Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Uncertainty-Aware Drone Dispatch: Recording-Energy Prediction and Charging-Constrained Simulation

Samer Abaddi

eng.sameralabbadi@gmail.com

IDentity Research https://orcid.org/0000-0003-4612-7384

Research Article

Keywords: drone dispatch, energy prediction, ridge regression, grouped conformal calibration, discrete-event simulation, battery charging

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11593565/v1

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

Additional Declarations: The authors declare no competing interests.

Uncertainty-Aware Drone Dispatch: Recording-Energy Prediction and Charging-Constrained Simulation

Samer Abaddi

Independent researcher, IDentity Research, Research Department, Amman, Jordan (962)

ORCID iD 0000-0003-4612-7384

Correspondence eng.sameralabbadi@gmail.com

Abstract

How much service does a drone depot give up when it allows for error in a mission's energy forecast? We analysed 181 of 209 recordings from a DJI Matrice 100 flown on a common route in Pennsylvania. A ridge regression predicted whole-recording electrical energy from planned speed, payload and altitude. We estimated a one-sided allowance from the largest underprediction on each of 4 calibration dates. Across 34 recordings on 4 reserved dates, mean absolute error was 0.650 Wh and root mean squared error was 0.755 Wh. Every recording on each date fell below its nominal 80% upper estimate. A separate discrete-event simulation compared dispatch rules based on point and upper energy estimates from the development data. In 864 paired replications across 18 hypothetical depot scenarios, the upper rule completed fewer requests in every scenario and reduced reserve breaches in 16. For the same 2,259 requests in the reference scenario, upper estimates reduced completions from 977 to 955 and reserve breaches from 5 to 0. An open-access workbench lets readers vary the available charge and inspect the resulting dispatch decision.

Keywords. drone dispatch, energy prediction, ridge regression, grouped conformal calibration, discrete-event simulation, battery charging

Introduction

A drone can complete a delivery and still use energy the dispatcher meant to keep in reserve. In a worked example from this study, a vehicle has 35 Wh available. Including the same reserve and ground-idle allowance, the request needs 32.88 Wh under the point estimate and 35.15 Wh under the calibrated upper estimate. The first rule dispatches the vehicle. Under the second, the request waits. The energy estimates come from recorded missions and the battery state is hypothetical. At a depot serving requests in arrival order, that delay can pass to the requests behind it.

What should count as mission energy in that calculation? Rotor power varies through ascent, cruise and descent, while total energy also depends on how long each phase lasts. In an analysis of 188 quadcopter flights, median power differed by only about 5% between flights at 4 and 12 m/s. Slower flights over the same distance nevertheless consumed more energy because they took longer. The authors' phase-specific model achieved a mean absolute relative error of 2.1% on flights held back from fitting (Rodrigues et al., 2022). That statistic describes the typical size of an error.

The measurement window matters as much as the choice of predictor. The campaign described by Rodrigues et al. (2021) produced 209 records from a DJI Matrice 100 flown in Pennsylvania in 2019. The investigators varied commanded speed, payload and cruise altitude. The released archive contains timestamped voltage and current as well as the planned settings (Rodrigues et al., 2020). The electrical record extends onto the ground, and our audit found no independent observation of surface contact at takeoff or landing against which to verify a cutoff. The response used here is therefore the signed electrical energy integrated over the entire recording. Any additional ground idle in the simulation sits outside that window.

Prior research already links energy estimates to operating decisions. Iyengar et al. (2021) measured charging and discharging behaviour and used those measurements to schedule work across drones. Chagas et al. (2025) treated requests as stochastic arrivals and used chance constraints to plan routes under uncertain energy consumption. Ren et al. (2026) predicted energy status and arrival times to select skyway routes and recharging schedules. In electricity-network inspection, Yi et al. (2026) incorporated charger queues into a discrete-event model. These studies inform the choice of route, schedule or infrastructure. With the route, reserve and queue discipline already fixed, a different question arises. How many requests can the depot serve, and how often is the reserve breached, if only the energy allowance changes?

For that comparison, the relevant prediction error is one-sided. A positive difference between recorded and predicted energy can encroach on the reserve. The relevant unit of replication also needs care. Flights recorded on the same day may share conditions, so a random split by flight can place related observations on both sides of an evaluation. Conformal prediction uses the rank of errors on separate calibration data to set a prediction limit when the calibration and future scores are exchangeable (Angelopoulos & Bates, 2023). Grouped observations complicate that assumption (Dunn et al., 2023). Here, each calibration date contributes its largest signed difference between recorded and predicted energy. Coverage can then be reported for every recording and for each complete date. Extending that result to a new season or fleet would require evidence that the dates are comparable.

The study therefore sets two objectives. The first is to assess point accuracy and upper-bound coverage on reserved dates. The second is to quantify the service and reserve consequences of choosing between the two energy allowances. The analysis is restricted to 181 eligible recordings from one route, R1. A fixed ridge regression uses commanded speed, its reciprocal, payload and planned altitude, all available before dispatch. It cannot use the eventual recording duration to decide whether a request fits. The depot experiment uses predictions from 147 development recordings on 12 dates. A separate, one-time assessment uses 34 recordings on four other dates to evaluate the energy model.

The study makes three contributions.

- **A mission-energy allowance with an explicit scope.** The analysis connects a traceable recording window to a date-level calibration rule. With four calibration dates, the rule yields a finite upper allowance at a nominal 80% level. At 90% and 95%, the required rank exceeds the number of date scores. The number of dates therefore sets a limit on the finite allowances this calibration rule can support.
- **A controlled comparison of dispatch rules.** In each of 18 hypothetical depot scenarios, both rules face the same 48 request streams, with recording energy and duration kept together. The upper rule completed fewer requests in all 18 scenarios. It registered fewer reserve breaches in 16 scenarios and the same number in the other two.
- **A public, interactive dispatch workbench.** A data-centre dashboard brought metered loads and forecasts together for capacity planning (Abaddi, 2026). The AERIS companion website applies that display principle to drone dispatch. It lets readers select a supported flight plan, change available charge or reserve, and see whether one request fits. They can then inspect saved policy comparisons and replay selected queue and charger events to see how those decisions unfold over time.

Figure 1 maps the two uses of the flight records. Materials and Methods traces each branch from the electrical measurements to either a held-out prediction or a simulated dispatch decision. The Results section reports prediction error and coverage before the paired service outcomes. The Discussion and Limitations sections examine how those outcomes depend on the route, the calibration dates and the assumed charging process. The Conclusion returns to the operational trade-off between completing requests and preserving reserve.

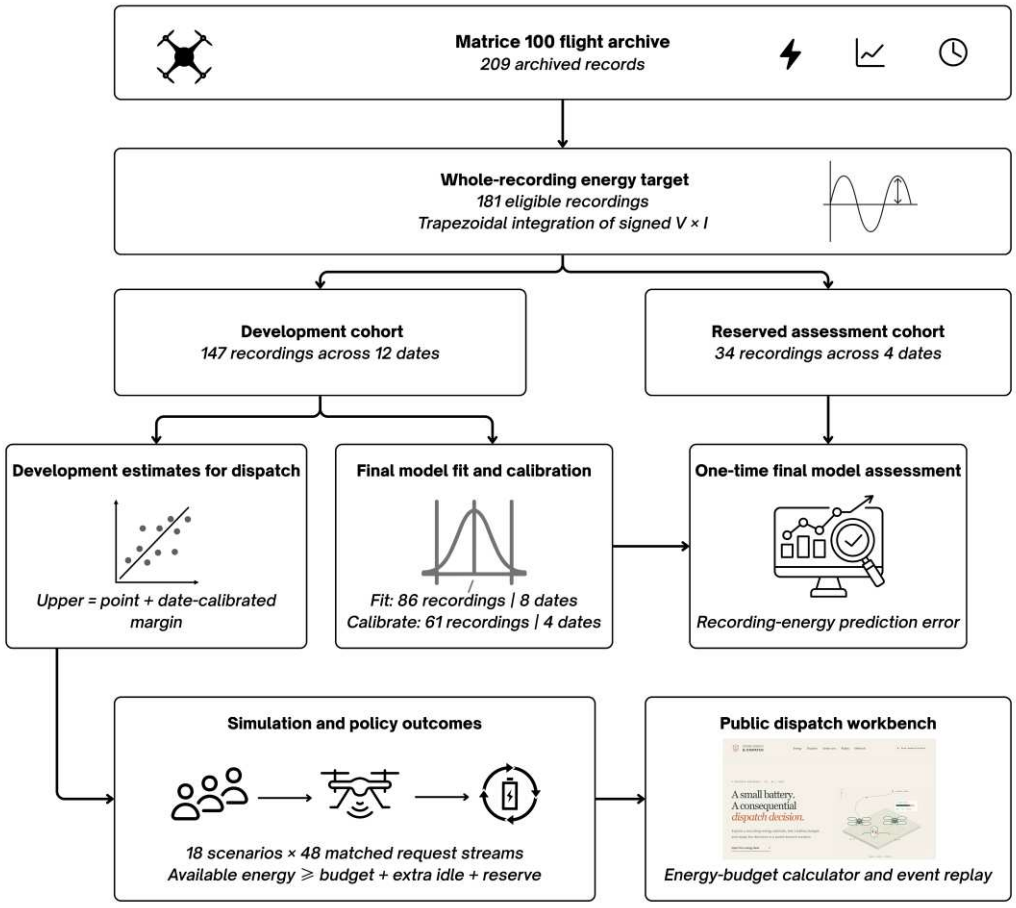


Figure 1. From Recorded Mission Energy to Dispatch Decisions and Model Evaluation

Materials and Methods

A dispatcher must commit a battery before the flight log reveals what the mission consumed. We preserved that order in two analyses. One estimated whole-recording energy and tested predictions once on reserved dates. The other compared two dispatch rules on identical requests using development predictions. The final model and its evaluation outcomes were kept separate from simulation inputs.

