

Powering Compute

What Load Flexibility, Financial Cover and Price Forecasts Are Worth to an AI Data Centre

Nicolò Angileri

Founder & Lead of Research, Mediterranean Quantitative Finance Society

ORCID: 0009-0004-3971-1114

Abstract. An AI training data centre buys electricity at an hourly price, and part of its load can wait. Using eleven years of archived day-ahead prices from five European bidding zones and a production GPU cluster trace, this paper measures what that flexibility is worth, how it combines with forward and cap cover, and what a price forecast adds to it. A receding-horizon scheduler that sees only the prices the auction has published lowered the realised cost of an unhedged 100 MW facility in all 15 zone-periods of 2024–2026, by 4.1 to 15.6 per cent, with day-level bootstrap intervals that stay above 3.7 per cent. In simulation the same schedule removes 6.25 per cent of the optimal 95 per cent conditional value at risk of the annual bill in Sicily, the dearest zone, and composes with the hedge rather than replacing it. Letting work wait beyond the published horizon is worth little: perfect foresight of two further days adds at most 3.2 per cent to the realised saving, and a neural quantile forecaster, a linear autoregression and a seasonal naive all make the schedule worse, because their errors are as large as the cross-day differences they would exploit. The financial hedge halves the tail, but largely by construction, since no forward quote was accessible for Sicily and the curve is modelled from the same simulation; a price cap is dominated by the forward. Three further results are negative: the linear model forecasts better than the network, a calibrated two-factor process generates better years than a neural stochastic differential equation, and quantum amplitude estimation, verified exact here, would need two-qubit error rates three orders of magnitude below today’s.

JEL codes: Q41, G32, C53, C63, L86. **Keywords:** electricity price risk, data centres, demand response, conditional value at risk, probabilistic forecasting, value of information.

1. Introduction

What this paper is for

Computing is becoming one of the large electricity loads. United States data centres used 176 TWh in 2023, 4.4 per cent of the country’s electricity, and are projected to use between 325 and 580 TWh in 2028 (Shehabi et al., 2024); the growth is driven by the training and serving of large models (International Energy Agency, 2025; de Vries, 2023). Operators of that capacity sign multi-year electricity contracts, and they sign them without a public quantitative frame that treats together the two things that matter most: the tail of the price distribution, and the fact that a large part of a training cluster’s load does not care exactly when it runs.

The literature has each half (Section 2). What is missing is the number an operator would actually use: how many euro per megawatt of installed capacity, per year, the exposure costs, and how much of that cost and of its tail can be removed by a hedge, by flexibility, by a better view of future prices, or by all of them together. This paper produces that number for a specific and deliberately awkward place, from data that was downloaded and archived with checksums, and it states which half of the answer is robust.

What it studies

The object of study is the annual electricity bill of a data centre with 100 MW of IT capacity, taken as a random variable. Three things determine its distribution: the hourly *price* in a bidding zone; the hourly *load*, which the operator partly controls; and the *hedge*, a set of monthly baseload

forwards and a price cap that settle against the same prices. The primary case is Sicily, a bidding zone with structurally high prices caused by network constraints, compared throughout with northern Italy, France, southern Sweden and south-western Norway so that no conclusion rests on one series.

What it finds

1. **Flexibility pays, on real prices.** A receding-horizon scheduler that sees only the prices the day-ahead auction has published, and moves work only when the saving beats a latency penalty, lowered the realised cost of the unhedged facility in all 15 zone-periods of 2024–2026, by 4.1 to 15.6 per cent and by 8.3 on average, with day-level bootstrap intervals whose lowest bound is 3.7 per cent (Section 9). The value runs against the price level: it is largest where the daily cycle is wide relative to the level, not where power is dear.
2. **Seeing further is worth little, and real forecasts lose.** On realised Sicilian prices, perfect foresight of two further days adds at most 3.2 per cent to what the published prices already deliver, and a neural quantile forecaster, a linear autoregression and a seasonal naive all make the schedule worse, because their errors are as large as the cross-day price differences they would exploit (Section 10).
3. **The hedge halves the tail, mostly by construction.** In simulation the jointly optimal design takes the 95 per cent conditional value at risk of the Sicilian bill from 2,714,550 to 1,286,157 EUR per MW. Because no forward quote was accessible and the curve had to be modelled from the same simulation, most of that reduction is an identity, and the paper sizes it rather than claiming it (Section 8.2). Flexibility composes with the hedge, adding 6.25 per cent of the optimal tail, and a price cap is dominated by the forward.
4. **The Sicilian tail is not heavy; it is long.** The dearest of the five zones has a generalised Pareto shape indistinguishable from zero, -0.019 , and is expensive through persistence: episodes above the threshold last 17.8 hours on average (Section 5.3).

What the paper does not claim

The neural models do not beat their classical counterparts, and the paper says so in the abstract; the quantum estimator offers no practical advantage at any size that can be run, and its appendix is a resource estimate. The forward curve is modelled, and every result that depends on it carries the assumption with it. The out-of-sample test covers two and a half years with a nine-month hole in the Italian source, and every long-horizon risk statement rests on seven annual observations, which is the binding constraint on how much anyone should believe about a once-in-twenty-years bill. Five defects found in a previous version of this work were corrected before the results reported here, and Appendix D lists them with the numbers they had produced.

How it is organised

Section 2 places the paper. Section 3 accounts for every tool and what removing it would cost. Sections 4 and 5 describe the data and the price model, Section 6 the scheduler and Section 7 the simulation. Section 8 reports flexibility and cover in simulation, Section 9 out of sample, and Section 10 the value of a forecast. Sections 11 to 13 give the sensitivity, the limitations and the verdicts on the pre-registered hypotheses.

2. Related literature

Four literatures meet in this problem, and each has solved its own part of it.

Data centres as electricity consumers. That computing load can be moved in time and space to follow electricity prices is an old observation in the systems literature. Qureshi et al. (2009) showed that routing requests between data centres in different wholesale markets cuts the electricity bill of an internet-scale system, Rao et al. (2010) formulated the placement of load across several electricity markets as an optimisation, Liu et al. (2013) studied workload shifting against the coincident peak charges of a single site, and Wierman et al. (2014) surveyed what a data centre can offer as demand response. The scale of the question has since changed, as the introduction recalls (Shehabi et al., 2024; International Energy Agency, 2025; de Vries, 2023), and Masanet et al. (2020) caution how uncertain such estimates are. Operational work has moved from prices towards carbon: the scheduler of Radovanović et al. (2023) shifts flexible compute in time against a carbon-intensity forecast at fleet scale. Demand response in general is surveyed by Albadi and El-Saadany (2008). What this literature seldom does is treat the annual bill as a random variable with a tail, price the hedge that an operator would buy against it, or test the schedule on realised prices out of sample.

Electricity price models. The price side is mature. Mean reversion with jumps and seasonality is the standard reduced form (Lucia and Schwartz, 2002; Cartea and Figueroa, 2005; Benth et al., 2008); a slow non-stationary level under a fast stationary deviation is the two-factor structure of Schwartz and Smith (2000), and extreme value theory has been applied to electricity price changes since at least Byström (2005). The review of Weron (2014) covers the forecasting side, which has since moved to probabilistic forecasts (Nowotarski and Weron, 2018). Two findings from that literature matter here. Linear autoregressions on well chosen lags and calendar terms are strong: their high-dimensional versions estimated with the lasso are among the most accurate day-ahead models (Ziel and Weron, 2018), and in the open benchmark of Lago et al. (2021) deep networks beat them only with careful design. And a forecast can be valued in money by the decisions it drives, as Kath and Ziel (2018) do for trading on the German quarter-hourly markets. Section 10 applies both ideas to a decision that has not been valued this way: how long a unit of compute should wait.

Hedging load with forwards and options. A consumer whose volume is uncertain cannot hedge price risk with forwards alone, and options on the same underlying help (Oum et al., 2006; Oum and Oren, 2010). Conditional value at risk, linear in the positions once scenarios are drawn (Rockafellar and Uryasev, 2000), is the usual objective in electricity procurement (Carrión et al., 2007; Conejo et al., 2010). The price of that cover is the forward premium, which theory ties to the skewness of the spot and to the hedging pressure of buyers (Bessembinder and Lemmon, 2002) and which is empirically large in some markets (Longstaff and Wang, 2004). The premium matters here because the one used below had to be assumed.

The gap. The procurement literature takes the consumer’s volume as exogenous risk; the data-centre literature makes the volume a decision but rarely prices the tail; the forecasting literature scores forecasts by their accuracy or their trading value but not by what they are worth to a scheduler that can wait. This paper puts the three together for one kind of consumer, measured on archived market data, with the schedule restricted to the information the day-ahead market actually publishes, and it reports each answer with the test that could have refuted it.

3. Research design

A paper that contains extreme value theory, a convolutional network, a C++17 core and a quantum circuit owes the reader an account of why each of them is there and what would be lost by deleting it. This section is that account, placed before the data so that the rest can be read against it. Figure 1 is the chain in one view; Table 1 makes the same claim checkable, one row per tool, ending in the column that matters most: what would change if that tool were removed.

Table 1. One row per tool. The last two columns are the ones that matter for a reader deciding whether the machinery is load-bearing: what number it produces, and what would change if it were removed.

Tool	Question it answers	Chosen over, and why	Runs on	What it produces, and what removing it would cost
Peaks over threshold, runs declustering	How heavy is the price tail and how long do expensive episodes last?	A Gaussian or lognormal fit, which gives the observed extreme hours essentially zero probability.	Hourly prices, training block, five zones.	$\xi = -0.019$ in Sicily, episodes of 17.8 h (§5.3). Removing it removes the measure the shifting window is read against.
Moving block bootstrap	Are the estimates distinguishable from noise?	Asymptotic standard errors, which assume an independence neither hours nor days have.	Hourly prices; simulated paths; realised days.	Intervals on ξ , on the simulated gain, and on every out-of-sample saving (§9). Removing it leaves the main claim without an interval.
Copulas, four families by AIC	Does siting compute in another zone break the joint tail?	Linear correlation, which cannot tell co-movement in the middle from co-movement in the tail.	Daily aggregates, five zones and drivers.	Upper tail dependence 0.620 even against the least correlated zone (§5.4). Removing it removes the siting answer.
Two-factor model, simulated method of moments	What is the dispersion of a yearly average, which is what the bill is?	The pre-registered one-factor process, which reproduces 31 per cent of it, and is reported.	The deseasonalised residual, training block.	81 per cent of the observed annual dispersion (§5.2). It generates every simulated path.
Receding-horizon scheduler	What can an operator do with the prices the auction has published?	A reallocation of the whole year, which uses prices nobody has; a calendar-day greedy, which strands work at midnight.	Simulated paths; realised prices.	The operational saving, 4.1 to 15.6 per cent out of sample (§9). Removing it removes the paper's main result.
Quantile network, 170,601 parameters	What is a view of the unpublished hours worth to the scheduler?	Nothing: scored against a linear expert and a seasonal naive, in CRPS and in euro, on the same days.	Windows of 168 h in, 72 h out; test block, once.	Test CRPS 0.11124 against 0.10890 for the linear expert; as a view of the unpublished hours it lowers the realised saving at every window, while perfect foresight is worth at most +1,764 EUR per MW-year (§10). <i>Removing it costs nothing.</i>
Neural stochastic differential equation	Can a learned generator represent a year better than the calibrated one?	Judging a generator by one-step likelihood, which is what makes it fail.	61,368 hourly transitions.	Variance-curve log error 0.963 against 0.069; it loses (App. A.2). <i>No other number depends on it.</i>
C++17 core through pybind11	Can the study afford tens of millions of scheduled path-years?	NumPy alone, which is kept as the exact reference.	Simulated price paths.	A factor of 138 (§6). Removing it makes the sweeps unaffordable rather than wrong.
Rockafellar-Uryasev programme, HiGHS	What to buy, how much, and what each euro of premium buys?	Minimising variance, which charges for cheap years; searching the hedge space, which only approximates.	Twenty thousand simulated years.	The frontier; 94.1 per cent forward and 0.0 per cent cap (§8). Removing it leaves the exposure measured but undecided.
Iterative amplitude estimation, Qiskit	What would a quantum tail estimate cost on this instance?	Citing the quadratic speed-up, true and irrelevant at reachable sizes.	The discretised cost distribution.	Error decay slope -0.873 ; a required two-qubit error rate of 2.9×10^{-6} (App. B). <i>No other number depends on it.</i>

