lithiation	0.160	−1.382	0.469	0.000	19
Persistence	delithiation	0.020	0.000	0.094	0.000	19
Persistence	delithiation	0.160	0.000	0.316	0.000	19
Mean replay / u	delithiation	0.020	0.325	0.077	0.036	19
Mean replay / u	delithiation	0.160	0.592	0.202	0.065	19
Mean replay / s offset	delithiation	0.020	0.162	0.086	0.000	19
Mean replay / s offset	delithiation	0.160	0.436	0.237	0.000	19
Mean replay / s original	delithiation	0.020	−0.319	0.108	0.377	19
Mean replay / s original	delithiation	0.160	−0.723	0.415	0.402	19
Pixel replay / u	delithiation	0.020	0.331	0.077	0.036	19
Pixel replay / u	delithiation	0.160	0.602	0.199	0.065	19
Pixel replay / s offset	delithiation	0.020	−0.024	0.095	0.000	19
Pixel replay / s offset	delithiation	0.160	0.365	0.252	0.000	19
Pixel replay / s original	delithiation	0.020	−0.674	0.121	0.377	19
Pixel replay / s original	delithiation	0.160	−1.043	0.452	0.402	19
Trend	delithiation	0.020	−0.102	0.098	0.000	19
Trend	delithiation	0.160	−1.056	0.453	0.000	19

Table 64: Per-direction U-Net models. Each direction model is scored on its own direction's windows; the whole is the SSE-recombined pair (never an average of skills) and needs both direction models of one seed. Comparator (Mixed): the mixed-direction model on identical windows and common seeds. Difference = per-direction minus mixed-direction skill, 95% paired particle-cluster interval. Particle fold 0 is a single identity fold. failed: training stopped because the direction filter left no validation windows for that seed.

Split	Scored on	Δ	Per-direction	Mixed	Difference [95% CI]	Seeds	Particles
C/20→C/2	lithiation	0.020	0.337	0.434	-0.097 [-0.107, -0.075]	2	22
C/20→C/2	delithiation	0.020	0.654	0.629	0.025 [0.001, 0.035]	2	22
C/20→C/2	whole	0.020	0.470	0.516	-0.046 [-0.061, -0.038]	2	22
C/20→C/2	lithiation	0.160	0.580	0.603	-0.023 [-0.113, 0.013]	2	22
C/20→C/2	delithiation	0.160	0.819	0.782	0.036 [-0.016, 0.064]	2	22
C/20→C/2	whole	0.160	0.663	0.666	-0.002 [-0.074, 0.029]	2	22
C/2→C/20	lithiation	0.020	0.362	0.333	0.029 [0.011, 0.056]	2	22
C/2→C/20	delithiation	0.020	0.426	0.443	-0.017 [-0.027, -0.007]	2	22
C/2→C/20	whole	0.020	0.388	0.379	0.010 [-0.003, 0.029]	2	22
C/2→C/20	lithiation	0.160	0.581	0.561	0.020 [-0.004, 0.061]	2	22
C/2→C/20	delithiation	0.160	0.561	0.585	-0.024 [-0.087, 0.055]	2	22
C/2→C/20	whole	0.160	0.572	0.572	0.000 [-0.019, 0.037]	2	22
(C/20+C/2) →2C	lithiation	0.020	0.500	0.513	-0.012 [-0.018, 0.047]	2	3
(C/20+C/2) →2C	delithiation	0.020	0.502	0.510	-0.008 [-0.070, -0.005]	2	3
(C/20+C/2) →2C	whole	0.020	0.502	0.511	-0.009 [-0.040, -0.007]	2	3
(C/20+C/2) →2C	lithiation	0.160	0.690	0.634	0.057 [0.054, 0.085]	2	3
(C/20+C/2) →2C	delithiation	0.160	0.628	0.510	0.118 [0.112, 0.223]	2	3
(C/20+C/2) →2C	whole	0.160	0.655	0.565	0.091 [0.087, 0.144]	2	3
C/2: 25 →6–15°C	lithiation	0.020	0.453	0.427	0.026 [-0.017, 0.076]	2	21
C/2: 25 →6–15°C	delithiation	0.020	0.340	0.482	-0.142 [-0.610, 0.108]	1	22
C/2: 25 →6–15°C	whole	0.020	0.424	0.432	-0.009 [-0.131, 0.136]	1	22
C/2: 25 →6–15°C	lithiation	0.160	0.645	0.363	0.282 [0.187, 0.493]	2	21
C/2: 25 →6–15°C	delithiation	0.160	0.905	0.883	0.023 [-0.013, 0.062]	1	21
C/2: 25 →6–15°C	whole	0.160	0.743	0.446	0.297 [0.143, 0.671]	1	21
C/2: 25 →35–45°C	lithiation	0.020	0.363	0.372	-0.009 [-0.103, 0.049]	2	17
C/2: 25 →35–45°C	delithiation	0.020	0.592	0.552	0.040 [-0.087, 0.106]	1	16

Table 65: Per-direction U-Net models. (continued).

Split	Scored on	Δ	Per-direction	Mixed	Difference [95% CI]	Seeds	Particles
C/2: 25 →35–45°C	whole	0.020	0.475	0.447	0.028 [-0.149, 0.148]	1	17
C/2: 25 →35–45°C	lithiation	0.160	0.588	0.665	-0.077 [-0.160, -0.041]	2	17
C/2: 25 →35–45°C	delithiation	0.160	0.741	0.639	0.101 [-0.011, 0.290]	1	16
C/2: 25 →35–45°C	whole	0.160	0.677	0.674	0.003 [-0.148, 0.124]	1	17
New particle, fold 0	lithiation	0.020	0.430	0.453	-0.023 [-0.033, 0.039]	2	5
New particle, fold 0	delithiation	0.020	0.644	0.624	0.020 [0.016, 0.051]	2	5
New particle, fold 0	whole	0.020	0.520	0.525	-0.005 [-0.012, 0.041]	2	5
New particle, fold 0	lithiation	0.160	0.608	0.698	-0.089 [-0.094, -0.017]	2	5
New particle, fold 0	delithiation	0.160	0.752	0.827	-0.075 [-0.081, -0.009]	2	5
New particle, fold 0	whole	0.160	0.663	0.747	-0.084 [-0.089, -0.024]	2	5
Later cycles acq. 14	lithiation	0.020	0.588	0.585	0.003 [-0.007, 0.028]	2	15
Later cycles acq. 14	delithiation	0.020	0.749	0.758	-0.009 [-0.013, 0.003]	2	14
Later cycles acq. 14	whole	0.020	0.665	0.668	-0.003 [-0.009, 0.016]	2	15
Later cycles acq. 14	lithiation	0.160	0.806	0.806	0.000 [-0.019, 0.067]	2	15
Later cycles acq. 14	delithiation	0.160	0.952	0.954	-0.003 [-0.016, -0.001]	2	14
Later cycles acq. 14	whole	0.160	0.845	0.845	-0.000 [-0.014, 0.050]	2	15
C/3→1C,2C	lithiation	0.020	0.835	0.838	-0.003 [-0.025, 0.011]	2	24
C/3→1C,2C	delithiation	0.020	0.780	0.776	0.003 [-0.003, 0.015]	2	24
C/3→1C,2C	whole	0.020	0.803	0.802	0.001 [-0.008, 0.009]	2	24
C/3→1C,2C	lithiation	0.160	0.944	0.945	-0.001 [-0.018, 0.009]	2	24
C/3→1C,2C	delithiation	0.160	0.863	0.874	-0.010 [-0.021, 0.008]	2	24
C/3→1C,2C	whole	0.160	0.893	0.900	-0.007 [-0.016, 0.004]	2	24
(C/3+1C) →2C	lithiation	0.020	0.833	0.831	0.002 [-0.015, 0.024]	2	24
(C/3+1C) →2C	delithiation	0.020	0.792	0.792	0.001 [-0.010, 0.012]	2	24
(C/3+1C) →2C	whole	0.020	0.810	0.809	0.001 [-0.008, 0.015]	2	24
(C/3+1C) →2C	lithiation	0.160	0.948	0.945	0.003 [-0.005, 0.012]	2	24

One model per current direction. Table 64 trains one autoregressive U-Net per current direction and recombines the two on the test half-cycles; Table 67 does the same for the direct U-Net. Failed runs are listed with their reason rather than omitted.

Direct prediction on the primary NMC splits. Table 68 adds the direct U-Net (seeds 0–4, the recipe of the graphite direct U-Net) on the two primary NMC rate splits, scored on the windows of the step-by-step U-Net and of the main-table best baseline, with the per-direction U-Nets of Table 66 for 2C to C/3. At $\Delta = 0.16$, direct prediction beats step-by-step prediction on both splits, with no

Table 66: Per-direction U-Net models. (continued).

Split	Scored on	Δ	Per-direction	Mixed	Difference [95% CI]	Seeds	Particles
(C/3+1C) →2C	delithiation	0.160	0.849	0.854	-0.005 [-0.014, 0.007]	2	24
(C/3+1C) →2C	whole	0.160	0.888	0.890	-0.002 [-0.007, 0.004]	2	24
2C→C/3	lithiation	0.020	0.701	0.690	0.010 [-0.016, 0.047]	2	24
2C→C/3	delithiation	0.020	0.680	0.219	0.460 [0.395, 0.583]	2	24
2C→C/3	whole	0.020	0.693	0.520	0.173 [0.141, 0.228]	2	24
2C→C/3	lithiation	0.160	0.826	0.845	-0.020 [-0.122, 0.051]	2	24
2C→C/3	delithiation	0.160	0.859	0.562	0.297 [0.220, 0.538]	2	24
2C→C/3	whole	0.160	0.844	0.696	0.147 [0.101, 0.269]	2	24

Table 67: Per-direction direct lead-time models (validation-selected recipe). Each direction model is scored on its own direction's windows; the whole is the SSE-recombined pair (never an average of skills) and needs both direction models of one seed. Comparator (Mixed): the mixed-direction model on identical windows and common seeds. Difference = per-direction minus mixed-direction skill, 95% paired particle-cluster interval. Particle fold 0 is a single identity fold. failed: training stopped because the direction filter left no validation windows for that seed.

Split	Scored on	Δ	Per-direction	Mixed	Difference [95% CI]	Seeds	Particles	Failed runs
C/20→C/2	lithiation	0.020	0.381	0.405	-0.024 [-0.053, -0.010]	1	22	seed 1
C/20→C/2	delithiation	0.020	0.632	0.554	0.078 [0.035, 0.098]	1	22	seed 0
C/20→C/2	whole	0.020	failed	failed	failed	0	0	delithiation seed 0 lithiation seed 1
C/20→C/2	lithiation	0.160	0.761	0.759	0.002 [-0.008, 0.026]	1	22	seed 1
C/20→C/2	delithiation	0.160	0.853	0.850	0.003 [-0.016, 0.011]	1	22	seed 0
C/20→C/2	whole	0.160	failed	failed	failed	0	0	delithiation seed 0 lithiation seed 1
C/2→C/20	lithiation	0.020	0.291	0.246	0.045 [0.023, 0.081]	1	22	seed 1
C/2→C/20	delithiation	0.020	0.369	0.351	0.018 [0.001, 0.036]	1	22	seed 0
C/2→C/20	whole	0.020	failed	failed	failed	0	0	delithiation seed 0 lithiation seed 1
C/2→C/20	lithiation	0.160	0.659	0.656	0.003 [-0.035, 0.016]	1	22	seed 1
C/2→C/20	delithiation	0.160	0.677	0.674	0.003 [-0.009, 0.009]	1	22	seed 0
C/2→C/20	whole	0.160	failed	failed	failed	0	0	delithiation seed 0 lithiation seed 1
New particle, fold 0	lithiation	0.020	0.335	0.372	-0.036 [-0.041, 0.012]	2	5	–
New particle, fold 0	delithiation	0.020	0.477	0.533	-0.056 [-0.062, 0.002]	2	5	–
New particle, fold 0	whole	0.020	0.395	0.440	-0.045 [-0.050, 0.006]	2	5	–
New particle, fold 0	lithiation	0.160	0.677	0.691	-0.014 [-0.016, 0.014]	2	5	–
New particle, fold 0	delithiation	0.160	0.823	0.845	-0.022 [-0.023, 0.004]	2	5	–
New particle, fold 0	whole	0.160	0.733	0.750	-0.017 [-0.019, 0.006]	2	5	–
Later cycles acq. 14	lithiation	0.020	0.515	0.542	-0.028 [-0.078, -0.013]	2	15	–
Later cycles acq. 14	delithiation	0.020	0.743	0.680	0.062 [-0.056, 0.099]	2	14	–
Later cycles acq. 14	whole	0.020	0.625	0.609	0.016 [-0.067, 0.040]	2	15	–
Later cycles acq. 14	lithiation	0.160	0.888	0.900	-0.011 [-0.029, -0.003]	2	15	–
Later cycles acq. 14	delithiation	0.160	0.965	0.956	0.009 [0.005, 0.018]	2	14	–
Later cycles acq. 14	whole	0.160	0.909	0.915	-0.006 [-0.020, 0.000]	2	15	–

resolved advantage over the best baseline. At $\Delta = 0.02$, it trails step-by-step prediction on both splits, but beats the baseline only from C/3 to 1C,2C.

Table 68: Direct prediction on primary NMC. Primary NMC, direct and step-by-step U-Nets with seeds 0–4. Best-baseline identity is frozen to the corresponding all-direction main-table row at each horizon, then rescored on common windows. The chosen baseline is named in each row. Per-direction U-Nets use both seeds 0–1 from Table 66; whole-direction scores pool squared errors, not direction skills. Each contrast retains all seeds in each arm. Direction rows restrict the same matched population. The last column is direct minus the row's comparator; 0.32 is secondary if evaluated. Skill: seed mean ± sample SD. Differences: 95% paired particle-cluster bootstrap intervals (2,000 resamples). Identical test windows within each split and horizon across all available arms and their seeds. An arm is pending until all its planned seeds are evaluated. Resolved means the interval excludes zero; five or fewer test particles are descriptive. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Split	Δ	Forecaster	Skill	Direct − comparator
C/3→1C,2C	0.02	Direct U-Net	0.790 ± 0.012	–
		Main-table best: Scaled population replay	0.752	+0.037 [0.007, 0.073]
		Step-by-step U-Net	0.809 ± 0.007	-0.019 [-0.025, -0.002]
C/3→1C,2C	0.16	Direct U-Net	0.910 ± 0.004	–
		Main-table best: Particle-mean ridge	0.881	+0.029 [-0.002, 0.058]
		Step-by-step U-Net	0.900 ± 0.004	+0.009 [0.006, 0.016]
C/3→1C,2C	0.32	Direct U-Net	0.933 ± 0.003	–
		Main-table best: Particle-mean ridge	0.890	+0.042 [0.027, 0.057]
		Step-by-step U-Net	0.913 ± 0.006	+0.019 [0.008, 0.038]
2C→C/3	0.02	Direct U-Net	0.422 ± 0.209	–
		Main-table best: AR(2)	0.585	-0.163 [-0.482, -0.041]
		Step-by-step U-Net	0.543 ± 0.081	-0.121 [-0.415, -0.008]
		AR(2)	0.585	-0.163 [-0.482, -0.041]
		Per-direction U-Nets, whole	0.693 ± 0.006	-0.271 [-0.589, -0.152]
		Per-direction U-Net, lithiation	0.701 ± 0.015	-0.218 [-0.343, -0.159]
		Per-direction U-Net, delithiation	0.680 ± 0.041	-0.366 [-1.121, -0.112]
2C→C/3	0.16	Direct U-Net	0.834 ± 0.034	–
		Main-table best: AR(2)	0.821	+0.013 [-0.063, 0.042]
		Step-by-step U-Net	0.712 ± 0.081	+0.122 [0.078, 0.162]
		AR(2)	0.821	+0.013 [-0.063, 0.042]
		Per-direction U-Nets, whole	0.844 ± 0.014	-0.010 [-0.102, 0.041]
		Per-direction U-Net, lithiation	0.826 ± 0.040	+0.102 [0.032, 0.258]
		Per-direction U-Net, delithiation	0.859 ± 0.009	-0.110 [-0.315, -0.038]

Table 69: Direct prediction on primary NMC. (continued) Conventions as in Table 68.

Split	Δ	Forecaster	Skill	Direct − comparator
2C→C/3	0.32	Direct U-Net	0.905 ± 0.030	–
		Main-table best: AR(2)	0.848	+0.057 [0.026, 0.086]
		Step-by-step U-Net	0.824 ± 0.046	+0.081 [0.041, 0.103]
		AR(2)	0.848	+0.057 [0.026, 0.086]
		Per-direction U-Nets, whole	0.909 ± 0.012	-0.004 [-0.074, 0.023]
		Per-direction U-Net, lithiation	0.915 ± 0.032	+0.043 [0.016, 0.101]
		Per-direction U-Net, delithiation	0.904 ± 0.003	-0.040 [-0.174, 0.005]

Table 70: Same-condition repeat comparisons and transfer gaps. Transfer gap = skill(same condition) minus skill(shifted condition). Cross-condition records are filtered to the identical test acquisition; C/2 uses cycle 0. Both models then share exact windows and seed IDs. Intervals resample paired particles. These are descriptive repeat comparisons, not an irreducible or causal rate-error floor.

Repeat split	Forecaster	Horizon	Same	Shifted	Gap	95 percent low	95 percent high	n particles
C/2 repeat to 10	U-Net (d, s)	0.020	0.529	0.492	0.037	0.021	0.046	19
C/2 repeat to 10	U-Net (d, s)	0.160	0.668	0.565	0.103	0.075	0.188	19
C/2 repeat to 10	U-Net none	0.020	0.523	0.478	0.046	−0.032	0.079	19
C/2 repeat to 10	U-Net none	0.160	0.209	0.527	−0.318	−0.364	−0.227	19
C/2 repeat to 10	SOC mean replay	0.020	0.087	0.026	0.061	−0.035	0.204	19
C/2 repeat to 10	SOC mean replay	0.160	0.593	0.545	0.049	−0.098	0.228	19
C/2 repeat to 10	SOC pixel replay	0.020	−0.126	−0.325	0.199	−0.037	0.311	19
C/2 repeat to 10	SOC pixel replay	0.160	0.651	0.527	0.124	−0.083	0.282	19
C/2 repeat to 11	U-Net (d, s)	0.020	0.521	0.463	0.058	−0.018	0.077	16
C/2 repeat to 11	U-Net (d, s)	0.160	0.730	0.444	0.286	0.226	0.545	16
C/2 repeat to 11	U-Net none	0.020	0.461	0.375	0.085	0.009	0.214	16
C/2 repeat to 11	U-Net none	0.160	0.559	0.403	0.157	0.021	0.210	16
C/2 repeat to 11	SOC mean replay	0.020	0.302	0.218	0.084	0.035	0.257	16
C/2 repeat to 11	SOC mean replay	0.160	0.544	0.459	0.085	0.051	0.292	16
C/2 repeat to 11	SOC pixel replay	0.020	0.327	−0.009	0.336	0.195	0.391	16
C/2 repeat to 11	SOC pixel replay	0.160	0.574	0.410	0.164	0.139	0.286	16
C/2 repeat to 14	U-Net (d, s)	0.020	0.538	0.493	0.045	−0.035	0.062	15
C/2 repeat to 14	U-Net (d, s)	0.160	0.757	0.712	0.045	0.032	0.112	15
C/2 repeat to 14	U-Net none	0.020	0.518	0.490	0.027	−0.025	0.043	15
C/2 repeat to 14	U-Net none	0.160	−0.569	0.518	−1.087	−1.292	−0.517	15
C/2 repeat to 14	SOC mean replay	0.020	−0.214	0.189	−0.403	−0.509	−0.080	15
C/2 repeat to 14	SOC mean replay	0.160	0.218	0.651	−0.433	−0.705	−0.224	15
C/2 repeat to 14	SOC pixel replay	0.020	−0.902	−0.080	−0.822	−1.026	−0.264	15
C/2 repeat to 14	SOC pixel replay	0.160	−0.002	0.653	−0.655	−0.901	−0.311	15

Table 71: Same-condition repeat comparisons and transfer gaps. (continued). Transfer gap = skill(same condition) minus skill(shifted condition). Cross-condition records are filtered to the identical test acquisition; C/2 uses cycle 0. Both models then share exact windows and seed IDs. Intervals resample paired particles. These are descriptive repeat comparisons, not an irreducible or causal rate-error floor.

Repeat split	Forecaster	Horizon	Same	Shifted	Gap	95 percent low	95 percent high	n particles
C/20 repeat 03→06	U-Net (d, s)	0.020	0.404	0.487	−0.083	−0.111	−0.009	15
C/20 repeat 03→06	U-Net (d, s)	0.160	0.559	0.774	−0.215	−0.344	−0.162	15
C/20 repeat 03→06	U-Net none	0.020	0.285	0.483	−0.198	−0.311	0.007	15
C/20 repeat 03→06	U-Net none	0.160	0.221	0.422	−0.201	−0.415	−0.097	15
C/20 repeat 03→06	SOC mean replay	0.020	−0.072	0.197	−0.269	−0.474	−0.174	15
C/20 repeat 03→06	SOC mean replay	0.160	0.234	0.692	−0.458	−0.569	−0.374	15
C/20 repeat 03→06	SOC pixel replay	0.020	−0.506	0.138	−0.645	−0.851	−0.548	15
C/20 repeat 03→06	SOC pixel replay	0.160	0.158	0.728	−0.570	−0.724	−0.464	15
C/20 repeat 06→03	U-Net (d, s)	0.020	0.066	0.376	−0.311	−0.656	−0.073	21
C/20 repeat 06→03	U-Net (d, s)	0.160	0.405	0.476	−0.072	−0.206	0.037	21
C/20 repeat 06→03	U-Net none	0.020	0.057	0.366	−0.309	−0.786	−0.043	21
C/20 repeat 06→03	U-Net none	0.160	−0.532	−0.307	−0.225	−0.622	−0.052	21
C/20 repeat 06→03	SOC mean replay	0.020	0.039	0.147	−0.108	−0.200	−0.003	21
C/20 repeat 06→03	SOC mean replay	0.160	0.184	0.354	−0.170	−0.264	−0.048	21
C/20 repeat 06→03	SOC pixel replay	0.020	−0.135	0.013	−0.148	−0.347	0.079	21
C/20 repeat 06→03	SOC pixel replay	0.160	0.130	0.271	−0.141	−0.213	−0.045	21

Same-condition repeats. Table 70 compares, on identical windows and seeds of one held-out acquisition, a model trained at the test condition with the model of the rate split tested there. At C/2 and 25 °C each held-out acquisition is forecast from the other two such acquisitions, against four C/20 acquisitions for the rate split; at C/20 and 25 °C one acquisition forecasts the other, against eight C/2 acquisitions. The same-condition U-Net (d, s) leads at C/2 on all three held-out acquisitions at $\Delta s = 0.16$ (by 0.10, 0.29 and 0.04) and trails at C/20, where it has one training acquisition; the U-Net without conditioning and the replay baselines change sign across acquisitions, so the size of a repeat gap depends on the acquisition as much as on the condition. These are descriptive comparisons, not a floor on the error of a rate change. Table 72 crosses a new particle with a new rate on the 15 particles of one C/2 acquisition (seed means).

Table 72: Particle and rate generalisation on matched targets. 0.02 / 0.16 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –.

Forecaster	Same particle / same rate	New particle / same rates	Same particle / new rate
U-Net / ds	0.54 ± 0.01 / 0.76 ± 0.12 15/1/685 ; 15/1/482	0.58 ± 0.00 / 0.76 ± 0.04 5/5 of 5 folds 15/1/685 ; 15/1/482	0.49 ± 0.03 / 0.70 ± 0.06 15/1/685 ; 15/1/482
U-Net / none	0.52 ± 0.07 / -0.57 ± 1.06 15/1/685 ; 15/1/482	0.56 ± 0.01 / 0.49 ± 0.04 5/5 of 5 folds 15/1/685 ; 15/1/482	0.49 ± 0.03 / 0.52 ± 0.16 15/1/685 ; 15/1/482

Table 73: Particle and rate generalisation on matched targets. 0.02 / 0.16 charge horizons. Entries are seed mean ± sample SD; a single seed has no SD. All available rows and seeds use identical windows within each split, fold and horizon. New-particle means require all five folds and weight folds equally; partial means are withheld. Counts are particles/acquisitions/windows; fold counts and sources are in the CSV. SD describes training seeds, not experimental replication. Missing, undefined or inapplicable entries are –.

Forecaster	New particle / new rate
U-Net / ds	0.49 ± 0.02 / 0.63 ± 0.04 5/5 of 5 folds 15/1/685 ; 15/1/482
U-Net / none	0.46 ± 0.01 / 0.39 ± 0.03 5/5 of 5 folds 15/1/685 ; 15/1/482

Table 74: Optical-marker and switch-map diagnostics on all splits. Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
NMC811 voltage window	U-Net (d, s)	0.356	0.038	7052.000	312.667	480.000
NMC811 voltage window	U-Net none	0.613	0.017	6883.500	499.500	480.000
NMC811 voltage window	U-Net (d, u, s)	0.379	0.033	7118.500	207.500	480.000
NMC811 voltage window	SOC mean replay	0.209	–	7721.000	83.000	480.000
NMC811 voltage window	SOC pixel replay	0.209	−0.025	7721.000	83.000	480.000
NMC811 voltage window	Damped trend	0.695	–	7618.000	220.000	480.000
NMC811 voltage window	History-aligned SOC replay	0.160	–	7702.000	88.000	480.000
NMC811 ageing	U-Net (d, s)	0.370	0.095	6400.000	110.000	480.000
NMC811 ageing	U-Net none	0.528	0.060	6460.000	141.500	480.000
NMC811 ageing	U-Net (d, u, s)	0.381	0.101	6332.000	130.000	480.000
NMC811 ageing	SOC mean replay	0.231	–	2451.000	684.000	480.000
NMC811 ageing	SOC pixel replay	0.226	−0.007	2466.000	677.000	480.000
NMC811 ageing	Damped trend	0.555	–	6980.000	92.000	480.000
NMC811 ageing	History-aligned SOC replay	0.254	–	3311.000	590.000	480.000
NMC811 particle fold 0	U-Net (d, s)	0.317	0.084	4528.333	130.333	336.000
NMC811 particle fold 0	U-Net none	0.532	0.050	4434.500	189.500	336.000
NMC811 particle fold 0	U-Net (d, u, s)	0.302	0.097	4535.500	149.000	336.000
NMC811 particle fold 0	SOC mean replay	–	–	–	–	0
NMC811 particle fold 0	SOC pixel replay	–	–	–	–	0
NMC811 particle fold 0	Damped trend	0.632	–	4861.000	99.000	336.000
NMC811 particle fold 0	History-aligned SOC replay	–	–	–	–	0
NMC811 particle fold 1	U-Net (d, s)	0.395	0.097	4714.000	123.000	336.000
NMC811 particle fold 1	U-Net none	0.564	0.065	4789.000	142.000	336.000
NMC811 particle fold 1	U-Net (d, u, s)	0.361	0.105	4735.000	133.500	336.000

F FURTHER ELECTRODES: RESULTS

This appendix tests transfer to additional electrodes and ageing regimes.

These are retrospective, descriptive transfer results: one flake per graphite cell (moving-block intervals over blocks of ten consecutive forecast origins within a half-cycle) and five test particles in the second NMC electrode. Graphite cells B and C are never trained on; every model is a main-electrode checkpoint of Sections 6.1 and 6.2, evaluated with the cell's own Q_{ref}. Primary and capacity-matched results in Table 92 concern different physical origins and targets and are reported together; neither the within-flake intervals nor the calibration sensitivity establish electrode-population replication. The capacity-matching sensitivity was declared before any cross-electrode neural score was inspected, after the history-only baselines had been inspected; the evaluated acquisition supplies the matched capacity. On cell B at C/2 with the capacity-matched scale, C/2-trained charge/time models score 0.45/0.42 at $\Delta = 0.16$; this comparison requires no rate transfer.

A6: fixed physical support and timing sensitivity. Table 93 reports the subsequent capacity contrast on identical histories, masks and physical targets, and the timing contrast on common physical origins. Capacity contrasts use a lead of 0.16 in the matched-capacity coordinate; timing contrasts use 0.16 in the primary coordinate. For B C/2 with the C/20 source, the charge U-Net's capacity gains are $+0.500$, $+0.569$ and $+0.493$ in seeds 0–2, but the 14 pooled origins split 9/5 by half-cycle. Ten-origin blocks have no alternative resamples, so all three intervals are unestimable. Timing changes over all four evaluated horizons have maximum absolute pooled skill changes 0.0147 on B C/2 and 0.0201 on C formation 1 (rounded); C formation 2 and B formation have no com-

Table 75: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
NMC811 particle fold 1	SOC mean replay	–	–	–	–	0
NMC811 particle fold 1	SOC pixel replay	–	–	–	–	0
NMC811 particle fold 1	Damped trend	0.603	–	5088.000	94.000	336.000
NMC811 particle fold 1	History-aligned SOC replay	–	–	–	–	0
NMC811 particle fold 2	U-Net (d, s)	–	–	–	–	0
NMC811 particle fold 2	U-Net none	–	–	–	–	0
NMC811 particle fold 2	U-Net (d, u, s)	–	–	–	–	0
NMC811 particle fold 2	SOC mean replay	–	–	–	–	0
NMC811 particle fold 2	SOC pixel replay	–	–	–	–	0
NMC811 particle fold 2	Damped trend	–	–	–	–	0
NMC811 particle fold 2	History-aligned SOC replay	–	–	–	–	0
NMC811 particle fold 3	U-Net (d, s)	–	–	–	–	0
NMC811 particle fold 3	U-Net none	–	–	–	–	0
NMC811 particle fold 3	U-Net (d, u, s)	–	–	–	–	0
NMC811 particle fold 3	SOC mean replay	–	–	–	–	0
NMC811 particle fold 3	SOC pixel replay	–	–	–	–	0
NMC811 particle fold 3	Damped trend	–	–	–	–	0
NMC811 particle fold 3	History-aligned SOC replay	–	–	–	–	0
NMC811 particle fold 4	U-Net (d, s)	–	–	–	–	0
NMC811 particle fold 4	U-Net none	–	–	–	–	0
NMC811 particle fold 4	U-Net (d, u, s)	–	–	–	–	0
NMC811 particle fold 4	SOC mean replay	–	–	–	–	0
NMC811 particle fold 4	SOC pixel replay	–	–	–	–	0
NMC811 particle fold 4	Damped trend	–	–	–	–	0

mon physical origins at any horizon. These missing comparisons do not establish timing robustness. Timing can also displace physical targets, unlike the capacity contrast.

The second NMC electrode's clock and baseline comparisons are reported in Tables 94 and 95, with paired descriptive intervals on their common windows.