Flight records and eligibility

We used version 1 of Carnegie Mellon University's public quadcopter archive (Rodrigues et al., 2020). The flights took place at Penn Hills, Pennsylvania, United States, from April to October 2019 using a DJI Matrice 100 (Rodrigues et al., 2021). We verified the published checksums and

matched telemetry to mission plans by flight identifier. Before allocating dates, we checked all 209 recordings (257,896 rows, 21 sensor columns) for schema consistency, finite values, duplicate rows and times, positive voltage and strictly increasing time.

We retained route R1 with commanded speeds of 4, 6, 8, 10 or 12 m/s. Eligible payloads were 0, 250 or 500 g, and planned altitude had to be a single value of 25, 50, 75 or 100 m. The data descriptor reports 14 hover or idle recordings (Rodrigues et al., 2021). For reproducibility we followed the archived route labels, which identified 9 ground tests and 4 hover tests for exclusion. Fifteen additional recordings had a different route or an unsupported setting, including one 750 g payload. Applying these rules left 181 recordings on 16 dates. Eligibility did not depend on energy consumption, prediction error or a timing-gap flag.

Electrical measurement window

We calculated energy in watt-hours from the first to the last logged sample. In Equation (1), E_i is the electrical energy for recording i, n_i is its sample count, P_{ij} is voltage times signed current in watts, and t_{ij} is elapsed time in seconds. The divisor converts watt-seconds to watt-hours.

$$E_i = \frac{1}{3600} \sum_{j=1}^{n_i - 1} \frac{P_{ij} + P_{i,j+1}}{2} \left(t_{i,j+1} - t_{ij} \right) \quad (1)$$

Integration used the actual time intervals. We retained gaps longer than 0.5 s, assuming linear change in power across each gap. In each eligible recording, samples with a three-axis speed above 2 m/s had to show positive median current and more than 99% positive readings. Negative readings remained in the integral.

Collection dates and measurement provenance

We examined altitude and velocity rules on 34 development recordings, fixed the rules, and checked them on the remaining 113. An audit of the original sensor driver and CSV exporter then compared four prespecified raw Robot Operating System logs with 4,798 corresponding rows in the released tables. The exporter used an elapsed battery-message clock rounded to 0.01 s. The velocity vector came from sensor-frame odometry. The comparisons corroborated the current sign and were consistent with the reported clock rounding. Independent, synchronised evidence of takeoff or ground contact was unavailable, so the target retained logged ground intervals. The prediction limits therefore concern energy as recorded. Per-flight sensor bias and battery health could not be quantified.

To select final evaluation dates, we permuted the sorted list of eligible dates using a seeded NumPy RandomState (seed 20261030) and reserved the first four. The other 12 dates supplied all development data. The date came from campaign metadata because the sensor software clock did not reliably identify calendar dates. Figure 2 shows why the unit of evaluation matters. A date contributed between 1 and 21 eligible recordings, and reserved dates were interspersed with development dates. The design evaluates transfer between sampled dates within this campaign.

Prediction model and date-level calibration

For each flight we used commanded speed, payload, planned altitude and reciprocal commanded speed. On a common route the reciprocal term represents a duration dependence available at dispatch. Measured duration, distance, wind, electrical readings, collection date, local time and flight identifier were excluded from the predictor. A scaler centred and scaled each feature using the fitting subset alone. Ridge regularisation shrinks the coefficients of the correlated speed terms. The model minimised Equation (2), where T denotes the fitting records, z_i the standardised features, b the intercept and β the coefficients. The intercept was unpenalised. The penalty was fixed at 1 and the solver used singular value decomposition (Hoerl & Kennard, 1970).

$$\min_{b,\beta}\left\{\sum_{i\in T}(E_i - b - z_i^{\top}\beta)^2 + \lambda\|\beta\|_2^2\right\}, \quad \lambda = 1 \tag{2}$$

The initial development benchmark compared ridge with the training mean and median in five folds. Whole dates were assigned in descending order of recording count to the least populated fold. For the later uncertainty analysis, one of the 12 development dates was withheld in each outer fold. Of the remaining 11 dates, four calibrated the primary upper bound and seven fitted the model. Calibration dates were chosen by sorting a fixed SHA-256 hash of the seed, the omitted date and each candidate date. This produced one point estimate and one nominal 80% upper estimate for each replayed recording, with its entire date excluded from both fitting and calibration. Baseline development outcomes had already been inspected. Only the outer-fold estimates were used for dispatch.



Figure 2. Eligible recordings by collection date and final predictive role. The two development columns also supplied the dispatch replays. Source data are from Rodrigues et al. (2020).

A large underprediction on any flight can defeat an allowance intended to cover an entire date. For each calibration date d we therefore retained the largest signed residual M_d, using the fitted point prediction $\hat{E}_i$. Equation (3) gives the rank for a nominal level p and m calibration dates. In Equation (4), $M_{(k)}$ is the kth smallest date score. We used the exact integer rank and retained negative margins. The rank follows split conformal calibration (Angelopoulos & Bates, 2023). Using one score per date accommodates dependence among flights recorded together (Dunn et al., 2023).

$$M_d = \max_{i \in d} \left(E_i - \hat{E}_i \right), \quad k = \lceil (m+1)p \rceil \tag{3}$$

$$U_i(p) = \begin{cases} \hat{E}_i + M_{(k)} & \text{if } k \leq m \\ +\infty & \text{if } k > m \end{cases} \tag{4}$$

With four calibration dates, the 80% limit uses the fourth score. The 90% and 95% limits require a fifth and are unbounded. The prespecified comparator pooled calibration flights for a 90% upper estimate. Sensitivities varied the number of dates (3, 5 or 9), used symmetric absolute-residual intervals or added a training-error allowance. The date-maximum 80% rule was designated in advance. Nominal coverage depends on exchangeable date scores and comparable sampling within dates. Neither condition was established for future operations.

One-time predictive evaluation

For the final model we split the 147 development recordings into 86 fitting records on eight dates and 61 calibration records on four dates. The four calibration dates were the first in a separate, prespecified hash ordering of the development dates. Before the final run, the source-processing code was rehearsed on development recordings, and the input identities, runtime and code digests were checked. A single scaler and ridge model were then fitted, the upper margin was obtained from the four calibration-date scores, and predictions were assessed once on the 34 reserved recordings. Final outcomes were used for evaluation only.

For the point predictions we reported mean absolute error, root mean square error and mean signed bias in Wh. Bias was defined as predicted minus recorded energy. For upper estimates we reported the proportion finite, the margin above the prediction, signed headroom above the observation and the magnitude of any overrun. Equation (5) distinguishes the fraction of recordings covered from the fraction of dates for which every recording was covered. Here N counts recordings, D counts dates, and I_i equals 1 when recorded energy is at or below its upper estimate and 0 otherwise. Unbounded limits were reported separately from finite-bound coverage. Counts were reported against their actual denominators of 34 recordings and four dates, without treating flights on the same date as independent binomial trials.

$$c_{\mathrm{rec}} = \frac{1}{N} \sum_i I_i \,, \quad c_{\mathrm{date}} = \frac{1}{D} \sum_d \min_{i \in d} I_i \tag{5}$$

Dispatch simulation and matched requests

The discrete-event simulation represented a single depot. Each arrival needed an energy estimate for admission and an energy realisation for the ensuing mission. For the latter, the simulator kept

the recorded energy and duration of the same development flight together. Each stream drew a random starting position within one collection date, then cycled through its records in flight-identifier order. The date containing one recording necessarily replayed that same mission. Interarrival times were independent exponential draws at the scenario's fixed hourly rate. They represent hypothetical traffic. The archive contains no customer arrivals or charging histories.

Four random streams were generated for each of the 12 fixed development dates. The seed was 20261030 plus 1000 times the zero-based chronological date index plus the within-date index. Within each comparison, both policies received identical ordered requests, verified by a content hash. Sharing the random inputs allows paired comparisons of dispatch decisions (Glasserman & Yao, 1992). The design comprised 18 scenarios, 48 streams per scenario and two policies, giving 864 matched pairs and 1,728 policy runs. Table 1 gives the reference case. Supplementary Table S1 specifies every departure from it.