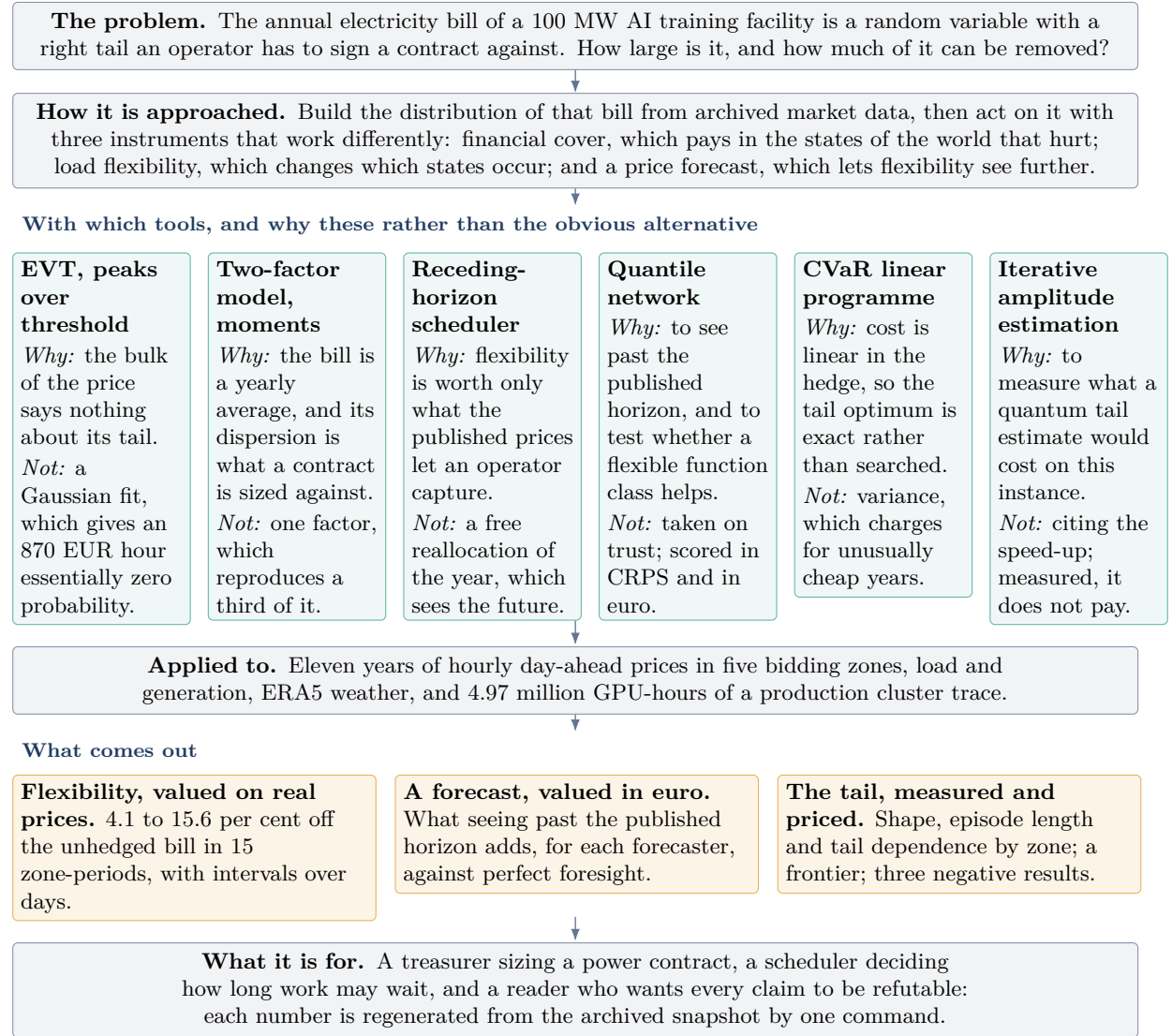


Figure 1. The chain of the paper, from the question to the use of the answer. The middle row is where a choice of method was available; each box states the alternative that was not taken and why.

Three of the objects in that chain are not only described but shipped: the fitted forecaster, the cost simulation and the amplified quantum state can be opened in a browser and driven by hand (Appendix C). The neural generator and the quantum estimator produce no number that any other result depends on; they are here because a reader is entitled to ask whether the modern machinery would have done better, and the only honest way to answer is to build it, score it against the same baselines, and report the margin. The forecaster is different from the version of this paper that preceded it: it now has a job, and Section 10 reports what that job is worth.

4. Data

The snapshot is 111 files and 180 MB. Two limitations of the source are recorded rather than smoothed over. Italian zonal prices are absent from the feed between the first of January and the twenty-ninth of September 2025, which removes 6,551 hours from each Italian series; Norwegian prices have a comparable hole in the same year, 5,279 hours. Zonal load for the Italian bidding zones is not published by this feed, so the system variables used are national. Every analysis that runs on realised prices cuts the series at these holes and never joins across them.

The Italian price series are licensed by their provider for private and internal use only, and the generation and load series carry no licence statement, so neither is redistributed with the code. The

Table 2. Sources. All series were downloaded and archived; each file in the snapshot carries its endpoint, request parameters, download timestamp, size and SHA-256 in `data/manifest.json`. The last column says what the repository redistributes.

What	Source	Period	Resolution	In the repository
Day-ahead prices, five zones	energy-charts (ENTSO-E)	2016–2026	hourly; quarter-hourly from Oct. 2025	France, Sweden, Norway (CC BY 4.0); Italy restored by script
Load, generation by source	energy-charts	2016–2026	quarter-hourly, aggregated to hourly	restored by script
Temperature, wind, irradiance	Open-Meteo ERA5 (Hersbach et al., 2020)	2016–2026	hourly, six sites	yes (CC BY 4.0)
GPU cluster jobs and tasks	Alibaba PAI trace (Weng et al., 2022)	two months	event level	restored by script

Table 3. Hourly day-ahead prices, 2016–2026, in EUR/MWh. The spread column is the mean daily peak-to-trough range, which is what intraday flexibility can monetise.

Zone	Obs.	Missing	Mean	S.d.	$q_{0.99}$	Max	Hours > 200	Excess kurt.	Spread
Sicily	86,352	6,551	107.13	90.22	479.9	870.0	10.79%	9.7	78.7
North Italy	86,352	6,551	103.96	94.57	502.3	871.0	10.69%	10.3	58.2
France	92,879	24	79.72	90.25	469.8	2,987.8	7.90%	31.4	68.2
Sweden SE4	92,903	0	59.69	66.42	354.1	800.0	3.69%	19.7	69.9
Norway NO2	87,624	5,279	65.10	72.08	404.3	898.2	4.16%	17.5	33.7

download script restores them, and the manifest verifies every restored file against the checksum of the snapshot on which this paper was computed.

4.1 What the prices look like

Table 3 contains the first result and it contradicts the expectation the project started from. Sicily is the most expensive zone, spending 10.79 per cent of all hours above 200 EUR/MWh and never once printing a negative price in eleven years, but it is not the zone with the fattest tail: its excess kurtosis is 9.7 against 31.4 in France. What distinguishes Sicily is the width of the daily cycle, 78.7 EUR/MWh on average, which is 86 per cent of the mean daily price level. That is the quantity a schedule can capture, and it is the largest of the five.

Tail dependence between zones is high. The empirical probability that the northern Italian daily price is in its top decile given that Sicily is in its own is 0.906, and even against the least correlated zone, southern Sweden, it is 0.620. Siting compute elsewhere in Europe lowers the level; it does not break the joint tail.

4.2 What the load looks like

The Alibaba PAI trace contains 1,261,050 task records of which 737,137 carry a usable interval and a positive GPU request, together 4.97 million GPU-hours over sixty-two days, averaging 3,318 GPUs in use. By GPU-hours the trace is almost entirely training: the roles `worker`, `PyTorchWorker` and `tensorflow` account for 99.5 per cent. The trace therefore cannot settle the split between deferrable and non-deferrable work by label, and this paper does not pretend that it can: the deferrable share is a stated parameter, set at 54.4 per cent by the proxy that a task reserving one or more whole GPUs is dedicated batch training, and swept in Section 11.

Two features of the trace enter the model directly. The diurnal profile has a peak-to-trough range of 30.4 per cent of its mean, which gives the load shape. And job durations have a median of 10 minutes, a ninetieth percentile of 3.5 hours and a ninety-ninth of 19.3 hours, which is what makes a shifting window of a few hours physically meaningful and one of several days a stretch.

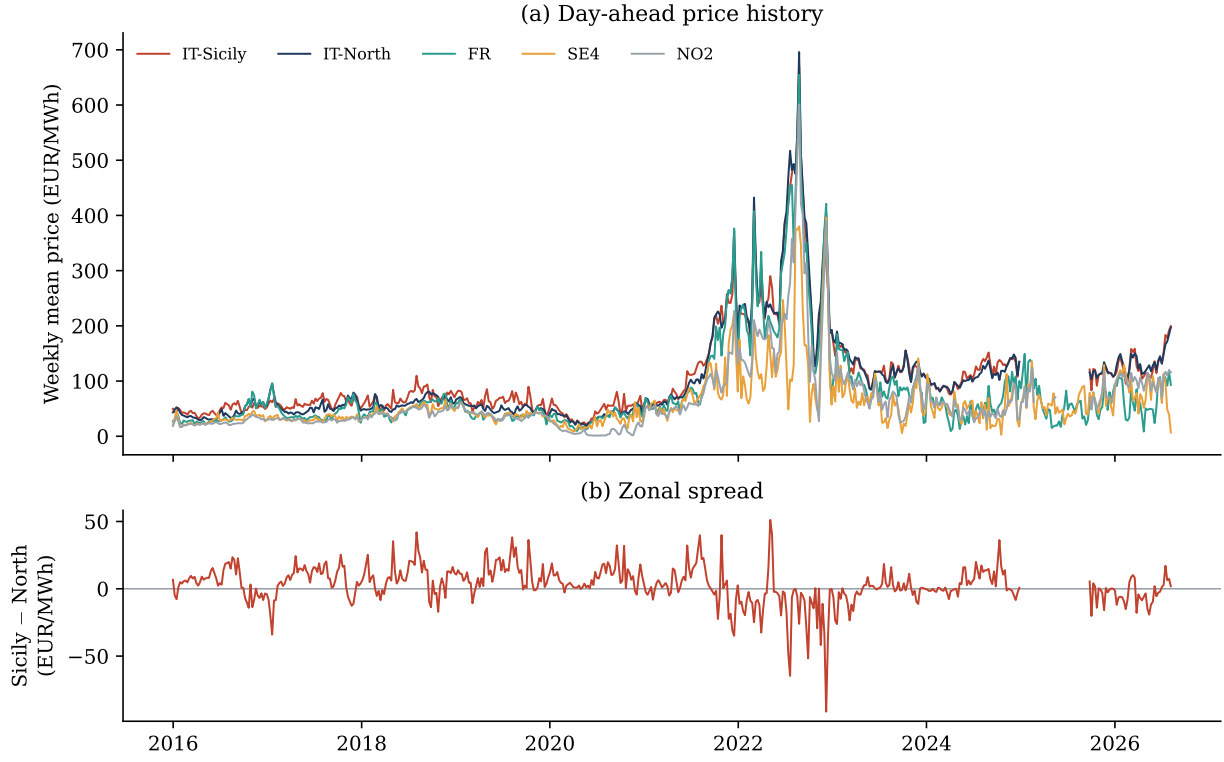


Figure 2. Weekly mean day-ahead prices by zone, and the Sicily-minus-north spread. The visible break in 2025 is the coverage gap in the source, left as a gap.

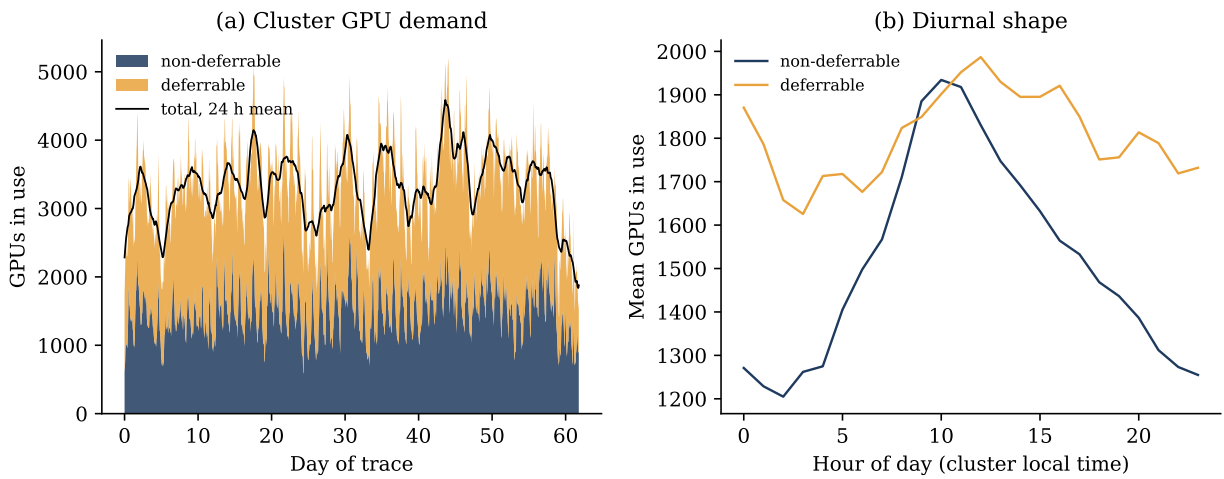


Figure 3. The compute load. Left, GPU occupancy over the sixty-two days of the trace, split by the resource-request proxy for deferrability. Right, the mean diurnal shape in the cluster's local time, which is what the cost engine reuses.

5. Statistical model of the price

5.1 Seasonality

The deterministic component is fitted on $y_t = \text{asinh}(p_t/s)$ with s the median absolute price of the training block. The transform behaves like a logarithm in the upper tail, so spike amplitude stays proportional to the level, and it remains defined where prices go negative, which a shifted logarithm cannot guarantee out of sample. The regressors are annual and semi-annual harmonics, day-of-week and national holiday dummies, and intraday harmonics interacted with the annual cycle; the number of intraday harmonics is chosen by BIC on the training block.