Identical origins (A7). Amendment A7 replaces the sparse A6 intersection by one prespecified list of stored-frame origins per acquisition valid under both compared arms (501 on cell B at C/2, 1,552 on B's formation cycle, 546 and 437 on C's formation cycles), with physical targets and history fixed for the capacity comparison (timing shifts move targets by up to 43 s). Table 96 gives the paired differences at the capacity-matched $\Delta = 0.16$ (descriptive moving-block intervals, ten-origin blocks). The capacity correction raises every model; it raises the C/20-trained charge model four times as much as the time model, so the forward clock lead appears on identical targets ($+0.22$ to $+0.45$, two of three seeds' intervals above zero; -0.05 with the primary capacity), whereas the C/2-trained models, which need no rate transfer, gain similarly under both clocks. Shifting the charge record by each acquisition's fitted optical lag changes charge and direct skills at $\Delta = 0.16$ by at most 0.019 and 0.022 (charge 0.027 at 0.08), including C's second formation cycle (-900 s lag).

F.1 NMC811 AGEING AND VOLTAGE WINDOW

We compare ageing and voltage-window transfer on the NMC811 cycles. Training on cycles 3–4 and 28–29 at 4.2 V and testing on cycles 99–117, direct U-Net scores 0.541 and 0.887 at $\Delta = 0.02$ and 0.16 (seeds 0–1), ahead of the best baseline by $+0.128$ [0.116, 0.143] and $+0.038$ [0.028, 0.058], respectively. The best baselines are AR(2) (0.413) and particle-mean ridge (0.849); the autoregressive U-Net also leads them, by $+0.200$ [0.188, 0.217] and $+0.018$ [0.009, 0.030] (three seeds; conditional particle intervals on this cell). When the upper cut-off rises to 4.4 and 4.6 V, half-cycles reach 1.26–1.31 Q_{ref}; voltage and cycle number are confounded. Direct U-Net scores -0.022 and 0.500, trailing AR(2) at 0.02 (-0.322 [-0.483, -0.198]) and unresolved at 0.16 (-0.023 [-0.112, 0.049]). The autoregressive U-Net trails AR(2) at 0.16 (-0.108 [-0.212, -0.039]). On new-particle fold 0, direct-minus-best is $+0.153$ [0.143, 0.171] at 0.02 and unresolved at 0.16 ($+0.023$ [-0.005, 0.039]); no five-fold mean is reported (Table 38).

Table 76: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
NMC811 particle fold 4 Later cycles	History-aligned SOC replay	–	–	–	–	0
acq. 14 Later cycles	U-Net (d, s)	0.020	0.306	125.200	27.800	84.000
acq. 14 Later cycles	U-Net none	0.089	0.156	156.333	74.333	84.000
acq. 14 Later cycles	U-Net (d, u, s)	0.016	0.300	133.333	24.667	84.000
acq. 14 Later cycles	SOC mean replay	0.342	–	165.000	75.000	84.000
acq. 14 Later cycles	SOC pixel replay	0.342	0.100	165.000	75.000	84.000
acq. 14 Later cycles	Damped trend	0.159	–	210.000	68.000	84.000
acq. 14 Later cycles	History-aligned SOC replay	0.142	–	127.000	71.000	84.000
2C Later cycles	U-Net (d, s)	0.031	0.092	59.000	7.667	36.000
2C Later cycles	U-Net none	0.053	0.137	61.000	17.000	36.000
2C Later cycles	U-Net (d, u, s)	0.037	0.128	60.000	8.333	36.000
2C Later cycles	SOC mean replay	0.090	–	38.000	38.000	36.000
2C Later cycles	SOC pixel replay	0.090	0.024	38.000	38.000	36.000
2C Later cycles	Damped trend	0.098	–	93.000	18.000	36.000
2C	History-aligned SOC replay	0.030	–	43.000	31.000	36.000
Formation	U-Net (d, s)	–	–	–	–	0
Formation	U-Net none	–	–	–	–	0
Formation	U-Net (d, u, s)	–	–	–	–	0
Formation	SOC mean replay	–	–	–	–	0
Formation	SOC pixel replay	–	–	–	–	0
Formation	Damped trend	–	–	–	–	0
Formation	History-aligned SOC replay	–	–	–	–	0
C/2 repeat to 10	U-Net (d, s)	0.077	0.058	36.667	19.333	19.000
C/2 repeat to 10	U-Net none	0.162	0.036	33.000	31.333	19.000

Dependence sensitivity (A8). Ten-origin blocks span about 270 s on cell B at C/2, much less than the 0.16 horizon (about 1,712 s), so overlapping forecasts could make the A7 intervals optimistic. Recomputing them from the saved errors with blocks spanning the horizon, 60- and 120-origin blocks and whole half-cycles (Table 97; declared before computing) leaves every point estimate unchanged: the C/20-trained capacity gain stays above zero in all three seeds under every scheme, the matched-capacity charge-minus-time lead in two of three (three with whole half-cycles; positive in all six seed-by-half-cycle comparisons), and the primary normalisation still fails.

NMC811 by cycle. Table 98 splits the NMC811 comparisons by test cycle. On the ageing split both U-Nets beat the pooled best baseline in all four late cycles at both horizons, and leaving any cycle out keeps the gain positive (autoregressive $+0.014$ to $+0.021$, direct $+0.035$ to $+0.040$ at 0.16); against the best baseline chosen per cycle, both lose to charge-state replay on cycle 99 at 0.16. On the voltage-window split the signs are mixed (two of four cycles), and cut-off and cycle number remain confounded.

F.2 Cross-material transfer

Graphite and NMC share the stored signal (the log change of optical contrast relative to a pre-onset reference) and the charge clock, but differ in pixel size, charge scale and physics: graphite converts through moving phase fronts, whereas NMC changes smoothly as a solid solution. We asked whether anything learned on one material transfers to the other (exploratory; seeds 0–1; Table 99).

Table 77: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
C/2 repeat to 10	U-Net (d, u, s)	–	–	–	–	0
C/2 repeat to 10	SOC mean replay	0.226	–	34.000	20.000	19.000
C/2 repeat to 10	SOC pixel replay	0.226	0.078	32.000	19.000	19.000
C/2 repeat to 10	Damped trend	0.245	–	45.000	19.000	19.000
C/2 repeat to 10	History-aligned SOC replay	0.150	–	39.000	41.000	19.000
C/2 repeat to 11	U-Net (d, s)	0.029	0.019	34.000	14.000	32.000
C/2 repeat to 11	U-Net none	0.074	0.056	45.333	33.000	32.000
C/2 repeat to 11	U-Net (d, u, s)	–	–	–	–	0
C/2 repeat to 11	SOC mean replay	0.134	–	38.000	16.000	32.000
C/2 repeat to 11	SOC pixel replay	0.134	0.157	39.000	17.000	32.000
C/2 repeat to 11	Damped trend	0.145	–	62.000	20.000	32.000
C/2 repeat to 11	History-aligned SOC replay	0.070	–	27.000	39.000	32.000
C/2 repeat to 14	U-Net (d, s)	0.031	0.286	49.667	14.333	28.000
C/2 repeat to 14	U-Net none	0.102	0.182	49.000	27.000	28.000
C/2 repeat to 14	U-Net (d, u, s)	–	–	–	–	0
C/2 repeat to 14	SOC mean replay	0.078	–	39.000	26.000	28.000
C/2 repeat to 14	SOC pixel replay	0.079	0.224	40.000	25.000	28.000
C/2 repeat to 14	Damped trend	0.158	–	68.000	20.000	28.000
C/2 repeat to 14	History-aligned SOC replay	0.196	–	41.000	51.000	28.000
Mixed to cold	U-Net (d, s)	0.043	0.052	164.000	33.667	121.000
Mixed to cold	U-Net none	0.068	0.044	167.000	75.000	121.000
Mixed to cold	U-Net (d, u, s)	0.046	0.014	166.333	58.333	121.000
Mixed to cold	SOC mean replay	0.086	–	81.000	129.000	121.000
Mixed to cold	SOC pixel replay	0.088	0.046	83.000	132.000	121.000

Table 78: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
Mixed to cold	Damped trend	0.115	–	231.000	63.000	121.000
Mixed to cold	History-aligned SOC replay	0.083	–	101.000	190.000	121.000
G particle fold 0	U-Net (d, s)	0.041	0.178	160.800	40.600	85.000
G particle fold 0	U-Net none	0.110	0.094	177.333	78.000	85.000
G particle fold 0	U-Net (d, u, s)	0.041	0.207	162.333	55.333	85.000
G particle fold 0	SOC mean replay	–	–	–	–	0
G particle fold 0	SOC pixel replay	–	–	–	–	0
G particle fold 0	Damped trend	0.177	–	208.000	58.000	85.000
G particle fold 0	History-aligned SOC replay	–	–	–	–	0
G particle fold 1	U-Net (d, s)	0.037	0.091	87.000	38.667	63.000
G particle fold 1	U-Net none	0.114	0.059	96.000	71.000	63.000
G particle fold 1	U-Net (d, u, s)	0.031	0.141	80.667	41.667	63.000
G particle fold 1	SOC mean replay	–	–	–	–	0
G particle fold 1	SOC pixel replay	–	–	–	–	0
G particle fold 1	Damped trend	0.185	–	124.000	47.000	63.000
G particle fold 1	History-aligned SOC replay	–	–	–	–	0
G particle fold 2	U-Net (d, s)	0.050	0.112	161.333	47.667	87.000
G particle fold 2	U-Net none	0.086	0.038	156.000	93.667	87.000
G particle fold 2	U-Net (d, u, s)	0.047	0.118	157.000	43.667	87.000
G particle fold 2	SOC mean replay	–	–	–	–	0
G particle fold 2	SOC pixel replay	–	–	–	–	0
G particle fold 2	Damped trend	0.175	–	207.000	66.000	87.000
G particle fold 2	History-aligned SOC replay	–	–	–	–	0
G particle fold 3	U-Net (d, s)	0.045	0.130	133.000	33.333	74.000

Applied unchanged, a charge-clock U-Net trained on graphite (C/20→C/2) is far worse than persistence on the NMC rate split (seed-mean skill -16 and -41 at $\Delta = 0.02$ and 0.16, with large differences between seeds): it predicts graphite-sized changes per unit charge where NMC changes little. The NMC-trained model applied to graphite is also worse than persistence (-0.04 and -0.14), predicting small, smooth changes and missing graphite's transitions. Starting training on the target material from the other material's weights gives no resolved gain over training from scratch (differ-

Table 79: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
G particle fold 3	U-Net none	0.103	0.116	138.667	75.667	74.000
G particle fold 3	U-Net (d, u, s)	0.026	0.153	130.667	33.000	74.000
G particle fold 3	SOC mean replay	–	–	–	–	0
G particle fold 3	SOC pixel replay	–	–	–	–	0
G particle fold 3	Damped trend	0.174	–	170.000	51.000	74.000
G particle fold 3	History-aligned SOC replay	–	–	–	–	0
G particle fold 4	U-Net (d, s)	0.042	0.037	99.000	47.667	66.000
G particle fold 4	U-Net none	0.109	0.059	104.000	58.667	66.000
G particle fold 4	U-Net (d, u, s)	0.032	0.090	96.667	37.000	66.000
G particle fold 4	SOC mean replay	–	–	–	–	0
G particle fold 4	SOC pixel replay	–	–	–	–	0
G particle fold 4	Damped trend	0.172	–	129.000	49.000	66.000
G particle fold 4	History-aligned SOC replay	–	–	–	–	0
C/20→2C	U-Net (d, s)	0.068	0.079	183.667	28.000	81.000
C/20→2C	U-Net none	0.095	0.069	182.000	55.333	81.000
C/20→2C	U-Net (d, u, s)	0.101	0.095	186.333	57.333	81.000
C/20→2C	SOC mean replay	0.066	–	93.000	114.000	81.000
C/20→2C	SOC pixel replay	0.069	0.007	97.000	117.000	81.000
C/20→2C	Damped trend	0.131	–	230.000	49.000	81.000
C/20→2C	History-aligned SOC replay	0.106	–	92.000	113.000	81.000
C/20→C/2	U-Net (d, s)	0.064	0.093	591.600	252.400	347.000
C/20→C/2	U-Net none	0.096	0.072	612.000	325.000	347.000
C/20→C/2	U-Net (d, u, s)	0.130	0.070	648.000	384.667	347.000
C/20→C/2	SOC mean replay	0.085	–	300.000	679.000	347.000

Table 80: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
C/20→C/2	SOC pixel replay	0.086	0.065	296.000	682.000	347.000
C/20→C/2	Damped trend	0.165	–	771.000	247.000	347.000
C/20→C/2	History-aligned SOC replay	0.087	–	336.000	691.000	347.000
Rate shift / fold 0	U-Net (d, s)	0.070	0.155	151.667	59.000	79.000
Rate shift / fold 0	U-Net none	0.131	0.073	149.333	126.000	79.000
Rate shift / fold 0	U-Net (d, u, s)	–	–	–	–	0
Rate shift / fold 0	SOC mean replay	–	–	–	–	0
Rate shift / fold 0	SOC pixel replay	–	–	–	–	0
Rate shift / fold 0	Damped trend	–	–	–	–	0
Rate shift / fold 0	History-aligned SOC replay	–	–	–	–	0
Rate shift / fold 1	U-Net (d, s)	0.081	0.091	88.333	69.333	59.000
Rate shift / fold 1	U-Net none	0.143	0.020	99.000	66.667	59.000
Rate shift / fold 1	U-Net (d, u, s)	–	–	–	–	0
Rate shift / fold 1	SOC mean replay	–	–	–	–	0
Rate shift / fold 1	SOC pixel replay	–	–	–	–	0
Rate shift / fold 1	Damped trend	–	–	–	–	0
Rate shift / fold 1	History-aligned SOC replay	–	–	–	–	0
Rate shift / fold 2	U-Net (d, s)	0.081	0.138	151.667	60.333	79.000
Rate shift / fold 2	U-Net none	0.118	0.053	151.667	78.000	79.000
Rate shift / fold 2	U-Net (d, u, s)	–	–	–	–	0
Rate shift / fold 2	SOC mean replay	–	–	–	–	0
Rate shift / fold 2	SOC pixel replay	–	–	–	–	0
Rate shift / fold 2	Damped trend	–	–	–	–	0
Rate shift / fold 2	History-aligned SOC replay	–	–	–	–	0

ences within ± 0.012 at both horizons on both materials), although the two warm-started seeds agree more closely than the two seeds trained from scratch. What transfers between the two materials is therefore the approach, such as measuring progress in charge, rather than the learned weights.

Table 81: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
Rate shift / fold 3	U-Net (d, s)	0.057	0.148	123.333	44.667	68.000
Rate shift / fold 3	U-Net none	0.082	0.038	122.333	72.333	68.000
Rate shift / fold 3	U-Net (d, u, s)	–	–	–	–	0
Rate shift / fold 3	SOC mean replay	–	–	–	–	0
Rate shift / fold 3	SOC pixel replay	–	–	–	–	0
Rate shift / fold 3	Damped trend	–	–	–	–	0
Rate shift / fold 3	History-aligned SOC replay	–	–	–	–	0
Rate shift / fold 4	U-Net (d, s)	0.061	0.033	98.333	50.667	62.000
Rate shift / fold 4	U-Net none	0.094	0.025	85.333	77.333	62.000
Rate shift / fold 4	U-Net (d, u, s)	–	–	–	–	0
Rate shift / fold 4	SOC mean replay	–	–	–	–	0
Rate shift / fold 4	SOC pixel replay	–	–	–	–	0
Rate shift / fold 4	Damped trend	–	–	–	–	0
Rate shift / fold 4	History-aligned SOC replay	–	–	–	–	0
C/2→C/20	U-Net (d, s)	0.063	0.142	49.600	26.200	28.000
C/2→C/20	U-Net none	0.243	0.136	58.333	46.333	28.000
C/2→C/20	U-Net (d, u, s)	0.061	0.161	54.333	25.000	28.000
C/2→C/20	SOC mean replay	0.411	–	64.000	34.000	28.000
C/2→C/20	SOC pixel replay	0.411	0.026	63.000	34.000	28.000
C/2→C/20	Damped trend	0.324	–	67.000	24.000	28.000
C/2→C/20	History-aligned SOC replay	0.371	–	64.000	41.000	28.000
(C/20+C/2) →2C	U-Net (d, s)	0.059	0.101	172.000	42.000	81.000
(C/20+C/2) →2C	U-Net none	0.091	0.102	183.000	43.333	81.000
(C/20+C/2) →2C	U-Net (d, u, s)	0.055	0.108	174.333	33.333	81.000

Table 82: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
(C/20+C/2) →2C	SOC mean replay	0.055	–	84.000	118.000	81.000
(C/20+C/2) →2C	SOC pixel replay	0.058	0.018	83.000	112.000	81.000
(C/20+C/2) →2C	Damped trend	0.131	–	230.000	49.000	81.000
(C/20+C/2) →2C	History-aligned SOC replay	0.076	–	82.000	143.000	81.000
C/20 repeat 06→03	U-Net (d, s)	–	–	–	–	0
C/20 repeat 06→03	U-Net none	–	–	–	–	0
C/20 repeat 06→03	U-Net (d, u, s)	–	–	–	–	0
C/20 repeat 06→03	SOC mean replay	–	–	–	–	0
C/20 repeat 06→03	SOC pixel replay	–	–	–	–	0
C/20 repeat 06→03	Damped trend	–	–	–	–	0
C/20 repeat 06→03	History-aligned SOC replay	–	–	–	–	0
C/20 repeat 03→06	U-Net (d, s)	0.077	0.074	61.333	31.667	28.000
C/20 repeat 03→06	U-Net none	0.276	0.041	60.667	46.333	28.000
C/20 repeat 03→06	U-Net (d, u, s)	0.066	0.073	57.333	27.667	28.000
C/20 repeat 03→06	SOC mean replay	0.034	–	40.000	65.000	28.000
C/20 repeat 03→06	SOC pixel replay	0.033	0.131	40.000	65.000	28.000
C/20 repeat 03→06	Damped trend	0.324	–	67.000	24.000	28.000
C/20 repeat 03→06	History-aligned SOC replay	0.147	–	44.000	67.000	28.000
C/2 same condition	U-Net (d, s)	0.026	0.293	47.667	11.333	28.000
C/2 same condition	U-Net none	0.124	0.030	53.333	33.333	28.000
C/2 same condition	U-Net (d, u, s)	0.037	0.250	48.333	20.000	28.000
C/2 same condition	SOC mean replay	0.078	–	39.000	26.000	28.000
C/2 same condition	SOC pixel replay	0.079	0.224	40.000	25.000	28.000
C/2 same condition	Damped trend	0.158	–	68.000	20.000	28.000

Table 83: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
C/2 same condition C/20: 25	History-aligned SOC replay	0.196	–	41.000	51.000	28.000
→15,45°C C/20: 25	U-Net (d, s)	–	–	–	–	0
→15,45°C C/20: 25	U-Net none	–	–	–	–	0
→15,45°C C/20: 25	U-Net (d, u, s)	–	–	–	–	0
→15,45°C C/20: 25	SOC mean replay	–	–	–	–	0
→15,45°C C/20: 25	SOC pixel replay	–	–	–	–	0
→15,45°C C/20: 25	Damped trend	–	–	–	–	0
→15,45°C C/2: 25	History-aligned SOC replay	–	–	–	–	0
→6–15°C C/2: 25	U-Net (d, s)	0.031	0.051	162.000	35.200	121.000
→6–15°C C/2: 25	U-Net none	0.075	0.032	173.333	75.000	121.000
→6–15°C C/2: 25	U-Net (d, u, s)	0.046	0.008	152.667	49.000	121.000
→6–15°C C/2: 25	SOC mean replay	0.081	–	138.000	53.000	121.000
→6–15°C C/2: 25	SOC pixel replay	0.082	0.078	139.000	56.000	121.000
→6–15°C C/2: 25	Damped trend	0.115	–	231.000	63.000	121.000
→6–15°C C/2: 25	History-aligned SOC replay	0.102	–	135.000	144.000	121.000
→35–45°C C/2: 25	U-Net (d, s)	0.037	0.063	99.800	45.000	63.000
→35–45°C C/2: 25	U-Net none	0.194	0.058	129.667	118.333	63.000
→35–45°C C/2: 25	U-Net (d, u, s)	0.043	0.171	109.333	53.000	63.000
→35–45°C C/2: 25	SOC mean replay	0.286	–	140.000	49.000	63.000
→35–45°C C/2: 25	SOC pixel replay	0.286	−0.040	140.000	51.000	63.000
→35–45°C C/2: 25	Damped trend	0.256	–	155.000	57.000	63.000
→35–45°C	History-aligned SOC replay	0.296	–	147.000	71.000	63.000
N particle fold 0	U-Net (d, s)	0.248	0.027	225.600	22.200	36.000
N particle fold 0	U-Net none	0.294	−0.014	243.500	33.500	36.000

Table 84: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
N particle fold 0	U-Net (d, u, s)	0.211	0.088	228.000	22.000	36.000
N particle fold 0	SOC mean replay	–	–	–	–	0
N particle fold 0	SOC pixel replay	–	–	–	–	0
N particle fold 0	Damped trend	0.424	–	269.000	22.000	36.000
N particle fold 0	History-aligned SOC replay	–	–	–	–	0
N particle fold 1	U-Net (d, s)	0.239	0.046	236.400	12.000	36.000
N particle fold 1	U-Net none	0.287	0.008	245.500	21.500	36.000
N particle fold 1	U-Net (d, u, s)	0.238	0.002	234.500	16.500	36.000
N particle fold 1	SOC mean replay	–	–	–	–	0
N particle fold 1	SOC pixel replay	–	–	–	–	0
N particle fold 1	Damped trend	0.353	–	271.000	19.000	36.000
N particle fold 1	History-aligned SOC replay	–	–	–	–	0
N particle fold 2	U-Net (d, s)	0.218	0.006	268.800	20.800	36.000
N particle fold 2	U-Net none	0.264	0.062	259.500	39.000	36.000
N particle fold 2	U-Net (d, u, s)	0.215	0.019	283.500	19.500	36.000
N particle fold 2	SOC mean replay	–	–	–	–	0
N particle fold 2	SOC pixel replay	–	–	–	–	0
N particle fold 2	Damped trend	0.373	–	309.000	18.000	36.000
N particle fold 2	History-aligned SOC replay	–	–	–	–	0
N particle fold 3	U-Net (d, s)	0.210	0.053	231.200	13.600	36.000
N particle fold 3	U-Net none	0.253	−0.002	197.500	53.000	36.000
N particle fold 3	U-Net (d, u, s)	0.197	0.041	235.500	13.500	36.000
N particle fold 3	SOC mean replay	–	–	–	–	0
N particle fold 3	SOC pixel replay	–	–	–	–	0

Table 85: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker	Switch-map			
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
N particle fold 3	Damped trend	0.415	–	269.000	21.000	36.000
N particle fold 3	History-aligned SOC replay	–	–	–	–	0
N particle fold 4	U-Net (d, s)	0.175	0.015	256.200	16.400	36.000
N particle fold 4	U-Net none	0.191	−0.020	201.500	60.000	36.000
N particle fold 4	U-Net (d, u, s)	0.199	0.006	250.500	24.500	36.000
N particle fold 4	SOC mean replay	–	–	–	–	0
N particle fold 4	SOC pixel replay	–	–	–	–	0
N particle fold 4	Damped trend	0.459	–	296.000	21.000	36.000
N particle fold 4	History-aligned SOC replay	–	–	–	–	0
2C→C/3	U-Net (d, s)	0.197	0.080	510.200	19.000	48.000
2C→C/3	U-Net none	0.314	0.094	531.500	19.500	48.000
2C→C/3	U-Net (d, u, s)	0.201	0.035	518.000	14.500	48.000
2C→C/3	SOC mean replay	0.236	–	582.000	1.000	48.000
2C→C/3	SOC pixel replay	0.236	−0.124	581.000	1.000	48.000
2C→C/3	Damped trend	0.571	–	573.000	14.000	48.000
2C→C/3	History-aligned SOC replay	0.242	–	585.000	1.000	48.000
(C/3+1C) →2C (C/3+1C)	U-Net (d, s)	0.175	0.052	157.000	43.200	48.000
(C/3+1C) →2C (C/3+1C)	U-Net none	0.212	0.000	157.500	42.000	48.000
(C/3+1C) →2C (C/3+1C)	U-Net (d, u, s)	0.180	0.048	165.500	46.000	48.000
(C/3+1C) →2C (C/3+1C)	SOC mean replay	0.067	–	82.000	126.000	48.000
(C/3+1C) →2C (C/3+1C)	SOC pixel replay	0.066	0.195	80.000	120.000	48.000
(C/3+1C) →2C (C/3+1C)	Damped trend	0.305	–	195.000	44.000	48.000
(C/3+1C) →2C	History-aligned SOC replay	0.140	–	20.000	237.000	48.000
C/3→1C,2C	U-Net (d, s)	0.220	0.072	412.800	61.600	96.000

Table 86: Optical-marker and switch-map diagnostics on all splits. (continued). Marker error is the mean dominant optical-derivative marker error in charge units; correlation is the median within-half-cycle pixel Spearman correlation, then averaged over seeds. Multiple markers are matched one-to-one only within 0.03 in s; missed/extra are seed-mean counts. Within a constant-current half-cycle, absolute differences in u equal differences in s. Finite-metric counts and fallback fractions are in the CSV. Saved segment correlations cannot reconstruct pixel-pooled model correlation or new model switch maps.

		Optical-marker		Switch-map		
Split	Forecaster	charge error	correlation	Missed	Extra	n segments
C/3→1C,2C	U-Net none	0.251	0.028	385.000	150.500	96.000
C/3→1C,2C	U-Net (d, u, s)	0.277	0.038	429.000	67.000	96.000
C/3→1C,2C	SOC mean replay	0.214	–	241.000	267.000	96.000
C/3→1C,2C	SOC pixel replay	0.210	0.024	241.000	261.000	96.000
C/3→1C,2C	Damped trend	0.299	–	515.000	68.000	96.000
C/3→1C,2C	History-aligned SOC replay	0.199	–	116.000	465.000	96.000

Table 87: Appendix protocol consistency. Legacy segment-validation/interpolated-origin scores versus corrected block-validation/causal-origin scores. Only overlapping configurations and common particle folds are included. Forecasters are matched within each protocol; origin snapping changes targets, so old and new scores are not an exact paired causal effect. Short / long = 0.02 / 0.16. A sign-agreement check does not establish unchanged model rankings. Second lines give sample seed SD where available; uncensored scores, counts, support fractions and sources are in the CSV. – denotes missing.

Split / model	Legacy	Corrected	Same skill sign
Later cycles acq. 14 / s-aligned population replay	0.00 / 0.00	0.34 / 0.69	no / no
Later cycles acq. 14 / s-aligned replay	0.00 / 0.00	0.42 / 0.75	no / no
Later cycles acq. 14 / s-aligned pixel replay	0.00 / 0.00	0.48 / 0.88	no / no
Later cycles acq. 14 / trend	0.08 / −0.89	0.05 / −0.89	yes / yes
Later cycles 2C / s-aligned population replay	−0.03 / 0.18	0.19 / 0.67	no / yes
Later cycles 2C / s-aligned replay	0.13 / 0.24	0.30 / 0.76	yes / yes
Later cycles 2C / s-aligned pixel replay	0.04 / 0.21	0.30 / 0.79	yes / yes
Later cycles 2C / trend	0.01 / −0.72	0.01 / −0.72	yes / yes
Later cycles 2C / U-Net charge, (d, u, s)	0.69 / 0.86 ±0.00 / ±0.05	0.69 / 0.83 ±0.02 / ±0.11	yes / yes
New particle / s-aligned population replay	0.18 / 0.51	0.17 / 0.52	yes / yes
New particle / trend	−0.16 / −0.72	−0.19 / −0.73	yes / yes
New particle / U-Net charge, fixed-scale inputs	0.53 / 0.73 ±0.01 / ±0.00	0.51 / 0.64	yes / yes
New particle / U-Net charge, no inputs	0.51 / 0.54 ±0.01 / ±0.01	0.50 / 0.51 ±0.01 / ±0.08	yes / yes
New particle / U-Net charge, (d, u, s)	0.54 / 0.73 ±0.00 / ±0.03	0.52 / 0.70 ±0.01 / ±0.06	yes / yes
New particle / U-Net charge, state + fixed-scale inputs	0.55 / 0.69 ±0.00 / ±0.02	0.51 / 0.69	yes / yes
C/20→C/2 / s-aligned population replay	0.20 / 0.50	0.19 / 0.50	yes / yes
C/20→C/2 / s-aligned replay	0.16 / 0.60	0.16 / 0.60	yes / yes
C/20→C/2 / s-aligned pixel replay	−0.11 / 0.60	−0.11 / 0.61	yes / yes
C/20→C/2 / trend	−0.02 / −0.76	−0.06 / −0.76	yes / yes
C/20→C/2 / U-Net charge, no inputs	0.51 / 0.49 ±0.01 / ±0.03	0.48 / 0.47 ±0.02 / ±0.09	yes / yes
C/20→C/2 / U-Net charge, (d, u, s)	0.38 / 0.52 ±0.03 / ±0.01	0.37 / 0.34 ±0.01 / ±0.07	yes / yes
C/20→C/2 / U-Net frame, no inputs	0.45 / 0.23 ±0.03 / ±0.00	0.42 / 0.25 ±0.06 / ±0.05	yes / yes
C/20→C/2 / U-Net frame, (d, u, s)	0.24 / 0.22 ±0.00 / ±0.07	0.26 / 0.27 ±0.03 / ±0.04	yes / yes
C/20→C/2 / U-Net time, no inputs	0.24 / 0.11 ±0.10 / ±0.00	0.27 / 0.09 ±0.02 / ±0.03	yes / yes

Table 88: Appendix protocol consistency. Legacy segment-validation/interpolated-origin scores versus corrected block-validation/causal-origin scores. Only overlapping configurations and common particle folds are included. Forecasters are matched within each protocol; origin snapping changes targets, so old and new scores are not an exact paired causal effect. Short / long = 0.02 / 0.16. A sign-agreement check does not establish unchanged model rankings. Second lines give sample seed SD where available; uncensored scores, counts, support fractions and sources are in the CSV. – denotes missing. (continued).