Battery, queues and charging

Table 1. Reference depot and battery settings (*S00*)

Input	Reference value
Requests and resources	Mean of 6 requests/h, 2 drones, 1 charger
Arrival horizon and deadline	8 h of arrivals, up to 30 min to dispatch
Nominal and usable energy	99.9 Wh nominal (TB47D), 80% usable (79.92 Wh)
Starting stock and reserve	100% of usable energy at opening, 15% reserve (11.988 Wh)
Idle outside the recording	12 W for 120 s (0.40 Wh)
Charging power and efficiency	60 W bulk setting, 100 W wall cap, 90% efficiency
Taper and realised demand	At 80% of usable capacity, charge at 50% of bulk power. Recorded-energy multiplier 1.0.

Note. The nominal pack ratings follow the manufacturer (DJI, 2026).

Let C be nominal pack energy times the assumed usable fraction and R the specified fraction of C held in reserve. In vehicle v, B_v denotes the current usable stock. For request i and policy π, $Q_{i\pi}$ is either the point estimate or its nominal 80% upper estimate. The additional idle draw uses power P for τ seconds outside the logged window. Equation (6) gives the minimum energy an idle vehicle must have before the request can be dispatched. The reserve is a policy threshold. The extra idle term is drawn from the battery only once.

$$B_v \geq Q_{i\pi} + E_{\text{idle}} + R, \quad E_{\text{idle}} = \frac{P\tau}{3600} \tag{6}$$

An unbounded estimate or a requirement exceeding full usable capacity was rejected on arrival. Other requests entered a strict first-come, first-served queue. The head request could depart through its deadline, 30 min after arrival. Among eligible idle vehicles the engine selected the most highly charged, breaking ties by vehicle identifier. If none could serve the head request, idle vehicles with depleted packs joined the charger queue. Later, smaller requests could not bypass it. A request still waiting at its deadline was deferred permanently.

After dispatch, the engine first debited additional idle energy, then spread recorded energy, multiplied by the scenario's demand factor, uniformly over its paired recording duration. The multiplier affected realised demand alone. A breach occurred when an attempt ended below reserve. A shortfall occurred when energy demanded exceeded energy supplied. Stock stopped at zero, unmet Wh were counted, and the vehicle remained booked for the full preparation and recording duration. A completed job could therefore breach reserve without suffering a shortfall.

Charging, closure and policy contrasts

Each drone had one modelled pack. Charging followed a first-in, first-out queue, ran to full charge without interruption, and used the two battery-side power levels in Equation (7). The charger setting and wall cap are P_{set} and P_{wall}. Here η is wall-to-battery efficiency and γ scales bulk power above the stored-energy threshold. Electricity drawn from the grid was stored energy gained divided by efficiency. Parked vehicles drew zero power by assumption. With no waiting request, vehicles started no new charge until closeout.

$$P_{\text{bulk}}=\min(P_{\text{set}},\eta P_{\text{wall}}), \quad P_{\text{taper}}=\gamma P_{\text{bulk}} \tag{7}$$

The production engine advanced to the earliest event. At the same timestamp it handled service, preparation or charge completions before arrivals, then deadlines, then closure, using insertion order within each priority. The queue was reconsidered after every event. Arrivals stopped at eight simulated hours. The engine recorded an eight-hour snapshot, resolved the remaining queue and attempts, and restored every pack to full. Keeping the post-arrival charging time and energy in the ledger prevented a policy from appearing to save energy simply by finishing the day with depleted packs.

The four primary outcomes were completed requests, end-of-attempt reserve breaches, energy shortfalls and deadline deferrals, each divided by all requests offered within the stream. We also retained attempted-mission breach rates, wait from arrival to dispatch for attempted jobs, delay to completion for successful jobs, charger utilisation during the arrival horizon, grid Wh and time to closeout. A change in mean completion delay can reflect a change in which requests finished.

For each scenario and outcome, the paired difference was the upper-policy value minus the point-policy value for the same stream. In Equation (8), D is the number of fixed date strata, L the streams per date, and the within-date mean and sample variance of the paired differences are $\bar{\delta}_d$ and s_d^2. Dates received equal weight. The approximate 95% Monte Carlo interval was the contrast plus or minus 1.96 standard errors. We left an interval undefined if either policy had an undefined metric in any pair. These intervals describe variation in synthetic streams conditional on the scenarios and recorded dates. Only four streams estimated the variance within each date.

$$\hat{\Delta}=\frac{1}{D}\sum_{d=1}^{D}\bar{\delta}_d\,, \quad \text{SE}_{\text{MC}}=\frac{1}{D}\sqrt{\sum_{d=1}^{D}\frac{s_d^2}{L}} \tag{8}$$

Implementation, verification and research materials

The analysis ran in CPU-only Google Colab notebooks. The recorded environment was 64-bit x86 Linux (kernel 6.6.159+, glibc 2.39), CPython 3.13.18, NumPy 2.1.3, pandas 2.2.3 and scikit-learn

1.6.1. Matplotlib 3.10.0 produced the original study plots. The production simulation used Python’s heapq event queue on a standard-memory runtime. The final evaluation reported two logical CPUs. The core run records do not identify a processor model or exact RAM allocation.

Using timestamps in the saved execution logs, the development calibration stage took 6.02 s from input verification to completion. The simulation took 16.33 s from its passed pre-run tests through the scenario runs and output generation. The final evaluation took 1.38 s from completion of the pre-opening checks through reconstruction and scoring. These are elapsed intervals for the recorded execution stages. They exclude cloud setup and transfers and are distinct from simulated mission or charging times.

Eleven deterministic cases tested the event engine before production. Every run checked terminal request counts, charger exclusivity, nonnegative stock and full replenishment. Battery, grid and demand ledgers each had to balance within 10^{-7} Wh. A second implementation in SimPy 4.1.2 checked 18 synthetic cases. Of 322 scalar comparisons, 75 had hand-calculated expectations for both engines and 247 compared the engines directly. Seventeen synthetic tests exercised the final evaluator. These checks concern implementation behaviour. Operational validity would require independent depot observations (Sargent, 2013).

The private AERIS repository preserves the analysis code, fixed parameters, scenario definitions and saved outputs at commit 3300ff4. The 142 release files were checked against their saved digests. Full reproduction in a fresh environment remains untested. The original telemetry is available under CC BY 4.0.

The public companion website is a static JavaScript application hosted on Cloudflare Pages. Users select a development date and a flight plan observed on it to obtain the point and upper energy estimates. They can vary battery stock, reserve and extra idle draw to test admission for one request. Its policy comparisons read saved outputs. Event replay uses stream 0 from the reference and combined-adverse scenarios under each policy. The private AnyLogic Cloud model (visual version 2, saved project format 8.9.10) processed 3 synthetic requests in 17 simulated seconds. A 10 s CF2X flight in gym-pybullet-drones 2.2.0 with PyBullet 3.2.7 took 38.43 s for simulation and frame capture. Both demonstrations were separate from the predictive evaluation and the dispatch experiment.

Results

All 34 reserved recordings fell below their calibrated upper estimates. Using upper estimates to dispatch drones reduced completions in all 18 scenarios. That comparison used predictions for 147 development recordings on separate dates.

What the electrical measurements supported

Of the 209 source recordings (Rodrigues et al., 2020), 181 across 16 dates met the eligibility criteria. Development energy ranged from 12.99 to 36.22 Wh (median 20.78 Wh), with a median duration of 182.5 s. Gaps longer than 0.5 s occurred in 16 of 147 development recordings and 2 of

34 reserved recordings. At most 1.67% of a development recording's energy was integrated across flagged gaps. The interpolation error is unknown.

The elevated-flight core accounted for a median of 91.03% of whole-recording energy. On the 113 confirmation recordings, the contact rule abstained on 55, flagged its candidate boundaries on 55 and returned provisionally consistent boundaries on 3. In four raw logs, all 4,798 released rows matched the source odometry fields. The raw logs supported the reported velocity frame and current sign. Ground contact remained unresolved. Prediction errors below refer to whole-recording energy.

Prediction from the flight plan

Across five date-grouped folds, mean absolute error (MAE) fell from about 4 Wh for both constant predictors to 0.623 Wh for ridge regression (Table 2). Ridge MAE was 0.583 Wh on 131 recordings without flagged gaps and 0.945 Wh on 16 with gaps. The groups differed in size and could also differ in flight plan.

Table 2. Error in predicting whole-recording energy from the flight plan

Predictor	Assessment	Records	MAE	RMSE	Bias
Training mean	Five grouped folds	147	3.990	4.902	+0.012
Training median	Five grouped folds	147	3.977	4.920	-0.443
Fixed ridge	Five grouped folds	147	0.623	0.825	-0.001
Fixed ridge	Outer dates	147	0.646	0.862	-0.062
Final ridge	Reserved dates	34	0.650	0.755	+0.194

Note. Errors are in Wh. Bias is prediction minus observation. Rows 1–3 share folds. Row 4 supplied the dispatch replays.

Calibration across recording dates

The development upper estimate covered 143 of 147 recordings and every recording on 10 of 12 dates. A mean allowance of 3.098 Wh above the point estimate produced that coverage. The four overruns occurred on just two dates, 2 of 12 recordings on 25 June and 2 of 8 on 29 June. They ranged from 0.021 to 1.265 Wh. Flight-level coverage was 97.28%. Complete-date coverage was 83.33% because a single overrun counted against the whole date.