No deterministic trend is included. A trend fitted through 2016–2022 is dominated by the 2021–2022 price crisis, and using it to simulate a future year pins that year to the level of the crisis peak; the low-frequency movement belongs instead to the non-stationary factor of the next section, which can be started from where the market actually is. This is one of the recorded deviations from the pre-registration (Appendix D).

5.2 Mean reversion, jumps, and why one factor is not enough

On the deseasonalised residual, three specifications are estimated by exact maximum likelihood of the hourly transition density: an Ornstein-Uhlenbeck process without jumps, and the same with compound Poisson jumps of Gaussian or of two-sided exponential size, the latter evaluated through the scaled complementary error function so that the tails do not overflow. Selection is by BIC, as pre-registered. Gaussian jumps are selected in 4 of the five zones, Sicily among them, and two-sided exponential jumps in northern Italy.

The selected single-factor model fits the hourly data well and then fails at the only thing the cost engine needs. Simulated from it, the standard deviation of the annual average deseasonalised price is 31 per cent of what the data show in Sicily, and between 32 and 41 per cent in the other continental zones. The empirical variance-time curve explains why: the variance of the one-year moving average of the Sicilian price is 55 per cent of its hourly variance, which no process with a multi-day half-life can produce.

The model used from here on is therefore two-factor, in the manner of [Schwartz and Smith \(2000\)](#): a stationary short-term deviation plus a non-stationary equilibrium level, with the intraday spike process estimated separately on the deviation from the daily mean, jumps included in the family BIC selected. Its three parameters are calibrated by simulated method of moments to the variance-time curve and to the autocorrelation of the daily level rather than by maximum likelihood, because one-step likelihood on two nearly collinear factors assigns the daily noise to the wrong one. It reproduces 81 per cent of the observed annual dispersion in Sicily and between 79 and 83 per cent elsewhere.

One caveat constrains how this is read. With a sample of this length the three parameters are only weakly identified individually: repeating the calibration with different random seeds recovers the target moments every time while the implied half-life moves substantially. What the procedure delivers, and all that is claimed from it, is a generator whose dispersion at horizons from a day to a year matches the data.

5.3 Extremes

Table 4 sharpens the picture from Table 3. The Sicilian shape parameter, -0.019 with an interval of $[-0.227, 0.240]$, is indistinguishable from zero: the tail is exponential, not heavy. France and Norway have genuinely heavy tails, with intervals excluding zero. And the Sicilian extremal index of 0.056 says that exceedances arrive in clusters lasting 17.8 hours on average, which is a direct constraint on the operational side: shifting load inside a day does not escape an episode of that length, and the shifting window has to be read against it.

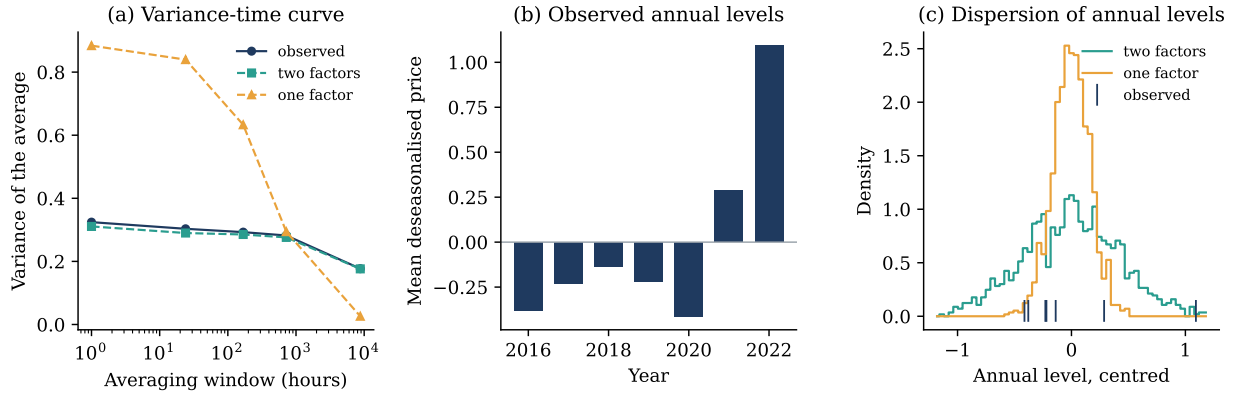


Figure 4. Why one factor is not enough. Left, the variance of h -hour averages against h : observed, and simulated from each generator. Centre, the observed annual levels. Right, the dispersion of annual levels each generator produces, against the observed ones marked as ticks.

Table 4. Peaks over threshold on the training block, runs declustering with a six-hour run length, threshold by the stability rule. Intervals are percentile intervals from 1,000 moving block bootstrap replications with a block length of one week. Expected shortfall in EUR/MWh.

Zone	Threshold	Clusters	Extremal index	Mean length	Shape ξ	95% interval	ES _{0.999}
Sicily	357.7	138	0.056	17.8 h	-0.019	[-0.227, 0.240]	734
North Italy	374.2	115	0.047	21.3 h	-0.111	[-0.359, 0.052]	785
France	125.5	149	0.014	70.0 h	0.499	[0.189, 0.685]	2,742
Sweden SE4	280.4	151	0.098	10.2 h	-0.088	[-0.367, 0.172]	764
Norway NO2	131.0	106	0.014	69.5 h	0.587	[0.066, 0.701]	3,050

Applying the pre-registered threshold rule required one adjustment recorded here: with a twenty-four-hour run length the Sicilian series leaves too few independent clusters at the ninetieth percentile for the pre-registered minimum, so the run length was set to six hours and the minimum stated in clusters rather than raw exceedances.

5.4 Dependence

Four copula families are fitted to daily aggregates and ranked by AIC. The Sicilian price against national residual load selects a Student- t copula with Kendall's $\tau = 0.200$; against the renewable share of load it selects a Gaussian copula with $\tau = -0.220$. Across zones the selected families are asymmetric: Gumbel against northern Italy, survival Clayton against France and Norway, Student- t against Sweden. Model-implied upper tail dependence is reported next to the empirical estimate throughout, and where they disagree, as they do for the driver pairs, the disagreement is shown rather than reconciled.

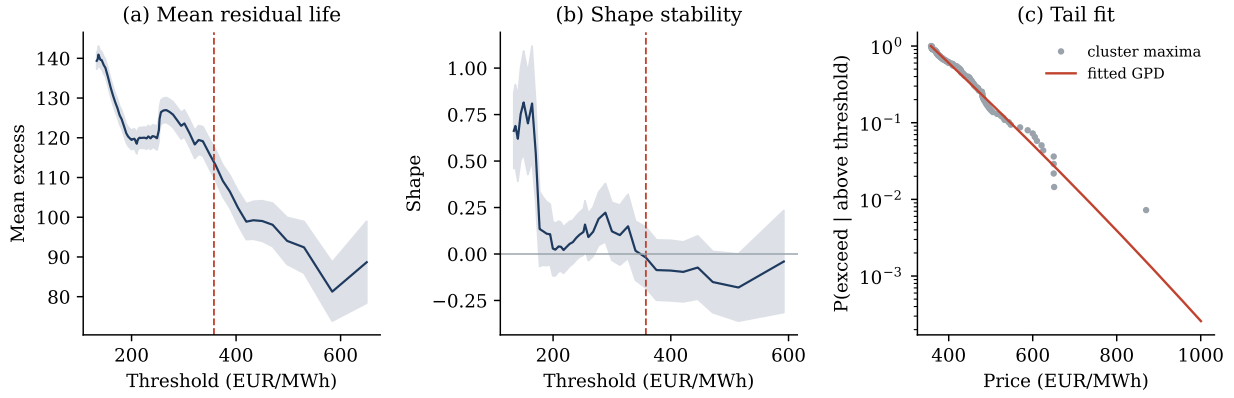


Figure 5. Threshold diagnostics for Sicily. Left, the mean residual life plot with its confidence band. Centre, the stability of the shape parameter, with the selected threshold dashed. Right, the fitted tail against the declustered cluster maxima; the near-straight line on a logarithmic scale is an exponential tail.

6. The scheduler

The value of flexibility is only as credible as the schedule that realises it, and a schedule is only as credible as the information it is allowed to use. This section states both, because a scheduler that sees the whole year would report a value no operator can capture.

The work. The hourly load of the facility is the diurnal profile of the cluster trace scaled to a mean draw of 80 per cent of a 100 MW connection at a power usage effectiveness of 1.2. A share $\delta = 54.4\%$ of each hour’s load is deferrable: it is a unit of work tied to that hour, its *origin*, which may run in any hour within w hours of the origin, the shifting window, as long as the hourly draw never exceeds the connection. Work run away from its origin costs a latency penalty λ per MWh, 2 EUR in the base case, which stands for the value of a later answer and for the operational overhead of moving it.

What the operator knows, and when. The day-ahead auction clears around midday and publishes prices to the end of the following day. At each gate closure the operator therefore knows thirty-six hours of prices: the remainder of today and all of tomorrow. The schedule is decided once a day at that moment, on a *receding horizon*: every unit of work that has not yet run and whose origin lies inside the published horizon is placed, including work an earlier decision had already moved, and only the next twenty-four hours are committed. What is left beyond them is re-planned at the next gate, when twenty-four more hours of prices are known.

The rule. At each decision the pending units are taken dearest first, by the price of the hour they currently occupy plus the penalty if that hour is not their origin, and each is moved to the cheapest reachable hour with headroom, provided the move saves more than the penalty it incurs. A unit that finds nowhere cheaper stays where it is, which is what makes the schedule feasible by construction: the unshifted profile is always a valid plan. Three properties follow and are tested in the repository: every unit runs exactly once and within its window; no hour exceeds the connection; and no move is ever made at a loss, so the value of flexibility cannot turn negative however large the penalty. The earlier version of this engine decided each move on the price difference alone and charged the penalty afterwards, which is why it could report a penalty at which flexibility destroyed value; that is a property of a myopic rule, not of the facility, and it is gone.

Seeing beyond the published horizon. A unit whose window reaches past the published horizon can wait for a cheaper hour only if the scheduler has a view on the hours not yet priced. With a forecast, the scheduler plans every unit it knows about over the published hours and

the forecast ones, using the forecast's expected price where no price is published, and commits nothing beyond the next twenty-four hours: the next gate publishes the true prices and everything uncommitted is planned again. A good forecast lets work wait for a cheap night two days away, or run now because its own hour will be dear; a bad one does the same on a wrong view. What that is worth is measured in Section 10.

Two implementations. The algorithm exists twice, in NumPy as the reference and in a C++17 core through pybind11 for the Monte Carlo. They run the same operations in the same order on the same pre-drawn shocks, so the test that compares them is exact rather than statistical: on every configuration tested, including a window wider than the published horizon and a penalty of zero, the energy moved agrees to the last bit and the annual cost to a relative 10^{-12} . The core is 138 times faster on the same evaluation. The interactive page of Appendix C runs a third, JavaScript, port of the same rule.

7. Simulation engine

The engine draws hourly price paths from the calibrated generator, schedules the deferrable load against them with the rule of Section 6, and settles the hedges. Jumps are drawn in the family the intraday model selected: normal sizes come from a Box-Muller transform of two pre-drawn uniforms, so the C++ core and the NumPy reference produce the same path. An earlier version represented Gaussian jumps as two-sided exponential ones, which doubled their variance and gave the simulated paths more intraday dispersion than the calibration had matched; since intraday dispersion is exactly what flexibility monetises, that approximation flattered the headline and has been removed (Appendix D).

Variance reduction is measured rather than asserted. Antithetic sampling removes 0.7 per cent of the variance on this payoff, which is to say nothing, and is off by default. A control variate on the annual mean of the simulated state, whose expectation is known in closed form, removes 97.7 per cent, taking the standard error of the mean annual cost from 102,158 to 15,400 EUR at twenty thousand paths.

Validating the simulator against the market

The generator is calibrated on 2016–2022, a window containing the price crisis, and started from the median deseasonalised level of the last observed month. Its simulated median price is 6 per cent above the median of the last observed year and its simulated mean is 23 per cent above, because the calibration window carries more upside risk than the calm year just past. This comparison is computed in the experiment itself; it is the check that caught a deterministic trend that had been pinning every simulated year to the level of the 2022 peak.

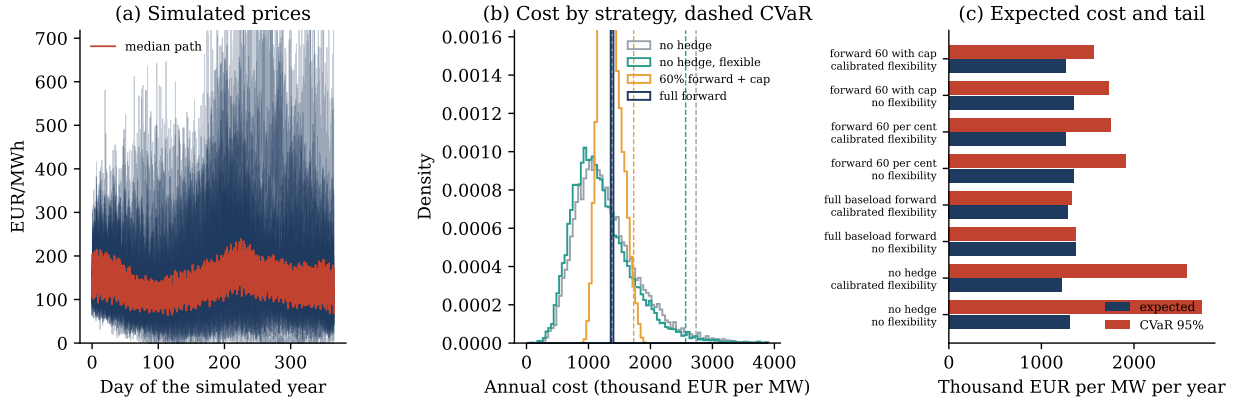


Figure 6. Left, simulated price paths over a year. Centre, the simulated distribution of the annual cost under four strategies, with the tail measure dashed; the fully hedged distribution is a spike that runs off the top of the axis. Right, expected cost and tail cost side by side for every strategy tested.