Split / model	Legacy	Corrected	Same skill sign
C/20→C/2 / U-Net time, (d, u, s)	0.20 / 0.16 ± 0.00 / ± 0.03	0.15 / 0.09 ± 0.09 / ± 0.05	yes / yes
C/2→C/20 / s-aligned population replay	0.14 / 0.36	0.14 / 0.36	yes / yes
C/2→C/20 / s-aligned replay	0.12 / 0.39	0.12 / 0.38	yes / yes
C/2→C/20 / s-aligned pixel replay	0.02 / 0.35	0.02 / 0.34	yes / yes
C/2→C/20 / trend	-0.35 / -0.84	-0.37 / -0.88	yes / yes
C/2→C/20 / U-Net charge, no inputs	0.40 / 0.04 ± 0.00 / ± 0.18	0.38 / 0.05 ± 0.00 / ± 0.08	yes / yes
C/2→C/20 / U-Net charge, (d, u, s)	0.40 / 0.54 ± 0.00 / ± 0.00	0.38 / 0.50 ± 0.01 / ± 0.07	yes / yes
C/2→C/20 / U-Net frame, no inputs	0.16 / -0.52 ± 0.03 / ± 0.09	0.10 / < -1 ± 0.04 / ± 0.70	yes / yes
C/2→C/20 / U-Net frame, (d, u, s)	-0.27 / -0.05 ± 0.10 / ± 0.13	-0.43 / -0.05 ± 0.28 / ± 0.04	yes / yes
C/2→C/20 / U-Net time, no inputs	< -1 / < -1 ± 5.61 / ± 0.24	< -1 / < -1 ± 3.51 / ± 0.06	yes / yes
(C/20+C/2) →2C / s-aligned population replay	0.09 / 0.30	0.09 / 0.30	yes / yes
(C/20+C/2) →2C / s-aligned replay	0.11 / 0.39	0.11 / 0.39	yes / yes
(C/20+C/2) →2C / s-aligned pixel replay	0.09 / 0.39	0.09 / 0.39	yes / yes
(C/20+C/2) →2C / trend	-0.03 / -0.58	-0.02 / -0.58	yes / yes
(C/20+C/2) →2C / U-Net charge, (d, u, s)	0.49 / 0.72 ± 0.02 / ± 0.01	0.49 / 0.68 ± 0.01 / ± 0.03	yes / yes
C/20: 25 →15,45°C / s-aligned population replay	0.07 / 0.30	0.07 / 0.30	yes / yes
C/20: 25 →15,45°C / s-aligned replay	-0.03 / 0.20	-0.03 / 0.20	yes / yes
C/20: 25 →15,45°C / s-aligned pixel replay	-0.34 / 0.09	-0.33 / 0.09	yes / yes
C/20: 25 →15,45°C / trend	-0.44 / -0.87	-0.47 / -0.93	yes / yes
C/20: 25 →15,45°C / U-Net charge, no inputs	0.37 / 0.20 ± 0.01 / ± 0.01	0.35 / 0.10 ± 0.02 / ± 0.03	yes / yes
C/2: 25 →6–15°C / s-aligned population replay	0.17 / 0.30	0.16 / 0.33	yes / yes
C/2: 25 →6–15°C / s-aligned replay	0.17 / 0.52	0.17 / 0.53	yes / yes
C/2: 25 →6–15°C / s-aligned pixel replay	0.04 / 0.53	0.05 / 0.54	yes / yes
C/2: 25 →6–15°C / trend	-0.12 / < -1	-0.21 / < -1	yes / yes

Table 89: Appendix protocol consistency. Legacy segment-validation/interpolated-origin scores versus corrected block-validation/causal-origin scores. Only overlapping configurations and common particle folds are included. Forecasters are matched within each protocol; origin snapping changes targets, so old and new scores are not an exact paired causal effect. Short / long = 0.02 / 0.16. A sign-agreement check does not establish unchanged model rankings. Second lines give sample seed SD where available; uncensored scores, counts, support fractions and sources are in the CSV. – denotes missing. (continued).

Split / model	Legacy	Corrected	Same skill sign
C/2: 25 →6–15°C / U-Net charge, fixed-scale inputs	0.50 / 0.70 ±0.02 / ±0.04	0.42 / 0.63 ±0.07 / ±0.03	yes / yes
C/2: 25 →6–15°C / U-Net charge, no inputs	0.46 / 0.38 ±0.02 / ±0.05	0.41 / 0.30 ±0.02 / ±0.12	yes / yes
C/2: 25 →6–15°C / U-Net charge, (d, u, s)	0.48 / 0.64 ±0.03 / ±0.01	0.42 / 0.47 ±0.00 / ±0.31	yes / yes
C/2: 25 →6–15°C / U-Net charge, state + fixed-scale inputs	0.49 / 0.68 ±0.02 / ±0.02	−0.30 / 0.58 ±1.04 / ±0.13	no / yes
C/2: 25 →6–15°C / U-Net_ref_charge_state	0.01 / 0.68 ±0.56 / ±0.00	0.38 / 0.52 ±0.10 / ±0.04	yes / yes
C/2: 25 →35–45°C / s-aligned population replay	0.14 / 0.54	0.14 / 0.54	yes / yes
C/2: 25 →35–45°C / s-aligned replay	0.19 / 0.54	0.18 / 0.54	yes / yes
C/2: 25 →35–45°C / s-aligned pixel replay	0.12 / 0.56	0.11 / 0.56	yes / yes
C/2: 25 →35–45°C / trend	−0.07 / −0.57	−0.10 / −0.58	yes / yes
C/2: 25 →35–45°C / U-Net charge, fixed-scale inputs	0.45 / 0.66 ±0.11 / ±0.08	0.48 / 0.68 ±0.01 / ±0.07	yes / yes
C/2: 25 →35–45°C / U-Net charge, no inputs	0.53 / −0.12 ±0.00 / ±0.02	0.49 / −0.26 ±0.02 / ±0.26	yes / yes
C/2: 25 →35–45°C / U-Net charge, (d, u, s)	0.50 / 0.71 ±0.01 / ±0.02	0.47 / 0.68 ±0.02 / ±0.05	yes / yes
C/2: 25 →35–45°C / U-Net charge, state + fixed-scale inputs	0.48 / 0.69 ±0.03 / ±0.02	0.47 / 0.70 ±0.03 / ±0.02	yes / yes
C/2: 25 →35–45°C / U-Net_ref_charge_state	0.44 / 0.70 ±0.03 / ±0.03	0.45 / 0.66 ±0.01 / ±0.10	yes / yes
NMC: new particle / fick_charge_state	0.46 / 0.56	0.44 / 0.55	yes / yes
NMC: new particle / s-aligned population replay	0.47 / 0.69	0.48 / 0.68	yes / yes
NMC: new particle / trend	0.29 / 0.39	0.39 / 0.52	yes / yes
NMC: new particle / U-Net charge, no inputs	0.78 / 0.86	0.59 / 0.62 ±0.28 / ±0.33	yes / yes
NMC: new particle / U-Net charge, (d, u, s)	0.77 / 0.88	0.77 / 0.90 ±0.01 / ±0.01	yes / yes
2C→C/3 / fick_charge_state	0.30 / 0.45	0.35 / 0.54	yes / yes
2C→C/3 / s-aligned population replay	0.29 / 0.42	0.26 / 0.42	yes / yes
2C→C/3 / s-aligned replay	0.15 / 0.30	0.13 / 0.30	yes / yes
2C→C/3 / s-aligned pixel replay	0.13 / 0.30	0.10 / 0.30	yes / yes
2C→C/3 / trend	0.18 / 0.39	0.03 / 0.53	yes / yes

Table 90: Appendix protocol consistency. Legacy segment-validation/interpolated-origin scores versus corrected block-validation/causal-origin scores. Only overlapping configurations and common particle folds are included. Forecasters are matched within each protocol; origin snapping changes targets, so old and new scores are not an exact paired causal effect. Short / long = 0.02 / 0.16. A sign-agreement check does not establish unchanged model rankings. Second lines give sample seed SD where available; uncensored scores, counts, support fractions and sources are in the CSV. – denotes missing. (continued).

Split / model	Legacy	Corrected	Same skill sign
(C/3+1C) →2C / fick_charge_state	0.54 / 0.61	0.53 / 0.61	yes / yes
(C/3+1C) →2C / s-aligned population replay	0.54 / 0.67	0.50 / 0.66	yes / yes
(C/3+1C) →2C / s-aligned replay	0.57 / 0.81	0.51 / 0.81	yes / yes
(C/3+1C) →2C / s-aligned pixel replay	0.57 / 0.83	0.52 / 0.83	yes / yes
(C/3+1C) →2C / trend	−0.12 / −0.30	0.40 / 0.10	no / no
C/3→1C,2C / fick_charge_state	0.51 / 0.57	0.54 / 0.61	yes / yes
C/3→1C,2C / s-aligned population replay	0.60 / 0.67	0.56 / 0.67	yes / yes
C/3→1C,2C / s-aligned replay	0.57 / 0.69	0.54 / 0.69	yes / yes
C/3→1C,2C / s-aligned pixel replay	0.55 / 0.70	0.52 / 0.70	yes / yes
C/3→1C,2C / trend	0.28 / 0.34	0.62 / 0.51	yes / yes

Table 91: Protocol consistency of model comparisons. Entries are skill differences at 0.02 / 0.16 on each protocol's matched support. A positive value favours charge/state U-Net. Direction is preserved in 33/38 available horizon-specific checks; changes to ties count as changes. Coverage is partial, and seed counts differ across protocols; these are descriptive comparisons.

Split / state minus comparator	Legacy	Corrected	Same ranking
Later cycles 2C / SOC mean	0.566 / 0.616	0.394 / 0.063	yes / yes
Later cycles 2C / SOC pixel	0.652 / 0.648	0.389 / 0.033	yes / yes
New particle / none	0.025 / 0.189	0.022 / 0.187	yes / yes
C/20→C/2 / SOC mean	0.217 / −0.082	0.209 / −0.265	yes / yes
C/20→C/2 / SOC pixel	0.492 / −0.085	0.485 / −0.268	yes / yes
C/20→C/2 / none	−0.134 / 0.026	−0.108 / −0.134	yes / no
C/20→C/2 / 60 s state	0.176 / 0.355	0.218 / 0.243	yes / yes
C/2→C/20 / SOC mean	0.280 / 0.156	0.260 / 0.125	yes / yes
C/2→C/20 / SOC pixel	0.383 / 0.191	0.361 / 0.162	yes / yes
C/2→C/20 / none	0.006 / 0.504	−0.002 / 0.453	no / yes
(C/20+C/2) →2C / SOC mean	0.381 / 0.327	0.373 / 0.288	yes / yes
(C/20+C/2) →2C / SOC pixel	0.399 / 0.328	0.391 / 0.290	yes / yes
C/2: 25 →6–15°C / SOC mean	0.313 / 0.121	0.250 / −0.056	yes / no
C/2: 25 →6–15°C / SOC pixel	0.439 / 0.117	0.375 / −0.063	yes / no
C/2: 25 →6–15°C / none	0.020 / 0.259	0.016 / 0.179	yes / yes
C/2: 25 →35–45°C / SOC mean	0.314 / 0.164	0.285 / 0.141	yes / yes
C/2: 25 →35–45°C / SOC pixel	0.388 / 0.146	0.360 / 0.122	yes / yes
C/2: 25 →35–45°C / none	−0.025 / 0.827	−0.017 / 0.938	yes / yes
NMC: new particle / none	−0.006 / 0.024	0.184 / 0.277	no / yes

Table 92: Main-electrode U-Nets transferred to independent graphite flakes without retraining. Skill at $\Delta = 0.02$ / 0.16 (means of three seeds, two for direct); persistence has skill 0. Nominal scale and primary Q_{ref}, except the last block's capacity match, declared before cross-electrode neural scores but after history-only baselines. A retrospective sensitivity adds six baselines fitted only on each source split's seed-0 training partition ($Q_{\text{ref}} = 280$): AR(2) (A), damped trend (D), particle-mean ridge (R), population replay in u (U) or s (S), and scaled replay (G; fixed gain rule using observed history). Best transfer is the maximum at each horizon; superscripts identify methods. Last two columns: paired differences at 0.16 on matched origins, range over seeds above the counts (+ above zero/− below zero) of three descriptive 95% moving-block intervals (10 origins, 1,000 resamples). Longer baseline history excludes 2–3 origins in C's first cycle. Intervals condition on the selected baseline; no selection adjustment or electrode-population inference. Other scale/capacity variants are in the released summary.

Test flake	Train	Charge	Time	Direct	Trend	Best transfer	Charge − time at 0.16	Charge − best at 0.16
C, formation 2 (C/20)	C/20	0.11 / 0.41	−0.01 / 0.29	0.09 / 0.56	−0.82 / −1.08	0.02^{S} / 0.47^{R}	−0.11 to 0.51 (1+/0−)	−0.16 to 0.02 (0+/0−)
	C/2	0.03 / 0.50	−2.85 / −0.72	0.02 / 0.57		0.08^{S} / 0.45^{S}	0.96 to 1.72 (3+/0−)	0.03 to 0.08 (0+/0−)
C, formation 1 (C/20)	C/20	−0.10 / −0.63	−0.54 / −0.63	−0.30 / −0.60	−0.59 / −0.93	-0.00^{G} / 0.04^{A}	−0.20 to 0.12 (0+/0−)	−1.00 to −0.37 (0+/3−)
	C/2	−0.09 / −0.61	−6.67 / −7.28	−0.49 / −0.52		0.02^{R} / -0.01^{G}	4.64 to 8.64 (3+/0−)	−0.79 to −0.36 (0+/3−)
B, formation (C/20)	C/20	−0.04 / −0.37	−0.30 / −0.53	−0.13 / −0.47	−0.50 / −0.33	0.04^{A} / 0.07^{A}	−0.25 to 0.56 (1+/1−)	−0.49 to −0.36 (0+/1−)
	C/2	−0.17 / −0.62	−4.51 / −6.31	−0.16 / −0.28		0.07^{R} / 0.03^{A}	2.06 to 8.89 (3+/0−)	−0.82 to −0.55 (0+/3−)
B, C/2	C/20	0.18 / −0.18	0.06 / 0.03	0.02 / −0.22	−0.19 / −0.29	0.15^{A} / 0.07^{A}	−0.28 to −0.16 (0+/0−)	−0.30 to −0.22 (0+/1−)
	C/2	0.16 / −0.14	0.05 / −0.24	−1.38 / 0.06		0.13^{D} / 0.04^{A}	−0.04 to 0.18 (1+/0−)	−0.27 to −0.11 (0+/0−)
B, C/2, Q_{ref} matched	C/20	0.35 / 0.47	0.14 / 0.13	0.26 / 0.39	−0.30 / −0.37	0.15^{R} / 0.43^{R}	0.22 to 0.45 (2+/0−)	−0.04 to 0.14 (1+/0−)
	C/2	0.29 / 0.45	0.20 / 0.42	0.04 / 0.52		0.16^{R} / 0.39^{R}	−0.03 to 0.13 (0+/0−)	−0.06 to 0.11 (2+/0−)

Table 93: A6 results and physical support. Skill-change ranges span seeds 0–2 for autoregressive U-Nets and 0–1 for direct U-Net. Support is pooled origins [counts by nonempty half-cycle]. CI status describes the saved 95% moving-block intervals (ten origins per block): D, descriptive and estimable; B, unestimable because there are no alternative blocks; O, unestimable because there are no common physical origins. Ranges are seed ranges, not intervals. A missing contrast is not zero.

Acquisition	Source	Origins	Charge U-Net	Time U-Net	Direct U-Net	CI
Capacity: matched minus primary, $\Delta_{\mathrm{matched}} = 0.16$, identical physical targets						
B C/2	C/20	14 [9,5]	[0.4932, 0.5695]	[0.0893, 0.1765]	[0.4393, 0.6118]	B
	C/2	14 [9,5]	[0.4826, 0.5800]	[0.4577, 0.6488]	[0.3116, 0.3873]	B
B formation	C/20	4 [3,1]	[−0.2201, 0.6512]	[0.0379, 0.2846]	[−0.0246, 0.0290]	B
	C/2	4 [3,1]	[−0.0839, 0.9657]	[0.9734, 1.6215]	[0.0069, 0.0657]	B
C formation 1	C/20	14 [11,3]	[0.0034, 0.1489]	[0.0041, 0.0456]	[0.0048, 0.0081]	D
	C/2	14 [11,3]	[−0.0045, 0.1918]	[−0.1465, 1.8520]	[0.0057, 0.0105]	D
C formation 2	C/20	1 [1]	[−0.0228, 0.0267]	[−0.0113, −0.0013]	[−0.0285, −0.0192]	B
	C/2	1 [1]	[−0.0139, 0.0100]	[−4.4531, 1.3891]	[−0.0292, −0.0201]	B
Timing: lag-shifted minus unshifted, $\Delta_{\mathrm{primary}} = 0.16$, common physical origins						
B C/2	C/20	84 [45,39]	[−0.0128, −0.0044]	[−0.0026, −0.0005]	[−0.0076, −0.0019]	D
	C/2	84 [45,39]	[−0.0075, −0.0045]	[−0.0079, −0.0054]	[−0.0059, −0.0052]	D
C formation 1	C/20	97 [71,26]†	[0.0002, 0.0007]	[0.0002, 0.0006]	[0.0001, 0.0005]	D
	C/2	97 [71,26]†	[−0.00003, 0.0004]	[−0.0078, 0.0201]	[0.0002, 0.0003]	D
B formation	Both	0	n/e	n/e	n/e	O
C formation 2	Both	0	n/e	n/e	n/e	O

†The time-clock U-Net has 99 origins [72,27] on C formation 1; the charge and direct U-Nets have 97 [71,26]. "Both" denotes both training sources.

Table 94: Controlled clock comparison, second NMC electrode. As Table 1: charge-clock skill (seed mean$_{\pm \mathrm{SD}}$, seeds 0–2) and charge minus the named clock on identical windows and common seeds, with a 95% paired particle-cluster interval (positive favours charge). Time step 12 s (the charge per step of the NMC 30 s grid at the slowest and fastest rates); frame clock on the single-rate splits only; (d, s, r) on the interpolation split only. Second NMC electrode (NMC-2): one field, five test particles; its intervals are descriptive (five or fewer particles). n/r: not run; n/e: not evaluated.

NMC-2	NMC-2: 0.5C →1C,2C,3C		NMC-2: 3C→0.5C		NMC-2: (0.5C+3C) →1C,2C	
Δ	.02	.16	.02	.16	.02	.16
Charge (d, s) skill	$.57_{\pm.06}$	$.74_{\pm.14}$	$-1.03_{\pm1.14}$	$-1.09_{\pm1.64}$	$.75_{\pm.02}$	$.85_{\pm.01}$
vs time	+.08 [-.13,.17]	+.22 [-.21,.39]	+19.9 [11.9,26.2]	+34.0 [27.9,45.3]	+.07 [.01,.10]	+.15 [.07,.18]
vs frame	−.03 [-.19,.04]	+.12 [-.25,.26]	+49.8 [37.6,79.6]	+310.2 [231.5,487.7]	n/r	n/r
Charge (d, s, r) skill	n/r	n/r	n/r	n/r	$.71_{\pm.05}$	$.80_{\pm.02}$
vs time	n/r	n/r	n/r	n/r	−.01 [-.03,.00]	−.03 [-.08,-.01]

Table 95: Second NMC electrode against strong baselines. As the NMC panel of the main table (same rows, windows rule, best-baseline rule and marks). Second NMC electrode (NMC-2): one field, five test particles; its intervals are descriptive (five or fewer particles). n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

NMC-2	NMC-2: 0.5C →1C,2C,3C		NMC-2: 3C→0.5C		NMC-2: (0.5C+3C) →1C,2C	
Δ	.02	.16	.02	.16	.02	.16
Persistence	.00	.00	.00	.00	.00	.00
Trend	−.85	−.36	−.26	−.14	−.78	−.13
AR(2)	.60	.87	.58	.80	.60	.86
SOC replay	.36	.73	.12	.08	.43	.74
History-aligned replay	.20	.80	.09	.15	.39	.73
Population replay	.41	.70	.12	.07	.46	.81
U-Net, autoregressive (d, s)	$.57_{\pm.06}$	$.74_{\pm.14}$	$-1.03_{\pm1.14}$	$-1.09_{\pm1.64}$	$.75_{\pm.02}$	$.85_{\pm.01}$
U-Net, direct lead-time (d, s)	$.20_{\pm.28}$	$.41_{\pm.57}$	$-.18_{\pm.11}$	$.55_{\pm.28}$	n/r	n/r
U-Net − best baseline	$-.03^{\mathrm{A}}$ [-.19,.03]	$-.13^{\mathrm{A}}$ [-.42,-.02]	-1.61^{A} [-2.18,-1.25]	-1.89^{A} [-2.20,-1.68]	$+.15^{\mathrm{A}}$ [.11,.24]	$-.02^{\mathrm{L}}$ [-.10,.05]
Direct U-Net − best baseline	$-.40^{\mathrm{A}}$ [-.48,-.28]	$-.46^{\mathrm{A}}$ [-.51,-.40]	$-.76^{\mathrm{A}}$ [-.91,-.53]	$-.25^{\mathrm{A}}$ [-.29,-.15]	n/r	n/r

Table 96: A7 paired differences on identical origins (and identical physical targets for the capacity rows, cell B at C/2; timing rows: each acquisition, targets shifted), $\Delta = 0.16$: range over seeds and the number of seeds whose descriptive interval excludes zero (+ above / − below). Source: results/xcell_a7/paired_contrasts.csv.

Contrast	Source	Charge U-Net	Time U-Net	Direct U-Net
Matched − primary capacity, B C/2	C/20	$+0.46$ to $+0.55$ (3+)	$+0.10$ to $+0.14$ (3+)	$+0.44$ to $+0.61$ (2+)
	C/2	$+0.42$ to $+0.54$ (3+)	$+0.47$ to $+0.62$ (3+)	$+0.32$ to $+0.38$ (2+)
Charge − time, B C/2, primary	C/20	-0.15 to $+0.01$ (0)		
Charge − time, B C/2, matched	C/20	$+0.22$ to $+0.45$ (2+)		
Timing-shifted − unshifted, 4 acquisitions	both	≤ 0.019 in magnitude	≤ 0.11	≤ 0.022

Table 97: A7 dependence sensitivity on B C/2 at the matched-capacity $\Delta = 0.16$ physical horizon. Horizon-spanning blocks contain 59 origins (median span about 1740 s; maximum forecast duration 1712 s). Point ranges and interval envelopes are across seeds, not intervals for a seed mean. Positive counts give intervals wholly above zero for 10 / horizon-spanning / 60 / 120 origins / whole half-cycles, respectively (three seeds for AR and C−T, two for direct; 1,000 paired resamples, seed 0). C−T means charge minus time. HC signs count positive/negative half-cycle differences across seeds. Whole-half-cycle intervals use only two blocks; individual half-cycle intervals for that method are non-estimable (NE). Capacity means matched minus primary; these rows have identical physical targets. All acquisitions, four horizons, timing contrasts (potentially different targets), failures and NE cases remain in the released sensitivity results. These are descriptive single-flake checks.

Train	Contrast	Point seed range	Horizon block 95% envelope	Positive intervals 10/H/60/120/HC	HC signs positive/negative
C/20	Capacity, AR	[+0.459, +0.549]	[+0.325, +0.829]	3/3/3/3/3	6+/0-
C/20	Capacity, direct	[+0.442, +0.606]	[+0.309, +0.915]	2/2/2/2/2	4+/0-
C/20	C−T, primary	[-0.150, +0.006]	[-0.587, +0.262]	0/0/0/0/0	3+/3-
C/20	C−T, matched	[+0.220, +0.445]	[-0.286, +0.620]	2/2/2/2/3	6+/0-
C/2	Capacity, AR	[+0.417, +0.538]	[+0.160, +0.924]	3/3/3/3/3	6+/0-
C/2	Capacity, direct	[+0.323, +0.379]	[+0.076, +0.818]	2/2/2/2/2	4+/0-
C/2	C−T, primary	[-0.021, +0.202]	[-0.135, +0.399]	1/1/1/1/1	4+/2-
C/2	C−T, matched	[-0.033, +0.128]	[-0.183, +0.315]	1/0/0/1/1	3+/3-

Table 98: NMC811 cycle sensitivity from saved errors, with three autoregressive (AR) seeds and two direct seeds averaged at each window. Top: each test cycle on common windows; Best is the strongest of 11 saved baselines on that cycle and horizon (retrospective conservative selection); signed differences give model minus Best. Bottom: pooled, equal-cycle and leave-one-cycle-out (LOCO) differences keep the full-split strongest baseline fixed (A: AR(2); L: particle-mean ridge). LOCO ranges are four three-cycle pooled estimates, not confidence intervals. Per-cycle signs against the fixed comparator and reselected-baseline LOCO results are also in the released cycle-sensitivity results. All cycles are from one cell; age and voltage cutoff remain confounded.

Split	Cycle	Δ	AR	Direct	Best	AR−Best	Direct−Best
Ageing	99	0.02	0.588	0.526	0.414[A]	+0.174	+0.113
Ageing	100	0.02	0.614	0.549	0.415[A]	+0.199	+0.135
Ageing	116	0.02	0.603	0.525	0.402[A]	+0.201	+0.123
Ageing	117	0.02	0.647	0.561	0.423[A]	+0.224	+0.138
Ageing	99	0.16	0.820	0.843	0.852[R]	-0.032	-0.009
Ageing	100	0.16	0.858	0.875	0.828[L]	+0.030	+0.047
Ageing	116	0.16	0.891	0.916	0.883[L]	+0.009	+0.034
Ageing	117	0.16	0.892	0.910	0.873[L]	+0.020	+0.038
High voltage	127	0.02	0.314	0.076	0.274[A]	+0.040	-0.199
High voltage	128	0.02	0.290	0.005	0.309[A]	-0.020	-0.304
High voltage	157	0.02	0.293	-0.165	0.315[A]	-0.021	-0.480
High voltage	158	0.02	0.328	-0.041	0.311[A]	+0.017	-0.353
High voltage	127	0.16	0.583	0.637	0.561[L]	+0.022	+0.076
High voltage	128	0.16	0.577	0.606	0.553[A]	+0.024	+0.053
High voltage	157	0.16	0.108	0.264	0.525[A]	-0.418	-0.262
High voltage	158	0.16	0.160	0.323	0.536[A]	-0.376	-0.213

Split	Δ	Model	Fixed best	Pooled	Equal-cycle	LOCO range
Ageing	0.02	AR	A	+0.200	+0.200	[+0.191, +0.209]
Ageing	0.02	Direct	A	+0.128	+0.127	[+0.123, +0.132]
Ageing	0.16	AR	L	+0.018	+0.018	[+0.014, +0.020]
Ageing	0.16	Direct	L	+0.038	+0.038	[+0.035, +0.040]
High voltage	0.02	AR	A	+0.006	+0.004	[-0.008, +0.015]
High voltage	0.02	Direct	A	-0.322	-0.334	[-0.374, -0.278]
High voltage	0.16	AR	A	-0.108	-0.167	[-0.196, -0.037]
High voltage	0.16	Direct	A	-0.023	-0.066	[-0.099, +0.031]

A: AR(2); L: particle-mean ridge; R: SOC replay.

Table 99: Cross-material check. Exploratory. Autoregressive charge-clock U-Nets with conditioning (d, s), $k = 4$, the standard recipe and seeds 0–1 on each material's rate split. Applied unchanged: the other material's trained model evaluated on this material's test windows (frames are log contrast in both datasets; charge is scaled by each dataset's Q_{ref}). Initialised: trained on this material starting from the other material's weights. Skill: seed mean ± SD on identical windows; differences from the model trained on this material with 95% paired particle-cluster bootstrap intervals (2,000 resamples). Two seeds: descriptive. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Test split	Forecaster	$\Delta = 0.02$	$\Delta = 0.16$	− native, 0.02	− native, 0.16
NMC, C/3→1C,2C	Trained on this material	0.792 ± 0.034	0.882 ± 0.030	–	–
	Trained on graphite, applied unchanged	-16.292 ± 18.134	-41.080 ± 48.669	-17.084 [-25.805, -14.049]	-41.961 [-73.774, -32.416]
	Initialised from graphite, trained on this material	0.803 ± 0.000	0.887 ± 0.000	+0.012 [-0.008, 0.020]	+0.006 [-0.008, 0.012]
Graphite, C/20→C/2	Trained on this material	0.516 ± 0.009	0.666 ± 0.049	–	–
	Trained on NMC, applied unchanged	-0.037 ± 0.016	-0.145 ± 0.045	-0.553 [-0.589, -0.521]	-0.811 [-0.853, -0.729]
	Initialised from NMC, trained on this material	0.519 ± 0.009	0.662 ± 0.001	+0.003 [-0.003, 0.008]	-0.004 [-0.057, 0.017]

G REPRODUCIBILITY OF THE OPTICAL SIGNATURES

This appendix measures how optical signatures agree across acquisitions.

How reproducible is the physics that a forecaster has to learn? We measure agreement between repeated observations of the same particles at three levels: the electrode-level optical marker of a staging transition in Coulomb-counted charge (L1), the order in which particles pass that marker (L2), and the within-particle pattern of optical switching (L3). The same quantities, computed between a model rollout and the observation, place the models against *repeat agreement* (App. G.8). The analysis is retrospective and exploratory; it re-tests earlier descriptions of these movies instead of inheriting them. All statements are associations within one electrode per material: acquisition order, cycling history, cut-off voltage and imaging are not varied independently of rate and temperature, and pairs of half-cycles are dependent, not replicates (Hurlbert, 1984).

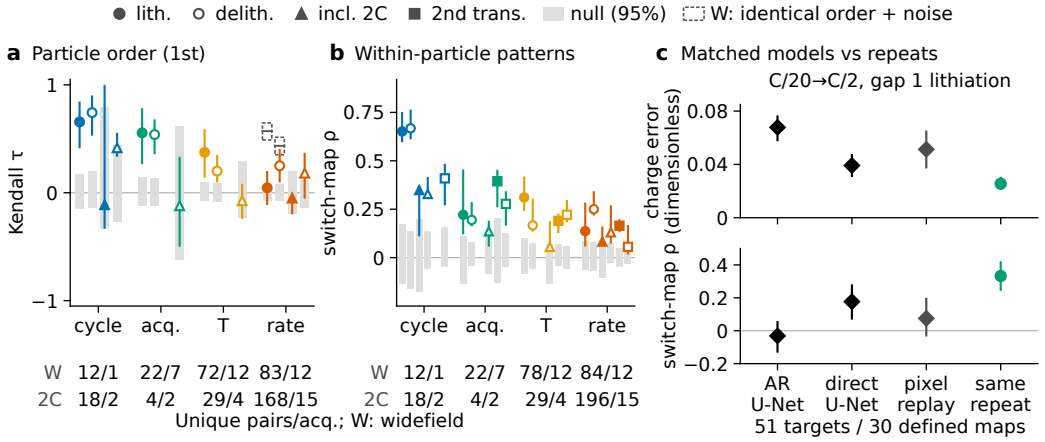

Figure 4: **Transfer errors and repeat agreement, full version of Fig. 3.** (a) Agreement of particle order at the first transition (Kendall τ_b rank correlation) and (b) agreement of switch maps (Spearman ρ) between pairs of recordings: another cycle in the same acquisition; another acquisition at the same rate and temperature; another temperature; another rate. Circles: widefield pairs; triangles: with a magnified 2C acquisition; squares: second transition; filled/open: lithiation/delithiation; counts: pairs/acquisitions (pairs share recordings). Grey: central 95% range with no agreement; dashed: order expected from marker noise alone. (c) The same 51 timing targets (19 particles; 30 with defined maps), C/20→C/2: timing error in dimensionless charge and switch-map correlation for the autoregressive and direct U-Nets, cross-rate pixel replay and same-condition repeats; bars: 95% particle-bootstrap CIs (Appendix H.2).