The pooled-flight 90% comparator used a smaller mean allowance of 1.464 Wh and covered 138 of 147 recordings, including 7 complete dates. The training-error allowance averaged 0.550 Wh and covered only one complete date. With four calibration dates, the date-based 90% and 95% bounds were unbounded. Setting aside nine dates made the 90% bound finite and covered 11 of 12 complete dates, but left two fitting dates per fold and raised point RMSE to 1.048 Wh. No tested allocation yielded a finite 95% date-based bound. The symmetric date-based 80% interval averaged 6.218 Wh wide and covered 144 recordings and 10 complete dates. The pooled 90% interval was narrower (3.192 Wh) and covered 137 recordings and 5 complete dates. All allocations are reported in Supplementary Table S2.

One-time evaluation on reserved dates

The final model used 86 fitting recordings and 61 calibration recordings. Equation (9) gives predicted recording energy in Wh. The four standardised predictors are commanded speed, payload, planned altitude and reciprocal speed, respectively. The two speed terms vary together when commanded speed changes.

$$\hat{E}_i = 21.296 + 0.574z_{v,\,i} + 1.516z_{m,\,i} + 3.227z_{h,\,i} + 3.783z_{1/v,\,i} \tag{9}$$

The date-maximum calibration scores were 1.569, 2.005, 1.457 and 0.384 Wh in chronological order. The 24 June score therefore set the common upper allowance at 2.005 Wh. At nominal 90% and 95%, the required fifth score exceeded the four available, so neither upper estimate was finite. On the reserved recordings, point MAE was 0.650 Wh, root mean square error (RMSE) was 0.755 Wh and mean prediction minus observation was +0.194 Wh. The final evaluation used a different fit and different dates from the development comparisons in Table 2.

Where the upper estimate came closest

All 34 upper estimates were finite and covered the measured energy on all four evaluation dates (Figure 3). Headroom, the upper estimate minus recorded energy, averaged 2.199 Wh, with a median of 2.353 Wh and a range of 0.520 to 3.438 Wh. Flight 238 on 24 July came closest to the upper limit. Its point estimate fell 1.484 Wh below recorded energy. Adding the fixed allowance left 0.520 Wh of headroom, and the maximum upper overrun across the test was 0 Wh.

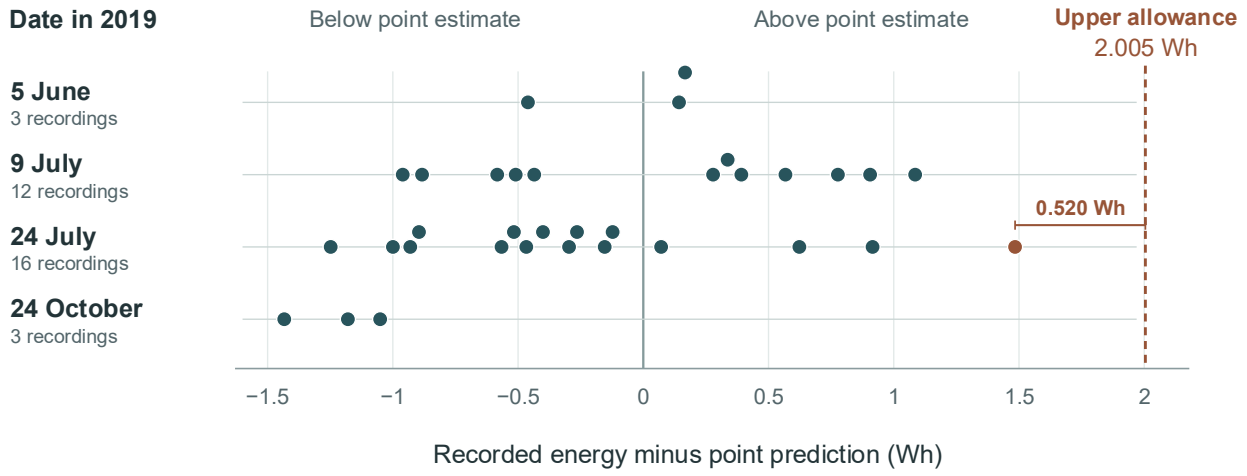


Figure 3. Prediction errors relative to the upper energy allowance on four held-out flight dates

Note. Each dot is one reserved recording. Positive values indicate underprediction. The bracket shows the smallest distance to the upper allowance.

The date with the largest MAE was 24 October (1.221 Wh, 3 recordings). All three point estimates on that date exceeded the observations. On 24 July, the date containing the closest approach to the upper limit, MAE was 0.622 Wh across 16 recordings. The other date-level MAEs were 0.257 Wh on 5 June and 0.643 Wh on 9 July. These four historical dates provide four opportunities to assess complete-date coverage.

Service and energy at the reference depot

The two policies faced the same 2,259 offered requests across the 48 reference streams. Point planning completed 977 and ended 5 attempts below reserve. Upper planning completed 955 and ended none below reserve. Neither policy exhausted a battery. The 22 fewer completions under upper planning were matched by 22 additional deadline deferrals. Every offered request in this scenario ended in completion or deferral. Reserve breaches were counted separately and could occur on completed requests.

Table 3. Service, charger use and grid energy under the two dispatch rules at the reference depot

Outcome	Point	Upper	Difference [95% MC interval]
Completed (%)	43.921	42.969	-0.952 [-1.566, -0.338]
Reserve breach (%)	0.198	0.000	-0.198 [-0.359, -0.037]
Energy shortfall (%)	0.000	0.000	0.000 [0.000, 0.000]
Deadline deferral (%)	56.079	57.031	+0.952 [0.338, 1.566]
Dispatch wait (min)	12.01	12.03	+0.02 [-0.44, 0.48]
Completion delay (min)	17.06	17.08	+0.02 [-0.44, 0.47]
Charger use (%)	87.52	87.97	+0.46 [-0.13, 1.04]
Grid energy (Wh)	465.19	456.17	-9.02 [-15.04, -3.00]
Closeout after arrivals (min)	120.13	113.79	-6.34 [-12.48, -0.19]

Note. MC denotes Monte Carlo.

The average stream completion rate fell by 0.952 percentage points under upper planning. Pooling all requests instead gives completion rates of 43.25% and 42.28% for point and upper planning. The two calculations weight streams differently. Table 3 retains the date-balanced comparison specified in Methods. Its intervals reflect random request streams conditional on the twelve development dates and the assumed depot.

The charger was occupied for nearly 88% of the arrival horizon under either rule. Upper planning drew 9.02 Wh less grid energy per stream on average and shortened mean time from the end of arrivals to full replenishment by 6.34 min. Fewer requests were served. Mean delay among completed jobs was about 17 min under both policies, although the set of completed requests changed.

Trade-offs across depots

Upper planning lowered the mean completion fraction in all 18 scenarios. The differences ranged from -0.133 percentage points under combined stress to -1.553 points with faster bulk charging. Nine of the 18 conditional intervals for completion included zero (Figure 4). End-of-attempt reserve breaches declined in 16 scenarios. Both rules recorded zero breaches in the no-extra-idle and tight-capacity cases.

Only three settings still produced upper-policy breaches. With a larger nominal pack there was 1 under upper planning versus 8 under point planning. Increasing realised mission demand by 10%

gave 8 versus 42. Combining stresses gave 2 versus 25. These are pooled event counts within each scenario. Figure 4 shows the corresponding changes in mean fractions per offered request.

The tight-capacity scenario shows why it matters how an unserved request leaves the queue. Of 2,259 offers, point planning rejected 108 immediately on capacity grounds and deferred 1,279 at the deadline. Upper planning rejected 312 immediately and deferred 1,077. Thus 202 fewer deadline deferrals accompanied 204 more capacity rejections and 2 fewer completions.