8. Flexibility and cover in simulation

8.1 The programme

Once the paths are drawn and the load is scheduled, the annual cost is linear in the hedge decisions: one more per cent of monthly forward cover moves every path by that month’s settlement, and one more per cent of cap cover by that path’s cap payoff. Minimising conditional value at risk over a linear cost is the programme of [Rockafellar and Uryasev \(2000\)](#),

$$\min_{h, \alpha, u} \quad \alpha + \frac{1}{(1 - \beta)N} \sum_{i=1}^N u_i \quad \text{s.t.} \quad u_i \geq C_i(h) - \alpha, \quad u_i \geq 0, \quad 0 \leq h \leq 1, \quad \mathbb{E}[C(h)] - \mathbb{E}[C(0)] \leq B, \quad (1)$$

a linear programme in twelve monthly forward covers, one cap cover, the value at risk level α and the tail slacks u , solved exactly by HiGHS; the budget constraint is written on the hedge positions alone, so that it holds with equality at no hedge rather than as the difference of two large sums. Sweeping the budget B traces the efficient frontier. Flexibility is not inside the programme, because the schedule changes the cost in a way that is not linear in any decision; it is an outer level, and the programme is solved once for each schedule.

8.2 What the financial result is, and what it is not

The headline of this section has to be read with its construction in view. There is no accessible forward quote for the Sicilian zone, so the forward curve is built as the simulated expected spot of each month multiplied by one plus a stated risk premium of five per cent, a level the empirical literature on forward premia makes plausible without pinning down ([Bessembinder and Lemmon, 2002](#); [Longstaff and Wang, 2004](#)). Buying that curve removes the variance of the annual bill *by construction*: the hedge settles against the same paths whose average defined its price, and what remains is the premium, the mismatch between the flat baseload volume and the shape of the compute load, and whatever the schedule does.

The arithmetic is visible in the numbers. For the rigid facility, full baseload cover takes the standard deviation of the annual cost from 525,533 to 1,405 EUR per MW, a factor of 374, and raises the expected cost by 5.0 per cent, which is the premium. The tail measure has almost nothing left to measure, which is why its bootstrap interval in Table 5 is 40 EUR wide. A statement of the form “the optimal hedge cuts the tail in half” is, in this construction, close to an identity: a forward priced off a distribution eliminates that distribution. Section 8.5 sweeps the premium so that the size of the mechanical part can be read off directly.

Table 5. Sicily, 100 MW of IT capacity at a power usage effectiveness of 1.2, twenty thousand simulated years. All figures are euro per MW of installed capacity per year. The forward curve is the simulated expected spot of each month plus a five per cent risk premium; the cap strikes at the 95th percentile of the simulated price for a premium of 1.15 times its expected payoff. Intervals are percentile intervals from 1,000 moving block bootstrap replications of the paths, described in Section 8.4.

Strategy	Expected cost	CVaR _{0.95}	95% interval	Change
No hedge, rigid load	1,305,585	2,714,550	[2,676,220, 2,754,150]	—
Full baseload forward, rigid	1,371,026	1,372,974	[1,372,953, 1,372,993]	−49.4%
CVaR optimum, rigid	1,370,898	1,371,928	[1,371,910, 1,371,950]	−49.5%
No hedge, flexible load	1,220,873	2,547,957	[2,511,599, 2,585,359]	−6.1%
CVaR optimum, flexible	1,282,473	1,286,157	[1,286,108, 1,286,209]	−52.6%

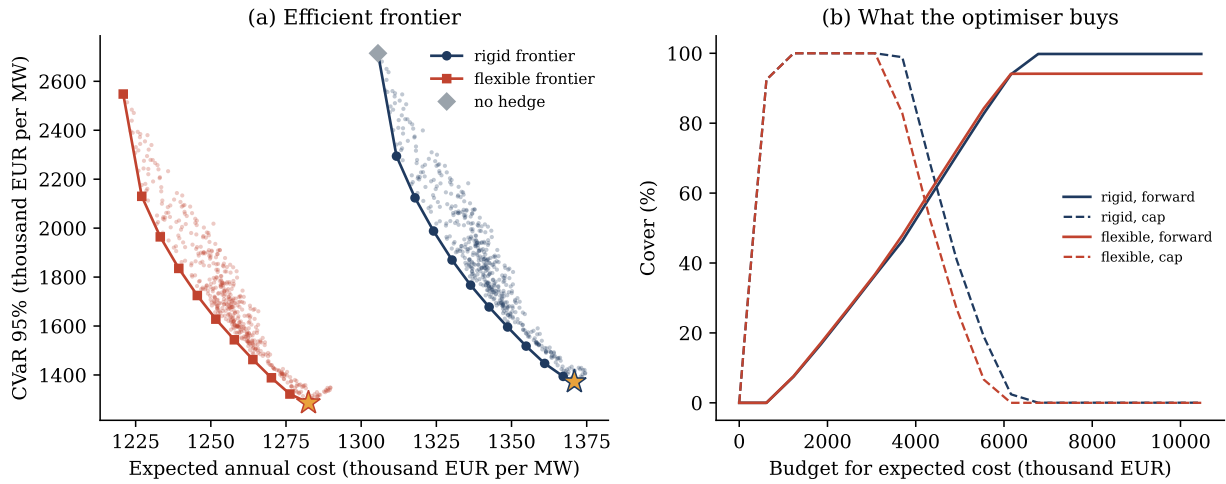


Figure 7. Left, the efficient frontier of expected cost against tail cost, for a rigid and a flexible facility, against a cloud of admissible but arbitrary hedges that it dominates; stars mark the minimum-tail solutions. Right, what the optimiser buys as the budget grows.

8.3 Results

Three things in Table 5 are worth stating plainly, in the order in which they deserve to be believed. Flexibility removes 6.5 per cent of the expected bill on its own, before any contract is signed, and it composes with the hedge rather than substituting for it: the flexible optimum lies below the rigid optimum on both coordinates, and at the tail optimum the schedule is worth 85,772 EUR per MW, 6.25 per cent of the rigid optimum’s tail. The optimiser buys 94.1 per cent forward cover and 0.0 per cent cap at the assumed loading, which Section 8.5 turns from a null result into a measurement by finding the loading at which the cap would be bought. And the financial hedge halves the tail at a cost of about the premium, but for the reason given in Section 8.2 that halving is closer to an accounting identity than to a discovery.

8.4 How precisely any of this is measured

What is resampled is the path, not the hour: a replication draws blocks of path indices and rescores every strategy on the same draw, with the hedge weights the programme chose on the whole sample held fixed, as the pre-registration requires of strategy parameters. Because the draw is shared, the difference between two strategies keeps its pairing and is estimated far more precisely than either level: the paired interval on the flexibility gain is [85,716, 85,824]. The unhedged tail is the imprecise quantity, with an interval 77,931 EUR per MW wide, because it is an average over the worst thousand of twenty thousand simulated years; the hedged tails are known to within a few tens of euro, which is not a triumph of estimation but the first symptom of the circularity above.

Table 6. Sweeping the two priced assumptions, Sicily, twenty thousand paths, everything else held at its base value. The left block moves the risk premium on the forward curve; the right block moves the loading on the cap, at the base premium and at a forward made deliberately expensive at fifteen per cent. Euro per MW.

Risk premium on the forward				Loading on the cap		
Premium	Optimal CVaR	vs. no hedge	Flex. gain	Loading	Cover, base	Cover, dear fwd
0.0%	1,224,501	−54.9%	6.28%	−90%	11.9%	100.0%
2.5%	1,255,340	−53.8%	6.27%	−75%	1.0%	7.0%
5.0%	1,286,157	−52.6%	6.25%	−50%	0.0%	1.0%
7.5%	1,316,942	−51.5%	6.24%	−25%	0.0%	0.1%
10.0%	1,347,702	−50.4%	6.23%	+0%	0.0%	0.0%
12.5%	1,378,434	−49.2%	6.22%	+15%	0.0%	0.0%
15.0%	1,409,137	−48.1%	6.21%	+50%	0.0%	0.0%

These intervals cover Monte Carlo sampling error and nothing else. The uncertainty in the generator’s own parameters is not in them, and that is why the operational claim of this paper rests on the realised prices of Section 9, with intervals over real days, rather than on this table.

Across zones the pattern holds but its composition changes. The optimal tail reduction ranges from 41.2 per cent in Sweden SE4 to 57.5 per cent in Norway NO2, while the contribution of flexibility to the optimum is 3.2 per cent of the tail in Norway, 4.2 in northern Italy, 5.4 in France, 6.3 in Sicily and 10.3 in southern Sweden. The value of hedging is general; the value of flexibility is specific, and it is largest where the price level is low relative to its intraday variation.

8.5 What the premium is worth, and what the cap is worth

The premium. Over the whole range a reader might argue for, the tail reduction moves from 54.9 to 48.1 per cent. It never approaches zero, because it is not really being bought: it is what happens when a position is settled against the distribution that priced it. Tripling the assumed premium costs the hedger 122,515 EUR per MW of expected cost, and the ranking of the strategies never changes. The value of flexibility, in the last column of the left block, moves only from 6.28 to 6.21 per cent of the optimal tail, because it is a physical saving that never touches the curve.

The cap. The optimiser does not buy the cap at a positive loading, and not at a zero one either. Cover of more than one per cent of the volume appears only when the cap is sold at a discount of 90 per cent to its expected payoff, or of 75 per cent when the forward is made dear. The reason is not the price but redundancy: with monthly forwards available, the forward cover already removes almost all of the variance, and a cap struck at the 95th percentile of an already hedged distribution has almost nothing left to insure. The right way to state the null result is that the cap is dominated by the instrument next to it, not that it is overpriced.

9. Out of sample

This is the section that carries the most weight, because nothing in it is simulated. The operational question is answered on the unhedged facility, where the flexible and the rigid schedules face the same prices and differ in nothing else: the saving is the cost of the rigid load minus the cost of the flexible one, latency penalty included, on the prices that occurred. Each series is cut into its complete stretches, the scheduler runs on each stretch with the information the auction had published at each gate, and the interval is a moving-block bootstrap over days, blocks of a week drawn within each stretch.

The financial question is answered beside it. For each period every model is refitted on data ending the previous December, the forward curve and the cap premium are built from that vintage, the hedge is chosen by (1) on a simulation of the whole period and frozen, and it is then settled on the hours that exist.

Table 7. Out of sample, 2024 to August 2026. *Saving*: cost of the unhedged rigid load minus cost of the unhedged flexible load, as a share of the former, with a 95 per cent moving-block bootstrap interval over days. The last four columns are the price actually paid, EUR/MWh, hedge settlements and latency penalties included. Italian 2025 covers October to December only, because of the source gap.

