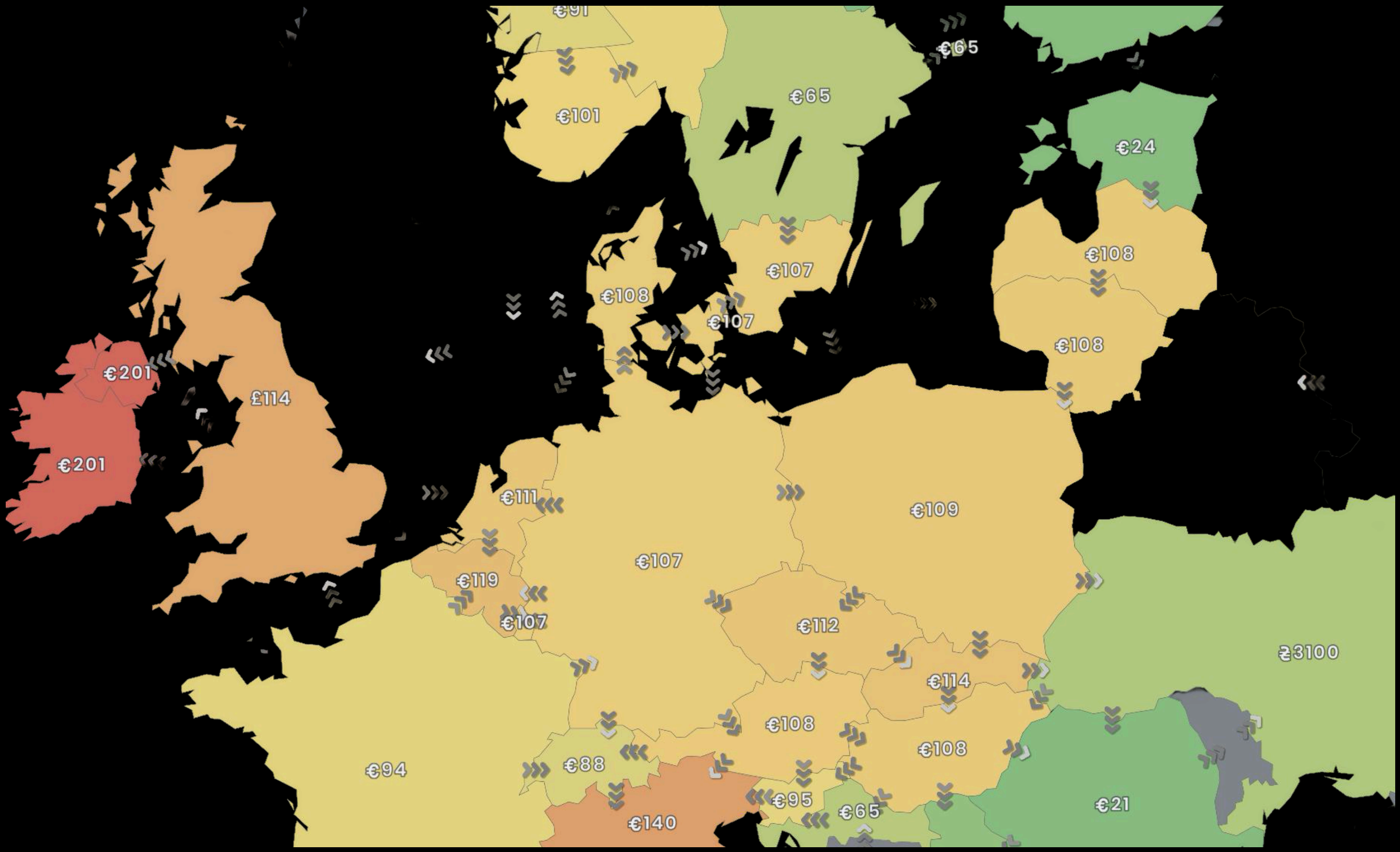
Evaluating BESS Forecast Value:

Revenue Capture, Rank Skill, and the Limits of MAE in European Day-Ahead Markets



RESEARCH REPORT

KEY FINDINGS

+€4,860/MW
Additional DAA Margin From The Electricity Maps Forecast Vs. Persistence, Over 8 Months

90.7%
Of maximum possible (perfect-foresight) revenue captured by the Electricity Maps forecast