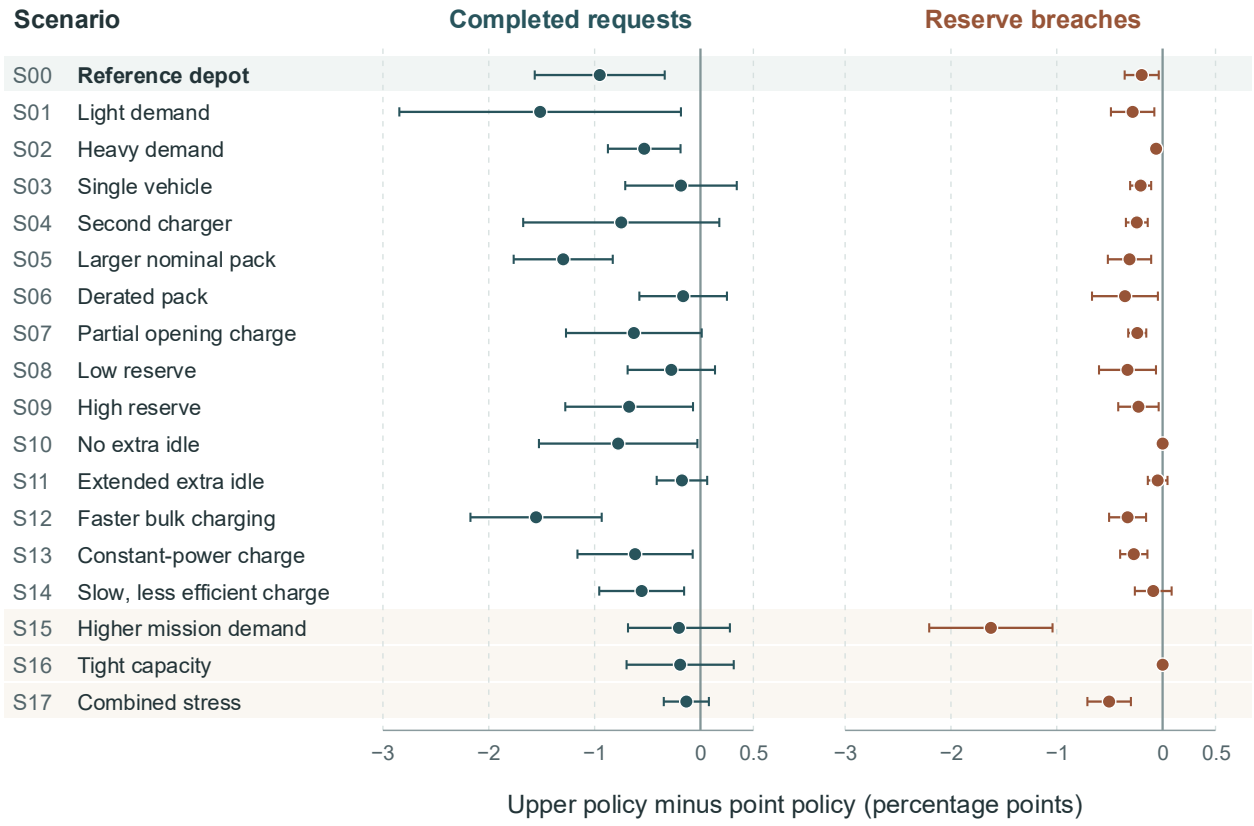


Figure 4. Changes in completion and reserve-breach rates across simulated depot scenarios

Note. Dots show the upper rule minus the point rule in percentage points per offered request, weighted equally across dates. Whiskers show approximate 95% Monte Carlo intervals conditional on the scenario and dates.

With combined stresses, point planning completed 392 of 4,519 requests and upper planning completed 386. The mean stream completion rates were 8.77% and 8.64%. The point policy incurred the experiment's only energy shortfall, leaving 0.908 Wh unsupplied. The upper policy had zero shortfalls but two reserve breaches. Mean breach rates per offered request were 0.54% and 0.04% for point and upper planning. Among attempted jobs the corresponding rates were 6.80% and 0.46%. Falling below the reserve threshold and running out of available energy were therefore distinct events in the same scenario.

Resource constraints also affected service within each rule. Under point planning, mean completion was 68.20% at the light arrival rate and 23.42% at the heavy rate. Adding a second charger to the reference depot changed mean completion from 43.92% to 57.03%. The corresponding upper-policy rates were 42.97% and 56.28%.

Execution and event checks

The Colab run completed all 1,728 Python policy runs. The two members of each of the 864 pairs had matching request records. Every terminal request was accounted for. Battery stock stayed nonnegative and the largest absolute energy-ledger residual was 1.76×10^{-12} Wh. Eleven deterministic engine checks passed. The separately coded SimPy benchmark agreed on all 322 scalar comparisons across 18 synthetic cases, including the 75 with hand-calculated expectations. The input rehearsal passed all 1,764 numerical comparisons on the 147 development records. All 17 synthetic tests of the final evaluator passed. Elapsed times were 6.02 s for development calibration, 16.33 s for simulation and output generation, and 1.38 s for final reconstruction and scoring. The private research release matched the saved file sizes and hashes at all 142 paths.

In the saved Python record for the three-request example, the vehicle served A from 5 to 8 s and B from 8 to 10 s. C reached its deadline at 10 s. Charging ended at 17 s with 12 Wh drawn from the grid and 10 Wh in the pack (Supplementary Figure S2). The private AnyLogic model replayed this case. Its native event log and test results are preserved in a separate author report. The CF2X flight visualisation completed 2,400 physics steps over 10 s, taking 38.43 s of wall time with frame capture. Position RMSE was 0.0929 m and maximum tracking error was 0.1114 m for that single trajectory. It supplied no electrical measurement for the Matrice 100 analysis.

Public workbench

The AERIS workbench reproduced all 147 development point and upper estimate pairs across its 12 date-specific models. The maximum absolute discrepancy for either estimate was 7.11×10^{-15} Wh. The browser also reproduced the 36 stored scenario-policy summaries. These checks establish numerical agreement with the archived analyses. They do not measure service availability over time or accuracy on additional flights.

For the supported 6 May plan at 4 m/s, zero payload and 25 m altitude, the point estimate was 20.50 Wh and the nominal 80% upper estimate was 22.76 Wh. The reference idle allowance added 0.40 Wh and the reserve added 11.988 Wh. At 35.00 Wh available, the point rule admitted the request and the upper rule waited. Figure 5 locates the narrow band of battery stock for which the same flight plan produces different decisions. The interactive companion lets the reader move the available stock while retaining the saved quotes and full-precision arithmetic.

Readers can inspect all 1,728 saved runs and replay four recorded event sequences containing 754 events. For stream 0 under combined stress, the upper-policy replay ended with 9 completions, 82 deferrals and 314.4 Wh of grid input from 91 offers (Supplementary Figure S1). That stream recorded zero breaches, while the 48-stream scenario recorded two. A single animation therefore omits a failure that is visible in the aggregate results. Public-browser tests recovered the estimate, selected results, replay and budget graphic. Playback reached closeout, and the 390-pixel mobile view had no horizontal overflow. Supplementary Figure S3 shows observed energy (Wh) and nominal 80% upper bounds

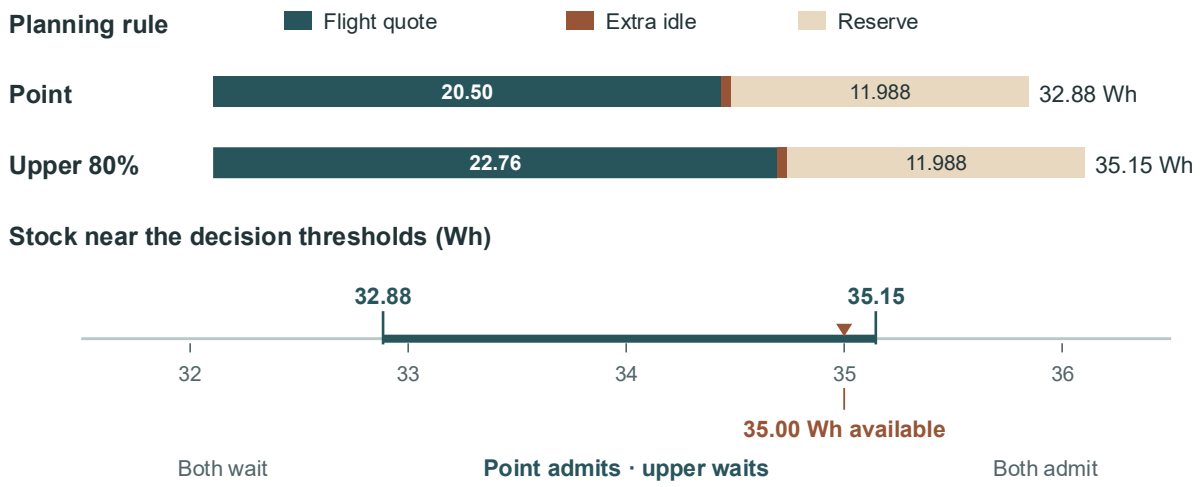


Figure 5. Battery stock required to dispatch the same flight under point and upper energy estimates

Discussion

The cost of a cautious energy estimate appears in the queue. AERIS achieved complete coverage of the reserved recordings, while its separate development-based simulation exchanged some completed requests for fewer reserve breaches. An energy allowance changes which requests fit, how long they wait and when charging occupies a vehicle. The study connects these consequences to a traceable measurement target and a fixed admission rule, making the operational cost of uncertainty visible.

What the energy predictions establish

The final mean absolute error of 0.650 Wh supports using planned speed, payload and altitude to estimate whole-recording energy within this campaign. The reciprocal-speed term provides a simple representation of duration dependence along the common route. All inputs are available at admission, when the eventual mission duration is still unknown. The fixed ridge model keeps the mapping from a flight plan to its energy quote inspectable within the sampled operating range.

The measurement window determines what that accuracy means. The retained target includes logged ground intervals and integration across timing gaps. It can therefore inherit changes in logging practice as well as changes in flight demand. The phase-specific model of Rodrigues and colleagues achieved 2.1% mean absolute relative error using the same underlying flight campaign (Rodrigues et al., 2022). Independent takeoff and landing references, together with synchronised electrical measurements, would allow future work to separate airborne demand from ground activity.