G.1 DATA, COMPARISON CATEGORIES AND UNCERTAINTY

We specify the comparison support and how uncertainty is estimated. The graphite analysis uses 258 sequences from 17 acquisitions (formation and C/20 to 2C), the NMC analysis the 72 sequences of the 24 accepted identities; 839 constant-current segments are extracted. The 2C acquisitions contain only the central flake and two neighbours, imaged in the magnified view, and cannot represent the population. We use the benchmark signal (background-subtracted log contrast), the stored masks and the stored quality flags (quality-1 frames retained, quality-2 frames missing), and the charge state $s = q/Q_{\mathrm{ref}}$ without per-half-cycle normalisation, calibration or fitted offsets. Traces are sampled every 0.005 in s only where the benchmark sampler reports valid data; gaps are never bridged.

Every comparison pairs two half-cycles of the same direction and transition and falls in one of four mutually exclusive categories: the same acquisition (another cycle); another acquisition at the same nominal rate and temperature; the same rate at another temperature; and another rate. "Same rate and temperature" is a label, not an identical protocol (for instance, one C/2, 25 °C repeat has a 1 mV and the other two a 5 mV lower cut-off; Table 3). The formation cycle and an ambiguous second constant-current bout in one 2C acquisition are excluded from the headline estimates but kept in the coverage records. Intervals are 95% percentile bootstraps over physical particle identities

(2,000 draws), resampling each particle coherently across its cycles and pairs; they are conditional on these acquisitions and this electrode and are not uncertainty for a new electrode or for a causal effect of rate. We report the number of distinct acquisitions separately from the number of pairs, and a leave-one-acquisition-out (LOAO) range that removes every pair touching one acquisition; undefined values after omission are retained. An acquisition-balanced alternative first averages all cycles of each acquisition pair.

G.2 Optical markers and switch maps

We define the optical markers and maps used to compare acquisitions. Markers use the benchmark's long-horizon extraction with its smoothing and peak parameters (App. C); in addition, maxima within 0.015 of an observed block boundary are unresolved. Staging is assigned separately for each direction from the area-weighted population trace, before any comparison. In increasing s, the optical minimum precedes the strongest positive slope, which is labelled gap 1 (stage 2L↔2). A later derivative valley at least 0.06 after this peak and below 25% of its slope separates gap 2 (2↔1), but only if a later interior peak reaches 10% of the first slope and the remaining optical rise reaches 10% of the block's optical range; otherwise gap 2 is undefined. A particle may begin gap 1 before the population minimum, so its search extends to lower s, excludes dimming (negative-slope) candidates and stops at the population divider. This is a documented staging filter on the benchmark's optical candidates, not an independent identification of phases; every resolved population marker coincides with the unfiltered dominant marker on its interval.

A switch map assigns to each pixel the charge of its maximum absolute smoothed derivative within a resolved transition interval of at least 13 samples (0.06 in charge; gap 1 starts at the particle's preceding optical minimum), provided the derivative exceeds the benchmark's activity tolerance. It is a map of maximal optical change, not a nucleation time. Two maps are compared by the Spearman correlation over their finite shared pixels in the stored crop coordinates, without any fitted registration; a particle's weight is the geometric mean of its two projected areas (not a capacity), and half-cycle pairs are weighted equally. Of 327 attempted graphite particle-half-cycle transitions, 231 (lithiation) and 266 (delithiation) gap-1 markers and 126 and 150 gap-2 markers are resolved; for NMC, 63 and 31 of 72.

G.3 L1: optical-marker alignment in Coulomb-counted charge

We test how optical markers align in Coulomb-counted charge. L1 is an empirical property of the optical marker: charge conservation constrains the total stored charge, not the charge at which an optical marker occurs, and s is not a calibrated particle composition. Each L1 pair uses the same physical particles and exactly the same finite charge samples on both acquisitions, with fixed weights (equal, or the geometric mean of the two projected areas); the cohort never changes along a trace. The shared charge block is chosen by pooled optical response strength inside the union of the declared transition intervals, not by how well the markers agree, and both markers must lie inside their own interval. A pair needs at least three fully observed common particles and 0.06 of shared charge support. Table 100 reports gap 1. Within an acquisition and between same-rate repeats, lithiation markers align to about 0.015–0.024 in s; across temperatures and rates the mean offset is 0.030–0.046. With area weights, delithiation markers align less well than lithiation markers in every category.

Pairwise alignment does not establish a fixed marker for the whole electrode. A stricter audit uses one cohort, one charge block and one weight vector across a whole set of acquisitions: across all 16 regular graphite acquisitions only three identities are common and their shared lithiation support ends at $s = 0.258$, below every gap-1 event, and the 12 wide-field acquisitions share support only to $s = 0.327$. The earlier claim that gap 1 occurs at one fixed electrode charge state at every rate and temperature is therefore untestable on identical support and *not confirmed*. On the largest fixed cohort that does span gap 1 (13 particles, the three C/2, 25 °C acquisitions, common support 0.35–0.475), the area-weighted markers are 0.390, 0.415 and 0.425 (equal weights 0.375, 0.410, 0.405). Charge offsets are not calibrated. The formation acquisition does not start delithiated (raw gap-1 lithiation marker $s = 0.175$) and is excluded; a quasi-equilibrium voltage reference suggests a translation of about 0.28, which is a sensitivity, not a measured offset. No uncertainty budget exists for the initial charge anchor of the other acquisitions. Assuming independent acquisition

Table 100: L1: mean absolute difference of the gap-1 population optical marker between two half-cycles, on identical particles and charge support (95% particle-bootstrap intervals). A/pairs/attempted: distinct acquisitions, eligible half-cycle pairs and attempted pairs (area weights). LOAO: range of the area-weighted estimate when one acquisition is left out.

Direction	Category	$\lvert\Delta s\rvert$, area weights	$\lvert\Delta s\rvert$, equal weights	A/pairs/attempted	LOAO
Lith.	Same acq.	0.015 [0.009, 0.025]	0.017 [0.010, 0.037]	2/12/25	0.005–0.026
Lith.	Same rate, T	0.018 [0.012, 0.036]	0.024 [0.017, 0.042]	3/9/20	0.010–0.025
Lith.	Same rate, other T	0.036 [0.030, 0.058]	0.046 [0.028, 0.053]	11/18/105	0.023–0.050
Lith.	Other rate	0.030 [0.027, 0.052]	0.038 [0.032, 0.050]	10/57/256	0.027–0.034
Delith.	Same acq.	0.020 [0.005, 0.035]	0.036 [0.006, 0.039]	3/15/25	0.002–0.032
Delith.	Same rate, T	0.059 [0.049, 0.081]	0.049 [0.031, 0.069]	7/10/20	0.049–0.086
Delith.	Same rate, other T	0.044 [0.041, 0.066]	0.042 [0.035, 0.062]	13/29/105	0.035–0.055
Delith.	Other rate	0.053 [0.039, 0.065]	0.054 [0.037, 0.060]	13/94/256	0.043–0.058

offsets uniform on ±0.01 or ±0.03 (shared by all cycles of an acquisition), the across-rate gap-1 lithiation alignment becomes 0.031 [0.026, 0.037] and 0.036 [0.023, 0.053] (area weights); these are sensitivity intervals, not measurement intervals, and same-acquisition differences cancel a constant offset.

G.4 L2: PARTICLE ORDER, CONCURRENCY AND MARKER NOISE

We compare particle order and concurrency against marker-noise controls.

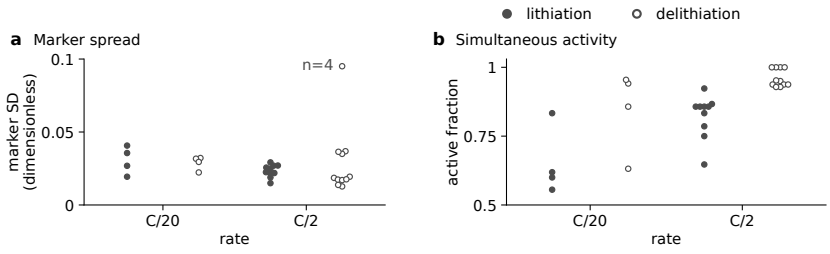

Figure 5: **Concurrency of the first optical transition.** Across-particle marker SD in dimensionless charge and the maximum fraction of a fixed cohort active together at C/20 and C/2; each point is an acquisition half-cycle, filled/open for lithiation/delithiation, with resolved-particle counts printed below ten; formation, the ancillary bout and 2C are excluded.

L2 compares the order of the particles' marker charges between two half-cycles by Kendall's τ_b (ties averaged); the null is 1,000 coherent permutations of particle identities (Table 101). For gap-1 lithiation, order agrees between same-rate repeats ($\tau_b = 0.555$ [0.265, 0.784]) and is at the null across rates (-0.006 [-0.113, 0.183]). Excluding the three-particle 2C acquisitions, the within-acquisition agreement rises to 0.657 [0.412, 0.845] (lithiation) and 0.743 [0.529, 0.901] (delithiation), but that category is then one acquisition, and the across-rate lithiation value is 0.045 [-0.114, 0.202]; restricted to comparisons at the same temperature, it is -0.110 [-0.298, 0.156]. Order is *not* a monotone function of condition distance in every cell: for gap-1 delithiation the same-rate, other-temperature agreement (0.082) is lower than the across-rate agreement (0.200).

Marker noise. A low τ_b alone does not show that the order changed: markers that are nearly simultaneous and noisy reshuffle ranks even if the underlying order is identical. We therefore simulate the τ_b expected under an *identical* latent order. Marker noise is estimated from same-rate, same-temperature acquisition differences after removing each pair's median charge shift (RMS difference$/\sqrt{2}$); this contains real repeat variation as well as measurement error. For gap-1 lithiation it is $\sigma = 0.0118$ at C/20 and 0.0125 at C/2. Missing rate-specific estimates (notably 2C lithiation) use the mean for that transition and direction. A shared latent rank profile is scaled so that signal plus noise variance matches each half-cycle's observed marker spread (negative residual variances truncated at zero); Gaussian noise is redrawn per half-cycle and reused for all of its pairs; charges are rounded to 0.005; 2,000 draws; noise multipliers 0.5, 1 and 1.5. A stricter control re-estimates the noise amplitudes and the latent profile after removing each acquisition. Table 102 shows that

Table 101: L2: Kendall τ_b of particle marker order between two half-cycles (mean over pairs; 95% particle-bootstrap interval). A/pairs/particles: distinct acquisitions, half-cycle pairs, physical particles. Null: central 95% of the identity-permutation null. LOAO: leave-one-acquisition-out range. Includes the 2C acquisitions.

Transition	Direction	Category	τ_b	A/pairs/particles	Null	LOAO
Gap 1	Lith.	Same acq.	0.273 [0.095, 0.849]	2/12/14	[−0.192, 0.338]	−0.111–0.657
Gap 1	Lith.	Same rate, T	0.555 [0.265, 0.784]	7/11/21	[−0.121, 0.140]	0.488–0.725
Gap 1	Lith.	Same rate, other T	0.375 [0.141, 0.590]	11/34/22	[−0.079, 0.094]	0.288–0.457
Gap 1	Lith.	Other rate	−0.006 [−0.113, 0.183]	12/88/22	[−0.115, 0.125]	−0.038–0.045
Gap 1	Delith.	Same acq.	0.525 [0.438, 0.876]	3/18/14	[−0.195, 0.313]	0.288–0.872
Gap 1	Delith.	Same rate, T	0.363 [0.240, 0.632]	9/15/20	[−0.181, 0.194]	0.151–0.539
Gap 1	Delith.	Same rate, other T	0.082 [−0.002, 0.336]	16/67/21	[−0.109, 0.128]	0.021–0.170
Gap 1	Delith.	Other rate	0.200 [0.024, 0.363]	16/163/22	[−0.094, 0.091]	0.127–0.315
Gap 2	Lith.	Same acq.	undefined	0/0/0	–	–
Gap 2	Lith.	Same rate, T	0.667 [0.167, 0.942]	2/1/15	[−0.417, 0.417]	0.667–0.667
Gap 2	Lith.	Same rate, other T	0.310 [0.133, 0.476]	7/8/22	[−0.137, 0.150]	0.267–0.402
Gap 2	Lith.	Other rate	0.192 [−0.050, 0.432]	7/12/22	[−0.116, 0.116]	0.119–0.328
Gap 2	Delith.	Same acq.	0.405 [−0.080, 0.810]	1/3/10	[−0.286, 0.391]	0.405–0.405
Gap 2	Delith.	Same rate, T	−0.202 [−0.528, 0.694]	4/4/14	[−0.348, 0.338]	−0.477–0.625
Gap 2	Delith.	Same rate, other T	0.392 [0.201, 0.600]	8/14/22	[−0.140, 0.152]	0.331–0.472
Gap 2	Delith.	Other rate	0.274 [0.114, 0.424]	8/24/20	[−0.097, 0.105]	0.213–0.359

gap-1 and gap-2 lithiation orders across rates remain below the identical-order expectation even at 1.5 times the estimated noise, for most or all refitted omissions, whereas delithiation does not. "Reshuffled across rates" is therefore supported, conditionally, for lithiation only. This is a declared sensitivity model, not a calibrated generative model of the measurement.

Table 102: Observed across-rate τ_b against the τ_b expected if the particle order were identical at both rates and markers carried the repeat-derived noise (mean and central 95% of 2,000 simulations; charge rounding 0.005). Below: refitted leave-one-acquisition-out omissions whose observed τ_b stays below the simulated interval.

		Observed	Noise ×1		Noise ×1.5	
Transition	Direction	τ_b	Expected τ_b	Below	Expected τ_b	Below
Gap 1	Lith.	−0.006	0.315 [0.196, 0.469]	16/16	0.138 [0.023, 0.269]	13/16
Gap 1	Delith.	0.200	0.369 [0.206, 0.541]	7/16	0.086 [−0.034, 0.231]	0/16
Gap 2	Lith.	0.192	0.824 [0.772, 0.872]	14/14 (2 undef.)	0.698 [0.624, 0.768]	14/14 (2 undef.)
Gap 2	Delith.	0.274	0.554 [0.424, 0.673]	14/16	0.283 [0.137, 0.425]	0/16

Concurrency. Concurrency is measured separately from rank concordance. The marker spread is the across-particle standard deviation of the marker charge (interquartile and 10–90% ranges are also stored); the maximum simultaneous active fraction is the largest fraction of a fixed cohort whose optical derivative is positive and above 25% of that particle's transition maximum on shared observed charge samples (the absolute derivative for NMC). This is an optical activity convention, not a synchronisation mechanism. For gap-1 lithiation at 25 °C (Table 103), the marker spread is 0.036 and 0.041 in charge for the two C/20 repeats against 0.022–0.026 for the three C/2 repeats, an order of magnitude less in minutes, and more particles are active at once (0.786–0.857 against 0.556–0.600). At C/20 the spread also depends on temperature (0.019 at 15 °C, where 0.833 of the cohort is active at once), so the rate association rests on the 25 °C repeats. It is compatible with more concurrent conversion at the faster rate, the particle-by-particle to concurrent transition reported for LFP electrodes (Li et al., 2014), which we use as an analogy only; it does not follow from the low τ_b.

The largest particle, the central flake, is not consistently first at C/20 or last at C/2: its gap-1 lithiation rank (1 = first) is 4 of 18 and 4.5 of 15 at C/20 and 25 °C, 14.5 of 18, 11 of 14 and 8.5 of 14 at C/2 and 25 °C, and ranges from 1 of 21 to 12 of 12 across temperatures at C/20.

G.5 L3: WITHIN-PARTICLE SWITCH-MAP AGREEMENT

We compare within-particle switching patterns with spatial null models. Table 104 gives the area-weighted switch-map correlation. Three nulls are reported: a rotation or flip of the second map

Table 103: Gap-1 lithiation concurrency by acquisition (first cycle): resolved particles, across-particle SD of the marker charge (95% particle-bootstrap interval) and in minutes, and maximum fraction of a fixed cohort active at once. Acquisitions without a resolved marker are omitted.

Acquisition	Rate, T	Particles	Marker SD (s)	Marker SD (min)	Max. active fraction
03	C/20, 25 °C	18	0.036 [0.026, 0.041]	46.6	0.556 [0.444, 0.778]
04	C/20, 45 °C	21	0.027 [0.019, 0.032]	35.1	0.619 [0.476, 0.810]
05	C/20, 15 °C	12	0.019 [0.012, 0.023]	25.3	0.833 [0.667, 1.000]
06	C/20, 25 °C	15	0.041 [0.026, 0.048]	53.2	0.600 [0.532, 0.867]
07	C/2, 15 °C	13	0.015 [0.009, 0.019]	1.9	0.923 [0.846, 1.000]
08	C/2, 15 °C	15	0.019 [0.012, 0.024]	2.4	0.867 [0.800, 1.000]
10	C/2, 25 °C	18	0.022 [0.013, 0.029]	2.8	0.833 [0.722, 1.000]
11	C/2, 25 °C	14	0.022 [0.015, 0.026]	2.8	0.786 [0.714, 1.000]
12	C/2, 35 °C	16	0.027 [0.019, 0.032]	3.5	0.750 [0.625, 0.938]
13	C/2, 45 °C	17	0.029 [0.019, 0.037]	3.8	0.647 [0.529, 0.882]
14	C/2, 25 °C	14	0.026 [0.016, 0.031]	3.3	0.857 [0.714, 1.000]
17	2C, 45 °C	3	0.003 [0.000, 0.003]	0.2	1.000 [1.000, 1.000]

within its mask (the seven non-identity square symmetries), and two nulls that preserve the exact mask geometry and the multiset of switch values: a smooth Gaussian random field whose width is fitted, independently of pair agreement, to each map's rank autocorrelation at lags 1, 2, 3, 4, 6 and 8 pixels (12 calibration fields per width, 200 fields per map), and the same field ranked separately within boundary-depth bands ($\leq$ 2, 2–4, 4–8 and $>$ 8 pixels), which preserves the coarse edge-associated distribution of switch values. The smooth-field fit identifies the autocorrelation of 723 of 731 maps, with median autocorrelation RMSE 0.147 (0.138 for the depth-stratified field) and 73.7% (75.0%) of maps within the diagnostic tolerance of 0.2; poor fits are retained, so exceeding these nulls is not a complete validation of the metric.

Table 104: L3: Spearman correlation of within-particle switch maps between two half-cycles (projected-area-weighted particle mean, equal weight per pair; 95% particle-bootstrap interval). Nulls: central 95% of the rotation/flip null and of the boundary-depth-stratified smooth-field null (the unstratified smooth-field null is centred on zero; see text). LOAO: leave-one-acquisition-out range.

Transition	Direction	Category	ρ	A/pairs/particles	Rotation/flip	Depth-stratified	LOAO
Gap 1	Lith.	Same acq.	0.500 [0.407, 0.719]	2/12/14	[−0.112, 0.128]	[−0.016, 0.147]	0.349–0.652
Gap 1	Lith.	Same rate, T	0.221 [0.120, 0.455]	7/11/20	[−0.137, 0.109]	[−0.040, 0.096]	0.197–0.363
Gap 1	Lith.	Same rate, other T	0.311 [0.241, 0.419]	11/34/22	[−0.085, 0.097]	[−0.004, 0.080]	0.267–0.338
Gap 1	Lith.	Other rate	0.105 [0.073, 0.214]	12/96/22	[−0.077, 0.073]	[0.002, 0.069]	0.076–0.137
Gap 1	Delith.	Same acq.	0.442 [0.419, 0.663]	3/18/14	[−0.011, 0.086]	[0.050, 0.125]	0.328–0.590
Gap 1	Delith.	Same rate, T	0.179 [0.152, 0.253]	9/15/20	[−0.020, 0.060]	[−0.006, 0.068]	0.133–0.273
Gap 1	Delith.	Same rate, other T	0.122 [0.098, 0.254]	16/73/21	[−0.061, 0.044]	[0.015, 0.059]	0.094–0.170
Gap 1	Delith.	Other rate	0.159 [0.115, 0.278]	16/184/22	[−0.026, 0.045]	[0.026, 0.067]	0.134–0.201
Gap 2	Lith.	Same acq.	undefined	0/0/0	–	–	–
Gap 2	Lith.	Same rate, T	0.395 [0.260, 0.454]	2/1/15	[−0.078, 0.204]	[−0.094, 0.170]	undefined
Gap 2	Lith.	Same rate, other T	0.190 [0.126, 0.228]	7/8/22	[−0.042, 0.066]	[−0.030, 0.051]	0.130–0.254
Gap 2	Lith.	Other rate	0.165 [0.131, 0.198]	7/12/22	[−0.031, 0.045]	[−0.039, 0.048]	0.110–0.271
Gap 2	Delith.	Same acq.	0.409 [0.270, 0.485]	1/3/10	[−0.045, 0.153]	[−0.033, 0.072]	undefined
Gap 2	Delith.	Same rate, T	0.278 [0.165, 0.343]	4/4/14	[−0.020, 0.127]	[−0.027, 0.060]	0.195–0.527
Gap 2	Delith.	Same rate, other T	0.221 [0.182, 0.297]	8/14/22	[−0.056, 0.057]	[−0.010, 0.060]	0.152–0.355
Gap 2	Delith.	Other rate	0.056 [0.017, 0.168]	8/24/20	[−0.032, 0.046]	[−0.008, 0.043]	0.006–0.186

Patterns agree best within an acquisition, but positive agreement across acquisitions is part of the result: gap-1 lithiation maps of same-rate repeats correlate at 0.221 [0.120, 0.455], and the single eligible gap-2 lithiation repeat pair at 0.395 [0.260, 0.454]. Excluding 2C, within-acquisition gap-1 agreement is 0.652 [0.596, 0.752] (lithiation) and 0.668 [0.611, 0.764] (delithiation), again from one acquisition. Across rates, gap-1 lithiation agreement is lowest (0.105), and 0.012 [-0.044, 0.245] when the temperature is matched. The categories do not form a monotone ladder: same-rate, other-temperature gap-1 lithiation agreement (0.311) exceeds the same-rate repeat value. An earlier whole-block definition (the longest jointly observable block after $u = 0.14$ of at least 0.10 in charge, median over particles, pixel-shuffle null) gives 0.136 for the two C/20, 25 °C repeats and 0.545 and 0.551 for two pairs of C/2, 25 °C repeats on lithiation; we reproduce these values exactly (maximum reconstruction error $< 4 \times 10^{-16}$). They differ from the transition-specific values because intervals, particle coverage and aggregation differ; the two definitions do not estimate one identical target.

Early optical change is enriched near the boundary for delithiation and for gap-2 lithiation: the fraction of rim pixels (within two pixels of the mask boundary) among the earliest 10% of switch ranks, divided by the rim's area fraction, is 1.555 [1.106, 1.944] and 1.430 [1.155, 1.710] for gap-1 and gap-2 delithiation and 1.352 [1.126, 1.548] for gap-2 lithiation, whose mean core–rim lag nevertheless overlaps zero. Point spread, projection, segmentation and spatially varying contrast can all create boundary associations, so this is not evidence of edge nucleation.

G.6 Robustness and acquisition weighting

We check sensitivity to extraction choices, registration and acquisition weighting. Smoothing width (0.015, 0.06), candidate height (0.125, 0.5 of the maximum) and candidate separation (0.015, 0.06) are varied one at a time with the population windows fixed, and the second map is shifted by each of the eight integer offsets within ± 1 pixel without wrap-around. For gap-1 lithiation the same-rate τ_b stays within 0.555–0.597 and the across-rate τ_b within -0.006–0.022; including the registration shifts, the same-rate ρ stays within 0.184–0.236 and the across-rate ρ within 0.078–0.129. Missed and extra events relative to the baseline candidate list are tabulated for every variant; they are not ground-truth recall or precision. Weighting acquisition pairs equally instead of half-cycle pairs changes gap-1 lithiation τ_b from 0.555 to 0.632 (same rate) and from -0.006 to 0.025 (other rate), and ρ from 0.221 to 0.285 and from 0.105 to 0.122.

G.7 Hysteresis

We quantify signed marker differences between lithiation and delithiation. Matching each particle's resolved lithiation and delithiation markers within an acquisition and cycle, the signed gap-1 difference (lithiation minus delithiation) at 25 °C is -0.026 [-0.043, -0.011] at C/20 and -0.011 [-0.031, 0.011] at C/2, while at C/2 and 15 °C it is $+0.068$ [0.056, 0.078]. Signs and magnitudes vary with rate and temperature, with one to three acquisitions per condition. A nonzero, rate-dependent optical-marker gap cannot separate kinetics from composition, optical path, cut-off, ageing or acquisition effects; the earlier claim that the hysteresis is kinetic rather than compositional is *not confirmed*.

G.8 Models against repeat agreement

We compare forecast errors with repeat agreement on identical targets. For a model rollout, the marker error and switch-map correlation are computed against the observed target; for the empirical repeat, the same quantities are computed between the target and another acquisition of the same condition. Both sides use identical targets: the same particles, pixels, direction, starting charge, horizon and every sample of the full half-cycle rollout. A repeat acquisition is eligible only if it covers every target charge sample, in the same direction and cycle, on the target's own mask; no rollout is shortened to make a repeat eligible. Each target's saved marker is reproduced from the native segments to within 2.3×10^{-16} in s. A metric is compared only on targets where it is finite on both sides; reference acquisitions and model seeds are averaged within a target, and target means are compared. A forecaster that conditions on the observed history can outperform another noisy acquisition, and a pairwise repeat difference contains two observation errors, so these values are neither error floors nor correlation bounds, and no "fraction of the achievable" is reported.

Table 105: Models against repeat agreement on identical full-rollout targets: optical-marker charge error (lower is better) and switch-map Spearman correlation, as target means with 95% particle-bootstrap intervals. Targets: eligible marker/correlation targets. Test / repeat acquisitions: 08, 11, 14 / 07, 10, 14 (C/20→C/2, lithiation), 07 / 08 (delithiation), 14 / 10 (same C/2) and 11 / 10, 14 (C/2 without 11). Seeds: two for the (d, s) U-Net, three for the (d, u, s) U-Net; replay is deterministic.

				Marker error (s)		Switch-map ρ	
Split	Forecaster	Dir.	Targets	Model	Repeat	Model	Repeat
C/20→C/2	U-Net (d, s)	Lith.	50/50	0.105 [0.097, 0.113]	0.030 [0.022, 0.040]	0.012 [−0.044, 0.076]	0.397 [0.309, 0.480]
C/20→C/2	U-Net (d, s)	Delith.	20/20	0.026 [0.009, 0.052]	0.022 [0.013, 0.034]	0.066 [−0.042, 0.175]	0.144 [0.042, 0.241]
Same C/2	U-Net (d, u, s)	Lith.	14/14	0.033 [0.024, 0.043]	0.037 [0.027, 0.048]	0.165 [0.034, 0.290]	0.555 [0.441, 0.663]
Same C/2	Pixel replay (s)	Lith.	13/13	0.021 [0.017, 0.025]	0.036 [0.026, 0.048]	0.376 [0.183, 0.539]	0.529 [0.418, 0.634]
C/2, without 11	Pixel replay (s)	Lith.	16/16	0.018 [0.010, 0.028]	0.024 [0.016, 0.033]	0.350 [0.183, 0.513]	0.375 [0.230, 0.517]

At a new rate (C/20→C/2, lithiation) the (d, s) U-Net's switch-map correlation (0.012) lies below the repeat agreement of the same targets (0.397); the transition-specific across-rate agreement of Table 104 (0.105) is a different estimand, computed on transition intervals rather than full rollouts, and is not compared with it here (App. H compares the models, cross-rate replay and repeats on identical transition support). Its marker error (0.105) is about three and a half times the repeat difference (0.030). Within the same rate, the (d, u, s) U-Net matches the repeat marker difference but not the repeat pattern agreement, whereas s-aligned pixel replay has a smaller marker error than the repeat difference and the highest pattern agreement of the listed forecasters. Replay on the same C/2 split loses one of 14 supported targets because its prediction has an undefined marker; both sides of that comparison use the remaining 13. Reading repeat agreement as an upper bound is already contradicted in the three-particle 2C transfer: the state-conditioned U-Net's delithiation switch-map correlation (0.313) exceeds the repeat value (0.278). Of the 154 requested split/model rows, 35 have an exact-target repeat comparison, 27 have no same-condition repeat covering the identical full support (including the same-rate C/20 split and C/2→C/20), and 92 would need full half-cycle rollouts of further models that were not run for this version; only the 35 evaluated rows are reported.

G.9 NMC

We assess which reproducibility comparisons the NMC data can support. The NMC subset has one accepted acquisition at each of C/3, 1C and 2C (nominal 25 °C), so there are no same-condition repeats, repeated cycles or temperature pairs: the corresponding categories, the marker-noise model and repeat agreement are undefined. Numerical derivative maxima of the smooth solid-solution response are not phase transitions and are sensitive to endpoints and glitches. Across rates, the lithiation marker order gives $\tau_b = 0.358$ [0.229, 0.730] (21 particles) and the switch maps $\rho = 0.231$ [0.102, 0.376]; delithiation gives 0.007 [-0.539, 0.666] and -0.025 [-0.145, 0.161] (11 particles). Doubling the smoothing width reduces the resolved lithiation markers from 63 to 22 of 72 and changes the lithiation τ_b to 0.172 (six particles), so we make no physical order claim for NMC.

G.10 Summary of verdicts

We collect the supported and unsupported optical-signature claims. Within this electrode, the lithiation gap-1 optical marker aligns in Coulomb-counted charge to about 0.02 between repeats and 0.03–0.05 across temperature and rate, but a single fixed electrode charge state at all rates and temperatures is not confirmed because no common support spans the event. Particle order repeats between same-rate acquisitions and, for lithiation, changes across rates by more than the estimated marker noise explains; at 25 °C, conversion is more concurrent at C/2 than at C/20 by marker spread and active fraction. Within-particle switch patterns agree best within an acquisition, less but positively across same-rate acquisitions, and least across rates for lithiation, with exceptions retained above. Kinetic rather than compositional hysteresis, and edge nucleation, are not confirmed.

H WHAT DIRECT PREDICTION LEARNS

This appendix separates mean-trajectory gains from spatial prediction.

Two diagnostics ask what the direct U-Net of Sec. 6.2 adds beyond aligning a particle's mean optical trajectory: an exact split of the image error into a particle-mean and a within-particle part, and a comparison of forecast optical transitions with cross-rate replay and same-condition repeats on identical transition support. Both measure optical prediction, not material state or mechanism, and their intervals describe particle variation within this electrode, not independent cells or condition interventions.