97.1%
Fill rate achieved by the Electricity Maps forecast across the period (orders that cleared)

JULY 2026

Electricity Maps

RESEARCH REPORT

Evaluating BESS Forecast Value: Revenue Capture, Rank Skill, and the Limits of MAE in European Day-Ahead Markets

A collaboration between **Electricity Maps** and **FlexUp**

Period: 1 October 2025 to 1 June 2026 (8 months)

KEY FINDINGS

+€4,860/MW	90.7%	97.1%
Additional DAA margin from the Electricity Maps forecast vs. persistence, over 8 months	Of maximum possible (perfect-foresight) revenue captured by the Electricity Maps forecast	Fill rate achieved by the Electricity Maps forecast across the period (orders that cleared)

PART ONE

Forecast quality for BESS operators

In Germany, BESS traders typically capture 20-30% of their annual trading margin at the day-ahead auction (DAA); the rest comes from intraday and balancing markets. In the DAA, Battery Energy Storage Systems earn revenue by arbitraging electricity prices: charge when prices are low, discharge when they are high.

The DAA is also the most transparent market for benchmarking forecasting performance. All bids are submitted before closure and settled against actual clearing prices. A better forecast means better timing, better bid sizing, and fewer rejected bids.

Many BESS operators lack a transparent, quantified view of how much their forecast quality is costing or earning them. The purpose of this paper is to provide that view.

Methodology

FlexUp simulated a time-arbitrage strategy in the DAA using a 1 MW / 2 MWh BESS in the DE-LU bidding zone, run with three different price forecasts over the eight-month period (1 October 2025 – 1 June 2026).

Three forecasts compared

- **Persistence forecast:** yesterday's prices used as today's forecast: a naive but widely-used baseline that assumes “today looks like yesterday.”
- **Electricity Maps forecast:** the day-ahead price forecast published at D-1 11:00, one hour before DAA gate closure.
- **Perfect foresight:** actual clearing prices used as the forecast, representing the theoretical maximum achievable margin.

Trading strategy

Each forecast is run through the same linear optimization: find the charge/discharge schedule that maximizes margin given expected prices, subject to the battery's physical constraints.

The optimizer sets bids and asks from the forecast's own price spread, accounting for round-trip efficiency; out-of-the-money bids are rejected, so forecast errors cause missed revenue directly through unfilled orders. Margins are calculated against actual clearing prices.

Two decisions drive DAA trading performance: when to schedule charges and discharges, and what bid price to set for buying and selling electricity. Fill rate measures the share of bids actually selected by the market, reflecting the quality of the forecast on the second decision. Capture rate measures the share of perfect-foresight revenue obtained, reflecting both decisions combined.

Specifications and limitations:

Power capacity	1.0 MW
Energy capacity	2.0 MWh
State-of-charge range	0% to 100%
Max. cycles per day	2.0
Round-trip efficiency (RTE)	85%

Results: 8-month DAA margin comparison

Across the eight-month period (1 October 2025 – 1 June 2026), the three forecasts produced the following total DAA margins per MW of installed capacity:

Persistence forecast (baseline)	**Electricity Maps forecast** (+€4,860 vs. baseline)	**Perfect foresight** (+€9,575 vs. baseline)
€41,302	€46,162	€50,877