Calibration depends on the date sample

Coverage of all 34 reserved recordings represents success on four held-out dates. Those dates were interspersed within the historical campaign, so the result assesses transfer between sampled campaign conditions. A new season or fleet presents a further generalisation question. The development comparison illustrates why the date denominator matters. The pooled nominal 90%

comparator covered 138 of 147 recordings but only seven of twelve complete dates. The date-maximum 80% rule covered ten complete dates with a larger mean allowance. Hierarchical prediction research likewise shows why observations within a group require explicit treatment when constructing uncertainty sets (Dunn et al., 2023).

The unbounded estimates at higher nominal levels are informative about the data requirement. With four calibration-date scores, the declared rank supports a finite nominal 80% allowance. A nominal 90% or 95% allowance requires a rank beyond those four scores. Equation (3) requires at least nine or nineteen calibration dates, respectively, for a finite bound. Additional flights on the same dates leave that shortage unresolved. This is a consequence of finite-sample calibration (Angelopoulos & Bates, 2023). More dates must also leave enough observations for fitting. Exchangeability of historical and future date scores remains an assumption, particularly when the number of flights per date changes. These requirements turn the requested coverage level into a concrete sampling-design choice.

Where reserve protection costs service

At the reference depot, reserve breaches fell from five to zero while completions fell by 22 across the same 2,259 offers (Table 3). Requests admitted under the two rules consequently differed. A lower breach rate per offered request can accompany reduced exposure because fewer missions are attempted. Delay among completed jobs and total grid draw also describe policies serving different sets of work. An efficiency comparison would need to account for delivered service, replenishment and any alternative delivery arrangements.

Under strict first-come, first-served admission, an energy-demanding head request can hold up smaller requests behind it. In the tight-capacity scenario, 202 fewer deadline deferrals accompanied 204 additional capacity rejections. A deferral-only measure would favour moving unserved work out of the queue earlier. Reporting all terminal outcomes exposes that substitution. Routing research already considers uncertain demand, energy and ground-service decisions jointly (Chagas et al., 2025). AERIS holds those choices fixed to isolate the consequence of changing the energy quote within its declared depot model.

Charger congestion also gives the allowance a time cost. Energy-predictive service planning has used expected consumption and arrivals to choose routes and recharge schedules (Ren et al., 2026). Adding a second charger raised point-policy completion from 43.92% to 57.03%. The full-charge, non-interruptible charging rule helps explain why available energy and available vehicles can diverge. Empirical work shows that charging profiles vary with battery age and charging conditions (Iyengar et al., 2021). Measured charging histories would therefore be particularly valuable for replacing the assumed taper and evaluating partial-charge or alternative queue policies.

The direction of the service trade-off was consistent across the 18 scenario point estimates, although nine of their completion-difference intervals included zero (Figure 4). Only four streams per date estimated within-date variability. Reserve breaches also persisted under upper planning in three scenarios, including increased realised demand. Together, these results locate the value of the

allowance within the assumed operating conditions. The prediction bound covers recorded energy, while the depot must also accommodate changes in demand and available battery stock. A deployed reserve policy would need evidence about those changes as well as prediction coverage.

Making the process inspectable

The Python experiment and the public AERIS workbench connect aggregate outcomes to individual decisions. Changing a battery budget recalculates admission for one supported request, while scenario plots and animations read saved runs. The combined-stress example makes the value of this connection concrete. Its displayed upper-policy stream contains zero breaches, although the corresponding scenario contains two. Reading the animation alongside the full summary reveals behaviour that one trajectory would miss. This pairing makes the website a useful research companion whose calculations and displayed cases can be examined directly.

Related battery-recycling research combined machine-learning forecasting, discrete-event simulation and an enterprise resource planning interface (Abaddi, 2025). The relevant connection is the translation of an analytical model into a process that users can inspect. In AERIS, the AnyLogic demonstration exposes the three-request queue mechanism, while the independently coded SimPy benchmark checks event behaviour against the production engine. Numerical agreement and conservation checks address implementation correctness, following the verification logic described by Sargent (2013). The separate PyBullet trajectory concerns a different aircraft model. Establishing physical accuracy and operator usability would require their own observations and evaluations.

Limits of transfer

The case is bounded by one aircraft platform, a common route and a historical US campaign. Application in Jordan would require local operating evidence. Ground-contact timing, electrical sensor bias and usable battery capacity remain unresolved at the individual-flight level. The simulation draws mission energy and duration together from these records, preserving their pairing while repeatedly reusing a small set of missions. Real demand could couple payload, route and arrival time in different ways. Wind, degradation and interruption of charging would also alter the operating problem. Transfer therefore requires a separately evaluated campaign and depot observations that can challenge the chosen measurement window, queue discipline and charging assumptions.

Implications for research and deployment

Methodologically, the two analysis paths keep the final predictive assessment separate from the policy experiment. This allows a clear attribution of evidence to either prediction on reserved dates or decisions made with development estimates. Future comparisons should retain that separation, report recording and date coverage together, and evaluate completion alongside reserve breaches and rejections. The Monte Carlo intervals quantify synthetic-stream variation conditional on the fixed dates and assumptions. Broader uncertainty would require variation in campaigns, demand processes and battery characteristics.

The theoretical implication concerns the relationship between statistical coverage and operational performance. Calibration addresses the recorded energy response. Reserve preservation additionally depends on accepted missions, available stock, extra idle and realised demand. Evaluating the decisions induced by predictions is also central to decision-focused optimisation (Elmachtoub & Grigas, 2022). This study contributes a controlled case of that relationship through a fixed predictor and two admission rules. Learning an operationally preferred rule would additionally require explicit costs for missed service, reserve breaches and resource occupation.

For operators and organisations evaluating deployment, the workbench offers a way to examine these trade-offs before committing resources. Its supported plans and hypothetical battery settings define the scope of that examination. Organisations commissioning a pilot can use the same reporting structure to specify required evidence on completed service, reserve breaches, rejected work and charging demand. Comparisons should use matched opening stock and account for energy needed to reach a common closing charge. The private source archive and frozen outputs provide a basis for reproducibility, with full reproduction in a fresh environment still requiring verification. The practical contribution is a decision process that can be interrogated from energy quote to queue outcome, with its evidence requirements visible.

Limitations

Energy was measured between the first and last logged samples. Without independently timed takeoff and landing, we could not separate ground and airborne consumption with confidence. Energy attributed to gaps in the log also depends on the integration rule.

The flights came from one aircraft type on a common route. With 4 calibration dates, only the nominal 80% upper estimate was finite. The 90% and 95% estimates were unbounded. The 4 reserved dates tested transfer within that campaign. In the depot experiment, demand, usable battery capacity and charging behaviour were scenario inputs. The reported simulation intervals quantify variation across generated request streams under those assumptions.

Conclusion

A depot can record fewer reserve breaches by completing fewer deliveries. The more conservative rule reduced service in every scenario, and 3 scenarios still recorded breaches under that rule. AERIS makes the trade-off traceable from the energy recorded for a mission to the decision to release a drone. Its public workbench lets readers select a flight plan and identify the range of battery charge over which the two rules disagree.

For a field pilot, we recommend reporting completed requests and reserve breaches against the same incoming demand. The next test should pair time-stamped requests and charger occupancy with synchronised flight and power records on new operating dates. Repeating that comparison on a second route or aircraft would test whether the service cost of protecting the reserve persists.

Declarations

Author contributions. Samer Abaddi conceived and designed the study, curated the data, developed the software and simulation, analysed and interpreted the results, and wrote the manuscript. As sole author, he takes responsibility for the entire work.

Funding. This study received no funding or in-kind support.

Competing interests. The author declares no financial or nonfinancial competing interests relevant to this work. The affiliation (IDentity Research) is the author's micro-startup.

Ethics and consent. The analysis used public aircraft telemetry and simulated requests. No human participants were enrolled, and no questionnaire responses or identifiable participant data were used. Human-participant ethics approval, consent to participate and consent for publication were therefore inapplicable.

Use of AI tools. The author used Grammarly AI and Microsoft Copilot to improve wording and organise the manuscript. The author takes full responsibility for the work in its current form.

Data and code availability

The source telemetry and flight settings are available from Carnegie Mellon University's KiltHub under the Creative Commons Attribution 4.0 International licence (Rodrigues et al., 2020).

The AERIS repository is private. At commit 3300ff4, it contained 142 files and its main branch had a history of 22 commits. It includes 12 Python files in the code directory, a development-check notebook, 4 derived development datasets, 3 model-parameter files, 22 result files, scenario definitions, tests, figures and the website source. A full rerun from raw data in a fresh environment remains unverified. Reviewers may request repository access from the corresponding author.

Supplementary videos

Four supplementary videos accompany the manuscript.

Video 1. The AnyLogic depot simulation and its synchronised Python flight animation.

Video 2. Navigation of the public workbench at drone-energy-dispatch.pages.dev.

Video 3. Playback of a saved dispatch run under a selected scenario and policy on the workbench.

Video 4. A snapshot of the private GitHub repository, its directory tree and study documentation.