Zone	Year	Days	Spot	Operational, unhedged		Price paid with a hedge			
				Saving	Interval	No hedge	Full fwd	Opt. rigid	Opt. flex.
Sicily	2024	365	112.1	5.87%	[5.28, 6.54]	111.7	119.5	119.5	112.4
Sicily	2025	92	111.5	4.44%	[3.75, 5.08]	111.2	147.1	146.9	139.9
Sicily	2026	217	130.5	5.16%	[4.69, 5.70]	128.8	98.5	98.6	93.3
North Italy	2024	365	107.3	4.74%	[4.32, 5.16]	107.3	108.0	108.1	102.9
North Italy	2025	92	115.3	4.44%	[3.71, 5.31]	115.6	149.0	149.1	142.6
North Italy	2026	217	134.0	4.11%	[3.73, 4.54]	133.1	98.9	98.9	94.5
France	2024	366	58.0	10.84%	[9.96, 11.81]	57.7	64.8	64.8	58.3
France	2025	363	61.1	11.45%	[10.75, 12.24]	60.4	87.7	87.7	79.4
France	2026	217	67.8	12.63%	[11.64, 13.75]	66.1	57.7	57.7	49.9
Sweden SE4	2024	366	49.7	15.62%	[13.56, 18.05]	49.8	62.6	62.5	53.9
Sweden SE4	2025	365	60.4	14.76%	[13.36, 16.33]	60.4	33.8	33.6	26.6
Sweden SE4	2026	217	86.3	10.50%	[9.66, 11.41]	85.3	57.3	57.2	50.6
Norway NO2	2024	366	50.1	7.02%	[5.41, 9.34]	50.1	66.5	66.5	62.5
Norway NO2	2025	142	72.8	8.31%	[6.62, 10.67]	73.6	40.7	40.6	35.6
Norway NO2	2026	217	100.9	5.05%	[4.50, 5.69]	99.9	65.8	65.8	62.0

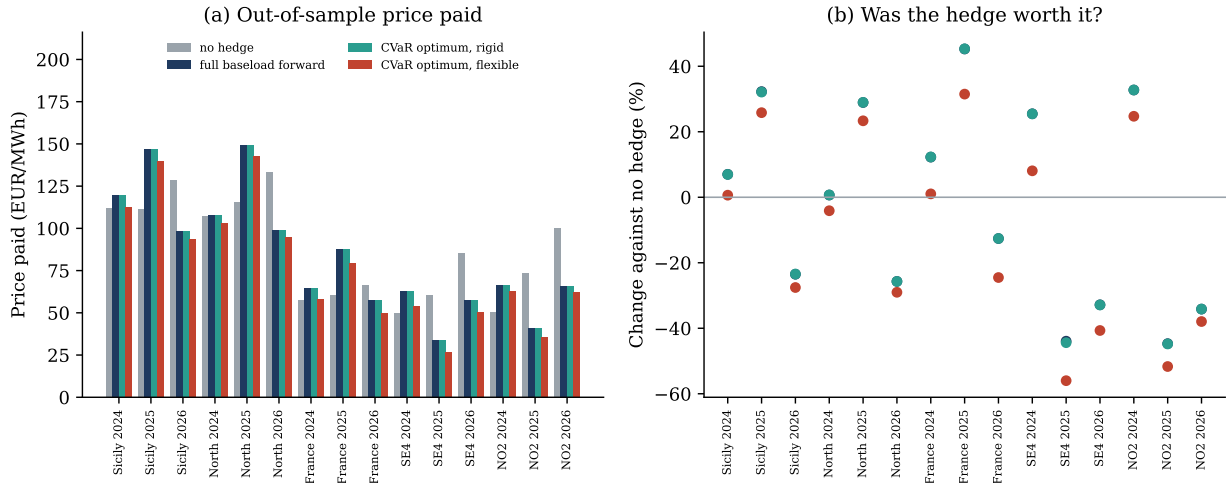


Figure 8. Out-of-sample outcomes. Left, the realised cost of each strategy in each zone and period. Right, the change against holding no hedge at all: above the line the hedge cost money, below it the hedge paid.

The schedule lowered the cost of the unhedged facility in all 15 zone-periods, by between 4.1 and 15.6 per cent and by 8.3 on average, moving between 15.9 and 17.7 per cent of the energy. The lowest lower bound of the fifteen intervals is 3.7 per cent. That the saving is positive is no surprise, since a mechanism that only moves work to an hour it has already seen priced lower, and only when the difference beats the penalty, cannot lose on the hours it moves, as long as its own load does not move the price (Section 12); what the table measures is the size, and the size is what an operator needs. The saving is largest in France and southern Sweden, the two zones whose average daily range exceeds their average price, and smallest in the two Italian zones.

The financial hedge behaved as insurance does. In Sicily the rigid optimum paid 119.5 euro per MWh in 2024 against 111.7 unhedged, when prices fell below the forward curve struck at the end of 2023, and 98.6 against 128.8 in 2026, when they rose. A paper that reported only the years in which the hedge paid would be misleading, and the paired columns of Table 7 are there so that the reader can see both.

Table 8. Test block, 2024–2026, scored once, all models on the same windows. Lower is better for the first two columns; coverage should match its nominal level; the last column is the Kolmogorov-Smirnov statistic of the probability integral transform against uniformity (Gneiting and Raftery, 2007).

Model	Pinball	CRPS	Skill vs. naive	Cov. 50%	Cov. 95%	PIT KS
Neural quantile network	0.03941	0.11124	0.034	0.541	0.975	0.062
Linear expert autoregression	0.03912	0.10890	0.054	0.528	0.973	0.099
Seasonal naive	0.04361	0.11511	0.000	0.352	0.695	0.179
Two-factor generator	0.05782	0.15078	−0.310	0.466	0.887	0.086

10. What a forecast is worth

The window sweep of Section 11 stops growing where the published horizon ends, and the same happens on real prices: in Sicily, with the published prices only, widening the shifting window from twelve to seventy-two hours raises the realised saving by 1.6 per cent. Work that could wait two or three days has nowhere to wait to, because the scheduler cannot see a price the auction has not published. A forecast of the unpublished hours is the obvious way through that ceiling, and it is the job the forecaster of this paper was built for. This section measures what the job is worth. It was added after the pre-registration and is the only place the forecasters meet the test block, so it is exploratory in the sense of Appendix D.

10.1 The forecasters on the test block

The frozen network, the linear expert autoregression and the two-factor model, both refitted on training plus validation as the network was, and the seasonal naive are scored once on 659 test windows issued between 2024 and 2026 (Table 8). The pre-registered comparison on the validation block, in Appendix A.1, went to the linear model; the test block confirms it. The linear expert has the lower pinball loss and the lower CRPS, 0.10890 against 0.11124, a skill over the seasonal naive of 5.4 per cent against 3.4. The network is the better calibrated of the two, with a probability integral transform closer to uniform, a Kolmogorov-Smirnov statistic of 0.062 against 0.099, but calibration is not accuracy, and on the proper scoring rules it loses. The network does not earn its complexity, which is what the benchmarks of Lago et al. (2021) and Ziel and Weron (2018) would lead one to expect.

10.2 The experiment

Each forecast is issued at the gate and predicts the next seventy-two hours; the auction has published the first thirty-six of them from its issue time, so the forecast adds 48 hours of view. The scheduler of Section 6 plans every unit of work it knows about over the published hours and those 48, using the expected price of the forecast where no price has been published, and commits only the next twenty-four hours; the gate after publishes the true prices and everything uncommitted is planned again. Settlement is always on the prices that occurred. Five policies run on the same 674 Sicilian days of 2024–2026: the published prices alone, the three forecasts, and perfect foresight, which plans the unpublished hours with the prices that did occur and is therefore the ceiling on what any forecast can be worth to this scheduler. The value of a forecast is its saving minus the saving of the published prices alone, a paired difference day by day, with a moving-block bootstrap interval over days.

10.3 What it shows

The ceiling is low. Even perfect foresight of the next 48 hours adds +228 EUR per MW per year at a twelve-hour window and +1,764 at seventy-two, which is 3.2 per cent of the saving the published prices already deliver, and from a twenty-four-hour window on its interval includes zero.

Table 9. Realised Sicilian prices, 674 days of 2024–2026, euro per MW per year. *Published:* saving against a rigid load with the published prices only. The other columns: what planning the next 48 unpublished hours with each forecast adds to that saving, with a 95 per cent paired moving-block bootstrap interval over days.

Window	Published	Seasonal naive	Linear expert	Neural network	Perfect foresight
12 h	53,546	−728 [−1,419, +25]	−879 [−1,657, −26]	−768 [−1,544, −28]	+228 [+46, +396]
24 h	54,343	−1,159 [−2,975, +779]	−988 [−2,612, +1,088]	−943 [−3,181, +783]	+561 [−208, +1,361]
36 h	54,407	−791 [−3,739, +2,776]	−1,111 [−3,959, +2,431]	−1,134 [−5,177, +2,172]	+1,083 [−541, +2,614]
48 h	54,411	−323 [−4,606, +4,574]	−1,167 [−4,952, +3,401]	−1,387 [−6,969, +3,183]	+1,391 [−1,119, +3,790]
72 h	54,409	−117 [−5,083, +5,804]	−1,074 [−5,206, +3,964]	−1,438 [−7,507, +3,494]	+1,764 [−1,922, +5,326]

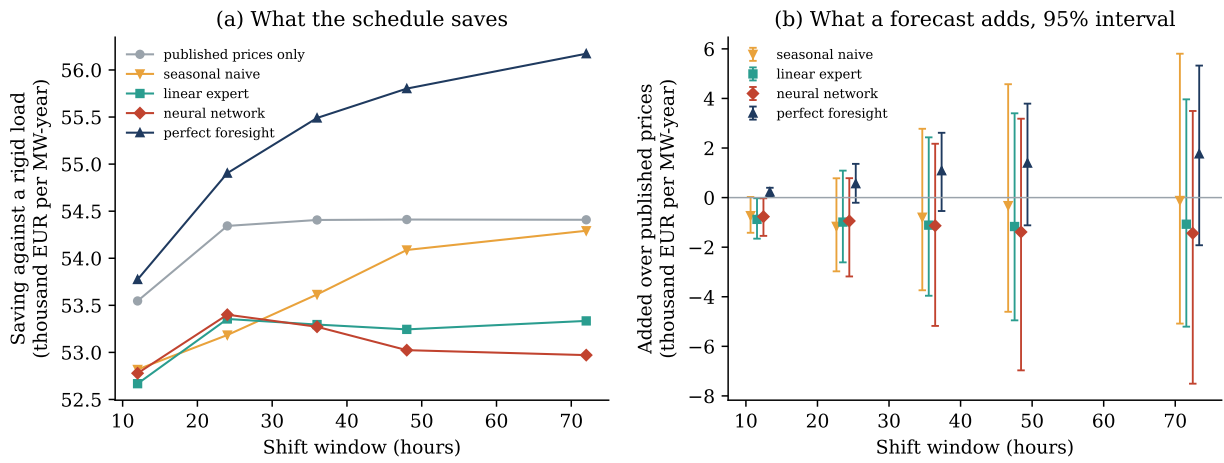


Figure 9. Left, the realised saving of each policy against a rigid load as the shifting window grows. Right, what each forecast adds over the published prices, with its 95 per cent interval over days.

The reason is in the prices: what a schedule monetises is the daily cycle, and the published horizon already contains a full day of it. Moving work by two or three days buys only the difference between daily levels, which are persistent, and that difference is small next to the cycle.

Real forecasts do worse than none. All three forecasts lower the saving at every window tested. At twelve hours the linear expert costs −879 and the network −768 EUR per MW per year, with intervals that exclude zero; at longer windows the point estimates stay negative and the intervals widen to include it. The mechanism is visible in the energy moved: acting on the forecasts moves 19.8 and 20.0 per cent of the energy at twelve hours, against 16.2 with the published prices and 16.4 with perfect foresight. The forecasts’ mean absolute error on the hours the scheduler uses them for is 18.6 and 19.4 EUR/MWh, while the mean absolute difference between the same hour one and two days apart, which is what waiting across days can capture, is 15.5 and 19.1. The error is as large as the signal, so the scheduler moves work on errors about as often as on information, and pays the latency penalty each time.

Accuracy does not rank the euro. The linear expert is the more accurate forecaster by pinball loss and CRPS, and the ranking does not carry over into money: which of the two loses less depends on the window, and neither gains. A forecast is worth what the decision it feeds is worth (Kath and Ziel, 2018), and for this decision the answer is that the scheduler is better off with the published prices alone. A scheduler that treated the forecast as uncertain rather than

plugging in its mean could at best recover part of a ceiling of +1,764, which bounds what the question is worth.

Looking further requires planning further. A variant that lets only the already published work wait into the forecast hours, without planning the work whose origin lies there, loses even with perfect foresight at a twelve-hour window, −484 EUR per MW per year: work that waits lands in hours whose own work has not been placed yet, and congests them. Whoever extends the horizon of a scheduler has to extend what it plans, not only where it may put things.

11. Sensitivity

Every assumption that carries a number is moved here, one at a time, and the effect on the simulated value of flexibility at the tail optimum is read off. Table 10 is the whole of it.

Two rows deserve a sentence. The window row says that in simulation the value of flexibility stops growing soon after twenty-four hours, and the reason is informational rather than physical: the scheduler cannot see a price the auction has not published. Whether a forecast can push that ceiling is a question for real prices, and Section 10 answers it. The latency row replaces a number the previous version of this paper reported, the charge at which flexibility would cost money; with a scheduler that does not make losing moves there is no such charge, and the honest statement is how fast the gain decays.

Table 10. Sicily, twenty thousand paths, one assumption moved at a time and everything else held at its base value. The quantity reported is the value of flexibility at the tail optimum, euro per MW. The last row is not a hedging quantity and is scored on its own terms.

Assumption	Range swept	Base value	Effect
Risk premium on the forward	0 to 15%	5%	Flexibility gain from 6.28 to 6.21 per cent of the optimal tail: invariant. The hedge's tail reduction moves from 54.9 to 48.1 per cent (Table 6).
Loading on the cap	−90 to +50%	+15%	No cover at any positive loading; more than one per cent of the volume only at a discount of 90 per cent. The cap is dominated, not overpriced.
Deferrable share	0 to 100%	54.4%	57,632, 85,772, 98,015 and 108,423 at 25, 54.4, 75 and 100 per cent. Concave: the first quarter of flexibility is worth the most.
Shifting window	4 to 48 h	12 h	72,286 at four hours, 85,772 at twelve, 87,104 at twenty-four and 87,225 at forty-eight: most of the value is reached by twelve hours and almost none is added beyond twenty-four, because nothing past the published horizon is visible.
Scheduling horizon	calendar day <i>vs.</i> receding	receding	The receding horizon adds +15.4 per cent to the gain at a four-hour window, +7.1 at twelve hours and +3.2 at thirty-six.
Latency penalty	0–75 EUR/MWh	2 EUR/MWh	88,611 at zero, 85,772 at the base value, 75,115 at 10, 34,918 at 50 and 19,703 at 75 EUR per MWh moved; halved at about 40. It cannot turn negative, because no move is made at a loss.
Confidence level	0.90 to 0.99	0.95	86,108, 85,772, 85,138. The conclusion does not live at one quantile.
EVT threshold	$q_{0.960}$ to $q_{0.995}$	$q_{0.960}$	Shape between −0.180 and −0.019; $ES_{0.999}$ between 732 and 747 EUR/MWh across the whole range.