H.1 PARTICLE-MEAN AND WITHIN-PARTICLE RESIDUAL SKILL

We separate forecast error into particle-mean and within-particle components. Within each masked window the error image is its mask-mean error plus a zero-mean spatial residual, so $\mathrm{SSE} = \mathrm{SSE}_{\mathrm{mean}} + \mathrm{SSE}_{\mathrm{resid}}$ exactly up to rounding, with $\mathrm{SSE}_{\mathrm{mean}}$ the squared mask-mean error times the number of mask pixels. Each component skill is one minus the pooled component SSE divided by the same component of the origin-persistence SSE; total skill uses the total persistence SSE. A forecaster that adds a uniform change inside the mask (s-aligned mask-mean replay, history-aligned replay, population-mean replay and the particle-mean ridge) keeps the origin's residual and has residual skill zero up to floating point, so the residual component isolates what a forecaster predicts about where inside the particle the contrast changes. Particle-mean ridge is a weighted ridge regression per direction, fitted on training windows only, of the mask-mean change on causal features that all vanish at zero horizon: the change of a learned direction-specific mean trajectory in s (Gaussian bumps evaluated at the origin and the target charge state), and the origin level and the last 0.04 of charge of mask-mean history (eight values relative to the origin), each multiplied by saturating functions of the horizon. Rows are weighted by scored pixels, features are RMS-scaled, and the ridge penalty is chosen by grouped K-fold SSE over training acquisitions (over particles where there is a single acquisition); the predicted change is added uniformly inside the mask.

For each split the records of every listed forecaster are intersected by sequence, window and horizon. Direct U-Net uses seeds 0–4 on both rate splits and fold 0, and 0–1 on later cycles; other learned models retain seeds 0–1. Seeds are combined by averaging each window's squared errors, not by ensembling images; 1,000 multinomial particle-cluster bootstrap draws (seed 0) give percentile 95% intervals that resample the same physical particle jointly across windows, acquisitions, seeds and methods. The best baseline is the baseline with the largest observed total skill on the exact support of a split, direction and horizon, including the particle-mean ridge; its label is held fixed during resampling, so the intervals are descriptive and selection-conditional. Paired differences use each component's own persistence denominator; contributions to the total gain divide both SSE reductions by the total persistence SSE and add to the total skill difference (Table 108), although their interval endpoints need not add. New-particle results use fold 0, where same-particle replay is unavailable and reverts to persistence. Per-source coverage is retained in the released records; common windows at $\Delta = 0.16$ number 5,005, 4,348, 2,102 and 1,434 on the forward and reverse rate splits, fold 0 and later cycles.

Table 106 gives every component skill at $\Delta = 0.02$ and 0.16. At $\Delta = 0.16$ the direct U-Net's residual skill is 0.389 [0.213, 0.464] from C/20 to C/2, 0.287 [0.190, 0.343] from C/2 to C/20, 0.315 [0.152, 0.473] on fold 0 and 0.640 [0.442, 0.721] on the later cycle, against 0.278 [0.054, 0.363], 0.144 [-0.025, 0.245], 0.347 [0.074, 0.391] and 0.524 [0.314, 0.592] for the autoregressive U-Net; the paired residual gain of direct over autoregressive prediction is 0.111 [0.061, 0.171] from C/20 to C/2 and not resolved on fold 0 (-0.032 [-0.043, 0.100]). Pixel replay of the same particle carries residual skill only where its reference is a repeat of the same condition (0.621 [0.548, 0.728] on the later cycle, against 0.027 [-0.077, 0.165] from C/20 to C/2 and -0.111 [-0.181, -0.061] in reverse). From C/2 to C/20 the direct U-Net's gain over history-aligned replay, the strongest baseline there, has similar mean (0.080 [0.041, 0.190]) and residual (0.083 [0.047, 0.097]) contributions. Residual contributions to the total gain over particle-mean ridge are 0.085 [0.052, 0.100] forward and 0.072 [0.020, 0.096] on fold 0. During lithiation from C/20 to C/2 the direct residual skill is 0.268 [0.038, 0.363] and the total gain over particle-mean ridge 0.077 [0.013, 0.104].

H.2 Optical transitions on identical support

We compare models, replay and repeats on the same optical transitions. The comparison of Sec. 6.3 puts four forecasts of the same optical transition on identical charge samples, target pixels and intervals: the direct U-Net and the autoregressive (d, s) U-Net (seeds 0 and 1, both required and averaged), s-aligned pixel replay from the other rate, and the same-condition repeat, an observation of the same particle at the same rate, temperature, direction and cycle in another acquisition, resampled at the target's charge samples on the target's mask. Targets are the resolved gap-1 and gap-2 optical intervals of App. G, sampled every 0.005 in charge without bridging quality gaps; no interval is shortened to make a repeat eligible. Three origins are reported. The fixed origin is the stored frame at or before $u = 0.14$, from which 80–88% of the C/20→C/2 targets extend beyond the direct model's longest trained lead of 0.32. The onset origin is the stored frame at or before the observed interval start. The pre-interval origin lies one charge step of 0.005 before the observed interval start, snapped back to a stored frame with $u \geq 0.14$, so that every lead lies in $[0.005, 0.32]$ (intermediate leads interpolate between the trained horizons); it was evaluated for the 82 forward-rate and 77 reverse-rate intervals with an available repeat, chosen by repeat availability alone. The last two origins are retrospective, since they use the observed interval. The marker is the largest positive-slope optical candidate of the particle-mean trace, and marker error is its absolute dimensionless charge distance from the observed marker (lower is better); switch maps use the benchmark definition and are compared by Spearman ρ on the target's finite pixels, which every source must cover exactly. Within a target, metrics are averaged over seeds or repeats, then equally over targets; intervals are 95% percentile intervals from 2,000 paired physical-particle bootstrap resamples (seed 270923), and a single-particle cell has no interval. Targets without a repeat, with fewer than 13 interval samples, with an unresolved observed marker or with an undefined prediction or map are excluded and counted: for the fixed origin from C/20 to C/2, 348 target–metric rows lack a repeat, 123 are too short, 15 have an unresolved marker and 105 an undefined map or marker (79 of them undefined direct-model maps), against 144 included rows; the pre-interval rule includes 127 rows and excludes 37 (31 undefined maps, 6 undefined markers). The new-particle fold has no training reference for its test particles, so same-particle replay and a four-way comparison are undefined there, and the later cycle has no included comparison. These restrictions condition the estimates and say nothing about the excluded targets; the repeat is an observed agreement, not a prediction bound.

On the in-range forward-rate lithiation support (Tables 109 and 110) the direct model is better than the autoregressive model on both metrics and better than cross-rate replay on timing, and worse than the repeat on both. With the fixed origin, where most targets extend beyond the trained range, its ρ falls to 0.012 and cross-rate replay is ahead (0.141), so the advantage is confined to leads the model was trained on. On the reverse-rate lithiation support (two acquisitions, eleven targets) both U-Nets' marker errors (0.020 and 0.014) are below the repeat difference (0.049) while their ρ (-0.079 and -0.023) is below the repeat's 0.262; the delithiation cell rests on one acquisition and, for ρ, four particles, where the repeat agreement is itself negative. These small cells are reported for completeness and support no claim. For forward-rate delithiation, the post-origin portions of matched intervals favour autoregression in pattern: $\rho = 0.27$ against 0.14 for direct prediction, a paired direct-minus-autoregressive difference of -0.12 [-0.23, -0.02]. This uses the marked row's support, not the full-rollout comparison of Appendix G.8. All marked delithiation intervals include a lead below the minimum trained horizon of 0.005, while none extends beyond 0.32; this is a diagnostic comparison, not evidence of a uniform direct-model disadvantage on trained leads.

Per-particle link. Table 111 asks whether the particles whose first-lithiation transition shifts most between rates, beyond their same-rate repeat difference, are those the models mispredict: across 13 particles the autoregressive model's timing error rises with the shift (Spearman 0.69 [0.23, 0.90], $p = 0.011$, not significant after Holm correction over the declared tests), the direct model's less clearly (0.41 [-0.24, 0.86]), and the same-rate placebo and switch-map versions are inconclusive, so rate specificity remains unresolved (declaration and code in the released notes).

H.3 Growth after onset

Because where conversion starts inside a particle may be stochastic, Table 112 and Fig. 6 examine growth conditional on each particle's own onset (declared before computing, exploratory; 219 of 555 switch maps pass the quality rules). (a) From all switched pixels of a particle, the charge from

onset (5% coverage) to 50% coverage keeps its across-rate order in lithiation (Kendall τ 0.35 [0.10, 0.73]; delithiation 0.31 [−0.03, 0.65]), whereas onset order does not (Sec. 6.3); onset-aligned curves differ across C/20 and C/2 by about 0.02 in charge quantiles, more than in the calibrated simulator (0.006–0.011), and a fixed-window collapse cannot be tested because no cross-rate pair spans the declared window. (b) Tracking connected regions individually (8-connected, fixed minimum size, censored at mergers), the first region's advance is measurable in 117 maps but its cross-rate order is unresolved; individual regions advance by a median 10.4 and 11.6 pixels per unit charge in 25 °C lithiation at C/20 and C/2, within the simulator's 8.1–16.7 under the same rules. A flux-limited front under a constant current share predicts linear aggregate growth in charge; these maps are consistent with a shared growth rate after onset but cannot establish it, and the model rollouts have too little eligible support for the same comparison.

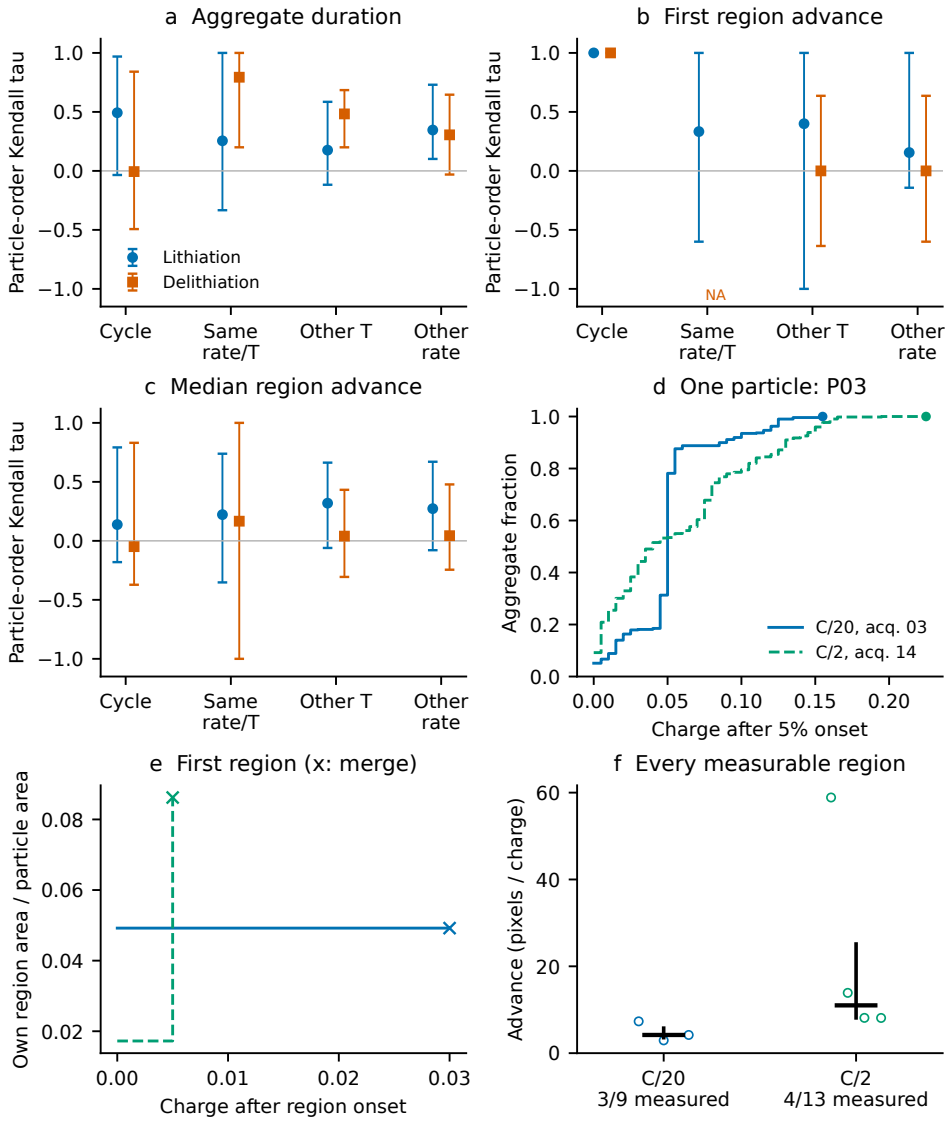

Figure 6: **Growth after onset.** Onset-aligned converted-area fraction and per-region advance rates (observed and calibrated simulator, same rules); details in the text and Table 112.

H.4 Example forecast

We explain how the displayed forecast window was selected.

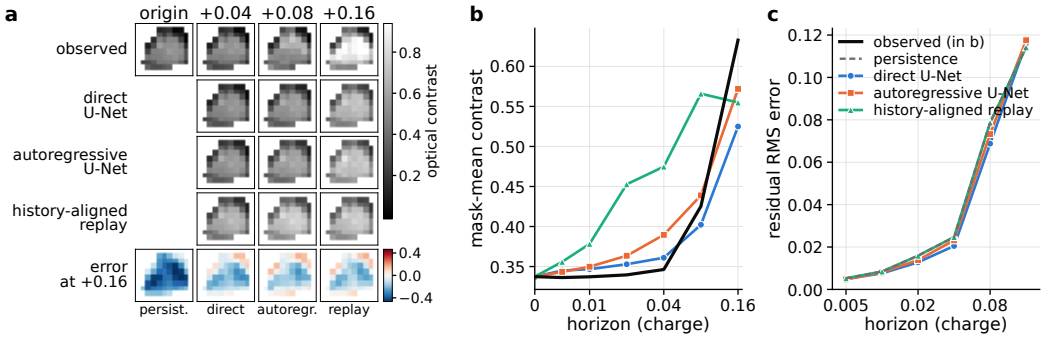

Figure 7: **Forecasting one test particle from C/20 training to C/2.** (a) One lithiation window chosen by a declared rule: among the 2,707 C/2 lithiation test windows scored at $\Delta = 0.04$, 0.08 and 0.16 by all three forecasters, the window at the lower median of the direct U-Net's per-window skill at $\Delta = 0.16$ (skill 0.75; quartiles 0.44, 0.75 and 0.87). Rows: observed optical contrast, direct U-Net, autoregressive U-Net and history-aligned replay (both U-Nets told direction and charge state, seed 0) at the origin ($s_0 = 0.48$, mid-lithiation) and $\Delta = 0.04$, 0.08 and 0.16 later in charge, on one grey scale spanning the pooled 0.5th–99.5th percentiles; pixels outside the mask are blank. Bottom row: signed error at $\Delta = 0.16$ inside the mask on one diverging scale (symmetric 99.5th-percentile limit) for persistence (the negative of the observed change) and each forecaster. (b) Mask-mean contrast against horizon, the component that replay and any mean-only forecaster predicts. (c) Root-mean-square error of the within-particle residual (forecast and target each minus their own mask mean); history-aligned replay adds a uniform change and so keeps exactly the persistence residual. Here the particle stays nearly unchanged until $\Delta \approx 0.04$ and then brightens abruptly: the direct U-Net tracks the quiet phase (skill 0.08 at 0.04 against -2.5 for the autoregressive model and -23.8 for replay, whose mean overshoots) but undershoots the jump (skill 0.75 at 0.16 against 0.83 and 0.81), and all three leave similar residual errors at 0.16.

Table 106: Particle-mean and within-particle residual skill: rate transfer. Skill is $1 - \mathrm{SSE}/\mathrm{SSE}_{\mathrm{persistence}}$, with separate persistence denominators for the particle mean, within-particle residual and total. Direct U-Net seeds 0–4 on both rate splits and fold 0, and 0–1 on later cycles, are averaged per window; other learned models use seeds 0–1. The 95% intervals resample paired particles (1,000 resamples) and are descriptive. Forecasters that add a uniform change inside the mask have residual skill zero (printed without an interval). These optical components do not identify material state or mechanism. New-particle results use fold 0, where same-particle replay reverts to persistence. Headers: particles and test acquisitions. n/e: derived records not evaluated in this version (never a zero).

	$\Delta = 0.02$			$\Delta = 0.16$		
Forecaster	Total	Mean	Residual	Total	Mean	Residual
C/20→C/2 (22 particles, 8 acquisitions)						
U-Net direct	0.464 [0.380,0.494]	0.712 [0.554,0.765]	0.244 [0.166,0.267]	0.792 [0.680,0.823]	0.904 [0.822,0.922]	0.389 [0.213,0.464]
U-Net autoregressive	0.515 [0.472,0.541]	0.786 [0.686,0.816]	0.277 [0.208,0.301]	0.666 [0.519,0.702]	0.774 [0.662,0.801]	0.278 [0.054,0.363]
ConvLSTM	0.473 [0.395,0.498]	0.754 [0.622,0.797]	0.224 [0.114,0.267]	0.719 [0.575,0.753]	0.842 [0.724,0.872]	0.277 [0.064,0.367]
Poseidon-T direct	0.457 [0.373,0.482]	0.732 [0.564,0.792]	0.214 [0.148,0.236]	0.625 [0.482,0.668]	0.732 [0.608,0.766]	0.242 [0.082,0.301]
Learned particle-mean predictor	0.343 [0.296,0.392]	0.732 [0.547,0.800]	0.000	0.681 [0.616,0.702]	0.870 [0.806,0.889]	0.000
History-aligned replay	0.248 [0.142,0.294]	0.530 [0.261,0.619]	0.000	0.655 [0.499,0.710]	0.837 [0.675,0.893]	0.000
s-aligned replay	0.163 [0.038,0.207]	0.347 [0.070,0.477]	0.000	0.602 [0.367,0.675]	0.769 [0.497,0.849]	0.000
Mean replay	-0.077 [-0.166,-0.042]	-0.164 [-0.313,-0.097]	0.000	0.017 [-0.118,0.071]	0.022 [-0.154,0.091]	0.000
s-aligned pixel replay	-0.114 [-0.181,-0.016]	0.350 [0.071,0.479]	-0.524 [-0.613,-0.273]	0.605 [0.360,0.678]	0.766 [0.497,0.847]	0.027 [-0.077,0.165]
Pixel replay	-0.503 [-0.561,-0.348]	-0.154 [-0.307,-0.087]	-0.810 [-0.897,-0.523]	-0.061 [-0.191,-0.009]	0.029 [-0.152,0.097]	-0.383 [-0.438,-0.247]
s-aligned population replay	0.193 [0.173,0.225]	0.412 [0.342,0.444]	0.000	0.502 [0.456,0.516]	0.641 [0.589,0.665]	0.000
Population replay	0.039 [-0.026,0.062]	0.084 [-0.052,0.139]	0.000	0.056 [-0.070,0.100]	0.071 [-0.094,0.126]	0.000
Scaled population replay	0.152 [0.107,0.171]	0.324 [0.200,0.372]	0.000	0.184 [0.033,0.236]	0.235 [0.044,0.297]	0.000
AR(2)	0.177 [0.152,0.190]	0.265 [0.218,0.281]	0.098 [0.070,0.109]	0.046 [0.032,0.088]	0.061 [0.044,0.120]	-0.011 [-0.018,0.001]
Trend	-0.058 [-0.147,0.050]	0.557 [0.295,0.650]	-0.601 [-0.697,-0.496]	-0.765 [-1.007,-0.700]	-0.330 [-0.629,-0.254]	-2.327 [-2.444,-2.112]
Damped trend	0.186 [0.156,0.216]	0.329 [0.262,0.353]	0.059 [0.028,0.090]	0.015 [0.010,0.017]	0.029 [0.020,0.031]	-0.035 [-0.048,-0.015]
History-mean persistence	-0.386 [-0.405,-0.356]	-0.499 [-0.521,-0.435]	-0.287 [-0.305,-0.255]	-0.031 [-0.032,-0.027]	-0.039 [-0.041,-0.032]	-0.001 [-0.014,0.006]
C/2→C/20 (22 particles, 4 acquisitions)						
U-Net direct	0.303 [0.192,0.374]	0.527 [0.386,0.629]	0.141 [0.007,0.217]	0.626 [0.506,0.695]	0.763 [0.650,0.837]	0.287 [0.190,0.343]
U-Net autoregressive	0.386 [0.280,0.462]	0.648 [0.519,0.745]	0.195 [0.044,0.290]	0.542 [0.399,0.633]	0.704 [0.585,0.785]	0.144 [-0.025,0.245]
ConvLSTM	0.382 [0.278,0.454]	0.634 [0.510,0.735]	0.198 [0.074,0.284]	0.588 [0.453,0.658]	0.747 [0.648,0.803]	0.194 [0.037,0.295]
Poseidon-T direct	0.312 [0.206,0.389]	0.540 [0.358,0.683]	0.146 [0.038,0.216]	0.458 [0.245,0.566]	0.586 [0.388,0.691]	0.142 [-0.024,0.246]
Learned particle-mean predictor	0.072 [-0.077,0.153]	0.171 [-0.151,0.404]	0.000	0.345 [0.165,0.433]	0.484 [0.249,0.600]	0.000
History-aligned replay	0.200 [0.118,0.238]	0.474 [0.240,0.619]	0.000	0.463 [0.292,0.545]	0.650 [0.408,0.763]	0.000
s-aligned replay	0.131 [0.059,0.166]	0.311 [0.111,0.438]	0.000	0.361 [0.171,0.465]	0.507 [0.232,0.650]	0.000
Mean replay	0.048 [-0.027,0.081]	0.115 [-0.054,0.209]	0.000	0.166 [0.096,0.200]	0.233 [0.127,0.283]	0.000
s-aligned pixel replay	0.038 [-0.068,0.083]	0.308 [0.105,0.435]	-0.158 [-0.252,-0.127]	0.327 [0.115,0.443]	0.504 [0.218,0.649]	-0.111 [-0.181,-0.061]
Pixel replay	-0.067 [-0.177,-0.012]	0.113 [-0.057,0.208]	-0.199 [-0.302,-0.140]	0.108 [0.025,0.150]	0.231 [0.121,0.281]	-0.197 [-0.274,-0.126]
s-aligned population replay	0.143 [0.104,0.167]	0.341 [0.209,0.436]	0.000	0.347 [0.132,0.463]	0.487 [0.202,0.640]	0.000
Population replay	0.072 [0.049,0.086]	0.172 [0.102,0.222]	0.000	0.202 [0.117,0.248]	0.284 [0.176,0.341]	0.000
Scaled population replay	0.167 [0.124,0.191]	0.397 [0.238,0.507]	0.000	-0.468 [-0.927,-0.220]	-0.657 [-1.424,-0.304]	0.000
AR(2)	-0.267 [-0.510,-0.128]	0.268 [-0.025,0.480]	-0.656 [-1.092,-0.490]	-0.004 [-0.063,0.022]	0.137 [0.069,0.176]	-0.355 [-0.491,-0.293]
Trend	-0.358 [-0.477,-0.303]	0.301 [-0.037,0.530]	-0.837 [-0.996,-0.781]	-0.947 [-1.150,-0.834]	-0.235 [-0.375,-0.170]	-2.704 [-3.187,-2.531]
Damped trend	0.069 [0.005,0.096]	0.356 [0.201,0.466]	-0.140 [-0.214,-0.111]	0.008 [0.000,0.012]	0.053 [0.030,0.063]	-0.103 [-0.129,-0.066]
History-mean persistence	-0.303 [-0.321,-0.268]	-0.455 [-0.491,-0.395]	-0.192 [-0.212,-0.136]	-0.032 [-0.037,-0.029]	-0.042 [-0.046,-0.037]	-0.007 [-0.015,0.000]

Table 107: Particle-mean and within-particle residual skill (continued): new particle and later cycles.

	$\Delta = 0.02$			$\Delta = 0.16$		
Forecaster	Total	Mean	Residual	Total	Mean	Residual
New particle, fold 0 (5 particles, 12 acquisitions)						
U-Net direct	0.450 [0.424,0.482]	0.742 [0.568,0.777]	0.246 [0.187,0.259]	0.751 [0.660,0.780]	0.879 [0.735,0.886]	0.315 [0.152,0.473]
U-Net autoregressive	0.532 [0.469,0.579]	0.823 [0.609,0.861]	0.329 [0.228,0.332]	0.740 [0.582,0.750]	0.855 [0.663,0.872]	0.347 [0.074,0.391]
ConvLSTM	0.539 [0.486,0.572]	0.838 [0.617,0.880]	0.330 [0.251,0.337]	0.749 [0.473,0.774]	0.862 [0.529,0.897]	0.364 [0.051,0.371]
Poseidon-T direct	0.443 [0.388,0.449]	0.730 [0.510,0.774]	0.243 [0.167,0.251]	0.558 [0.531,0.571]	0.700 [0.585,0.711]	0.074 [0.052,0.317]
Learned particle-mean predictor	0.313 [0.237,0.323]	0.763 [0.432,0.835]	0.000	0.645 [0.524,0.658]	0.835 [0.615,0.856]	0.000
History-aligned replay	0.000	0.000	0.000	0.000	0.000	0.000
s-aligned replay	0.000	0.000	0.000	0.000	0.000	0.000
Mean replay	0.000	0.000	0.000	0.000	0.000	0.000
s-aligned pixel replay	0.000	0.000	0.000	0.000	0.000	0.000
Pixel replay	0.000	0.000	0.000	0.000	0.000	0.000
s-aligned population replay	0.175 [0.141,0.185]	0.427 [0.217,0.458]	0.000	0.516 [0.374,0.524]	0.667 [0.431,0.681]	0.000
Population replay	0.110 [0.093,0.140]	0.269 [0.145,0.280]	0.000	0.314 [0.254,0.348]	0.407 [0.295,0.436]	0.000
Scaled population replay	0.197 [0.118,0.207]	0.481 [0.182,0.534]	0.000	0.088 [-0.405,0.115]	0.113 [-0.468,0.149]	0.000
AR(2)	0.192 [0.085,0.207]	0.374 [0.226,0.403]	0.065 [-0.143,0.083]	0.072 [0.069,0.092]	0.106 [0.095,0.131]	-0.045 [-0.087,-0.043]
Trend	-0.163 [-0.323,-0.141]	0.600 [0.070,0.713]	-0.694 [-0.991,-0.678]	-0.741 [-1.199,-0.707]	-0.235 [-0.699,-0.186]	-2.462 [-4.301,-2.400]
Damped trend	0.171 [0.054,0.188]	0.505 [0.240,0.561]	-0.060 [-0.194,-0.047]	0.013 [-0.021,0.017]	0.057 [0.009,0.063]	-0.137 [-0.151,-0.123]
History-mean persistence	-0.360 [-0.369,-0.286]	-0.508 [-0.531,-0.380]	-0.257 [-0.268,-0.151]	-0.030 [-0.033,-0.018]	-0.041 [-0.043,-0.024]	0.006 [0.001,0.007]
Later cycles, acq. 14 (15 particles, 1 acquisitions)						
U-Net direct	0.609 [0.549,0.631]	0.863 [0.776,0.884]	0.393 [0.274,0.433]	0.915 [0.871,0.933]	0.990 [0.971,0.993]	0.640 [0.442,0.721]
U-Net autoregressive	0.668 [0.597,0.687]	0.916 [0.816,0.944]	0.457 [0.324,0.504]	0.845 [0.786,0.861]	0.933 [0.884,0.941]	0.524 [0.314,0.592]
ConvLSTM	0.635 [0.554,0.657]	0.910 [0.786,0.941]	0.402 [0.255,0.456]	0.835 [0.737,0.860]	0.942 [0.834,0.962]	0.442 [0.324,0.495]
Poseidon-T direct	0.626 [0.555,0.644]	0.894 [0.779,0.920]	0.398 [0.258,0.448]	0.873 [0.805,0.902]	0.958 [0.907,0.971]	0.560 [0.305,0.655]
Learned particle-mean predictor	0.401 [0.375,0.474]	0.874 [0.686,0.919]	0.000	0.753 [0.654,0.779]	0.958 [0.844,0.976]	0.000
History-aligned replay	0.400 [0.363,0.498]	0.871 [0.697,0.914]	0.000	0.722 [0.571,0.759]	0.920 [0.725,0.959]	0.000
s-aligned replay	0.415 [0.377,0.510]	0.904 [0.724,0.950]	0.000	0.746 [0.593,0.785]	0.950 [0.747,0.991]	0.000
Mean replay	0.370 [0.333,0.455]	0.805 [0.608,0.851]	0.000	0.719 [0.567,0.758]	0.916 [0.726,0.956]	0.000
s-aligned pixel replay	0.484 [0.429,0.638]	0.904 [0.724,0.950]	0.126 [0.076,0.312]	0.880 [0.708,0.923]	0.950 [0.747,0.991]	0.621 [0.548,0.728]
Pixel replay	0.327 [0.255,0.494]	0.805 [0.608,0.851]	-0.079 [-0.163,0.159]	0.836 [0.680,0.883]	0.916 [0.726,0.956]	0.544 [0.472,0.683]
s-aligned population replay	0.344 [0.324,0.402]	0.748 [0.635,0.784]	0.000	0.690 [0.676,0.726]	0.878 [0.866,0.909]	0.000
Population replay	0.339 [0.322,0.386]	0.738 [0.597,0.786]	0.000	0.687 [0.669,0.713]	0.874 [0.853,0.897]	0.000
Scaled population replay	0.303 [0.274,0.354]	0.659 [0.496,0.710]	0.000	0.473 [0.301,0.516]	0.602 [0.384,0.658]	0.000
AR(2)	0.369 [0.336,0.455]	0.689 [0.566,0.720]	0.096 [0.057,0.164]	0.030 [0.014,0.064]	0.130 [0.087,0.139]	-0.334 [-0.408,-0.180]
Trend	0.051 [-0.015,0.234]	0.691 [0.457,0.748]	-0.493 [-0.546,-0.338]	-0.885 [-1.157,-0.830]	-0.412 [-0.803,-0.330]	-2.618 [-2.774,-2.201]
Damped trend	0.289 [0.262,0.369]	0.551 [0.452,0.576]	0.066 [0.045,0.137]	0.022 [0.018,0.032]	0.063 [0.051,0.067]	-0.128 [-0.157,-0.073]
History-mean persistence	-0.407 [-0.444,-0.391]	-0.518 [-0.532,-0.469]	-0.314 [-0.351,-0.298]	-0.033 [-0.036,-0.032]	-0.045 [-0.047,-0.041]	0.009 [-0.008,0.017]

Table 108: Direct U-Net gains decomposed by image component. Direct U-Net minus the learned particle-mean predictor and minus the strongest baseline of this decomposition (chosen on the same windows among the baselines and the learned particle-mean predictor), on identical windows; where the strongest baseline is the learned particle-mean predictor the comparison appears once. Mean and residual contributions divide by the total persistence SSE and add to the total skill gain; they are not differences of the separately normalised component skills. Paired particle-bootstrap 95% intervals (1,000 resamples); direct seeds 0–4 on both rate splits and fold 0, 0–1 on later cycles. New-particle results use fold 0.