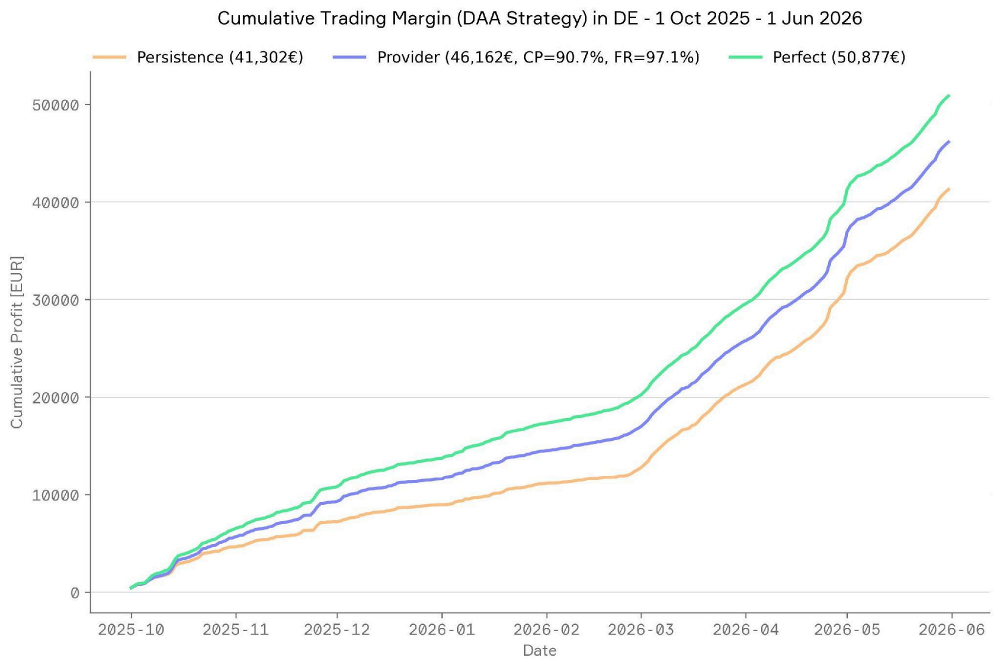


Figure 1 – Cumulative DAA trading margin by forecast, DE-LU, 1 Oct 2025 – 1 Jun 2026 (€/MW). The gap widens through the volatile spring – the period that, in Part Two, also drives the highest revenue capture.

The Electricity Maps forecast outperformed the persistence baseline by €4,860 per MW over the eight months, capturing 90.7% of the total possible perfect-foresight margin at a 97.1% fill rate.

Perfect foresight would have added a further €4,715/MW, the headroom that still separates a strong forecast from a flawless, but fictional one.

Example day: forecast accuracy and fill rates

A single day from the period (the 30th of November 2025) shows how forecast errors propagate into missed trading opportunities.

- **The persistence trader** expects an average price of €90/MWh against an actual €85, and gets the shape wrong, expecting a morning peak that did not materialize. Seven of the eighteen bids are rejected. Daily margin: €31.79, at 57% of optimal.
- **The Electricity Maps trader** correctly identifies the cheapest overnight charging window and the evening discharge peak. All 18 MTUs fill. Daily margin: €47.90, at 85% of optimal.
- **The perfect-foresight trader** schedules charges and discharges perfectly, at optimal bid prices. Daily margin: €56.26, at 100% of optimal.

Forecast	Daily margin	Fill rate	vs. optimal
Persistence	€31.79	61%	57%
Electricity Maps	€47.90	100%	85%
Perfect foresight	€56.26	100%	100%