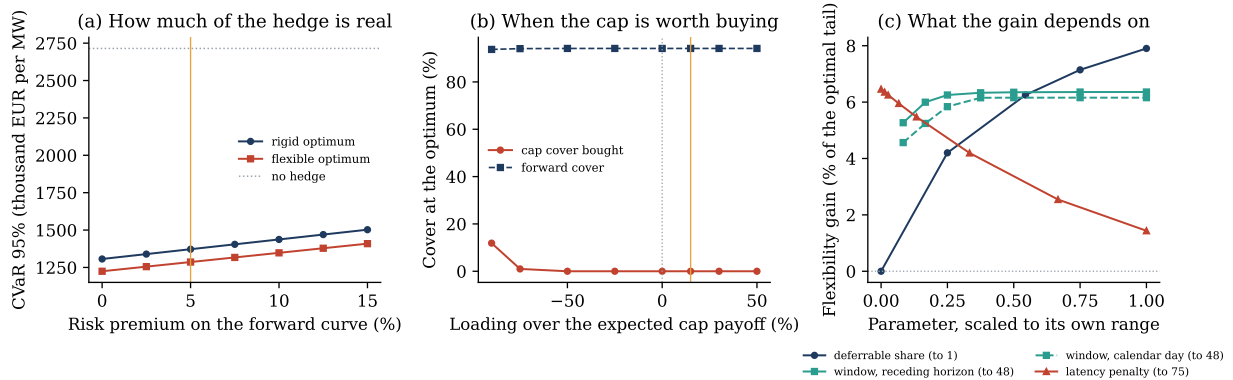


Figure 10. Left, the optimal tail as the assumed risk premium moves, against the unhedged tail: the reduction is nearly parallel to the premium, which is what an identity looks like. Centre, the cover the optimiser buys as the cap is made cheaper. Right, the operational gain against each of the parameters that govern it, each scaled to its own range.

12. Limitations

1. **The forward curve is modelled, not quoted.** Italian power futures were not accessible, so the curve is the simulated expected spot plus a stated risk premium. This is the single most consequential assumption in the paper, and it is not merely a source of noise: it makes the tail reduction of the financial leg largely circular, as Section 8.2 sets out and Section 8.5 quantifies. A quoted curve would change the level of the reduction and could change its sign relative to the premium actually charged. Nothing about the operational results depends on it.
2. **The facility takes prices as given.** Every value is per megawatt and is computed on prices that the facility’s own load does not move. That is reasonable for a facility small against its zone and doubtful for a campus of hundreds of megawatts in a zone the size of Sicily. A load that moves the price changes the result in two opposite directions: work shifted into cheap hours raises their price and narrows the spread the schedule captures, while the load withdrawn from dear hours lowers their price on everything the facility still consumes in them. Which effect dominates depends on the slope of the zonal supply curve hour by hour, which this paper does not estimate, so the operational saving cannot simply be multiplied by the size of the facility. If flexible load becomes common, the spreads it feeds on narrow for every facility, whatever its size.
3. **Every long-horizon risk number rests on seven annual observations.** The training block is 2016–2022, so the dispersion of an annual average, the quantity the two-factor model was built to reproduce and the one the whole cost distribution inherits, is calibrated against seven realisations, two of them from an energy crisis without modern precedent. The simulated tail of the annual bill is an order of magnitude, not a measurement, and no interval computed here repairs that.
4. **Two source gaps.** Italian and Norwegian prices are missing for most of 2025, which shortens the backtest and the forecasting test.
5. **National, not zonal, system variables** for Italy, because the feed does not publish zonal load.
6. **The deferrable share is a parameter.** The cluster trace is almost entirely training work and cannot separate deferrable from non-deferrable by label; the value used is the share of GPU-hours in tasks reserving at least one whole GPU, and it is swept.
7. **The schedule is a heuristic.** The greedy rule is feasible by construction and never moves work at a loss, but it is not the optimum of the placement problem, which is a transportation problem with window constraints. Since it is one feasible policy among those the same information allows, the values reported here are lower bounds on what that information is worth.
8. **One forecaster, one zone.** The forecast experiment of Section 10 is Sicilian, because the network was trained for Sicily only, and it was added after the pre-registration.
9. **The cluster trace is from one operator and two months,** and its timestamps are offsets, so only the time of day and the day of the week are recoverable.

13. Falsification

Four hypotheses were pre-registered so that the data could refute them.

H1, that flexibility reduces tail cost beyond what financial hedging alone achieves, survives the test it was written with. The rule is that H1 is falsified unless the flexible optimum beats the finance-only optimum by more than the width of the bootstrap confidence interval. The gain in Sicily is 85,772 EUR per MW against a widest compared interval of 101, 849 times the width, and the paired interval on the difference itself is [85,716, 85,824]. The criterion holds in all five zones, the closest being Norway NO2 at 35,206 against 62.

Two qualifications belong next to that verdict. The pre-registration says “out of sample”, and the criterion as written is not computable out of sample: the backtest yields one realised

outcome per zone and period, not a distribution with a tail. The criterion is therefore evaluated on the simulated distribution, where the margin is as large as it is partly because the hedged distributions are nearly deterministic, and the out-of-sample evidence is reported beside it as the separate statement it is: the operational saving of Section 9, positive in all 15 zone-periods with day-level intervals whose lowest bound is 3.7 per cent.

H2, that the neural scenario generator reproduces the tail better than the classical one, is *falsified*. The classical two-factor model is better on the variance-time curve by more than an order of magnitude and on the dispersion of the annual mean (Appendix A.2); the headline numbers use the classical generator.

H3, that the conclusions are specific to a high-price zone, is falsified in its strong form and refined: the tail reduction from hedging is general across all five zones, while the contribution of flexibility varies from 3.2 to 10.3 per cent of the optimal tail and is largest where prices are low relative to their intraday range.

H4, that amplitude estimation offers no practical advantage at reachable sizes, holds, and Appendix B quantifies by how much.

14. Conclusion

The finding that survives every objection raised against it in this paper is the operational one. A schedule that sees only what the day-ahead auction publishes, re-plans every day and never moves work at a loss, lowered the realised cost of an unhedged training facility in all 15 zone-periods of 2024–2026, by 4.1 to 15.6 per cent, and it did so without a view on any contract price or any future price. The result an operator can act on is where that value lives: in the width of the daily cycle rather than in its height, so that siting compute where power is cheap does not make flexibility less valuable, and in the next thirty-six hours rather than beyond them.

The second finding is the one the paper did not expect. A forecast of the unpublished hours, the natural way to let work wait longer, is worth little even when it is perfect and worth less than nothing when it is realistic: forecasts whose accuracy is competitive by the standard scores make the schedule worse, because their errors are the size of the differences they would have to predict. For this decision the right investment is in the intraday schedule and in the window, not in the forecaster, and the right way to judge a forecaster is by the decision it feeds.

The financial result is reported with its construction attached. For the Sicilian facility the annual bill has a mean of 1.31 and a 95 per cent conditional value at risk of 2.74 million euro per MW, and the jointly optimal design cuts that tail to 1.29 million. A forward curve built as the simulated expected spot plus a premium removes the variance of the thing it is priced from, and the honest content of that section is the shape of the frontier, the behaviour of the hedge out of sample, and the finding that a cap is dominated by the forward next to it.

The methodological findings are as useful as the numbers. A single mean-reverting factor, however well it fits hourly data, cannot represent the risk of an annual average. One-step likelihood is the wrong objective for a generator judged over a year, for the neural model as much as for the classical one. A linear autoregression with the right lags is as good a forecaster as a network with a hundred and seventy thousand parameters. And a quantum estimator that behaves exactly as the theory says can still be three orders of magnitude of hardware error rate away from being useful. Each of those branches was built to be tested, and each was.

Data and code availability

Every number in this paper is produced by a script in the repository and regenerated from the archived data snapshot, whose files carry their endpoint, request parameters, size and SHA-256 in `data/manifest.json`. `experiments/run_all.py` runs the experiments in dependency order, and `paper/build.py` writes every number quoted in the text from the result files before the paper is compiled, so the text cannot drift from the results. The pre-registration and every recorded deviation from it are in the repository. Series that cannot be redistributed are restored

by `experiments/download_data.py` and verified against the manifest.

Repository: <https://github.com/nicoloangileri/mqfs-compute-energy>.

Use of AI tools

This project was developed with extensive use of Anthropic’s Claude models, in `claude.ai` and in Claude Code, which wrote most of the code and drafted most of the text under the author’s direction. The author set the scope, the data sources and the rules the work is bound by: every published number comes from downloaded data and executed code, negative results are reported, and no citation is included that has not been checked against its source. The author reviewed the results and is responsible for the content. The version history of the repository records how the work was produced.

This working paper is an independent research output of the Mediterranean Quantitative Finance Society and is not affiliated with, endorsed by, or produced under the auspices of the University of Palermo. It is not investment advice. © Nicolò Angileri, Mediterranean Quantitative Finance Society.

References

- M. H. Albadi and E. F. El-Saadany. A summary of demand response in electricity markets. *Electric Power Systems Research*, 78(11):1989–1996, 2008. doi: 10.1016/j.epsr.2008.04.002.
- Fred Espen Benth, Jurate Saltyte Benth, and Steen Koekebakker. *Stochastic Modelling of Electricity and Related Markets*. World Scientific, Singapore, 2008. doi: 10.1142/6811.
- Hendrik Bessembinder and Michael L. Lemmon. Equilibrium pricing and optimal hedging in electricity forward markets. *Journal of Finance*, 57(3):1347–1382, 2002. doi: 10.1111/1540-6261.00463.
- Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. In *Quantum Computation and Information*, volume 305 of *Contemporary Mathematics*, pages 53–74. American Mathematical Society, 2002. doi: 10.1090/conm/305/05215.
- Hans N. E. Byström. Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance*, 14(1):41–55, 2005. doi: 10.1016/S1059-0560(03)00032-7.
- Miguel Carrión, Antonio J. Conejo, and José M. Arroyo. Forward contracting and selling price determination for a retailer. *IEEE Transactions on Power Systems*, 22(4):2105–2114, 2007. doi: 10.1109/TPWRS.2007.907397.
- Álvaro Cartea and Marcelo G. Figueroa. Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, 12(4):313–335, 2005. doi: 10.1080/13504860500117503.
- Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.
- Antonio J. Conejo, Miguel Carrión, and Juan M. Morales. *Decision Making Under Uncertainty in Electricity Markets*. International Series in Operations Research & Management Science. Springer, New York, 2010. doi: 10.1007/978-1-4419-7421-1.
- Alex de Vries. The growing energy footprint of artificial intelligence. *Joule*, 7(10):2191–2194, 2023. doi: 10.1016/j.joule.2023.09.004.
- Tilman Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.

- Dmitry Grinko, Julien Gacon, Christa Zoufal, and Stefan Woerner. Iterative quantum amplitude estimation. *npj Quantum Information*, 7:52, 2021. doi: 10.1038/s41534-021-00379-1.
- Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: 10.1002/qj.3803.
- International Energy Agency. Energy and AI. World energy outlook special report, International Energy Agency, Paris, 2025.
- Christopher Kath and Florian Ziel. The value of forecasts: Quantifying the economic gains of accurate quarter-hourly electricity price forecasts. *Energy Economics*, 76:411–423, 2018. doi: 10.1016/j.eneco.2018.10.005.
- Patrick Kidger, James Foster, Xuechen Li, and Terry Lyons. Neural SDEs as infinite-dimensional GANs. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5453–5463, 2021.
- Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. doi: 10.2307/1913643.
- Jesus Lago, Grzegorz Marcjasz, Bart De Schutter, and Rafał Weron. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983, 2021. doi: 10.1016/j.apenergy.2021.116983.
- Zhenhua Liu, Adam Wierman, Yuan Chen, Benjamin Razon, and Niangjun Chen. Data center demand response: Avoiding the coincident peak via workload shifting and local generation. *Performance Evaluation*, 70(10):770–791, 2013. doi: 10.1016/j.peva.2013.08.014.
- Francis A. Longstaff and Ashley W. Wang. Electricity forward prices: A high-frequency empirical analysis. *Journal of Finance*, 59(4):1877–1900, 2004. doi: 10.1111/j.1540-6261.2004.00682.x.
- Julio J. Lucia and Eduardo S. Schwartz. Electricity prices and power derivatives: Evidence from the Nordic power exchange. *Review of Derivatives Research*, 5(1):5–50, 2002. doi: 10.1023/A:1013846631785.
- Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. Recalibrating global data center energy-use estimates. *Science*, 367(6481):984–986, 2020. doi: 10.1126/science.aba3758.
- Jakub Nowotarski and Rafał Weron. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568, 2018. doi: 10.1016/j.rser.2017.05.234.
- Yumi Oum and Shmuel S. Oren. Optimal static hedging of volumetric risk in a competitive wholesale electricity market. *Decision Analysis*, 7(1):107–122, 2010. doi: 10.1287/deca.1090.0167.
- Yumi Oum, Shmuel Oren, and Shijie Deng. Hedging quantity risks with standard power options in a competitive wholesale electricity market. *Naval Research Logistics*, 53(7):697–712, 2006. doi: 10.1002/nav.20184.
- Asfandiyar Qureshi, Rick Weber, Hari Balakrishnan, John Guttag, and Bruce Maggs. Cutting the electric bill for internet-scale systems. In *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication*, pages 123–134, 2009. doi: 10.1145/1592568.1592584.