Split	Δ	Comparator	Total gain [95% CI]	From mean [95% CI]	From residual [95% CI]
C/20→C/2	0.020	Learned particle-mean predictor (strongest baseline)	0.120 [0.077,0.134]	-0.009 [-0.019,0.010]	0.130 [0.076,0.152]
C/20→C/2	0.160	Learned particle-mean predictor (strongest baseline)	0.111 [0.050,0.134]	0.026 [-0.008,0.038]	0.085 [0.052,0.100]
C/2→C/20	0.020	Learned particle-mean predictor	0.231 [0.152,0.336]	0.150 [0.078,0.295]	0.082 [0.003,0.133]
C/2→C/20	0.020	History-aligned replay (strongest baseline)	0.104 [0.049,0.142]	0.022 [-0.003,0.082]	0.082 [0.003,0.133]
C/2→C/20	0.160	Learned particle-mean predictor	0.281 [0.243,0.347]	0.198 [0.166,0.268]	0.083 [0.047,0.097]
C/2→C/20	0.160	History-aligned replay (strongest baseline)	0.163 [0.133,0.241]	0.080 [0.041,0.190]	0.083 [0.047,0.097]
New particle, fold 0	0.020	Learned particle-mean predictor (strongest baseline)	0.136 [0.127,0.202]	-0.009 [-0.023,0.108]	0.145 [0.066,0.151]
New particle, fold 0	0.160	Learned particle-mean predictor (strongest baseline)	0.105 [0.092,0.230]	0.034 [0.022,0.135]	0.072 [0.020,0.096]
Later cycles, acq. 14	0.020	Learned particle-mean predictor	0.207 [0.135,0.239]	-0.005 [-0.017,0.057]	0.213 [0.122,0.251]
Later cycles, acq. 14	0.020	s-aligned pixel replay (strongest baseline)	0.125 [-0.015,0.184]	-0.019 [-0.035,0.044]	0.144 [0.000,0.207]
Later cycles, acq. 14	0.160	Learned particle-mean predictor	0.162 [0.111,0.217]	0.025 [0.013,0.102]	0.137 [0.081,0.157]
Later cycles, acq. 14	0.160	s-aligned pixel replay (strongest baseline)	0.035 [-0.030,0.165]	0.031 [-0.001,0.177]	0.004 [-0.049,0.031]

Table 109: Optical transitions on identical support: marker charge error (lower is better) and switch-map Spearman ρ (higher is better) of the direct U-Net, the autoregressive U-Net, cross-rate s-aligned pixel replay and the same-condition repeat, as target means with 95% paired particle-bootstrap intervals, on full intervals except the marked (†) post-origin delithiation row. Support: targets / particles / acquisitions; beyond: fraction of targets whose longest lead exceeds the direct model's trained range. Reverse-rate rows rest on one or two acquisitions.

Split, direction, gap	Origin	Metric	Support	Beyond	Direct	Autoregressive	Cross-rate replay	Repeat
C/20→C/2, lith., gap 1	fixed	marker	55/19/4	0.80	0.037 [0.030, 0.045]	0.073 [0.065, 0.080]	0.038 [0.024, 0.052]	0.026 [0.021, 0.031]
		ρ	43/18/4	0.88	0.012 [−0.044, 0.069]	0.006 [−0.047, 0.070]	0.141 [0.040, 0.253]	0.342 [0.244, 0.444]
	onset	marker	49/19/4	0	0.036 [0.028, 0.045]	0.064 [0.054, 0.073]	0.053 [0.038, 0.067]	0.025 [0.020, 0.029]
		ρ	33/19/4	0	0.130 [0.021, 0.242]	−0.004 [−0.112, 0.076]	0.088 [−0.020, 0.206]	0.375 [0.283, 0.470]
	pre-interval	marker	51/19/4	0	0.039 [0.031, 0.048]	0.068 [0.057, 0.077]	0.051 [0.037, 0.065]	0.025 [0.021, 0.031]
		ρ	30/17/4	0	0.177 [0.068, 0.281]	−0.031 [−0.134, 0.059]	0.075 [−0.035, 0.200]	0.332 [0.242, 0.423]
C/20→C/2, delith., gap 1	pre-interval†	ρ	20/12/3	0	0.145 [0.001, 0.265]	0.269 [0.136, 0.387]	−0.038 [−0.142, 0.054]	0.319 [0.232, 0.429]
C/2→C/20, lith., gap 1	pre-interval	marker	11/8/2	0	0.020 [0.010, 0.031]	0.014 [0.008, 0.021]	0.041 [0.025, 0.061]	0.049 [0.031, 0.069]
		ρ	11/8/2	0	−0.079 [−0.156, 0.020]	−0.023 [−0.237, 0.164]	0.066 [−0.139, 0.222]	0.262 [0.084, 0.442]
C/2→C/20, delith., gap 1	pre-interval	marker	13/13/1	0	0.036 [0.029, 0.043]	0.034 [0.027, 0.042]	0.033 [0.024, 0.041]	0.034 [0.025, 0.043]
		ρ	4/4/1	0	0.214 [0.132, 0.288]	0.144 [0.001, 0.235]	0.016 [−0.063, 0.117]	−0.142 [−0.228, −0.055]
C/2→C/20, lith., gap 2	pre-interval	marker	5/5/1	0.80	0.020 [0.009, 0.038]	0.025 [0.009, 0.042]	0.040 [0.017, 0.065]	0.039 [0.035, 0.047]
		ρ	15/15/1	0.93	0.010 [−0.064, 0.082]	0.036 [−0.031, 0.097]	0.060 [−0.015, 0.142]	0.326 [0.199, 0.452]

Table 110: Paired differences, direct U-Net minus comparator, on the pre-interval C/20→C/2 gap-1 lithiation support of Table 109 (51 marker targets, 30 ρ targets), plus the labelled post-origin delithiation pattern contrast (20 targets). Negative marker differences and positive ρ differences favour direct prediction.

Comparator	Marker error	Switch-map ρ
Autoregressive U-Net	-0.029 [-0.034, -0.022]	$+0.208$ [0.057, 0.363]
Cross-rate pixel replay	-0.012 [-0.024, -0.000]	$+0.101$ [-0.076, 0.271]
Same-condition repeat	$+0.014$ [0.003, 0.024]	-0.156 [-0.296, -0.009]
Autoregressive U-Net (delith., post-origin)	—	-0.124 [-0.226, -0.022]

Table 111: Exploratory per-particle change versus forecast error, first lithiation transition. C/20 repeats 03/06 and C/2 test acquisitions 10/11/14 are all at 25 °C. Each row gives particle count n, exact target count N, Spearman ρ with a 95% particle-bootstrap interval (20,000 draws), and two-sided permutation p (99,999 shuffles). AR and direct U-Nets use identical targets and existing seed-0/1 mean errors. Intervals and tests are conditional on this electrode and these acquisitions.

Predictor / outcome	Model	n	N	ρ [95% CI]	p	p_{Holm}
Charge change / timing	AR	13	32	$+0.693\,[+0.228, +0.904]$	0.011	0.129
Charge change / timing	Direct	13	32	$+0.410\,[-0.236, +0.859]$	0.164	1.000
Charge repeat / timing	AR	13	32	$+0.396\,[-0.233, +0.837]$	0.183	1.000
Charge repeat / timing	Direct	13	32	$+0.075\,[-0.607, +0.670]$	0.804	1.000
Switch change / timing	AR	12	29	$-0.112\,[-0.700, +0.568]$	0.733	1.000
Switch change / timing	Direct	12	29	$-0.354\,[-0.775, +0.271]$	0.256	1.000
Switch repeat / timing	AR	12	29	$-0.140\,[-0.726, +0.512]$	0.669	1.000
Switch repeat / timing	Direct	12	29	$-0.189\,[-0.709, +0.460]$	0.551	1.000
Switch change / map error	AR	12	21	$+0.161\,[-0.597, +0.749]$	0.618	1.000
Switch change / map error	Direct	12	21	$+0.252\,[-0.423, +0.884]$	0.429	1.000
Switch repeat / map error	AR	12	21	$+0.084\,[-0.585, +0.773]$	0.800	1.000
Switch repeat / map error	Direct	12	21	$+0.014\,[-0.662, +0.673]$	0.974	1.000

With $s = q/Q_{\mathrm{ref}}$, charge change is $|s_{C/2} - (s_{03} + s_{06})/2| - |s_{03} - s_{06}|$; its repeat placebo is $|s_{03} - s_{06}|$. Switch change is the mean $1 - \rho$ to the two C/20 maps minus $1 - \rho_{03,06}$; its placebo is $1 - \rho_{03,06}$. Map error is $1 - \rho_{\mathrm{model,target}}$. Values are averaged within particle/acquisition, then over acquisitions; negative excess changes are retained. Holm adjustment covers all 12 displayed tests. Only resolved observed markers and existing matched full intervals are used. The origins are retrospective; shared target error and acquisition history can confound associations. A null placebo alone cannot establish rate specificity.

Table 112: Onset-conditioned optical growth (exploratory). (A) Particle-order Kendall τ_b of aggregate duration D_{50}, first-region advance v, and median individual-region advance. Values are estimates [95% particle-bootstrap intervals]; counts below give particles/half-cycle pairs. Full inherited categories include the separate magnified 2C cohort; widefield sensitivities are in the CSVs. (B) Matched-temperature C/20–C/2 error minus same-rate/T repeat error on common particles. No observed cross-rate pair covers the full prespecified 0.20-charge onset axis, so fixed-axis collapse is untestable. The simulator row gives its absolute distance across three draws. (C) Fixed-origin ($u = 0.14$) model error minus repeat error on identical targets, with both seeds required; these gap-1 forecasts can exceed the trained lead range. Pre-interval/in-range results are separate CSV sensitivities. Regions use 8-connectivity and $\max(5, \lceil 0.01N \rceil)$ pixels; parents end at mergers. Region velocities use digital perimeter and are not calibrated material speeds. Dashes denote unavailable support, not zero. For fixed-origin D_{50}, lithiation/delithiation support is 0/4 targets, 0/4 particles and 0/1 acquisitions.

A. Reproducibility across particles

Estimand	Direction	Other cycle	Same rate/T	Other T	Other rate
(a) D_{50}	Lith.	$0.49[-0.03, 0.97]$ 9/6	$0.26[-0.33, 1.00]$ 10/9	$0.18[-0.12, 0.59]$ 17/20	$0.35[0.10, 0.73]$ 15/27
(a) D_{50}	Delith.	$-0.01[-0.49, 0.84]$ 8/6	$0.79[0.20, 1.00]$ 6/2	$0.48[0.20, 0.68]$ 15/15	$0.31[-0.03, 0.65]$ 16/29
(b1) First v	Lith.	$1.00[1.00, 1.00]$ 6/3	$0.33[-0.60, 1.00]$ 6/4	$0.40[-1.00, 1.00]$ 11/7	$0.16[-0.14, 1.00]$ 11/9
(b1) First v	Delith.	$1.00[1.00, 1.00]$ 3/1	– 0/0	$0.00[-0.64, 0.64]$ 4/2	$0.00[-0.60, 0.64]$ 4/2
(b2) Median v	Lith.	$0.14[-0.18, 0.79]$ 9/6	$0.22[-0.35, 0.74]$ 10/9	$0.32[-0.06, 0.66]$ 14/13	$0.27[-0.08, 0.67]$ 13/21
(b2) Median v	Delith.	$-0.05[-0.37, 0.83]$ 8/6	$0.17[-1.00, 1.00]$ 5/2	$0.04[-0.31, 0.43]$ 15/15	$0.04[-0.24, 0.48]$ 15/28

B. Extra cross-rate error relative to repeats

Estimand (units)	Lithiation	Delithiation	Particles L/D
(a) Quantile distance (charge)	0.004 [-0.002, 0.009]	0.001 [-0.003, 0.014]	8/6
(b1) First v (pixels/charge)	0.27 [-30.50, 24.89]	-4.09 (CI –)	4/1
(b2) Median v (pixels/charge)	-1.07 [-12.96, 1.91]	-13.33 [-24.20, 17.58]	7/5
Sim. absolute distance (draw range)	0.006–0.011	0.006–0.008	3 draws

C. Model error minus repeat error

Estimand	Model	Lithiation	Delithiation	Targets L/D
(a) D_{50}	AR	–	0.013 [-0.001, 0.028]	0/4
(a) D_{50}	Direct	–	0.027 [0.013, 0.037]	0/4
(b1) First v	AR	–	–	0/0
(b1) First v	Direct	–	–	0/0
(b2) Median v	AR	–	-0.76 (CI –)	0/1
(b2) Median v	Direct	–	2.60 (CI –)	0/1

I CONTROLLED SIMULATOR: MODEL CARD, EXPOSURE LEDGER AND GATES

This appendix records the simulator's assumptions, exposures and failed gates.

The simulator is the controlled counterpart of the experiment: it can change the applied rate alone, with electrode geometry, disorder and observation noise held fixed. It is a deliberately constructed mechanism model, calibrated on movies that had already been inspected and whose observations shaped its design. Its results are therefore *consistency tests of a deliberately constructed model*, not independent validation of a material law, and it enters the main text only for claims whose pre-specified held-out gates pass (App. I.4). All four claim gates fail. The main text therefore contains only a pointer to this appendix; the simulator is retained as a conditional illustration, without a claim that it reproduces the observations.

I.1 MODEL CARD

We specify the constructed model and its limits as a physical comparator.

Scope. Mechanism claims only: the simulator is not a digital twin, its latent field is not a calibrated lithium concentration, and it contains no avalanche or depinning ingredient.

Geometry and observation. Particles use the measured atlas masks of the graphite field of view at the benchmark pixel size. Simulated images pass through a local optical map $g(c)$ (piecewise linear with knots at the phase compositions, fitted against the particle-mean responses in both directions; a shared map cannot match their direction-dependent optical response), point spread, noise and illumination drift, and are written in the benchmark data format so that every forecaster and evaluator runs unchanged. For the controlled rate comparison the observation is standardised: uniform charge sampling every 0.002, common noise and drift parameters at the reference rate, and identical random seeds at every rate, with the measured cadence retained only as provenance.

Thermodynamics and kinetics. One latent field $c \in (0, 1)$ per pixel evolves by an Allen–Cahn-reaction law with Butler–Volmer-type kinetics (Bazant, 2013; Ferguson & Bazant, 2014). The homogeneous chemical potential has a smooth solid-solution region at low c and two miscibility gaps, for gap 1 and gap 2, each with its own gradient-energy coefficient and exchange-rate prefactor (a smooth switch in c), so that the two transitions can differ in how staggered they are across particles. One electrode potential, shared by all particles, is solved at every right-hand-side evaluation so that the total reaction rate equals the applied galvanostatic current; integration is adaptive and explicit as in the base solver, which never clips or rescales concentrations. The implemented homogeneous free energy and chemical potential are

$$f(c) = c\log c + (1-c)\log(1-c) + \tfrac{1}{2}Hc^2 - \sum_{j=1}^{2} A_j w_j \log\cosh[(c-m_j)/w_j] - \mu_0 c, \quad (2)$$

$$\mu_h(c) = \log[c/(1-c)] + Hc - \sum_{j=1}^{2} A_j \tanh[(c-m_j)/w_j] - \mu_0, \quad (3)$$

$$\dot{c} = 2k(x,c)\sqrt{c(1-c)}\sinh[(\phi-\mu)/2], \qquad \langle \dot{c} \rangle = \mathrm{Da}. \quad (4)$$

Here $H = 18$ and the linear gauge μ_0 sets the chemical potential to zero at the midpoint of the two gap centres. The common-tangent binodals are $[0.4436, 0.5859]$ and $[0.7287, 0.8751]$. The gradient energy is $\frac{1}{2}\sum_{(x,y)} \kappa((c_x + c_y)/2)(c_x - c_y)^2$; its exact discrete derivative, including the derivative of κ, enters μ. The shared potential is solved analytically at each right-hand-side evaluation. Conservative adaptive Heun steps reject nonpositive or superunit states; no state clipping is used.

Disorder hierarchy. All randomness is Gaussian. At the particle level, a lognormal kinetic prefactor and a nucleation barrier are drawn independently and fixed for an electrode draw across all of its acquisitions, with a small per-acquisition jitter; within particles, a correlated random field with a fixed correlation length has three variance components, persistent for the draw, per acquisition and per half-cycle; one rim amplitude enhances nucleation at the particle boundary for gap 2 and

for delithiation. Each ingredient sits behind a switch, so that reduced models (single gap, no local optics, no hierarchy, no rim term) are available for the ablation.

Solid-solution counterpart. The NMC-like model uses the effective diffusivity curve $D(\theta)$ reported for these particles (Xu et al., 2022), the measured particle sizes and the measured currents, with no free diffusivity scale, and a Butler–Volmer rim flux. Because $D(\theta)$ comes from the same experiment, this is a conditional material illustration. The reaction and diffusion numbers, $\mathrm{Da_r} = R\,(I/Q_{\mathrm{ref}})/v_{\mathrm{BV}}$ and $\mathrm{Da_d} = R^2\,(I/Q_{\mathrm{ref}})/D_{\mathrm{eff}}$, are kept distinct; D and size are varied by $\times 0.5$ and $\times 2$ as sensitivities, not fitted.

Numerics and tests. Unit tests cover the galvanostatic residual, positivity, charge conservation per half-cycle, two common-tangent gaps, exact reproduction of the base model with all switches off, and determinism by seed. At the lowest and highest rates of the ladder the time step is halved and the grid refined by two in each coordinate with the same latent fields (App. I.4). The generation audit passes for 36/36 graphite and 30/30 NMC acquisitions. For graphite, all 48 refinement half-cycles complete with positivity enforced and maximum charge-conservation error 2.09×10^{-8}; 260/280 numerical signature comparisons satisfy both registered bounds. The unresolved 20 comparisons and missing S5 refinement block the corresponding claims. The cheap/full selected-point check passes 34/36 fitted cells at half tolerance. Seven targeted evaluation/reporting regression tests also pass.

Table 113: Selected simulator parameters and their roles. Roles follow the pre-fitting design: Stage A is fitted at C/20, Stage B at C/20 and C/2; anchors are reported beside calibrated values and never forced.

Group	Role	Value
Gap centres / strengths; widths / gradients	Fitted / fixed	$m_1, m_2 = (0.5148, 0.8041)$; $A_1, A_2 = (1.5782, 1.8524)$; $w_1, w_2 = (0.025, 0.035)$; $\kappa_1, \kappa_2 = (0.6, 1.2)$
Local optical map $g(c)$	Fitted, 4 values	Knots $(.015, .200, .4436, .5859, .7287, .8751, .995)$; values $(0, .0135, -.3574, .5463, .5463, .7530, .7530)$
Within-particle disorder and rim amplitude	Fitted	Length 0.9325 native pixels; persistent/acquisition/half-cycle SDs $(0.2069, 0.2474, 0.1492)$; rim 0.5110
Rate scale and potential cut-off	Fitted	$\mathrm{Da}(C/20) = 0.016807$; symmetric $\lvert\phi\rvert$ cut-off 8.4004
Particle kinetic/barrier spreads; acquisition jitter	Fitted	$(0.0300, 0.8052)$; jitter 0.001619
Gap-2 exchange-rate multiplier	Fitted	0.106106
NMC effective diffusivity, size, current and rim rate	Anchored / fixed	Measured $D(\theta)$, masks and currents; $v_{\mathrm{BV}} = 2.082 \times 10^{-10}$ m/s (0.003/s at 0.0694 μm native pixels)

I.2 Signatures and their roles

We distinguish signatures used for fitting from those held out for evaluation. The same code computes every signature for real and simulated data, and the reproducibility metrics are imported from the real-data analysis of App. G (a reconciliation on the fitted real cohort reproduced all 247 optical markers and 83 particle switch-map correlations exactly). Before any fitting, the signatures were designated as fitted or held out (Table 114); all 2C values are held out. Held-out results are opened only after the primary parameter vector is selected.

I.3 Exposure ledger

We record how prior observations influenced model design, fitting and selection. A number withheld from the objective may still have influenced the design. The ledger (Table 115), frozen before fitting and appended rather than reset when the procedure changed, records for each quantity whether it shaped the model, entered the fit or the selection among fits, or only evaluates it. No signature

Table 114: Signatures used for calibration and checking, with their designation before fitting. Intervals on real data are particle-bootstrap intervals (App. G); S2 is an acquisition-level quantity without a particle interval.

Signature	Definition	Role
Response	Particle-mean optical trace against s, both directions	Fitted (C/20, C/2)
S1	Charge collapse: unscaled same-particle trace prediction across rates on overlapping charge, same temperature	Held out
S2	Capacity truncation: charge passed per half-cycle	Fitted (C/20, C/2)
S3	Concurrency: fraction of particle markers within ± 0.02 of the half-cycle median, per transition (never a rank statistic)	Fitted (C/20, C/2)
S4	Blockiness: share of total absolute optical change in the most active 20% of equal charge windows, per transition	Fitted (C/20, C/2)
S5	Baseline skill against horizon (trend, same-particle and population s-aligned replay; $\Delta = 0.01$–0.32; causal origins)	Held out
S6	Switch-map agreement (L3) by category	Fitted within acquisition and same rate; held out across rates
S7	Particle-order concordance τ_b (L2) by category, with marker-noise and tie sensitivity	Held out
S8	Electrode-level population marker (L1)	Held out

family is claimed to be wholly unexposed. Two procedural changes are recorded there: a restricted optical map (tying the optical values across the intermediate and final single-phase branches) was compared with the unrestricted map on fitted quantities only, and the search changed from a serial simplex to a parallel quasi-random design followed by an evolution strategy, with the objective, development draw and cheap configuration unchanged. Subsequently, fitted-only boundary hits prompted a logged search-space extension to $m_1 \in [0.34, 0.56]$ and $m_2 \in [0.57, 0.88]$, before any held-out simulator value was examined. This extension is post hoc. Also recorded, from fitted C/20 data only, is a structural limitation: with one shared optical map and near-equilibrium filling, the particle-mean contrast inside a two-phase gap is linear in charge in both directions, whereas the measured response differs between lithiation and delithiation, so the fit trades the two directions off; no direction-dependent optical map was added.

Table 115: Exposure ledger (condensed; frozen before fitting). Design: the quantity shaped the model structure. Fit: it enters the objective. Selection: it chooses among parameter vectors.

Quantity	Design	Fit	Selection	Evaluation role and limitation
Response shape	yes	C/20, C/2	yes	Same-data construction; marker positions also constrain S8 indirectly
S1 charge collapse	yes	no	no	Held out numerically; exposed design-consistency test
S2 truncation	yes	C/20, C/2	yes	Fit diagnostic; 2C held out
S3 concurrency	yes	C/20, C/2	yes	Fit diagnostic; frozen as a concurrency statistic
S4 blockiness	yes	C/20, C/2	yes	Fit diagnostic; 2C held out
S5 baseline skill	yes	no	no	Held-out profile, not independent validation
S6 switch maps	yes	within acq., same rate	fitted cells	Across rates and 2C held out; the hierarchy was designed for the observation
S7 order	yes	no	no	Held out; ties, marker noise and S3 reported separately
S8 population marker	yes	indirectly	response only	Conservation fixes total latent storage, not this optical marker
Fixed electrode marker (prior)	yes	via response	no	Re-measured; the prior value is not a target
Within/across-rate order (prior)	yes	spreads via S3	fitted only	Low τ_b alone is not concurrency
Kinetic hysteresis (prior)	yes	both responses	yes	Conditional explanation; no direction-specific optics
Spatial repeats, rim nucleation (prior)	yes	S6 fitted, S4	yes	Cross-rate patterns held out but exposed in design
Transition-specific morphology (prior)	yes	response, S3, S4	yes	Deliberate construction, not discovery
Reaction-limited model preferred (prior)	yes	no	no	Gradients are interface parameters, not a detected curvature law
Earlier reaction-model fit (prior)	yes	no	no	Two same-data, model-dependent estimates of the rate scale
No avalanche signature (prior)	yes	no	no	Gaussian fields only; no avalanche claim
NMC $D(\theta)$, sizes, currents	yes	no	no	Conditional material illustration

I.4 GATES

We state the held-out tests required for each simulator claim. The inherited rule that the simulation is main-text grade if at least four of S1–S5 fall within the real 95% intervals (or within 25% where no interval exists) is kept unchanged but is a *fit diagnostic only*: three of these five signatures are fitted, so it cannot authorise any claim. Claims are instead governed by claim-specific held-out gates, fixed before fitting (Table 116), and a claim whose gate fails is never worded as "the simulator reproduces". Shared rules apply to every gate. (i) The primary vector minimises the recorded fitted-only objective J; all tested admissible vectors with $J \leq 1.10\, J_{\min} + 0.10$ form the acceptable band, which is a finite-search band, not a confidence interval, and places the real rates on the rate axis as bands from two same-data, model-dependent estimates (this calibration and the earlier reaction-model fit to the same movies); a global prefactor change of $\times 0.8$ and $\times 1.2$ and optical-knot changes of $\times 0.9$ and $\times 1.1$ are sensitivities that cannot select the optimum. (ii) Calibration uses a development electrode draw; three fresh draws with two repeat acquisitions each are evaluated afterwards, and a claim must meet its tolerance in at least two of three draws and in the ensemble mean, with a draw-mean range no wider than twice the tolerance; the range of three draws is not called a population interval. (iii) Compatibility with a real interval means that the simulated estimate lies within the stated absolute tolerance of that interval; literal inside-interval fractions are reported separately. (iv) At the lowest and highest rates, halving the time step and refining space by two with the same latent fields must shift each signature by at most half its tolerance and 25% of any claimed effect; otherwise that part of the rate map is numerically unresolved. Conservation and positivity must pass at both resolutions.

Table 116: Claim-specific gates, fixed before fitting (per direction and, where identifiable, per transition; missing or censored markers never count as passes).

Possible claim	Held-out evidence	Tolerance	Verdict
Simulated rate-transfer map agrees with the observed transfer profile	S1 at C/20→C/2 and the S5 trend and replay profiles at all six horizons, common causal-origin evaluation; same-draw and new-draw transfer reported separately; no learned-model rate-transfer claim before the learned sweeps	0.15 in R^2 or skill; $\geq$ 80% of evaluable cells per direction; no wrong sign where the real effect's interval excludes zero	FAIL; L 4/19, D 4/19
Within-rate versus across-rate particle-order contrast	S7 same-rate minus across-rate τ_b and same-acquisition τ_b, with permutation and marker-noise-only nulls; S3 reported separately	0.15 in τ_b and in the contrast; same sign where the real contrast excludes zero; a physical reshuffling claim needs the contrast to exceed the 95th percentile of noise-only degradation	FAIL; L 2/3, D 5/7
Cross-acquisition pattern agreement	Held-out across-rate S6 and its contrast with fitted same-rate S6; rotation/flip null and coverage	0.15 in ρ or in the contrast; same sign where the real interval excludes zero	FAIL; L 4/4, D 6/8
Consistent extrapolation to 2C	All finite 2C cells of S1–S8 per transition, undefined gaps reported	0.10 of Q_{ref} (S2), 0.05 in s (S8), 0.15 otherwise; $\geq$ 80% of cells per direction and $\geq$ 50% per family; response RMSE ≤ 0.10	FAIL; L 0/3, D 7/17

For the order gate, marker noise is estimated from paired repeat traces where available and otherwise from background-calibrated replicates; independent Gaussian noise added to an unchanged latent order (2,000 resamples, charges quantised to 0.005) gives the τ_b expected at noise scales 0, 0.5, 1 and 2 and tie grids 0.0025, 0.005 and 0.01, and the noise scale is never chosen to make a mechanism pass. Across-rate agreement is gated on *unpaired* simulated pairs (same draw, different repeat, different rate), the like-for-like analogue of two real acquisitions; *paired* pairs (same draw and repeat, every nuisance stream identical, rate changed alone) are the controlled intervention and are reported separately, never gated. The word "ceiling" is reserved for a conditional oracle that forward-simulates from the true state at the origin with the persistent and particle-level fields known and the others resampled (ensemble of eight); unless that oracle is run, no simulator ceiling is reported. Reduced models in the ablation are refitted on the same calibration subset with the same budget, or the ablation is labelled not refitted and illustrative.

I.5 GENERATION DESIGN

We specify the paired rate ladders and their controlled sources of variation. The rate ladder multiplies the calibrated C/20 rate scale by 0.32, 1, 3.2, 10.1428, 21.0118 and 46, where 10.1428 and

21.0118 place C/2 and 2C by their measured currents rather than their nominal rates. Each rung has three electrode draws, two repeat acquisitions per draw and two cycles per acquisition; for a given draw and repeat the initial state, all field streams and the noise and drift seeds are identical at every rung, and the same physical cut-off rule applies at every rate, so the rate is the only difference along a paired ladder. The NMC-like ladder places the three measured currents and reports their reaction and diffusion numbers separately, plus one rung below and one above (three draws, two repeats). Splits mirror the real ones: training at the C/20 or at the C/2 rung and testing all others, repeat 0 to repeat 1 at one rung, and training on two draws to test the third.

I.6 Calibration and consistency results

We report the fitted model and the outcomes of its consistency gates. A two-gap reaction model was fitted to the previously inspected graphite movies (J=2.651; 199 evaluations; 81 acceptable vectors). All four claim gates failed. Simulated C/20-to-C/2 charge collapse R^2 was 0.989/0.988 (lithiation/delithiation), versus 0.519/0.359 experimentally. The delithiation gap-1 spatial contrast (same rate minus across rate) had the wrong sign (+0.053 versus -0.103). At 2C, response RMSE was 0.257/0.215, above the 0.10 limit, with unmeasurable signature families. Only 260/280 numerical comparisons at the ladder endpoints resolved. A single composition-to-contrast map cannot match the direction-dependent optical response. Calibration and the earlier spatial fit are two same-data, model-dependent estimates. The model therefore provides a conditional illustration, without a physical reshuffling or forecasting ceiling claim; no conditional oracle was run.

The 81 acceptable objective values span $[2.650726, 3.011417]$, below the fixed threshold 3.015799. There are 19 fitted continuous parameters and 211 fitted scalar values; the response values are correlated and are not 211 independent observations. The particle kinetic-spread lower bound (0.03) is attained; the finite search does not establish identifiability. The selected rate convention gives $k_c = 0.004826$/min and gap-2 $\mathrm{Da}(C/20) = 0.1584$, versus 0.008261/min and 0.09254 after converting the earlier fit to the same midpoint convention. The effective interaction is 4.080 versus 5.5 previously; mean gap-2 drive is $+0.929/-0.964\,kT$ (lithiation/delithiation), and intermediate fractions are 0.364/0.329 versus the earlier 0.07 with different spatial sampling. These are two same-data, model-dependent estimates, not independent validation anchors.

Table 117: Measured-current placement of real rates. Bands span the acceptable tested vectors; they are not confidence intervals.

Rate	Current (μA)	Selected Da	Da band	Gap-2 Da band
0.05C	12.843	0.01681	[0.01235, 0.04298]	[0.1094, 0.3446]
0.5C	130.267	0.17047	[0.12528, 0.43594]	[1.1100, 3.4956]
2C	269.861	0.35314	[0.25952, 0.90310]	[2.2995, 7.2414]

Table 118: Ladder-end numerical checks: both the half-tolerance and quarter-effect conditions are required. Time-step halving passes all 140 cells. Spatial refinement fails 20/140. S5 refinement is unavailable.