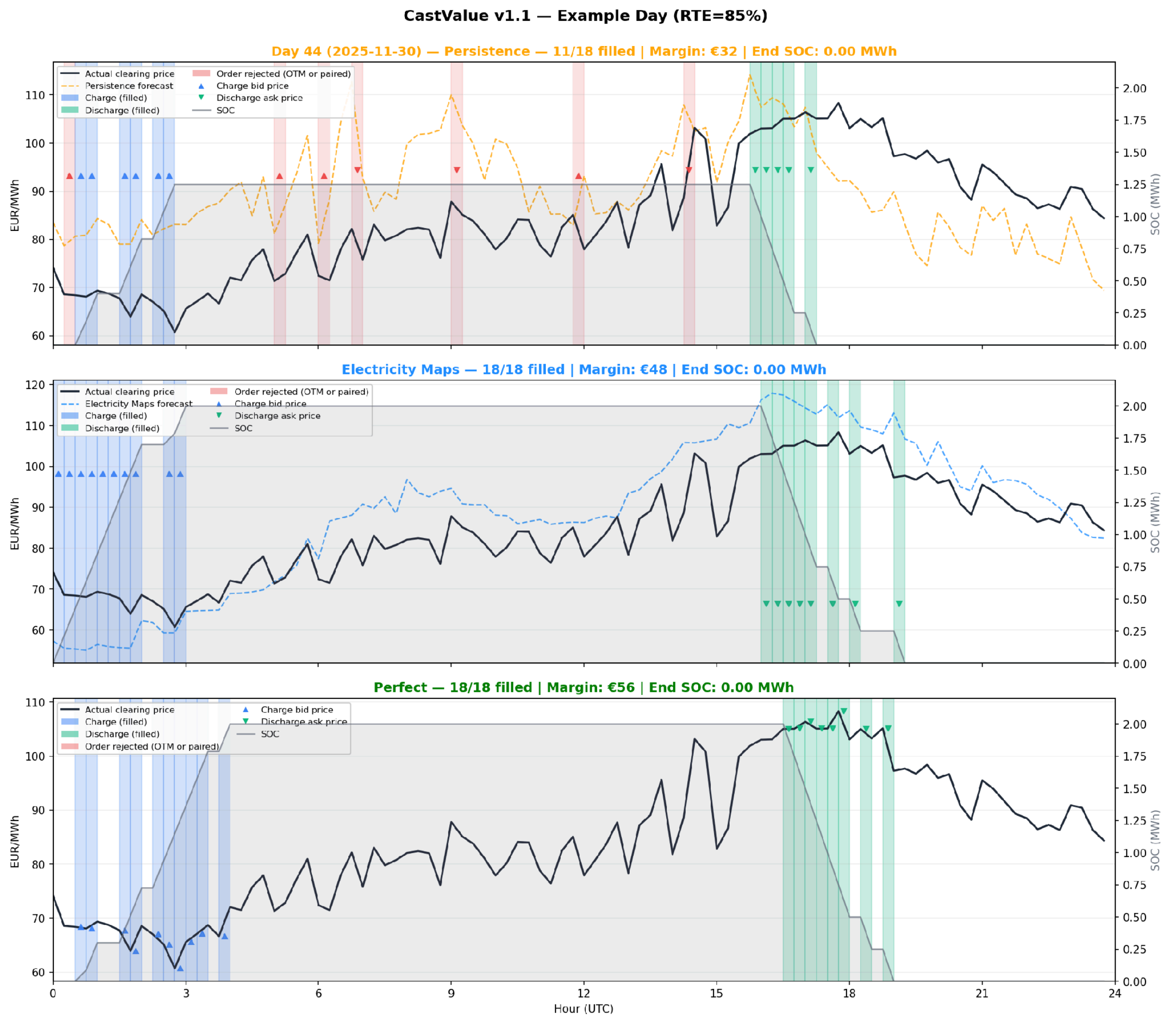


Figure 2 – Example day (30 Nov 2025). Actual price, each forecast's bids/asks, fills, and resulting state of charge.

On this example day, the persistence forecast achieved a 61% fill rate at 57% of optimal margin. The Electricity Maps forecast achieved a 100% fill rate at 85% of optimal. That gap compounds across a full year, not only in the quality of the schedule, but in whether orders execute at all.

PART TWO

Lower error doesn't mean higher profit

In Part One we quantified the value of a better forecast by simulating the actual decisions it drives: when to charge and discharge, and what bid prices to set. The result was a direct measure of revenue impact. In practice, forecasts are rarely evaluated this way. Most procurement decisions rely on regression metrics such as MAE or RMSE. The question Part Two addresses is whether those metrics are good proxies for revenue impact.

The same model extended across key European markets (Belgium, East and West Denmark, France, Germany, Great Britain, and the Netherlands) over the same eight months, surfaced a pattern worth highlighting.

Forecast error and captured revenue do not move together the way one would assume. The metric used almost everywhere to buy, sell, and contract forecasts: mean absolute error (MAE), turns out to be a poor, and at times actively misleading, guide to how much revenue a forecast lets you capture. This extends to all 34 European zones analyzed (Part Three).

The evidence

Across 56 zone-months (seven bidding zones over eight months), and consistent across all 34 zones in the full dataset, the months with the lowest forecast error were not the months with the highest capture.

Germany makes that vivid: its three highest-error months (March–May 2026, MAE €22–24/MWh) were its three best-capture months (94–96%), while its three lowest-error months were its three worst-capture months. The correlation runs in the opposite direction from what MAE implies. Read the way MAE is normally read, 'lower is better', it points in the wrong direction.