- Ana Radovanović, Ross Koningstein, Ian Schneider, Bokan Chen, Alexandre Duarte, Binz Roy, Diyu Xiao, Maya Haridasan, Patrick Hung, Nick Care, Saurav Talukdar, Eric Mullen, Kendal Smith, MariEllen Cottman, and Walfredo Cirne. Carbon-aware computing for datacenters. *IEEE Transactions on Power Systems*, 38(2):1270–1280, 2023. doi: 10.1109/TPWRS.2022.3173250.
- Lei Rao, Xue Liu, Le Xie, and Wenyu Liu. Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity-market environment. In *2010 Proceedings IEEE INFOCOM*, pages 1–9, 2010. doi: 10.1109/INFOCOM.2010.5461933.
- R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000. doi: 10.21314/JOR.2000.038.
- Eduardo Schwartz and James E. Smith. Short-term variations and long-term dynamics in commodity prices. *Management Science*, 46(7):893–911, 2000. doi: 10.1287/mnsc.46.7.893.12034.
- Arman Shehabi, Sarah J. Smith, Alex Hubbard, Alex Newkirk, Nuo Lei, Md Abu Bakar Siddik, Billie Holecek, Jonathan Koomey, Eric Masanet, and Dale Sartor. 2024 United States data center energy usage report. Technical Report LBNL-2001637, Lawrence Berkeley National Laboratory, Berkeley, CA, 2024.
- Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. WaveNet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
- Qizhen Weng, Wencong Xiao, Yinghao Yu, Wei Wang, Cheng Wang, Jian He, Yong Li, Liping Zhang, Wei Lin, and Yu Ding. MLaaS in the wild: Workload analysis and scheduling in large-scale heterogeneous GPU clusters. In *19th USENIX Symposium on Networked Systems Design and Implementation (NSDI 22)*, pages 945–960, 2022.
- Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, 2014. doi: 10.1016/j.ijforecast.2014.08.008.
- Adam Wierman, Zhenhua Liu, Iris Liu, and Hamed Mohsenian-Rad. Opportunities and challenges for data center demand response. In *International Green Computing Conference*, pages 1–10, 2014. doi: 10.1109/IGCC.2014.7039172.
- Stefan Woerner and Daniel J. Egger. Quantum risk analysis. *npj Quantum Information*, 5:15, 2019. doi: 10.1038/s41534-019-0130-6.
- Florian Ziel and Rafał Weron. Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics*, 70:396–420, 2018. doi: 10.1016/j.eneco.2017.12.016.

A. The two networks

Both networks are trained from random initialisation on the archived data of Section 4. Neither is a language model, neither is fine-tuned from a pretrained checkpoint, and no external weights are used. A model card in the repository records the data, features, objectives, optimiser settings, seeds and runtimes.

A.1 The probabilistic forecaster

The forecaster states a predictive distribution, not a point, for each of the next 72 hourly prices, issued daily at 12:00 UTC from the past week of prices and system state. It is a stack of dilated causal convolutions (van den Oord et al., 2016) with a monotone quantile head trained on the

pinball loss (Koenker and Bassett, 1978); the selected configuration has 96 channels, 5 blocks and 170,601 parameters. Nothing dated after the issue time enters the features: no weather forecast, no future load. Exogenous inputs are standardised with the mean and standard deviation of the training block only; an earlier version used the whole sample, which let the test period’s statistics into the inputs. Windows are built on a complete hourly grid, so a window that would straddle the nine-month gap in the Italian feed is dropped rather than joined across it, and a window is used for training or scoring only if every hour it predicts lies inside its block.

A negative result about a neural network is only worth reading if the network was trained competently, so the protocol is given in full in Table 11.

Table 11. Training protocol of the quantile forecaster. Every entry is produced by `experiments/forecast.py` and stored in `results/forecast_validation.json`; the parameter count is read off the frozen checkpoint.

Data	
Windows	2,547 training issue times, one per day of 2016–2022; 362 validation issue times in 2023. Each window is 168 hours of history in and 72 hours of prediction out, issued at 12:00 UTC with a 12-hour lead, which is when the day-ahead auction has cleared.
Target	The deseasonalised deviation of the Sicilian day-ahead price, on the inverse hyperbolic sine scale of Section 5.
Features	
Past window, 13 per hour	Deviation, transformed price, deterministic seasonal level, standardised residual load, temperature, wind speed at 100 m, and seven cyclical calendar encodings.
Future window, 8 per hour	Seasonal level and the same calendar encodings.
Deliberately excluded	<i>Anything dated after the issue time.</i> No weather forecast, no forecast or realised future load, no future price.
Objective and optimisation	
Loss	Pinball loss averaged over nine quantile levels (0.01 to 0.99) and 72 horizons. The head emits the lowest quantile plus non-negative increments, so the quantiles cannot cross.
Settings	AdamW, learning rate 10^{-3} , weight decay 10^{-4} , batch 64, gradient clipping at norm 1, dropout 0.1, seed 20260806.
Stopping	Early stopping on validation pinball loss, patience 15, cap of 200 epochs. The selected run stopped after 35 epochs with its best at epoch 20.
Model choice and freezing	
Candidates	Four, compared on the validation block by mean pinball loss as pre-registered (Table 12).
Refit	As pre-registered, the selected configuration is refitted on training plus validation, 2,912 windows, for the 20 epochs early stopping chose, and then frozen. The frozen network is the one scored on the test block and used by the scheduler.
Cost	461 seconds in total on the CPU of a laptop.

Table 12. The four candidate configurations on the validation block, (channels, blocks).

Configuration	Pinball	CRPS	Epochs	Seconds
(32, 4)	0.04124	0.11536	54	68
(64, 4)	0.04129	0.11584	31	72
(64, 5)	0.04090	0.11320	22	61
(96, 5)	0.04062	0.11407	35	163

On the validation block, the comparison the pre-registration names, the linear expert has the lower pinball loss and CRPS, 0.11287 against 0.11407; the network is not ahead on the rule that selects it, and Section 10 shows the test block agrees.

Table 13. Validation block, calendar 2023, identical windows for every model; the network is the one selected on this block, the baselines are fitted on the training block. Lower is better for the first two columns; coverage should match its nominal level; the last column is the Kolmogorov-Smirnov statistic of the probability integral transform.

Model	Pinball	CRPS	Skill vs. naive	Cov. 50%	Cov. 95%	PIT KS
Neural quantile network	0.04062	0.11407	0.149	0.536	0.971	0.038
Linear expert autoregression	0.04039	0.11287	0.158	0.498	0.968	0.058
Seasonal naive	0.05065	0.13407	0.000	0.394	0.748	0.171
Two-factor generator	0.09149	0.22670	-0.691	0.308	0.781	0.174

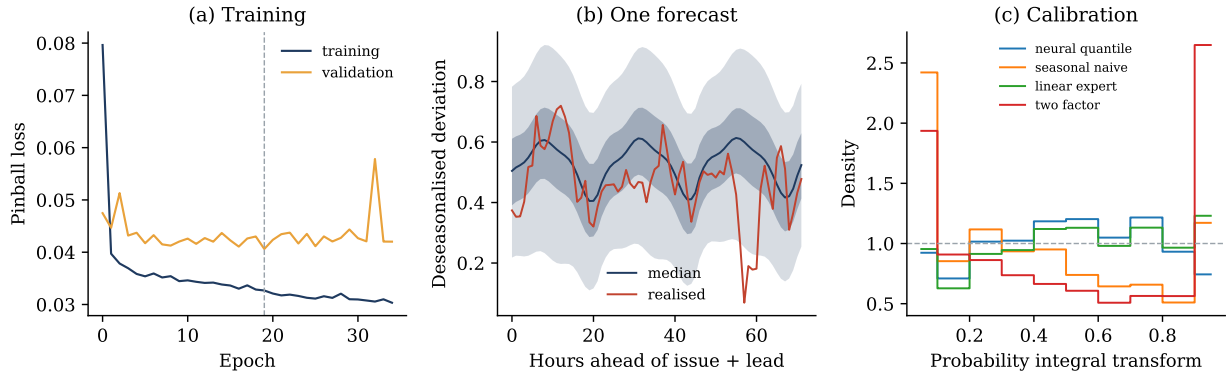


Figure 11. Left, the training curve of the selected network. Centre, one issued forecast with its 1–99, 5–95 and 25–75 per cent bands against what was realised. Right, the probability integral transform of every model: a calibrated forecast is flat at one.

A.2 The neural stochastic differential equation

The second network is the drift and diffusion of a stochastic differential equation with a learned Student- t innovation, estimated by exact maximum likelihood of the one-step Euler transition (Chen et al., 2018; Kidger et al., 2021). Its state carries a monthly exponentially weighted memory of the path, so that the level can wander as the equilibrium factor does in the classical model, learned rather than imposed. It is trained on the 61,368 hourly transitions of the Sicilian training block.

Table 14. The three variants of the neural generator, from `results/scenarios.json`. The dispersion column is the standard deviation of the annual mean deviation across simulated years, against 0.534 in the data and 0.420 from the classical two-factor model.

Variant	Epochs	Seconds	ν	Dispersion
Free drift, pure likelihood	109	30	2.77	30.138
Signed reversion, pure likelihood	120	33	2.56	0.169
Signed reversion, moment penalty	48	781	3.73	0.803

The variants are reported because the failures are the finding. With a free drift the one-step likelihood is excellent and the simulation explodes: a very slightly repelling drift is invisible one hour ahead and fatal over sixty thousand hours. Constraining the reversion speeds to be non-negative removes the explosion and replaces it with the opposite error, a generator whose annual average barely moves. Only the third variant, which adds a differentiable penalty matching the variance of h -hour averages to the same quantity in the data, is usable, and it still loses to the classical model: its mean absolute log error on the variance-time curve is 0.963 against 0.069. Hypothesis H2 is answered against the neural generator, and the classical one carries every simulated number in this paper. The lesson that carries beyond it is that one-step likelihood and

long-horizon realism are different objectives.

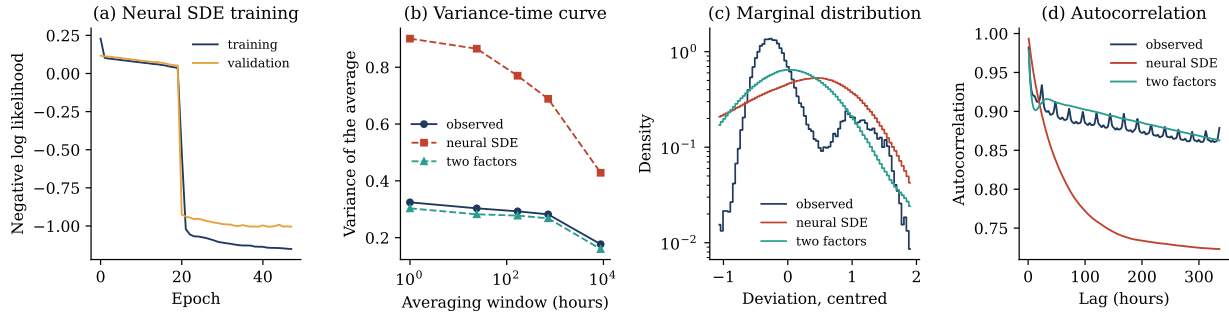


Figure 12. Generators against the market. Training curve of the neural stochastic differential equation, the variance-time curve of each generator against the observed one, the marginal distributions, and the autocorrelation out to two weeks.

B. Quantum tail estimation

What this appendix is: a verified implementation and a resource estimate on one specific instance, not a methodological contribution. The algorithm is due to Brassard et al. (2002) and Grinko et al. (2021); what is added here is only that it is run against the cost distribution this paper produces, with the gates counted after transpilation and the error rate that would be needed to survive them computed rather than cited. The conclusion that there is no practical advantage was the pre-registered expectation, hypothesis H4, and it is what the measurements give.

The tail quantities of Section 8 are expectations of a function of a random variable, and such an expectation can be written as the amplitude of one qubit (Brassard et al., 2002; Woerner and Egger, 2019). The discretised cost distribution is loaded as an amplitude encoding on n qubits, an integer comparator flips an objective qubit on the states above the threshold, and iterative amplitude estimation squeezes a confidence interval on that amplitude by running the Grover operator a growing number of times. The estimator is implemented in the repository rather than imported, so that the oracle calls can be counted.

Correctness. The amplitude prepared by the circuit equals the exact tail probability of the discretised distribution to 8.6×10^{-15} across registers of three to seven qubits, and the estimator’s interval covers the exact value at every level tested.

Scaling. Fitting the error against the number of evaluations of the distribution, on the same axis for both methods, gives a slope of -0.873 for amplitude estimation and -0.485 for plain Monte Carlo, against the theoretical -1 and $-1/2$.

Resources. After transpilation to a hardware basis on a linear coupling map, the state preparation needs 39 two-qubit gates at three qubits and 383 at seven; the Grover operator needs 236 and 3,753 respectively, at depths of 478 and 5,432. The growth is a factor of 2.0 per additional state qubit, which is the exponential cost of exact loading; projected to the twelve qubits a usable cost resolution would need, the Grover operator alone is 1.4×10^5 two-qubit gates.