Signature	Resolved / tested	Largest absolute shift
S1	8/8	0.01149
S2	16/16	0.00050
S3	30/32	0.14773
S4	30/32	0.08151
S6	75/80	0.09023
S7	70/80	0.09786
S8	31/32	0.07000

The inherited fit diagnostic passes 0/5 families (requires 4/5). Literal held-out coverage is 21/90 inside the real 95% particle intervals. The 2C lithiation gap-1 transition is censored for S3/S4 and pairwise maps; gap-2 population markers are absent in both directions. No real 2C S5 cell is evaluable on the inherited matched-cohort causal-origin support. These gaps block a general 2C claim even where a finite cell is compatible.

The S7 C/20-to-C/2 order contrasts (real/simulated) are 0.609/0.006 and 0.834/0.177 for lithiation gaps 1/2, and 0.176/0.027 and -0.350/0.354 for delithiation. At the declared marker-noise scale, the real noise-only 95th percentiles are 0.202, 0.097, 0.385 and 0.154, respectively. The two delithiation effects fail that noise test; simulated lithiation gap 1 also fails (0.006 versus 0.040). Among fresh draws, the respective simulator noise tests pass 0/3, 2/3, 1/3 and 2/3. No physical-reshuffling claim is supported.

For the separate S8 check, the delithiation gap-2 C/20-to-C/2 marker shift is -0.224 with shared-particle 95% interval $[-0.260, -0.145]$, versus $+0.030$ in the simulator (draw shifts 0.048, 0.005, 0.038). It fails the 0.05-charge tolerance; measured repeat SD is 0.0048 and the largest simulated repeat-mean spread is 0.0025. The 2C lithiation gap-1 shift also fails (-0.084 versus $+0.038$); four of six measurable shift checks are compatible. The three-draw range is not a population confidence interval.

Global kinetic-scale multipliers 0.8/1.2 and optical excursions 0.9/1.1 relative to the initial optical reference were evaluated without reselection. The four ablations (single-gap-like, two gaps, hierarchy, full) are *not refitted, illustrative*; they cannot establish model-family superiority. The NMC diffusivity and physical-size multipliers 0.5/2 were likewise not fitted. NMC draws have identical geometry and kinetics, with only a 2×10^{-4} initial-state perturbation and observation seeds changing; they are not independent latent-disorder draws.

Table 119: Largest absolute post-selection signature changes across the evaluated cells. The rate scale and optical excursions are varied without reselection. These are sensitivities, not fitted alternatives.

Case	S1	S2	S3	S4	S6	S7	S8
$k_0 \times 0.8$	0.036	0.016	0.395	0.036	0.233	0.232	0.031
$k_0 \times 1.2$	0.028	0.012	0.136	0.026	0.071	0.437	0.129
Optics $\times 0.9$	0.000	0.000	0.012	0.000	0.027	0.014	0.001
Optics $\times 1.1$	0.000	0.000	0.024	0.001	0.023	0.024	0.001

Table 120: Unrefitted illustrative ablations: finite real cells matched within the real interval plus signature tolerance. These single-draw counts do not apply the full fresh-draw, sign, noise or numerical gates and cannot rank model families under equal fitting effort.

Variant	S1	S2	S3	S4	S6	S7	S8
Single-gap-like	2/4	2/6	2/9	6/9	5/18	9/18	1/10
Two gaps	0/4	1/6	4/9	8/9	11/18	9/18	5/10
Hierarchy	0/4	1/6	7/9	9/9	15/18	11/18	4/10
Full	0/4	1/6	8/9	9/9	15/18	12/18	4/10

NMC refinement resolves 135/152 signature comparisons; maximum conservation error is 1.53e-08. NMC response RMSE is 0.010/0.007 at C/3, 0.036/0.016 at 1C and 0.053/0.019 at 2C (lithiation/delithiation). These optical agreements are conditional on the same-experiment effective diffusivity. They do not extend to all signatures: at 2C, NMC lithiation charge-collapse R^2 is 0.690 versus -0.353 experimentally (particle interval $[-0.867, 0.082]$), and all six population-marker comparisons fail the 0.05-charge compatibility tolerance. These are fixed-model checks, not a new fitted-model gate. Real/simulated finite C/20-to-C/2 map-pair counts are 154/505 and 30/514 for lithiation gaps 1/2, and 161/526 and 84/518 for delithiation. The real and simulated totals differ because of the acquisition design; they are not independent-particle sample sizes. The real-data reconciliation matches all 898 finite particle-map correlations exactly and all 102 finite order correlations within 8.4×10^{-17}.

I.7 Forecasting on the simulated movies

We evaluate forecasters on the constructed movies within the failed-gate limits. Completed benchmark evaluations are available for 23/32 planned U-Net runs. All forecasts use causal origins. Clock differences below use identical window/horizon intersections, with shared physical-particle resampling (2,000 replicates); training seeds are reported separately. The rate-transfer models train at one rate, with either the charge clock or the matched time clock. Repeat and new-draw arms use the

Table 121: Conditional NMC sensitivities: largest absolute signature changes from the fixed reference across evaluated cells. The large collapse-score changes are retained; optical agreement at the reference does not establish robustness or identify diffusivity.

Case	S1	S2	S3	S4	S6	S7	S8
$D \times 0.5$	12.180	0.430	0.533	0.057	0.460	1.144	0.611
$D \times 2$	4.255	0.438	0.333	0.036	0.888	1.611	0.715
Size $\times 0.5$	0.243	0.176	0.371	0.022	0.464	1.484	0.388
Size $\times 2$	13.829	0.239	0.592	0.055	0.477	0.855	0.503

Table 122: Lithiation: selected C/2 checks. S1 uses C/20 references; S5 is shown at horizon 0.08; S6/S7 are same-rate minus unpaired across-rate contrasts. The simulator column is the mean of three fresh-draw estimates; its range is not a confidence interval.

Signature	Transition / baseline	Real [95% CI]	Draw mean	Draw range	Compatible
S1	whole	0.519 [0.357, 0.672]	0.989	0.002	no
S2	whole	0.601 (no CI)	0.868	0.001	no
S3	gap1	0.728 [0.535, 0.853]	0.351	0.440	no
S4	gap1	0.556 [0.498, 0.616]	0.583	0.157	yes
S3	gap2	0.588 [0.353, 0.824]	0.493	0.058	yes
S4	gap2	0.400 [0.368, 0.437]	0.474	0.049	yes
S8	gap1	0.417 [0.401, 0.425]	0.481	0.030	no
S8	gap2	0.495 [0.490, 0.550]	0.735	0.031	no
S6	gap1	0.132 [0.004, 0.349]	0.002	0.002	yes
S7	gap1	0.609 [0.232, 0.997]	0.006	0.023	no
S6	gap2	0.406 [0.086, 0.541]	0.126	0.044	yes
S7	gap2	0.834 [0.149, 1.419]	0.177	0.200	yes
S5	population replay	0.449 [0.267, 0.490]	0.603	0.155	yes
S5	particle replay	0.282 [0.036, 0.380]	0.868	0.013	no
S5	trend	-0.347 [-0.548, -0.286]	0.015	0.089	no

Table 123: Delithiation: selected C/2 checks. S1 uses C/20 references; S5 is shown at horizon 0.08; S6/S7 are same-rate minus unpaired across-rate contrasts. The simulator column is the mean of three fresh-draw estimates; its range is not a confidence interval.

Signature	Transition / baseline	Real [95% CI]	Draw mean	Draw range	Compatible
S1	whole	0.359 [0.250, 0.476]	0.988	0.007	no
S2	whole	0.603 (no CI)	0.794	0.002	no
S3	gap1	0.750 [0.605, 0.878]	0.330	0.386	no
S4	gap1	0.471 [0.431, 0.513]	0.469	0.035	yes
S3	gap2	1.000 [0.975, 1.000]	0.511	0.136	no
S4	gap2	0.521 [0.490, 0.552]	0.485	0.021	yes
S8	gap1	0.427 [0.421, 0.460]	0.539	0.006	no
S8	gap2	0.574 [0.562, 0.576]	0.827	0.004	no
S6	gap1	-0.103 [-0.163, -0.027]	0.053	0.010	no
S7	gap1	0.176 [-0.060, 0.452]	0.027	0.068	yes
S6	gap2	0.314 [0.141, 0.412]	0.176	0.057	yes
S7	gap2	-0.350 [-0.736, 0.410]	0.354	0.675	no
S5	population replay	0.513 [0.477, 0.595]	0.588	0.163	yes
S5	particle replay	0.541 [0.474, 0.656]	0.861	0.015	no
S5	trend	-0.436 [-0.869, -0.382]	-0.019	0.126	no

charge clock only. The new-draw arm trains on draws 0 and 1 across the ladder and tests draw 2 on the same particle atlas, so it is not also a held-out-rate or new-geometry test. NMC has charge-clock evaluations only. These learned results describe the constructed simulator and do not override the failed optical-consistency or numerical gates. No conditional oracle was computed.

Table 124: Charge-clock minus time-clock skill at horizon 0.02, with direction and charge-state inputs. L/D denote lithiation/delithiation. Positive differences favour charge. Brackets are paired 95% particle intervals. Common/charge counts show the retained comparison support; an empty intersection is undefined.

Training rate	Test Da	Direction	Seed 0	Seed 1	Common/charge
C/20	0.00538	D	unavailable	2.825 [2.354, 3.436]	9240/9240
C/20	0.00538	L	unavailable	2.814 [2.349, 3.385]	10032/10032
C/20	0.05378	D	unavailable	0.252 [0.215, 0.289]	8976/8976
C/20	0.05378	L	unavailable	0.171 [0.123, 0.210]	9900/9900
C/20	0.17047	D	unavailable	0.369 [0.310, 0.426]	8184/8448
C/20	0.17047	L	unavailable	0.281 [0.178, 0.367]	9240/9504
C/20	0.35314	D	unavailable	0.102 [-0.002, 0.185]	5566/7942
C/20	0.35314	L	unavailable	0.079 [-0.092, 0.223]	6688/9064
C/20	0.77312	D	unavailable	undefined	0/7150
C/20	0.77312	L	unavailable	0.494 [0.451, 0.540]	1188/8470

Table 125: Charge-clock minus time-clock skill at horizon 0.16, with direction and charge-state inputs. L/D denote lithiation/delithiation. Positive differences favour charge. Brackets are paired 95% particle intervals. Common/charge counts show the retained comparison support; an empty intersection is undefined.

Training rate	Test Da	Direction	Seed 0	Seed 1	Common/charge
C/20	0.00538	D	unavailable	0.447 [0.325, 0.580]	7392/7392
C/20	0.00538	L	unavailable	0.314 [0.236, 0.375]	8184/8184
C/20	0.05378	D	unavailable	0.400 [0.302, 0.487]	7128/7128
C/20	0.05378	L	unavailable	0.326 [0.185, 0.467]	8052/8052
C/20	0.17047	D	unavailable	0.585 [0.493, 0.662]	6336/6600
C/20	0.17047	L	unavailable	0.553 [0.435, 0.667]	7392/7656
C/20	0.35314	D	unavailable	0.735 [0.663, 0.790]	3718/6094
C/20	0.35314	L	unavailable	0.704 [0.636, 0.778]	4840/7216
C/20	0.77312	D	unavailable	undefined	0/5302
C/20	0.77312	L	unavailable	0.239 [0.167, 0.320]	264/6622

Table 126: Charge-clock skill at horizon 0.16 for repeat, new-draw and conditional NMC transfer. Brackets are 95% particle intervals. There is no time-clock control for these arms.

Split	Test rate (C)	Direction	Seed 0	Seed 1
New draw	0.01611	D	0.845 [0.791, 0.884]	0.820 [0.753, 0.883]
New draw	0.01611	L	0.817 [0.758, 0.882]	0.826 [0.766, 0.881]
New draw	0.05	D	0.851 [0.798, 0.888]	0.823 [0.755, 0.886]
New draw	0.05	L	0.820 [0.764, 0.885]	0.829 [0.771, 0.885]
New draw	0.1589	D	0.857 [0.806, 0.895]	0.829 [0.754, 0.895]
New draw	0.1589	L	0.830 [0.776, 0.895]	0.840 [0.784, 0.894]
New draw	0.5	D	0.870 [0.815, 0.908]	0.843 [0.766, 0.909]
New draw	0.5	L	0.847 [0.794, 0.914]	0.860 [0.809, 0.911]
New draw	2	D	0.890 [0.830, 0.919]	0.850 [0.790, 0.890]
New draw	2	L	0.872 [0.819, 0.938]	0.886 [0.842, 0.934]
New draw	8.887	D	0.903 [0.830, 0.929]	0.931 [0.886, 0.946]
New draw	8.887	L	0.913 [0.859, 0.964]	0.930 [0.898, 0.959]
C/2 repeat	0.5	D	0.881 [0.834, 0.920]	0.873 [0.819, 0.916]
C/2 repeat	0.5	L	0.874 [0.841, 0.907]	0.852 [0.802, 0.897]
C/20 repeat	0.05	D	0.826 [0.768, 0.877]	0.817 [0.760, 0.866]
C/20 repeat	0.05	L	0.794 [0.729, 0.855]	0.792 [0.739, 0.844]
NMC rate	0.1067	D	0.973 [0.966, 0.980]	0.975 [0.968, 0.982]
NMC rate	0.1067	L	0.981 [0.977, 0.985]	0.981 [0.977, 0.985]
NMC rate	1	D	0.907 [0.880, 0.936]	0.912 [0.882, 0.942]
NMC rate	1	L	0.958 [0.946, 0.970]	0.962 [0.947, 0.975]
NMC rate	2	D	0.725 [0.634, 0.819]	0.733 [0.639, 0.830]
NMC rate	2	L	0.919 [0.888, 0.947]	0.907 [0.872, 0.939]
NMC rate	4	D	0.424 [0.239, 0.604]	0.463 [0.291, 0.632]
NMC rate	4	L	0.912 [0.880, 0.941]	0.885 [0.850, 0.915]

All six registered horizons and the unconditioned, direction/state and direction/progress/state input variants are retained in the evaluator-derived tables. The displayed horizons and direction/state panels are descriptive slices, not a selection criterion or a replacement for the failed real-data consistency gates.

J STEP-SIZE STUDY

This appendix tests sensitivity to the forecasting step size.

The default steps of the three clocks (0.005 of Q_{ref} in charge, one stored frame, 60 s) were never tuned. Most stored frames change little, and most optical change happens within short charge windows, so a model trained on every step spends capacity on near-identity targets and needs many rollout steps to reach a horizon. The step-size study retrains the forecasters over a grid of native steps for every clock, chooses each step on validation data only, and reports the full test matrix for every step. It also asks whether the charge-clock result of the main text has a step-size counterpart.

J.1 HYPOTHESES

We state the hypotheses registered before the step-size results were available. The following were registered before any result of this study existed. The rest of the paper had already inspected other experiments, so the registration is prospective for this study only; direct-prediction runs used as comparators are inherited and exploratory. *H1 (shape).* For each clock and model, skill against the autoregressive step has an interior, approximately unimodal maximum: very small steps give near-identity targets and long rollouts, very large steps lose temporal detail and training pairs. A monotone curve, a boundary optimum or reproducible separated maxima contradicts interior unimodality on the tested range; a noisy or flat curve is inconclusive, and no fitted peak is taken as evidence of its own existence. *H2 (invariance in charge).* The validation-optimal step, expressed in units of Q_{ref}, agrees within one candidate spacing between models trained at C/20 and at C/2, whereas the optimal time step should be about ten times longer at C/20 than at C/2 (the ratio of the measured currents) and the optimal frame step longer by the ratio of the charge per stored frame. The latter is not 10: frames are stored every 120 s at C/20 and every 30 s at C/2, and the median charge per stored frame is 0.0015 and 0.0039, so the predicted frame-step ratio is about 2.5. The primary comparison uses the two same-rate splits (Table 15), separately per model and clock; charge candidates agree when their indices differ by at most one, and frame or time candidates, converted to charge with the measured charge per step of the training cohort, agree when their log-charge separation is at most the larger local log-spacing of the grid at the two selected candidates. H2 is falsified on this grid by complete, interior, seed-stable optima more than one candidate spacing apart in charge; equal frame or time optima despite distinct charge rates also contradict its scaling prediction if their uncertainty excludes the predicted ratio. Flat, boundary-limited, incomplete or seed-unstable optima are inconclusive and are never counted as evidence of equality; the transfer splits are secondary checks. *H3 (large-step limit).* Direct one-pass prediction (App. D) reaches a horizon in a single step; it is compared at matching splits and horizons as an exploratory large-step comparator, not as the identical-objective limit of one-step training, and never enters step selection.

J.2 GRIDS AND NEAR-IDENTITY TARGETS

We specify candidate steps and measure how much their training targets change. Candidate native steps are 0.0025, 0.005, 0.01, 0.02 and 0.04 of Q_{ref} (charge), 1, 3, 10 and 30 stored frames (frame) and 60, 180, 600 and 1800 s (time). Every grid is built with the options of the default grids; rebuilding the graphite default reproduces it exactly. The NMC default differs only in stored temperature metadata (NaN versus 25), unused by these (d, s) models. Table 127 expresses each step in charge, minutes and stored frames at both rates. The near-identity fraction is the share of first-step training targets whose masked RMS change is below the optical noise propagated through that step's interpolation weights. The noise is the mask RMS of a per-pixel temporal MAD of valid second differences divided by $0.67448975\sqrt{6}$, which assumes a locally smooth signal and independent frame noise; for each pair of adjacent grid frames the threshold is this noise times the Euclidean norm of the difference of the two interpolation-weight vectors, which accounts for shared raw frames. Fractions are also reported at half and twice the threshold. The diagnostic is descriptive, was frozen before any result, and never selects a step or a checkpoint; it is a contrast between steps, not evidence that lithium does not move.

At the default steps, 11–12% (charge), 26–46% (frame) and 6–41% (time) of first-step training targets are below the noise threshold, falling to at most 1% at a charge step of 0.02; the default time step interpolates half-frames at C/20. At frame 30 only 34 of the 416 C/2 half-cycles keep

Table 127: Native steps expressed in charge (units of Q_{ref}), minutes and stored frames (medians over training segments at each rate), and the percentage of first-step training targets below the propagated noise threshold at $1\times$ ($2\times$) the threshold ($k = 4$, four unrolled steps, seed 0, same-rate split at that rate). C/2 segments: of 416 C/2 segments, those keeping at least eight grid points at that step; "–": no training support.

		C/20				C/2				
Clock	Step	Charge	Min	Frames	Near-id. %	Charge	Min	Frames	Near-id. %	C/2 segs
Charge	0.0025	0.0025	3.3	1.63	27 (61)	0.0025	0.32	0.64	19 (56)	416
Charge	0.005	0.0050	6.5	3.26	11 (41)	0.0050	0.64	1.28	12 (48)	416
Charge	0.01	0.0100	13.1	6.53	3 (20)	0.0100	1.29	2.56	3 (23)	416
Charge	0.02	0.0200	26.2	13.06	0 (6)	0.0200	2.58	5.13	1 (5)	416
Charge	0.04	0.0400	52.3	26.11	0 (2)	0.0400	5.16	10.26	0 (1)	416
Frame	1	0.0015	2.0	1.00	46 (72)	0.0039	0.50	1.00	26 (65)	416
Frame	3	0.0046	6.0	3.00	18 (52)	0.0117	1.51	3.00	3 (24)	416
Frame	10	0.0153	20.0	10.00	2 (16)	0.0390	5.03	10.00	0 (2)	416
Frame	30	0.0460	60.1	30.00	0 (3)	0.1170	15.09	30.00	–	34
Time	60 s	0.0008	1.0	0.50	41 (69)	0.0078	1.00	1.99	6 (34)	416
Time	180 s	0.0023	3.0	1.50	29 (63)	0.0233	3.00	5.96	0 (4)	416
Time	600 s	0.0076	10.0	4.99	4 (30)	0.0775	10.00	19.85	0 (0)	258
Time	1800 s	0.0229	30.0	14.97	0 (4)	–	–	–	–	0

eight grid points, none of them in the training or validation data of the same-rate C/2 split, and at 1800 s none do (600 s keeps 258), so the same-rate C/2 candidate sets are frames $\{1, 3, 10\}$ and times $\{60, 180, 600\}$ s, and an optimum at the largest feasible candidate is boundary-limited. For NMC (C/3, 1C, 2C), the charge steps correspond to 2.9–47.1, 1.6–26.2 and 1.2–19.6 stored frames, and the near-identity fraction at the default charge step is 3–5%.

J.3 Arms, selection and evaluation

We define the study arms, validation choices and common evaluation targets. Table 128 lists the arms. The main arm trains the (d, s) autoregressive U-Net with $k = 4$ on all thirteen grids and six graphite splits (the two same-rate splits, both rate-transfer splits, new-particle fold 0 and later cycles) with three seeds. FNO and Poseidon-T repeat the charge and frame grids; Walrus (head fine-tuning with the reduced evaluation of App. D) is run on two charge and two frame steps; the NMC arm repeats the charge grid on three NMC splits. Because the main arm's first-to-last input history spans $(k - 1)$ times the step, a matched-$k\Delta$ arm holds $k\Delta = 0.02$ of Q_{ref} (step 0.0025 with $k = 8$, 0.005 with $k = 4$, 0.01 with $k = 2$, 0.02 with $k = 1$). Its actual observed spans are 0.0175, 0.015, 0.010 and zero, respectively. This is a coupled step/history-count sensitivity analysis; it does not hold observed history duration fixed or isolate step size from history. Twenty-one combinations with no training or validation segment on their grid are not run (frame 30 on the same-rate C/2 and later-cycle splits; 1800 s on the same-rate C/2, C/2→C/20 and later-cycle splits), and the matched-$k\Delta$ arm reuses the identical main-arm runs at 0.005 and $k = 4$.

Table 128: Step-size arms (all charge-clock runs condition on (d, s); grouped validation; causal evaluation).

Arm	Grids	Splits, seeds	Runs
U-Net, $k = 4$	all 13	6 graphite, 0–2	219
FNO, $k = 4$	charge and frame	6 graphite, 0–1	104
Poseidon-T, full fine-tuning	charge and frame	6 graphite, 0	52
Walrus, head fine-tuning	charge 0.005, 0.02; frame 1, 10	C/20→C/2 and same-rate C/20, 0	8
U-Net, NMC	5 NMC charge grids	3 NMC, 0–1	30
U-Net, $k\Delta = 0.02$	charge 0.0025, 0.01, 0.02	6 graphite, 0–1	36

Each run is trained with its arm's standard validation loss for early stopping, identical across steps within an arm, because a common-lead-time validation inside training requires every validation horizon to be a whole number of native steps. The selection metric is computed afterwards by the test evaluator on the run's own validation segments: causal origins every 0.02 from $u = 0.14$ and charge horizons 0.005, 0.02 and 0.16, interpolating between the bracketing rollout frames exactly as at test;

V is the mean over the three horizons of $\sum \mathrm{SSE}_{\mathrm{model}} / \sum \mathrm{SSE}_{\mathrm{persistence}}$ over the mask. The primary V uses the validation windows common to every candidate step of the same arm, model, clock, split and seed; grids that drop short segments can change the validation membership, and a selection on such a candidate is flagged. Seeds are averaged with equal weight only when all prescribed seeds and candidates are complete, exact ties go to the smaller native step, and selection is separate per split, material, model, clock and history arm. No test score, stratum or rollout marker enters the selection; partial candidate sets may be tabulated but are never labelled the final optimum. Tests use horizons 0.005–0.32 with causal origins and full half-cycle rollouts, and are reported over all test windows, per direction (recombining SSE, never averaging skills), per training-defined transition window (App. K; available for the four transfer splits), per retrospective activity tertile, and for the full-rollout optical-marker error and switch-map correlation. Test windows pair across steps by their stored origin and horizon, which are identical across grids because causal origins snap to stored frames; paired differences against each clock's default step use matched windows and seeds and 2,000 particle-cluster bootstrap resamples (seed 340). These intervals condition on the selected steps and do not capture acquisition-level or selection uncertainty. The main-arm observed history spans three steps. The matched-$k\Delta$ arm also changes observed span, input count, interpolation, sample count and optimisation, so its differences do not causally isolate step size.

J.4 Results and registered verdicts

We report the available support and the verdict for each registered hypothesis.

Support audit. All available small outputs from the seven sweeps were fetched, including per-window records and validation parts. The expected matrix contains 486 rows (including 12 exact reuses): 435 complete, 21 prospectively infeasible, 13 with empty native validation, 10 with wholly empty scoring validation (nine unique runs and one reused row), four protocol exclusions and three unfinished Walrus runs at the snapshot. Twenty-two unique validation parts are provably empty: ten in trained runs and twelve without a checkpoint. One trained run retains a nonempty validation part alongside an empty part. The original-frame CPU audit reproduces 597 scored-part record counts; production training-window counts reproduce 64 completed controls. For example, acquisition 11, half-cycle 2 has usable frames only up to charge 0.0585, before the first scoring origin at 0.14, so every candidate origin fails history eligibility. The 13 absent checkpoints have zero native validation windows despite 378–2867 valid training windows. They cannot be recovered by evaluation alone. No seed, split, quality rule or origin was changed. Empty support remains missing selection evidence, never zero loss or a removed candidate. All expected cells, including missing cells and the reduced Walrus horizon protocol, remain in the released result tables.

H1: no universal interior optimum on the tested grid. Among the 126 main U-Net split/clock/horizon curves, 38 have the same boundary maximum in all three seeds, 12 have a reproducible interior unimodal shape, and 76 are inconclusive because of incomplete support, seed-variable shape or unresolved small secondary maxima. These are descriptive classifications of correlated curves, not independent hypothesis tests; no peak was fitted. Later-cycle short-horizon curves decrease towards coarser steps for all three clocks, contradicting universal interior unimodality on this grid. At same-rate C/20 and horizon 0.02, the frame curve peaks at three frames in every seed. At horizon 0.16, its seed winners are 3/3/10 frames; the time-clock winners are 1800/1800/600 s. Such noisy or boundary-limited cases do not establish an interior optimum. FNO has 35 boundary, 11 interior and 38 inconclusive curves; Poseidon-T's single-seed shapes remain descriptive. Figure 9 retains the failed small-step predictions as well as the useful ones.

H2: inconclusive. No applicable arm/clock supplies a complete, interior, seed-stable pair of rate-specific selections. U-Net's C/20 charge winners by seed are .0025/.005/.01; the frame winners are 1/1/3 and time winners 60/600/180 s. The charge losses at .005 and .01 nearly tie ($V = .637339$ and .637367); the selected frame step is a boundary. Every C/2 U-Net and FNO selection is incomplete. Poseidon-T's charge choices differ by three grid spacings (C/20:C/2 ratio .125), but the C/2 choice .04 is a boundary, so this does not falsify H2. Its frame choices have charge ratio 1.308, log separation .268 and allowed local log-spacing 1.204: numerical agreement, but with one seed stability is untested. The native ratio 3.333 is compared with the measured cadence-corrected prediction 2.549; the time-clock prediction is 10.143. None is evidence for charge invariance.

Table 129: Complete validation choices on the two same-rate splits. Entries are C/20 / C/2; "–" means the registered selection is incomplete, not equal to the other rate. A complete choice can still be boundary-limited or seed-unstable. Charge equivalents use measured training-cohort charge per step.

Model	Clock	Native choice	Charge equivalent	H2 verdict
U-Net	charge	.005 / –	.005 / –	inconclusive
U-Net	frame	1 / –	.001532 / –	inconclusive
U-Net	time	180 s / –	.002293 / –	inconclusive
FNO	charge	.0025 / –	.0025 / –	inconclusive
FNO	frame	3 / –	.004596 / –	inconclusive
Poseidon-T	charge	.005 / .04	.005 / .04	boundary-limited
Poseidon-T	frame	10 / 3	.015320 / .011713	one seed only

The complete C/20 U-Net time selection (180 rather than 60 s) improves skill from .3135 to .4141 at horizon .02, a paired gain of .1006 [.0549, .1549], and from .4039 to .5702 at .16, a gain of .1663 [.1477, .2557]. These use 3540 and 2931 matched window–seed pairs over 15 particles and all three seeds, with 100% coverage of default-step windows. Charge and frame selections at C/20 equal the defaults. Intervals are conditional on selection and do not quantify acquisition-level uncertainty. Figure 8 summarises the U-Net curves and the complete selections.

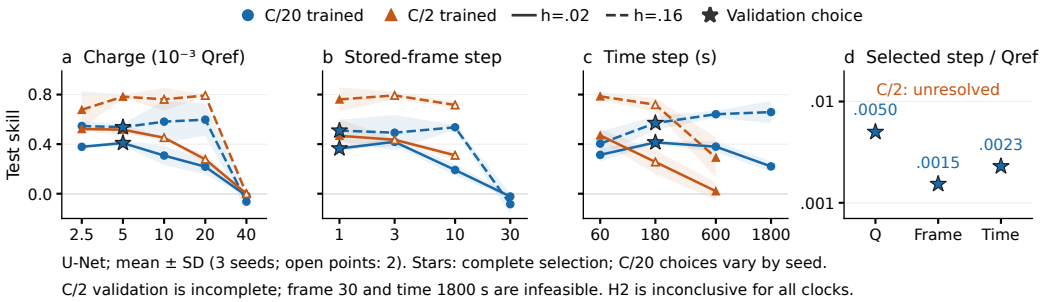

Figure 8: U-Net test skill against native step at charge horizons 0.02 (solid) and 0.16 (dashed), trained and tested at C/20 (circles) or C/2 (triangles); lines are equal-weight seed means and shading the seed SD (three seeds; open points two). Stars mark complete validation selections, which need not be interior or seed-stable optima; the last panel expresses those selections in charge units. C/2 has no complete selection, the C/20 frame choice lies on the grid boundary, and C/2 frame 30 and 1800 s have no training support.

History, architecture and material sensitivity. The matched-$k\Delta$ arm is not a fixed-duration control. At same-rate C/20 and horizon .16, changing $k = 4$ to $k = 8/2/1$ at steps .0025/.01/.02 changes skill by $-.1028$ [$-.1604, -.0801$], $+.1289$ [.0949, .1799] and $+.0694$ [.0274, .0964], respectively. Each uses 1954 identical window–seed pairs, seeds 0/1 and 15 particles (100% coverage); the .005/$k = 4$ comparison is identical by construction. C/2 comparisons at .01/.02 retain only one paired seed. These effects change sign and combine changes in step, observed span and input count (Fig. 10).

C/20 FNO chooses a boundary charge step (.0025) and an interior frame step (3), both varying by seed. Poseidon-T can fail at tiny steps (C/2 charge .0025 has skill -56.3 at horizon .02). Walrus arms are listed in Table 130: eight of twelve are complete, four were excluded because block validation removed a whole current direction, and the frame-1 arm from C/20 to C/2 awaits validation scoring. The C/20 charge arms favour .02 over .005 (skill .207 versus .027 at horizon .02), but two candidates and one seed cannot establish an interior maximum. NMC's selected charge steps are .02 for particle fold 0, .04 for 2C→C/3, and .0025 for C/3→1C,2C, with test skills .755/.920, .601/.877 and .799/.891 at horizons .02/.16. The transfer choices are boundaries, and the particle-fold seed winners differ (.02/.01).