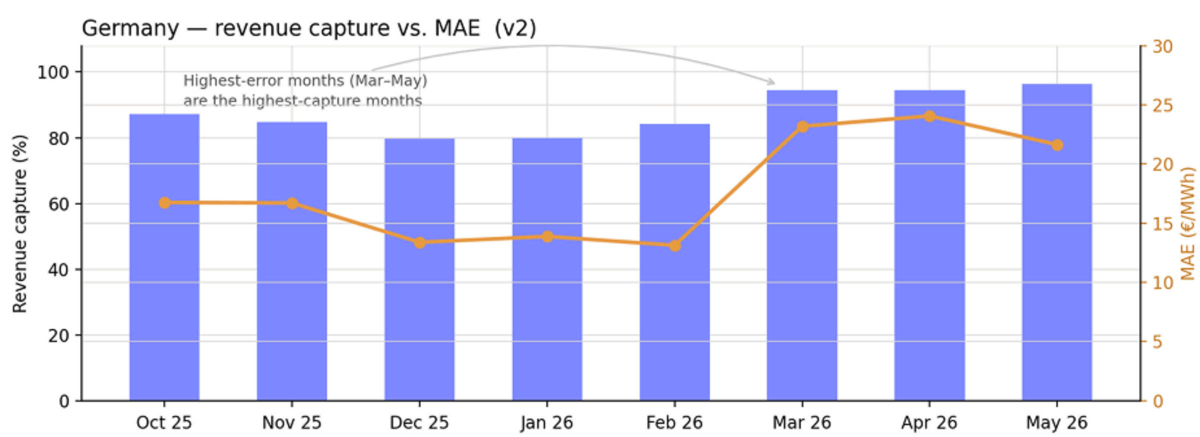


Figure 3 – Germany, monthly. Revenue capture (purple bars) against MAE. The highest-error months are the highest-capture months, a correlation that runs counter to the "lower is better" intuition.

MAE misleads

MAE answers one question: on average, how many €/MWh is the forecast off by? But a BESS arbitrage P&L does not depend only on that. First and foremost, it depends on whether the forecast got the order of the day right – whether the periods it called cheap really are the cheapest (charge) and the periods it called expensive really are the most expensive (discharge). Accuracy in level and accuracy in ranking are two different things, and they come apart in both directions:

- **High error, correct ranking → high capture.** A forecast can be biased or noisy in absolute terms but still rank the day correctly. Here, the optimizer charges in the right trough and discharges in the right peak. The example day from Part One demonstrates it: persistence got the shape wrong – smearing discharge across midday – and captured 57% of optimal, while the Electricity Maps forecast ranked market time units over the day correctly and captured 85%.
- **Low error, inverted ranking → lost capture.** A forecast can sit tight against the actual price curve yet invert two adjacent ranks around the evening peak. The optimizer then discharges an hour early, or charges in the wrong trough, and a single rank inversion at the day's extremes erases a disproportionate share of the spread, even though the MAE barely moved.

This is also why the relationship runs backwards across the panel. Error magnitude scales with volatility: volatile months have wide price spreads, which produce large absolute errors, but wide spreads are also where arbitrage is largest, and the cheap and expensive hours are most important to tell apart. In this dataset, higher MAE coincides with higher capture, a positive correlation that runs counter to the 'lower is better' intuition.

This does not mean that a higher-error forecast is always better. It means that selecting a forecast on MAE alone, without accounting for the decisions it will inform, risks prioritizing the wrong signal. A dispatch decision turns on order, not distance, and MAE measures distance.

The metric that works: Spearman ρ

Spearman's rank correlation measures exactly what the optimizer acts on. It ranks every settlement period of the day from cheapest to most expensive and asks how faithfully the forecast reproduces that ranking.