List of abbreviations

Term	*Meaning*	*Term*	*Meaning*
ACM	Association for Computing Machinery	**MAE**	Mean absolute error
AERIS	Audited Energy and Reserves In Simulation	**MC**	Monte Carlo
AI	Artificial intelligence	**NumPy**	Numerical Python
CC BY	Creative Commons Attribution	**ORCID**	Open Researcher and Contributor ID
CF2X	Crazyflie quadcopter in X configuration	**PyBullet**	Python interface to the Bullet engine
Colab	Google Colaboratory	**RAM**	Random-access memory
CPU	Central processing unit	**RMSE**	Root mean squared error
CPython	Reference Python interpreter in C	**SE_{MC}**	Monte Carlo standard error
CSV	Comma-separated values	**SHA-256**	Secure Hash Algorithm (256-bit digest)
DJI	Da-Jiang Innovations	**SimPy**	Discrete-event simulation library for Python
ERP	Enterprise resource planning	**UAV**	Unmanned aerial vehicle
glibc	GNU C Library	**US**	United States
ID	Identifier		

Units. g, gram · h, hour · m, metre · m/s, metres per second · min, minute · s, second · W, watt · Wh, watt-hour.

Identifiers. S00–S17 identify depot scenarios. TB47D and TB48D are DJI battery models. x86 identifies the processor architecture.

References

Abaddi, S. (2025). A proposed electric/hybrid batteries recycling hub with ERP implementation and simulation. *Journal of Simulation.* Advance online publication. https://doi.org/10.1080/17477778.2025.2534661

Abaddi, S. (2026). The optimization and load forecasting of power distribution units in data halls. *Applied Operations and Analytics, 2*(1), 1–25. https://doi.org/10.1080/29966892.2026.2649444

Angelopoulos, A. N., & Bates, S. (2023). Conformal prediction: A gentle introduction. *Foundations and Trends in Machine Learning, 16*(4), 494–591. https://doi.org/10.1561/2200000101

Chagas, G. O., Coelho, L. C., Laganà, D., & Beraldi, P. (2025). A dynamic drone routing problem with uncertain demand and energy consumption. *Transportation Research Part B: Methodological, 202*, Article 103335. https://doi.org/10.1016/j.trb.2025.103335

DJI. (2026). *Support for Matrice 100.* Retrieved September 1, 2026, from https://www.dji.com/support/product/matrice100

Dunn, R., Wasserman, L., & Ramdas, A. (2023). Distribution-free prediction sets for two-layer hierarchical models. *Journal of the American Statistical Association, 118*(544), 2491–2502. https://doi.org/10.1080/01621459.2022.2060112

Elmachtoub, A. N., & Grigas, P. (2022). Smart "predict, then optimize". *Management Science, 68*(1), 9–26. https://doi.org/10.1287/mnsc.2020.3922

Glasserman, P., & Yao, D. D. (1992). Some guidelines and guarantees for common random numbers. *Management Science, 38*(6), 884–908. https://doi.org/10.1287/mnsc.38.6.884

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Iyengar, S., Saxena, R. R., Pal, J., Chhaglani, B., Ghosh, A., Padmanabhan, V. N., & Venkata, P. T. (2021). Holistic energy awareness for intelligent drones. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 41–50). Association for Computing Machinery. https://doi.org/10.1145/3486611.3486651

Ren, G., Shahzaad, B., Alkouz, B., Lakhdari, A., & Bouguettaya, A. (2026). Energy-predictive planning for optimizing drone service delivery. *Expert Systems with Applications, 297*, Article 129251. https://doi.org/10.1016/j.eswa.2025.129251

Rodrigues, T. A., Patrikar, J., Choudhry, A., Feldgoise, J., Arcot, V., Gahlaut, A., Lau, S., Moon, B., Wagner, B., Matthews, H. S., Scherer, S., & Samaras, C. (2020). *Data collected with package delivery quadcopter drone* (Version 1) [Data set]. Carnegie Mellon University. https://doi.org/10.1184/R1/12683453.v1

Rodrigues, T. A., Patrikar, J., Choudhry, A., Feldgoise, J., Arcot, V., Gahlaut, A., Lau, S., Moon, B., Wagner, B., Matthews, H. S., Scherer, S., & Samaras, C. (2021). In-flight positional and energy use data set of a DJI Matrice 100 quadcopter for small package delivery. *Scientific Data, 8*, Article 155. https://doi.org/10.1038/s41597-021-00930-x

Rodrigues, T. A., Patrikar, J., Oliveira, N. L., Matthews, H. S., Scherer, S., & Samaras, C. (2022). Drone flight data reveal energy and greenhouse gas emissions savings for very small package delivery. *Patterns, 3*(8), Article 100569. https://doi.org/10.1016/j.patter.2022.100569

Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation, 7*(1), 12–24. https://doi.org/10.1057/jos.2012.20

Yi, Z., Feng, Z., Bohan, Z., Jianguo, M., & Yutao, Q. (2026). Research on collaborative optimisation of smart inspection and charging networks for distribution network UAVs based on digital twins. *Digital Twins and Applications, 3*(1), Article e70034. https://doi.org/10.1049/dgt2.70034

Supplementary material

Table S1. Frozen dispatch scenarios and their departures from the reference case

ID	Scenario	Inputs changed from S00
S00	Reference	All reference values in Table 1.
S01	Light demand	Arrival rate 3 requests/h.
S02	Heavy demand	Arrival rate 12 requests/h.
S03	Single vehicle	Fleet reduced to 1 drone.
S04	Second charger	2 chargers available.
S05	Larger nominal pack	TB48D, nominal pack energy 129.96 Wh.
S06	Derated pack	Usable capacity 60% of nominal.
S07	Partial opening charge	Starting stock 60%.
S08	Low reserve	Reserve 5%.
S09	High reserve	Reserve 25%.
S10	No extra idle	Extra idle power 0 W, extra idle time 0 s.
S11	Extended extra idle	Extra idle power 60 W, extra idle time 300 s.
S12	Faster bulk charging	Bulk setting 90 W.
S13	Constant-power charging	Taper power 100% of bulk power.
S14	Slow, less efficient charging	Bulk setting 30 W, wall-to-battery efficiency 85%.
S15	Higher mission demand	Recorded-energy multiplier 1.10×.
S16	Tight capacity	Usable capacity 35% of nominal.
S17	Combined stress	Usable capacity 60% of nominal, starting stock 60%, reserve 5%, arrival rate 12 requests/h, extra idle power 60 W, extra idle time 300 s, bulk setting 30 W, wall-to-battery efficiency 85%, recorded-energy multiplier 1.10×.

Note. Starting stock, reserve and taper threshold are fractions of usable capacity.

Table S2. Development coverage and energy allowance for selected fixed alternatives

Method	Fit / cal. dates	Finite	Recordings covered	Dates covered	Mean Wh
Training-error upper	7 / 4	147	114/147	1/12	0.550
Pooled upper 90%	7 / 4	147	138/147	7/12	1.464
Date-based upper 80%	8 / 3	0	—	—	—
Date-based upper 80%	7 / 4	147	143/147	10/12	3.098
Date-based upper 80%	6 / 5	147	143/147	10/12	3.273
Date-based upper 90%	7 / 4	0	—	—	—
Date-based upper 90%	2 / 9	147	145/147	11/12	4.299

Method	Fit / cal. dates	Finite	Recordings covered	Dates covered	Mean Wh
Date-based upper 95%	2 / 9	0	—	—	—
Date-based symmetric 80%	7 / 4	147	144/147	10/12	6.218
Pooled symmetric 90%	7 / 4	147	137/147	5/12	3.192

Table S3. Terminal request counts by scenario and policy

ID	Offers	Completed	Deferred	Capacity rejected	Reserve breach	Shortfall
S00	2,259	977 / 955	1,282 / 1,304	0 / 0	5 / 0	0 / 0
S01	1,142	765 / 750	377 / 392	0 / 0	4 / 0	0 / 0
S02	4,519	1,048 / 1,025	3,471 / 3,494	0 / 0	3 / 0	0 / 0
S03	2,259	731 / 728	1,528 / 1,531	0 / 0	5 / 0	0 / 0
S04	2,259	1,278 / 1,258	981 / 1,001	0 / 0	6 / 0	0 / 0
S05	2,259	1,008 / 979	1,251 / 1,280	0 / 0	8 / 1	0 / 0
S06	2,259	925 / 921	1,334 / 1,338	0 / 0	7 / 0	0 / 0
S07	2,259	865 / 851	1,394 / 1,408	0 / 0	6 / 0	0 / 0
S08	2,259	1,001 / 995	1,258 / 1,264	0 / 0	7 / 0	0 / 0
S09	2,259	928 / 914	1,331 / 1,345	0 / 0	5 / 0	0 / 0
S10	2,259	990 / 974	1,269 / 1,285	0 / 0	0 / 0	0 / 0
S11	2,259	799 / 795	1,460 / 1,464	0 / 0	1 / 0	0 / 0
S12	2,259	1,336 / 1,301	923 / 958	0 / 0	8 / 0	0 / 0
S13	2,259	1,174 / 1,161	1,085 / 1,098	0 / 0	7 / 0	0 / 0
S14	2,259	591 / 578	1,668 / 1,681	0 / 0	2 / 0	0 / 0
S15	2,259	887 / 882	1,372 / 1,377	0 / 0	42 / 8	0 / 0
S16	2,259	872 / 870	1,279 / 1,077	108 / 312	0 / 0	0 / 0
S17	4,519	392 / 386	4,126 / 4,133	0 / 0	25 / 2	1 / 0

Note. Cells show point / upper counts across 48 matched streams.