Table 130: Walrus step-size comparisons. Head-adapted Walrus (1,500 steps, one seed) on a sparse origin lattice, evaluated to $\Delta = 0.08$; test skill at each arm's native clock step (charge step in s, frame step in frames). Every planned arm is listed: validation scoring pending means the test evaluation exists but the validation selection does not; unfinished training is n/e; arms whose block validation held out a whole current direction are excluded (n/a) and replaced by the automatic direction-preserving validation. These rows do not select the main-table recipe. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

Split	Validation	Clock	Native step	$\Delta = 0.02$	$\Delta = 0.08$	Seeds	Windows	Status
C/20→C/2	block	Charge	0.005	0.134	-0.041	1	581	complete
C/20→C/2	block	Charge	0.02	0.301	0.449	1	581	complete
C/20→C/2	block	Frame	1	0.043	0.017	1	583	complete
C/20→C/2	block	Frame	10	0.224	0.238	1	580	complete
C/20 repeat	automatic	Charge	0.005	0.027	-1.818	1	302	complete
C/20 repeat	automatic	Charge	0.02	0.207	-0.076	1	302	complete
C/20 repeat	automatic	Frame	1	-0.016	-0.000	1	302	complete
C/20 repeat	automatic	Frame	10	0.223	0.217	1	302	complete
C/20 repeat	block (whole direction)	Charge	0.005	n/a	n/a	–	0	excluded: validation held out a whole direction
C/20 repeat	block (whole direction)	Charge	0.02	n/a	n/a	–	0	excluded: validation held out a whole direction
C/20 repeat	block (whole direction)	Frame	1	n/a	n/a	–	0	excluded: validation held out a whole direction
C/20 repeat	block (whole direction)	Frame	10	n/a	n/a	–	0	excluded: validation held out a whole direction

J.4.1 H3: EXPLORATORY DIRECT PREDICTION, NOT AN IDENTICAL-OBJECTIVE LIMIT

We compare direct prediction with the validation-selected autoregressive step. At horizon .16, direct U-Net minus the validation-selected autoregressive step is $+.188$ $[.158, .240]$ for C/20→C/2, $+.083$ $[.059, .126]$ for C/2→C/20, $+.043$ $[.031, .160]$ for new particles and $+.075$ $[.043, .092]$ for later cycles. These comparisons pair seeds 0/1 and identical physical windows, covering 100% of direct windows (10010, 9038, 4284 and 2868 window–seed pairs, respectively). For C/20→C/2, paired direct and selected-autoregressive skills are .7995 and .6116. Poseidon-T is mixed: direct minus selected autoregression is $+.100$ $[-.001, .149]$, $-.109$ $[-.219, -.035]$, $-.069$ $[-.095, .172]$ and $+.220$ $[.198, .304]$ in the same split order, using seed 0 and 99.98%, 84.89%, 96.69% and 99.93% coverage. Its lower reverse-transfer coverage is explicit.

Both recipes condition on direction and state coordinate s; direct prediction also receives requested lead time and trains on seven horizons in one pass. Its $k = 4$ history on the .005 grid spans .015 and its checkpoint uses common-horizon validation, capped at 256 windows. The autoregressive U-Net trains four native rollout steps; the step-size Poseidon-T trains one step, both using native checkpoint validation followed by common-horizon step selection. The step-size history spans 3Δ. These conditioning, history, loss and validation differences prevent identifying a literal large-step limit from the observed direct-prediction gains.

Compute. Observed allocated GPU-hours from SLURM are 39.513 (U-Net), 10.086 (FNO), 53.117 (Poseidon-T), 6.520 (matched-$k\Delta$), 4.155 (NMC), and 23.638 (Walrus including the corrected sweep), 137.029 in total at the accounting snapshot; Walrus tasks that finished later make that total a lower bound. Failed-task time is included; concurrent processes sharing a GPU are not counted as separate allocated GPUs. Requested array wall-time bounds are reported separately in the released accounting table. One evaluation-only job was later submitted for the unfinished Walrus evaluation; known empty-support candidates are not resubmitted.

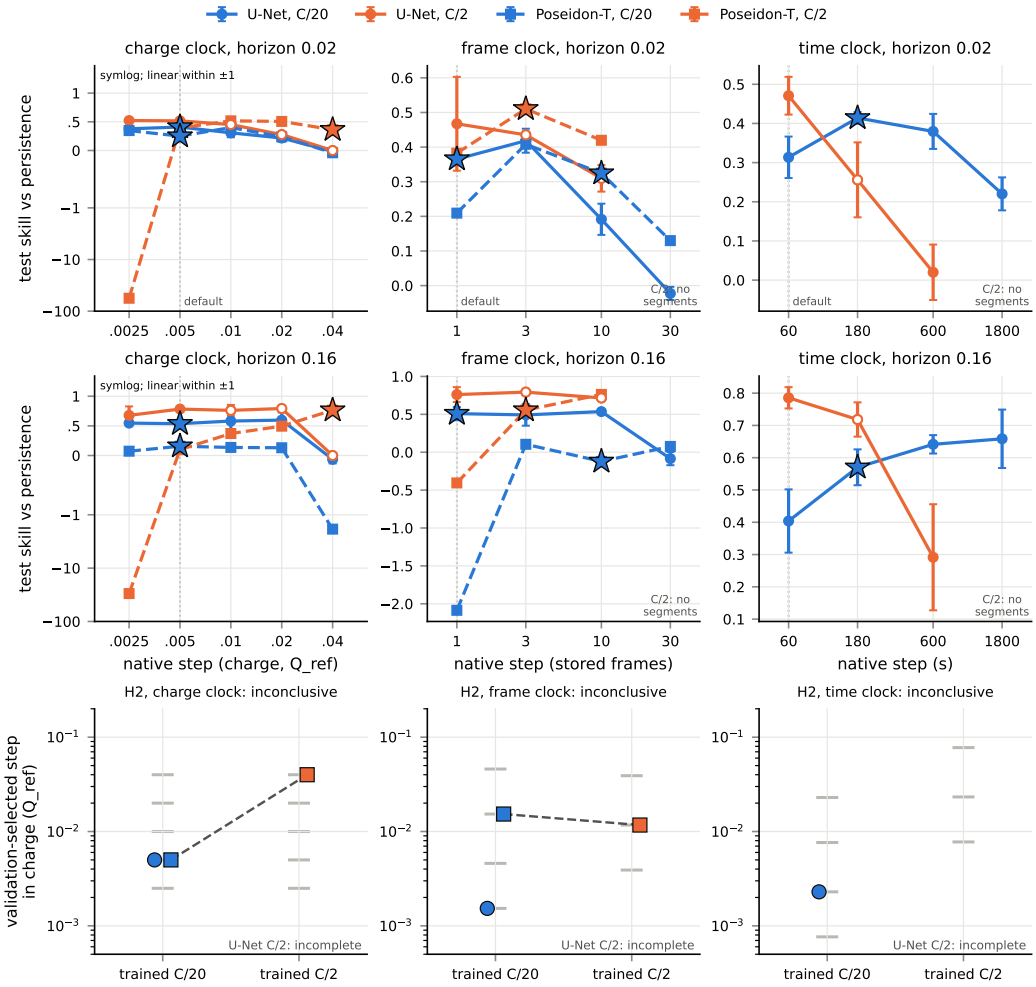

Figure 9: Step-size curves at horizons .02 and .16, with complete validation choices marked by stars. Blue/orange: C/20/C/2 training and same-rate testing. Solid circles: U-Net (mean ± SD, three prescribed seeds); dashed squares: Poseidon-T (one seed). Open ordinary points have incomplete test seed coverage. Charge panels use a labelled symmetric-log skill scale, linear within ±1, retaining extreme failures. Bottom: the same complete choices in charge units; grey ticks are candidates. Missing selections have no marker. All H2 comparisons remain inconclusive.

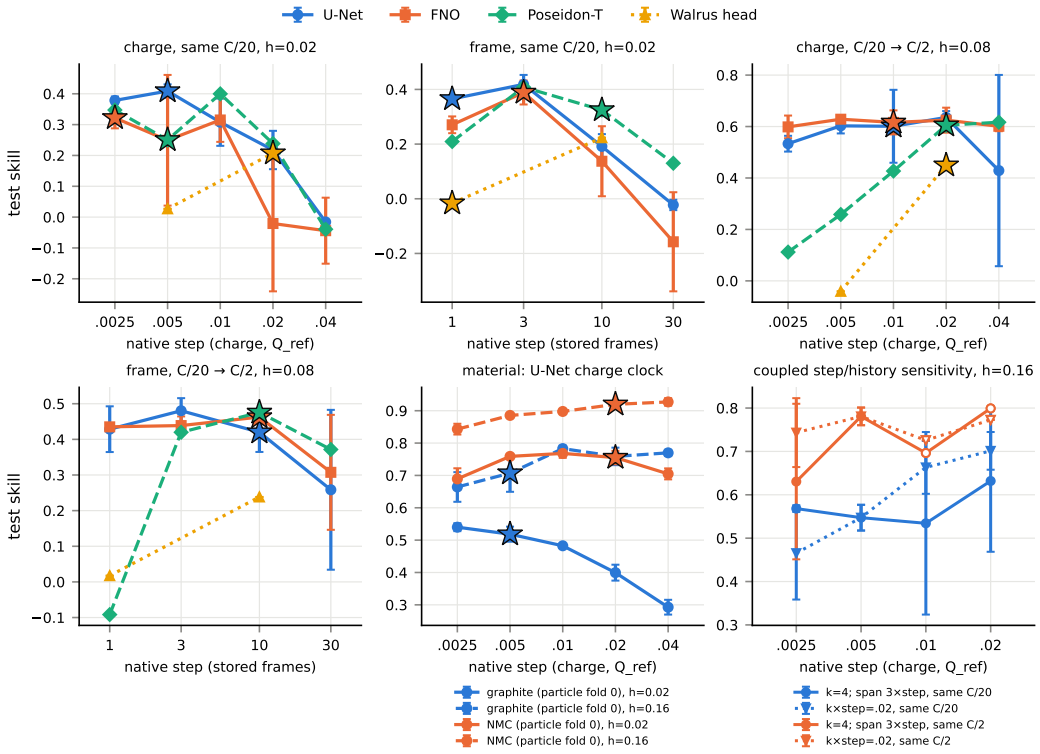

Figure 10: Architecture, material and coupled history sensitivity. First four panels: charge/frame clocks, same-rate C/20 at horizon .02 and C/20→C/2 at .08. Bottom middle: graphite and NMC particle fold 0 at .02/.16. Bottom right: $k = 4$ versus $k\Delta = .02$ on paired seeds and physical windows; the latter's actual observed spans are .0175/.015/.010/0. Open history-panel points have only one paired seed. Bands are not used; error bars denote seed SD. Missing runs remain absent. This arm does not hold observed history duration fixed.

K TRANSITION WINDOWS AND SPECIALIST MODELS

This appendix tests forecasts within fixed optical transition windows.

Graphite passes through three optical regimes in each half-cycle: a dimming region at low charge and two staging transitions (gap 1, 2L↔2, and gap 2, 2↔1; App. G). Earlier inspection of these movies suggested that they behave differently, so one forecaster for a whole half-cycle may be worse than forecasters specialised to one transition, and skill per transition is informative for every model. This is an exploratory hypothesis from previously inspected data; what is prospective is the analysis plan, fixed before any new run. The stratification itself is *prospective* in the sense of Sec. 4: the charge state s at a forecast origin follows from the current schedule, so selecting origins by s uses no future information, provided the window boundaries are fixed from training data alone.

K.1 FROZEN WINDOWS FROM TRAINING DATA

We construct transition windows using training data only. For each split and direction, windows are derived from the training acquisitions and training particle and cycle memberships only, after removing the union of the seed-0 and seed-1 validation partitions of both the pooled and the per-direction versions of the split; no test trace is read. The population trace is the projected-area-weighted mean of the background-subtracted log contrast over the original masks, on a uniform grid in $s = q/Q_{\text{ref}}$ with spacing 0.005. A grid point is used only if at least half of the eligible particle-segments are observed there, and each contiguous block then keeps only the segments observed throughout it (at least three), so that the cohort is fixed within a block; quality gaps are never bridged. The block with the strongest positive derivative carries the gap assignment of App. G (same code, same Savitzky–Golay smoothing of width 0.03 and order 2). T0 is the dimming region below the optical minimum, T1 is gap 1 and T2 is gap 2. Around each assigned derivative peak, the window is the contiguous region in which the absolute derivative exceeds 25% of that transition's peak, padded by 0.02 on each side, with overlapping pads split at the midpoint between adjacent cores; intervals are half-open, $[\text{low}, \text{high})$, and specific to the direction. Because quality gaps can separate the dimming region from the block that holds T1 and T2, T0 is located on the strongest dimming block below the gap-1 peak by the same rule; this clarification was made from training traces only, before any forecast result. Gap 2 is never invented: where the rule does not resolve it, T2 is recorded as unresolved and no threshold is relaxed. The window files record support, censoring and cohort membership and are frozen by checksum (Table 131).

Table 131: Frozen charge-state windows $[\text{low}, \text{high})$ in s for the dimming region (T0), gap 1 (T1) and gap 2 (T2), derived from training data only. *: the 25% region reaches the edge of the observed support (censored). Cohort: training acquisitions / particle-segments in the fixed cohort of the selected block. Splits as in Table 15.

Split	Direction	T0	T1	T2	Cohort
C/20→C/2	Lith.	[0.185, 0.3775)	[0.3775, 0.52)	[0.54, 0.755)	1 / 15
	Delith.	[0.31, 0.39)	[0.55, 0.705)*	[0.72, 0.99)	2 / 37
C/2→C/20	Lith.	[0.185, 0.375)	[0.375, 0.55)	[0.55, 0.645)*	7 / 158
	Delith.	[0.17, 0.3375)	[0.3375, 0.475)*	unresolved	7 / 142
New particle, fold 0	Lith.	[0.18, 0.3725)	[0.3725, 0.54)	[0.54, 0.635)*	10 / 163
	Delith.	[0.19, 0.395)	[0.395, 0.6)	unresolved	7 / 130
Later cycles	Lith.	[0.215, 0.3775)	[0.3775, 0.635)	unresolved	1 / 11
	Delith.	[0.195, 0.355)	[0.395, 0.5325)	[0.5325, 0.63)	1 / 11
C/20, C/2→2C	Lith.	[0.185, 0.375)	[0.375, 0.55)	[0.55, 0.635)*	10 / 214
	Delith.	[0.195, 0.4)	[0.4, 0.595)	unresolved	7 / 171
C/2, 25 °C→cold	Lith.	[0.165, 0.35)	[0.35, 0.46)	[0.46, 0.645)*	3 / 81
	Delith.	[0.195, 0.4)	[0.4, 0.535)	[0.535, 0.635)	1 / 45
C/2, 25 °C→hot	Lith.	[0.165, 0.35)	[0.35, 0.46)	[0.46, 0.645)*	3 / 81
	Delith.	[0.195, 0.4)	[0.4, 0.535)	[0.535, 0.635)	1 / 45
C/20, 25 °C→15, 45 °C	Lith.	[0.185, 0.365)*	[0.415, 0.555)*	[0.58, 0.825)	2 / 29
	Delith.	[0.165, 0.36)*	[0.48, 0.615)*	[0.615, 0.905)	2 / 29

K.2 ORIGIN-s ELIGIBILITY

We specify which forecast origins enter each transition-window comparison. A test window belongs to a transition window if its origin s, after causal snapping to a stored frame, lies in the interval of

its direction. Histories may start before the window and targets may leave it; a separately labelled sensitivity additionally requires the scheduled target s of each evaluated horizon to lie inside. No image activity, target error, final capacity or future switch time enters the selection. For training and validation of specialists the same rule is applied to the last history frame, before the condition statistics are fitted. Long half-cycle rollouts keep their origin at $u = 0.14$ and are filtered, not re-anchored, so the window evaluations use the four horizons $\Delta = 0.01$, 0.02, 0.04 and 0.08. Without a window, training and evaluation are unchanged (verified bit for bit against the previous code on synthetic data).

K.3 Specialist and general models

We compare models trained within a transition window with general forecasters. *Specialists* are trained and validated on one direction and one window and evaluated on the same window: U-Nets on the charge clock with (d, s) (seeds 0 and 1) and Poseidon-T with full fine-tuning and (d, s) (seed 0), each on the four splits C/20→C/2, C/2→C/20, new-particle fold 0 and later cycles, by two directions and three windows, with grouped validation and the default recipe of each model. Of the planned 48 U-Net and 24 Poseidon-T specialists, 42 and 21 are trainable; the others belong to unresolved T2 windows and remain in the matrix as such. Training and validation windows per specialist range from 216 to 6,879 and from 22 to 1,113 (U-Net). *General* models are the finished whole-half-cycle models of the main tables, re-evaluated on the same frozen windows: the main-recipe (d, s) U-Nets with retained weights (five seeds per split, three for the 2C split) and the general Poseidon-T (d, s) runs where finished, together with persistence, trend, AR(2), s-aligned mask-mean replay and history-aligned replay using the reference pools of the main tables. For finished runs whose weights were not retained, the saved per-window errors are filtered by the same frozen origin rule; predictions are unchanged, the reuse is labelled, and it is allowed only when the original evaluation used causal origins, the available-horizon policy, all four horizons and the same charge reference. Per-direction general models, where reported, are identified separately.

The primary contrast is specialist minus general mask skill on exactly matched origins, particles, direction, transition and horizon, reported separately for T0, T1 and T2, both directions and the four horizons, with paired particle-cluster bootstrap intervals (2,000 resamples, seed 31) that resample whole particle identities with all their windows. Seeds are reported separately and as seed-averaged paired differences; they are not bootstrap clusters. The full split × direction × transition × horizon × model × seed matrix is reported, including failures, empty strata, unresolved transitions and negative differences, together with matched and unmatched window counts and cells with zero persistence error. Recipe and checkpoint choices use validation data only; no window or recipe is tuned on held-out errors. Two confounds are known in advance: some inherited general Poseidon runs used block rather than grouped validation, which on the later-cycle split removed a direction from training, and the acquisition diversity of the training cohorts is small (Table 131), so the particle bootstrap does not license claims about new cells. As a material contrast, Poseidon-T (d, s) with the same one-step recipe is trained on the three NMC splits (C/3→1C and 2C, 2C→C/3, new-particle fold 0; seeds 0 and 1); no graphite transition labels are applied to NMC.

Table 132 reports every planned cell. For the U-Net, specialists fall below the general model on identical windows and seeds in 9 of 40 paired cells and above it in 4 (median difference -0.025, range -0.52 to $+0.08$); the Poseidon-T specialists beat the one-step general Poseidon-T, itself weaker than the U-Net, in 12 of 18 cells (median $+0.10$). Empty cells state why: no resolved window in the split, a general model not re-scored on the windows, no common seed, or no causal origins inside the window. Specialising to one transition therefore does not help a model that already sees the whole half-cycle.

L Extended related work

This appendix places the benchmark in its scientific and forecasting context.

Graphite staging connects lithium occupancy of interlayer galleries to distinct phases (Dahn, 1991). Guo et al. (2016) combined optical and Raman imaging with Cahn–Hilliard reaction dynamics to describe phase fronts in graphite crystals; Smith et al. (2017) developed a layered free-energy description with generalised Butler–Volmer kinetics. Lu et al. (2023b) connected operando optical

Table 132: Transition-window specialists: U-Net (d, s). Frozen optical transition windows T0 (low-s dimming), T1 (gap 1) and T2 (gap 2), per current direction, with causal origins inside the window. General models (trained on all training windows) are re-scored on the same frozen windows; specialists train and validate on one direction and window. Specialist and general skill on identical windows and common seeds; difference = specialist minus general, with a 95% paired particle-cluster bootstrap interval (2,000 resamples). Best baseline: the strongest of persistence, trend, AR(2), SOC replay and history-aligned replay on the same windows. n/r marks windows without a resolved training transition, general models that were not re-scored (checkpoint not retained) and pairs without a common seed; – marks windows that contain no causal forecast window. n/e: planned but not evaluated in this version (never a zero); n/r: no run planned for this cell; n/a: inapplicable; –: undefined (zero persistence SSE or no common windows).

			$\Delta = 0.02$				$\Delta = 0.08$				
Split	Window	Dir.	Specialist	General	Difference	Best base.	Specialist	General	Difference	Best base.	Seeds
C/20→C/2	T0	lith.	0.308	0.423	-0.115 [-0.298,-0.012]	0.170	0.288	0.341	-0.053 [-0.129,-0.008]	0.376	2
C/20→C/2	T0	delith.	0.308	0.494	-0.186 [-0.352,-0.163]	–	0.711	0.664	0.047 [-0.001,0.070]	0.562	2
C/20→C/2	T1	lith.	0.282	0.439	-0.157 [-0.187,-0.080]	0.196	0.677	0.683	-0.006 [-0.056,0.008]	0.539	2
C/20→C/2	T1	delith.	0.440	0.534	-0.095 [-0.180,-0.023]	0.414	0.469	0.437	0.032 [-0.038,0.153]	0.402	2
C/20→C/2	T2	lith.	0.392	0.419	-0.027 [-0.066,0.043]	0.321	0.294	0.415	-0.121 [-0.213,-0.070]	0.405	2
C/20→C/2	T2	delith.	0.548	0.503	0.044 [0.012,0.102]	0.406	0.387	0.308	0.079 [-0.018,0.513]	0.332	2
C/2→C/20	T0	lith.	0.423	0.409	0.014 [-0.060,0.087]	0.346	0.288	0.355	-0.067 [-0.265,0.035]	0.425	2
C/2→C/20	T0	delith.	0.271	0.441	-0.169 [-0.518,0.013]	–	0.437	0.507	-0.070 [-0.235,0.017]	–	2
C/2→C/20	T1	lith.	-0.106	0.411	-0.516 [-1.309,-0.276]	–	0.633	0.634	-0.000 [-0.026,0.052]	–	2
C/2→C/20	T1	delith.	0.619	0.631	-0.012 [-0.082,0.010]	–	0.510	0.551	-0.041 [-0.356,0.004]	–	2
C/2→C/20	T2	lith.	0.122	0.229	-0.107 [-0.203,0.065]	0.171	0.306	0.379	-0.073 [-0.183,0.032]	0.384	2
C/2→C/20	T2	delith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
New particle fold 0	T0	lith.	0.428	0.380	0.048 [-0.027,0.087]	0.090	0.414	0.410	0.004 [-0.294,0.068]	0.000	2
New particle fold 0	T0	delith.	0.636	0.617	0.019 [-0.044,0.023]	–	0.813	0.878	-0.065 [-0.069,0.012]	–	2
New particle fold 0	T1	lith.	0.459	0.492	-0.033 [-0.044,0.080]	–	0.640	0.788	-0.148 [-0.155,0.016]	–	2
New particle fold 0	T1	delith.	0.652	0.669	-0.017 [-0.023,0.045]	–	0.767	0.789	-0.022 [-0.028,0.039]	–	2
New particle fold 0	T2	lith.	0.320	0.399	-0.079 [-0.177,-0.055]	0.150	0.469	0.500	-0.031 [-0.251,0.004]	0.265	2
New particle fold 0	T2	delith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
Later cycles acq. 14	T0	lith.	0.589	0.579	0.010 [-0.010,0.063]	0.553	0.537	0.467	0.070 [0.014,0.169]	0.549	2
Later cycles acq. 14	T0	delith.	0.750	0.701	0.048 [-0.008,0.068]	0.526	0.918	0.911	0.007 [0.002,0.027]	0.763	2
Later cycles acq. 14	T1	lith.	0.529	0.587	-0.058 [-0.100,-0.038]	0.356	0.797	0.813	-0.017 [-0.027,0.006]	0.627	2
Later cycles acq. 14	T1	delith.	0.809	0.777	0.033 [0.011,0.063]	0.505	0.890	0.894	-0.004 [-0.016,0.038]	0.666	2
Later cycles acq. 14	T2	lith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
Later cycles acq. 14	T2	delith.	–	–	–	–	–	–	–	–	–

measurements with phase-field modelling across rates, resolving heterogeneity within and between particles. More recently, Han et al. (2026) observed rapid, localised intercalation events in dilute graphite stages and related them to static disorder. These results motivate testing whether particle-specific structure makes previous trajectories informative. They do not establish that transition locations or timings repeat across operating conditions. Our optical transition metrics assess reproducibility of the measured signal; identifying a staging mechanism or a local concentration field requires additional calibration and physical evidence.

Single-particle kinetics are embedded in a many-particle electrode. Reaction-limited intercalation theory and nonequilibrium charge-transfer kinetics provide the basis for Allen–Cahn reaction models (Singh et al., 2008; Bazant, 2013), while multiphase porous-electrode theory couples particle transformations through their environment (Smith & Bazant, 2017). Li et al. (2014) observed a current-induced change from sequential to concurrent particle intercalation, and Li et al. (2022) tracked changing population synchrony over cycling. Optical measurements also reveal communication between neighbouring NMC particles (Šedajová et al., 2026) and asymmetric flux in aged particles (Lun et al., 2025). These observations motivate distinct condition and cycle splits: nei-

Table 133: Transition-window specialists (continued): Poseidon-T, one-step fine-tuning.

			$\Delta = 0.02$				$\Delta = 0.08$				
Split	Window	Dir.	Specialist	General	Difference	Best base.	Specialist	General	Difference	Best base.	Seeds
C/20→C/2	T0	lith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/20→C/2	T0	delith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/20→C/2	T1	lith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/20→C/2	T1	delith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/20→C/2	T2	lith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/20→C/2	T2	delith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/2→C/20	T0	lith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/2→C/20	T0	delith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/2→C/20	T1	lith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/2→C/20	T1	delith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/2→C/20	T2	lith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
C/2→C/20	T2	delith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
New particle fold 0	T0	lith.	0.332	0.267	0.065 [0.047,0.156]	0.090	-8.448	0.075	-8.524 [-11.371,0.042]	0.000	1
New particle fold 0	T0	delith.	0.588	0.396	0.192 [-0.138,0.214]	–	0.774	0.459	0.315 [0.133,0.323]	–	1
New particle fold 0	T1	lith.	0.426	0.321	0.105 [0.094,0.194]	–	0.596	0.503	0.093 [0.085,0.367]	–	1
New particle fold 0	T1	delith.	0.674	0.427	0.248 [0.126,0.253]	–	0.691	-3.446	4.136 [0.334,4.503]	–	1
New particle fold 0	T2	lith.	0.191	0.261	-0.070 [-0.117,0.111]	0.150	0.160	0.201	-0.041 [-0.076,0.095]	0.265	1
New particle fold 0	T2	delith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
Later cycles acq. 14	T0	lith.	0.517	-0.214	0.731 [0.555,0.915]	0.553	0.142	0.015	0.127 [0.009,0.164]	0.549	1
Later cycles acq. 14	T0	delith.	0.698	0.637	0.061 [-0.007,0.082]	0.526	0.545	0.594	-0.049 [-0.081,0.083]	0.763	1
Later cycles acq. 14	T1	lith.	0.540	0.275	0.265 [0.170,0.292]	0.356	0.544	0.238	0.306 [0.203,0.357]	0.627	1
Later cycles acq. 14	T1	delith.	0.790	0.711	0.080 [0.050,0.103]	0.505	0.781	0.700	0.081 [0.034,0.106]	0.666	1
Later cycles acq. 14	T2	lith.	n/r	n/r	n/r	n/r	n/r	n/r	n/r	n/r	–
Later cycles acq. 14	T2	delith.	–	–	–	–	–	–	–	–	–

ther the cell current nor the charge passed uniquely specifies an individual particle's local flux. Re-observing the same particle controls its identity, but does not hold its neighbours, surface state or electrode environment fixed. Transfer within one observed electrode therefore cannot establish transfer to independent cells.

Hybrid physics–ML methods offer several ways to combine mechanistic structure with learnable components. Universal differential equations embed learned terms in differential equations (Rackauckas et al., 2020); APHYNITY augments incomplete physics with a constrained, minimum-norm learned correction (Yin et al., 2021b). Physics-informed neural networks penalise governing-equation residuals (Raissi et al., 2019), while PhyDNet separates physics-constrained latent dynamics from residual factors for video prediction (Le Guen & Thome, 2020). Related image inversion recovers constitutive behaviour from pattern evolution (Zhao et al., 2020) or correlative composition and strain measurements (Deng et al., 2022). These approaches address different inferential objectives despite sharing physical priors. In an optical forecasting benchmark, a successful physics-inspired model or learned correction demonstrates predictive usefulness under the evaluated split. It does not, by itself, identify a diffusivity, reaction law or missing mechanism from optical contrast. Such claims require an observation model and independent physical validation.

The choice of forecasting architecture also needs strong reference methods. U-Net was introduced for image segmentation (Ronneberger et al., 2015), and modern U-Net variants are competitive PDE surrogates (Gupta & Brandstetter, 2023). Recurrent networks have predicted microstructure evolution and cathode damage from phase-field simulation sequences (Yang et al., 2021; Zeng et al., 2024); their success motivates evaluation on measured images without establishing experimental transfer. Skill relative to a reference forecast follows established forecast-evaluation practice (Murphy, 1988). Same-particle replay is an analogue forecast (Lorenz, 1969): an earlier trajectory supplies a prediction indexed by charge. Comparing it with learned models tests whether additional flexibility improves on recurring particle behaviour. Persistence remains essential when images change slowly, and transition timing and spatial-change metrics complement average pixel error when most changes occur in brief intervals.

At cell level, battery ML already has substantial experimental precedents for prediction and transfer. Severson et al. (2019) predicted cycle life using early-cycle features, including voltage-indexed capacity differences, and Attia et al. (2020) combined early prediction with closed-loop optimisation of charging protocols. BatteryML organises models and data for degradation prediction (Zhang et al., 2024), while field-data studies emphasise changes in usage and operating conditions (Sulzer et al., 2021). Beyond scalar lifetime targets, Zhong et al. (2024) predicted discharge curves conditioned on composition, current and cycle, and Lu et al. (2023a) studied state-of-health transfer across battery types. Fluorescence-image learning also predicts discharge capacity (Negrete et al., 2025). These studies establish that learning electrochemical responses across conditions is already possible. Our contribution concerns spatially resolved future observations of re-identified particles, with condition transfer distinguished from particle transfer, rather than a general claim to introduce battery forecasting or transfer learning.

Finally, a charge coordinate should be distinguished from a sufficient physical state. Throughput ageing models retain temperature, rate or calendar-time effects (Wang et al., 2011; Schmalstieg et al., 2014); matching charge therefore need not remove history dependence. Likewise, neural controlled differential equations are invariant to reparameterising their control path, but including physical time as a channel preserves information about traversal speed (Kidger et al., 2020). Rate-dependent phase separation and morphology (Bai et al., 2011; Fraggedakis et al., 2020) motivate testing the limits of charge alignment. Our comparison measures the utility of the clock for forecasting under specified conditions, without assuming that a cell-level Coulomb count measures each particle's state of charge.