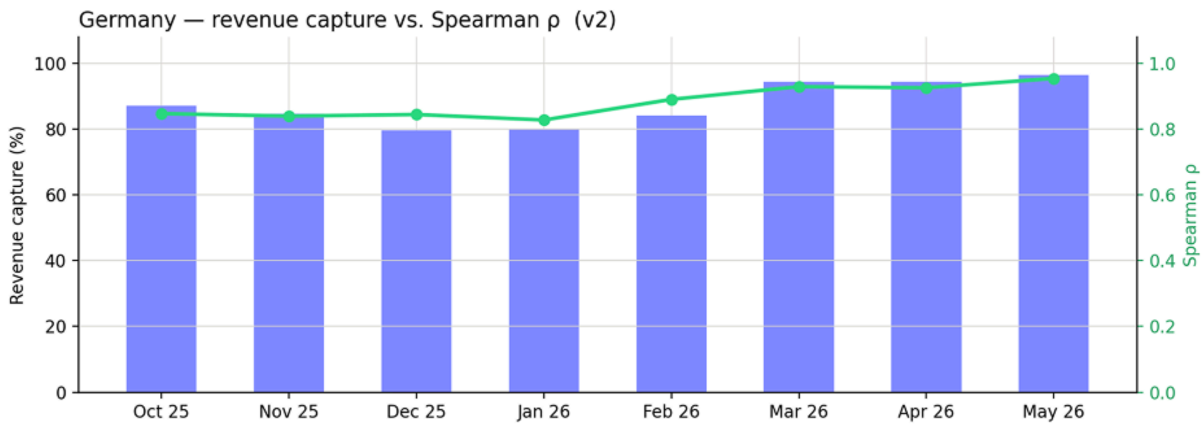


Figure 4 – Germany, monthly. Revenue capture (purple bars, left axis) against spread-weighted Spearman ρ (right axis). Capture moves with rank skill: the months where the forecast ranks hours correctly are the months where revenue is highest

A forecast that nails the order of the day scores near 1.0 regardless of how far off its absolute levels are; one that scrambles the order scores low even if its average error is small. We compute it daily and weight it by the day-ahead spread, so getting a wide-spread, high-value day right counts for more than a flat, low-value one.

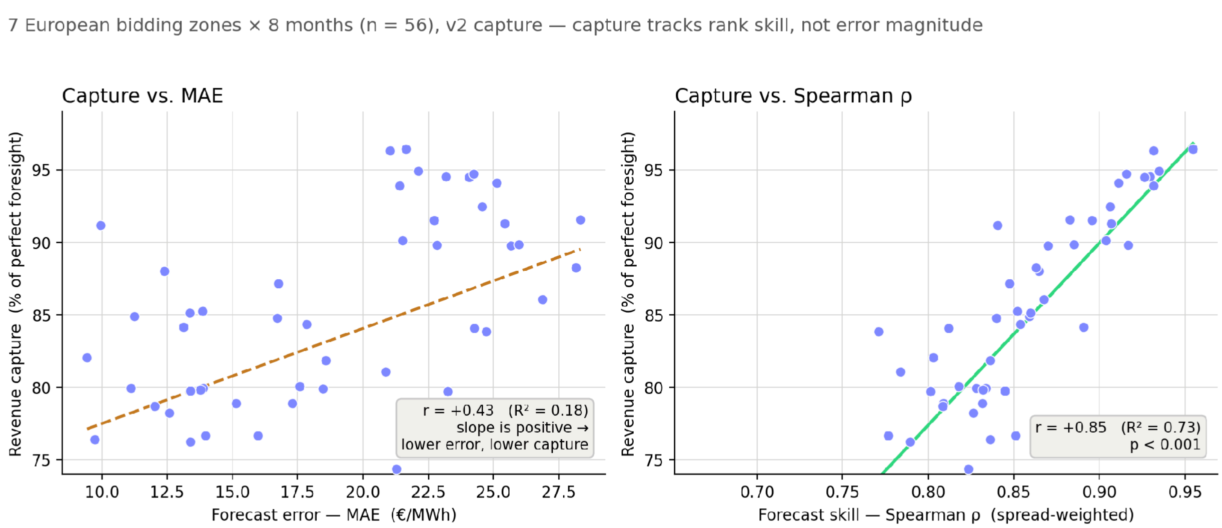


Figure 5 – Across all 56 zone-months (v2 capture), MAE relates to capture weakly and with the wrong sign (left, r = +0.43), while spread-weighted Spearman ρ tracks it closely (right, r = +0.85). Capture rises with rank skill in every market.

If MAE were a precise predictor of revenue capture, the left chart would slope downward: higher error, lower capture. The right chart shows what a useful predictor looks like: as spread-weighted Spearman ρ rises, capture rises with it.

As a reference point, we also include Pearson correlation, which measures how closely the forecast tracks actual price levels rather than their order. For MAE, lower is better. For Pearson and Spearman, higher is better.