Noise. A useful run at seven qubits, 64 Grover applications, is 240,192 two-qubit gates. Under a depolarising model it survives with probability 3.9×10^{-314} at the three-in-a-thousand error rate of current superconducting hardware, and would need 2.9×10^{-6} per two-qubit gate for an even chance of surviving.

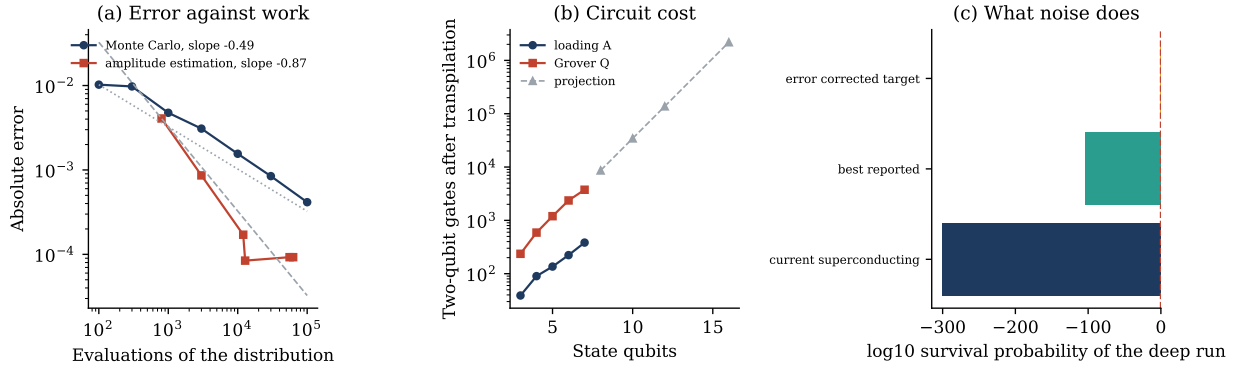


Figure 13. Left, measured error against work for both estimators with the theoretical slopes as dashed guides. Centre, transpiled two-qubit gate counts against register size, with the projection. Right, the survival probability of a useful run under three two-qubit error rates.

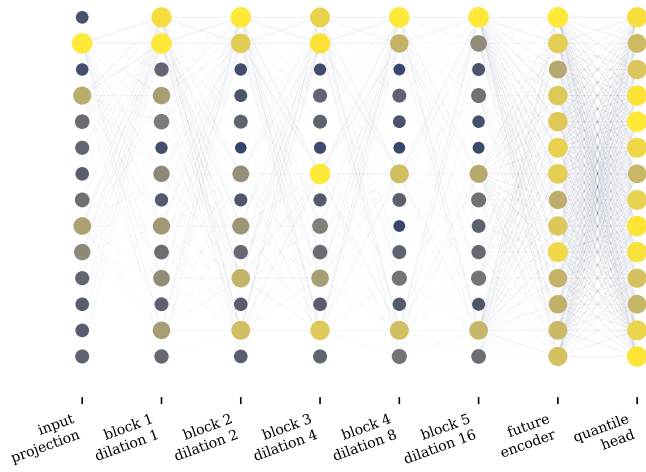
C. Interactive artefacts

Three self-contained pages accompany the paper. Each carries real fitted objects rather than pictures of them, recomputes something in the browser, and states how far its own recomputation lands from the reference computed offline. They are listed with the command that regenerates each, because an artefact that cannot be rebuilt is a screenshot. They are served at [nicoloangileri.github.io/mqfs-compute-energy](https://github.com/nicoloangileri/mqfs-compute-energy), and each is also a single file that works offline.

Table 15. The three pages, what they carry and what a reader can change. Each also writes the static counterpart used here, from the same arrays that feed the page, with the `--figure` flag.

File and command	What it carries	What the reader can do
results/ forecaster_brain.html python experiments/ brain_viz.py	The 170,601 weights of the frozen forecaster, three days of the test block it never saw, and the PyTorch prediction for each as a reference.	Redraw the last week of prices by hand and watch the channel activations and the predictive fan respond. The page runs the same forward pass in JavaScript and reports its largest deviation from the PyTorch reference.
results/ monte_carlo_3d.html python experiments/ mc_viz.py	The calibrated generator with its fitted jump family, the seasonal shape, the compute profile, the forward curve, and reference cost statistics from 20,000 paths through the compiled core.	Watch price paths grow through the simulated year in three dimensions while the cost distribution builds beside them, and move the forward, the cap and the flexible share while it runs. The page schedules with a JavaScript port of the receding-horizon rule and shows its running estimate against the offline reference.
results/ quantum_3d.html python experiments/ quantum_viz.py	The amplitudes of the loaded and amplified state, computed statevector by statevector in Qiskit for every Grover power shown.	Step the amplification and see probability rotate into the tail; run the iterative estimator live against the exact amplitude, next to plain Monte Carlo on the same axis.

(a) Channel activation through the network, issue of 2026-03-23



(b) The predictive fan of that pass

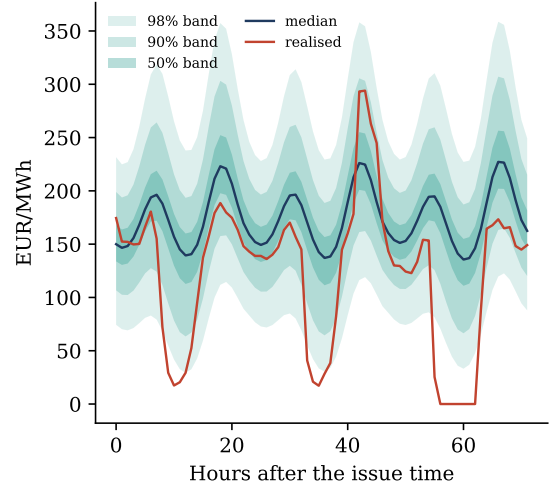
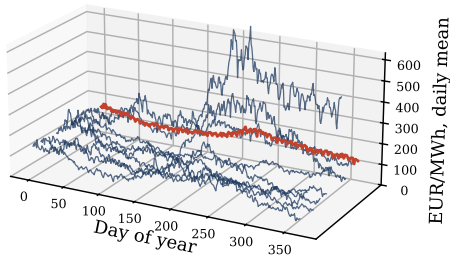


Figure 14. The forecaster as the page shows it, on the test-block day with the widest intraday range. Left, the mean absolute activation of each channel through the input projection, the dilated blocks, the future encoder and the quantile head. Right, the predictive fan of the same forward pass against what was realised.

(a) Nine paths and the median, in depth



(b) Cost distribution, dashed CVaR at 95%

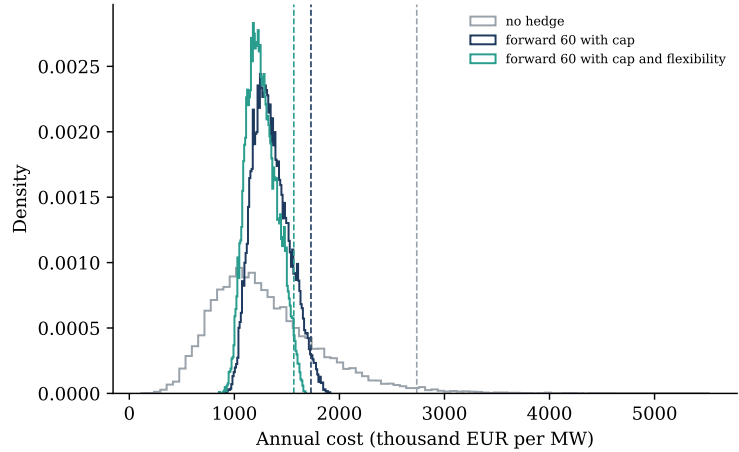


Figure 15. One frame of the live simulation. Left, simulated price paths over the year, laid out in depth, with the median in front. Right, the distribution of the annual bill under each of the three reference hedges the page starts from, with the tail measure dashed.

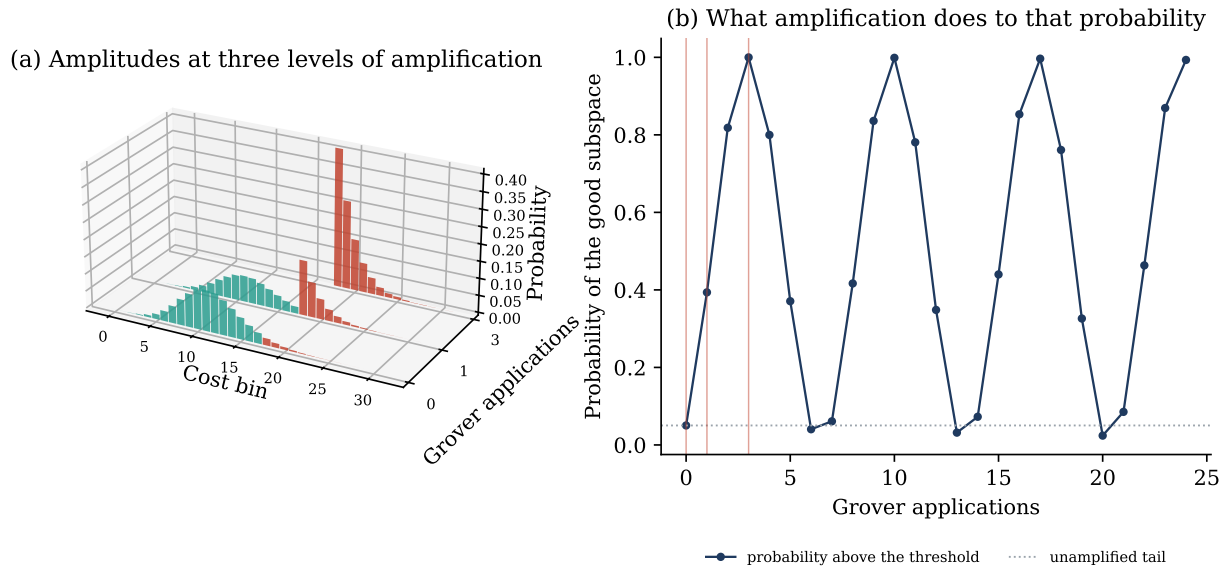


Figure 16. Amplitude amplification on the discretised cost distribution. Left, the probability of each cost bin at three levels of amplification, with the bins above the threshold in the second colour. Right, the probability of the good subspace as a function of the number of Grover applications, with the unamplified tail marked.

D. Deviations from the pre-registration, and corrections

The pre-registration was written on 6 August 2026, after the descriptive statistics and before any model was fitted. It fixes the sample split, the selection rules and the four hypotheses, and it says that a discovered error is fixed by re-running the whole pipeline and recording the fact. Everything below is recorded in `PREREGISTRATION.md` as well, with its date. The numbers of the version before the corrections are in the repository history (tag `v0.7`).

Before the first results. The linear trend listed in the seasonal specification was dropped, because a trend fitted through the 2021–2022 crisis pins every simulated year to the crisis level; a two-factor model was added to the pre-registered single-factor family, because no single factor reproduces the dispersion of an annual average; the extreme value run length was shortened from twenty-four to six hours, because the longer one leaves too few independent clusters. None of these decisions consulted the test block.

Where a criterion could not be computed as written. Hypothesis H1 asks for a conditional value at risk out of sample, which one realised year per zone cannot provide. It is evaluated on the simulated distribution, and the out-of-sample evidence is reported next to it as what it is.

Corrections made on 25 September 2026, after the backtest had been run once. Each of the following was a defect in the code, found by reading it, not by looking for a better number; each was fixed and the whole pipeline re-run.

1. **The jump family.** The engine drew Gaussian jumps, selected in four zones including Sicily, from a two-sided exponential law with twice the fitted variance, so its paths did not have the moments the calibration had matched. The previous version reported a simulated flexibility gain of 90,232 EUR per MW in Sicily; this one reports 85,772.
2. **The optimiser.** Of the 200 programmes solved in the previous hedging run, 92 ended *optimal_inaccurate* and 8 at the iteration limit of the interior point solver. The programme is now solved by HiGHS and any status other than optimal stops the run.
3. **The scheduler.** Moves were decided on the price difference alone and the latency penalty charged afterwards, and a decision could not re-plan work placed by the previous one nor reach back across the gate to hours it could already see. The receding-horizon rule of Section 6 replaces it.
4. **The out-of-sample metric.** The previous version measured the value of flexibility out of sample as the difference between the two hedged optima, which differ in their forward cover as well as in their schedule, and reported it as 4.2 to 18.5 per cent. The operational saving is now measured on the unhedged facility, where nothing but the schedule differs, with day-level intervals; realised series are also cut at their gaps, which the previous version joined across in Norway and France in 2025.
5. **The forecaster.** Exogenous inputs were standardised on the whole sample, windows were joined across the 2025 gap, training windows predicted up to three days of the following block, and the refit on training plus validation that the pre-registration asks for had not been performed. All four are fixed; the network was retrained and frozen before the test block was scored for the first time.

Added after the pre-registration. The value of a forecast to the scheduler, Section 10, was not pre-registered. It uses the test block, which the pre-registration reserves for a single scoring of each frozen model, and it is the first and only time the forecasters are scored there; its results should be read as exploratory.