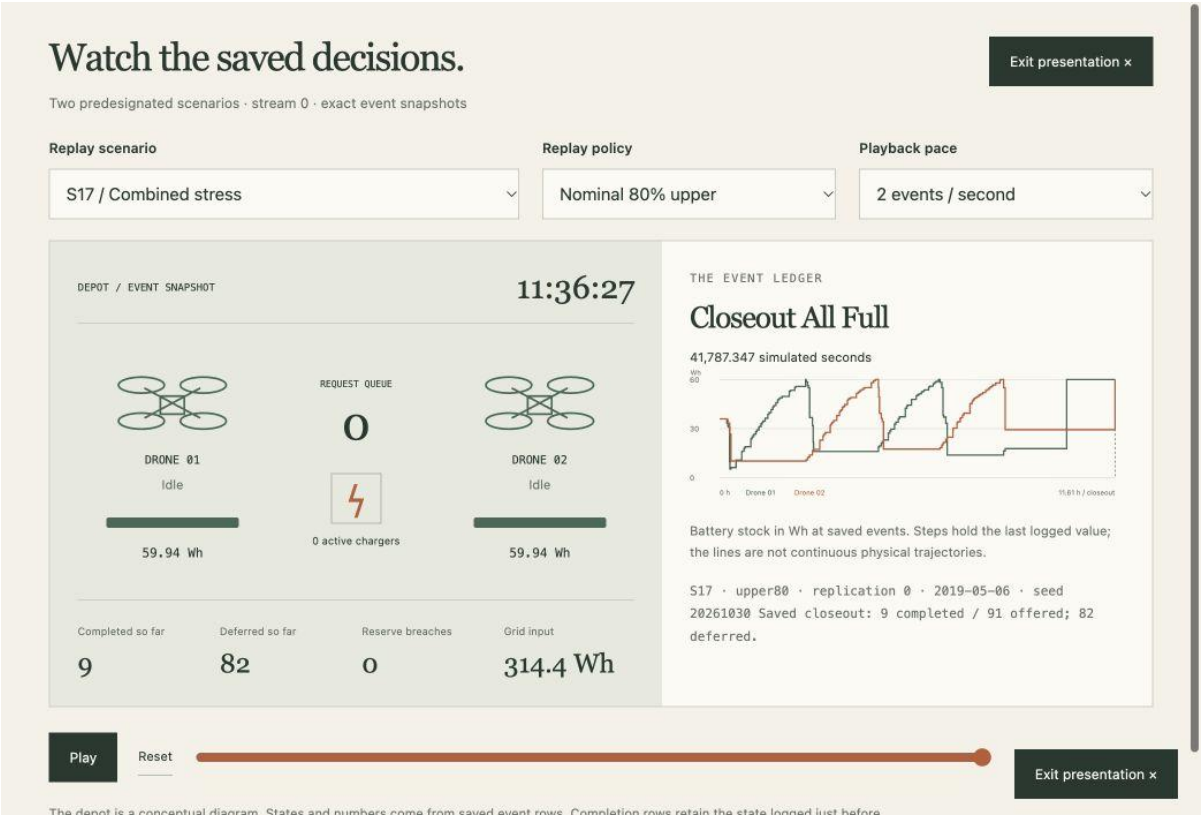


Figure S1. Depot closeout in a single replay under combined stress

Note. The public workbench shows the upper rule in stream zero. Battery traces connect consecutive recorded event states. It can be accessed through the link.

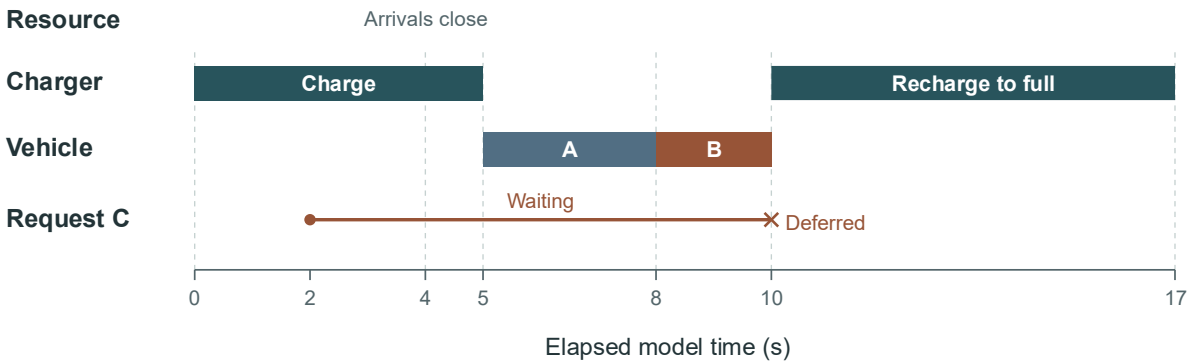


Figure S2. Service, charging and deadline deferral in the three-request benchmark

Note. The timeline follows the saved Python events. Request C reached its deadline when the final charging interval began.

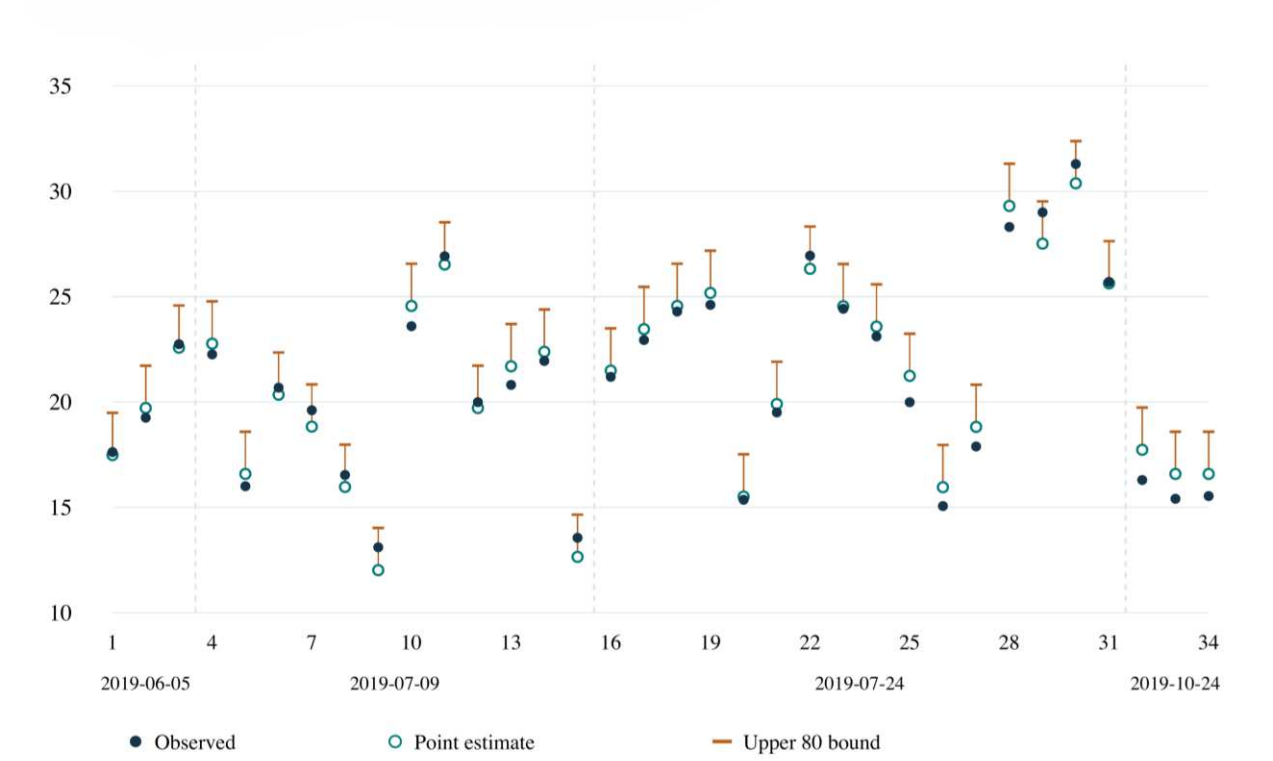


Figure S3. Observed energy (Wh) and nominal 80% upper bounds

Note. All 34 final observations, point estimates and nominal 80% upper bounds.

Supplementary Files

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

- ANYLOGICSIMULATIONRUNVIDEO1.mp4
- NAVIGATIONOFTHEPUBLICWEBSITEVIDEO2.mov
- MODELRUNONWEBSITEVIDEO3.mov
- GITHUBREPOSITORYVIDEO4.mov