Across the 56 zone-months, the three candidate metrics rank cleanly by how much of the revenue capture each one explains:

Predictor of revenue capture	Correlation r	Variance explained
MAE (€/MWh)	+0.43	18% (wrong-signed)
Pearson ρ (price level)	+0.51	26%
Spearman ρ, spread-weighted	**+0.85**	**73%**

MAE correlates with revenue capture in the wrong direction: a higher MAE means a worse forecast, yet higher revenue capture. Spearman and Pearson correlation both get the direction right, where higher values mean better forecasts and translate into higher revenue capture. But Spearman outperforms Pearson because a dispatch decision cares more about order than distance.

This is not an artifact of comparing markets at different price levels. Within a single zone, Spearman remains the dominant predictor (r = 0.85), while MAE stays positively, and therefore misleadingly, signed.

Choosing and contracting a forecast

- **If you trade BESS:** stop shortlisting forecast providers on MAE or RMSE. Two forecasts with identical MAE can differ materially in the €/MW they let you capture, and the one with lower-MAE is not reliably the better earner. Ask for spread-weighted, daily Spearman ρ against your delivery zone, and translate it into €/MW with your own battery's constraints.
- **If you provide forecasts:** optimize and report on the metric that maps to client P&L. Electricity Maps' day-ahead price forecast is built and evaluated for trading decisions, tuned to maximize rank skill, not merely to minimize average error, which is why its capture holds up even through volatile months.

PART THREE

Where the forecast is worth most

Capture rate is most meaningful when read against the persistence baseline in the same market. A 91% capture rate in Spain, where persistence already achieves 87%, represents a 4-percentage-point uplift over the naive baseline. A 63% capture rate in Finland, where persistence delivers only 51%, represents a 12-percentage-point uplift. Despite the lower absolute number, the Finland result reflects stronger forecast performance relative to the baseline available in that market.

In highly liquid, well-forecasted continental markets, even yesterday's prices are a reasonable guide to today's. The gap between a commercial forecast and persistence is narrower because the baseline is already high, not because the forecast is weak. In more volatile or less predictable markets, persistence consistently misreads the day, and a forecast that correctly ranks the hours earns a larger premium over that baseline.

Capture rate alone is an incomplete benchmark. The right question is: how much does a commercial forecast add over the persistence baseline in your specific delivery zone, and does that addition justify the cost?

Running the model across 27 bidding zones over the same period, the answer varies noticeably.

Markets where the Electricity Maps forecast adds the most value, Ireland (+16pp over persistence), Estonia (+14pp), Great Britain (+14pp), Finland (+12pp), and Poland (+12pp), tend to be markets with significant day-to-day variability that persistence consistently misreads. The forecast earns its keep by correctly ranking days that persistence gets wrong.

In markets like Spain, Germany, the Netherlands, and Portugal, Electricity Maps' capture rates reach 88–91%, but the uplift over persistence is narrower (4–10pp) because even a naive forecast gets the broad shape right in highly liquid, well-forecasted markets.

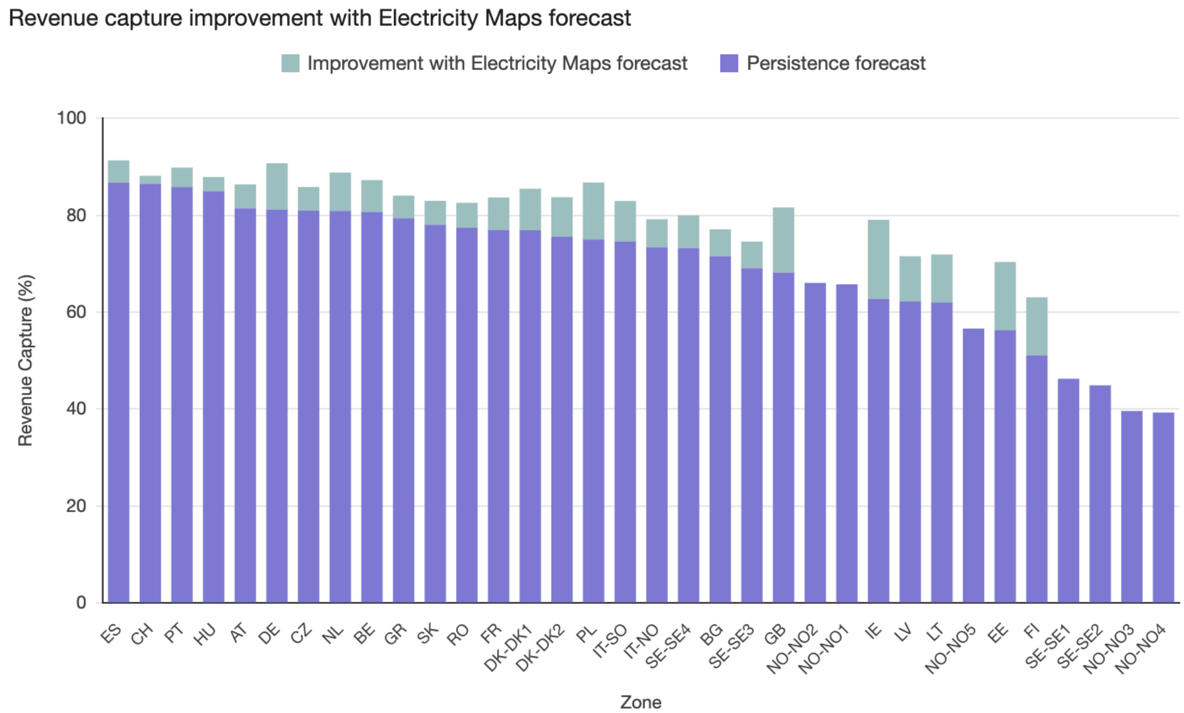


Figure 6 – Day-ahead revenue capture by bidding zone. Zones showing Electricity Maps' forecast results are markets where the EM model is live and validated. Zones showing persistence only represent markets not yet covered by the Electricity Maps forecast model.

Over 34 European zones where the Electricity Maps forecast is live, it delivers additional captured revenue for 27 of them, and most of the time, a significant improvement. This includes markets with different levels of renewables penetration and price volatility, highlighting a high and consistent performance. Seven remaining zones in Sweden and Norway show no improvement compared to persistence results. These are largely hydro-dominated Nordic markets that require a different modelling approach, and where the Electricity Maps forecast model is not yet tailored. Development here is ongoing.

The practical implication: before selecting a forecast provider, benchmark their performance against the persistence baseline in your specific delivery zone. A high capture rate alone is not a reliable guide to performance.

Finally, capture rate should be measured against the spreads available in each market. A forecast that captures 85% in Germany, where daily price spreads routinely exceed €100/MWh, delivers far more €/MW than the same capture rate in a market where spreads rarely move. Capture rate benchmarked against persistence tells you how well the forecast performs relative to the baseline. Translated into €/MW against actual spreads, it tells you what that performance is worth.

The three parts point to the same conclusion. The metric that predicts revenue capture is rank skill, not average error. And the market where that skill is worth most is the one with the widest spreads. Taken together, these considerations are important to determine what a forecast is worth.

Methodology and data: the dispatch model and formulation are documented at github.com/Flexup-energy/CastValue. The analysis applies the same v2 dispatch logic (including round-trip efficiency and out-of-the-money bid rejection) to Germany over eight months (Part One), the seven-zone / eight-month correlation panel (Part Two), and 27 zones for capture by market (Part Three).

Scope and limitations

This analysis covers physical time-arbitrage in the day-ahead auction only, under a single trading strategy with realistic frictions: 85% round-trip efficiency and out-of-the-money bid rejection. It does not model intraday continuous trading, intraday auctions, balancing markets, or cross-market arbitrage, all of which offer additional revenue and are impacted by a better forecast in various ways. State-of-charge measurement errors and line losses are not modelled.

On the MAE finding specifically: MAE remains a sensible metric where what matters is price level: settling an imbalance position, for example. It is also a useful metric when the forecast is used for cross-market arbitrage and anticipating what day-ahead spread can be expected ahead of the market closure. The claim here is more specific to time-arbitrage dispatch decisions, where rank skill predicts captured revenue and error magnitude does not. The relationship is correlational across this eight-month window and will be extended across more zones, more months, and additional markets.

About this report

FlexUp operates the CastValue platform for transparent, revenue-based benchmarking of day-ahead price forecasts across European power markets, collecting, comparing and ensembling forecasts from multiple providers to build better predictions. hello@flexup.pro

Electricity Maps has built the world's most comprehensive electricity data platform, providing historical, real-time, and 72-hour forecast data across 190+ countries – generation mix, carbon intensity, cross-border flows, and day-ahead prices, delivered via API with an enterprise SLA. electricitymaps.com