TECHNICAL WHITEPAPER · VOLUME 01 OF 02
01 · BEFORE THE GATE · SEPTEMBER 2026

FARSEES 2.0

96.9%

FORECAST ACCURACY

BEFORE THE GATE.

Forecasting price, load and generation for behind-the-meter assets.

This volume introduces Farsees 2.0, the second generation of Solship's forecasting engine for behind-the-meter systems, and reports its measured record on the Italian electricity market and on Solship's fleet.

8

energy market products forecast, each against a named baseline and a significance test

42%

less day-ahead consumption error than the previous generation, on a live industrial site

99

calibrated quantiles per signal, every 15 minutes, 37 hours ahead

00 EXECUTIVE SUMMARY

Solship AI manages behind-the-meter assets: batteries, photovoltaic plants and consumption loads on the customer's side of the grid connection, and lets them take part in the electricity markets. Every bid and every charging decision is taken before the relevant quantities are known, so market participation begins with a forecasting engine that produces three signals: what each market will pay, what the site will consume, and what its photovoltaic plant will generate.

THE MODEL	**Farsees 2.0, the second generation of Solship forecasting model.** This paper introduces Farsees 2.0, Solship's own model architecture: engineered grid features for the market products, an event layer, a horizon blend and calibration for the site signals, and a deep network of Solship's design carried in its research.
MARKET PRICES	**Farsees 2.0 performance in forecasting is tested against eight Italian energy markets** Farsees 2.0 cuts mean absolute error by 20% to 40% against each product's baseline on held-out windows, and every result passes a Diebold-Mariano significance test. Farsees 2.0 outperformed all published models including, Amazon's and Google's among them time series forecasting models.
SITE SIGNALS	**Consumption and generation, with zero shot architecture** On a live industrial site, replayed through the same optimizer on the same market prices, Farsees 2.0 cuts day-ahead consumption error by 42% and generation error by 18%. It runs on a new site without site-specific training. Across several datasets, consumption accuracy runs from 73.5% to 99.0% depending on site and layer.
WHAT IS NEXT	**Farsees 2.5 targets training the forecaster on the optimizer decision making.** The next generation of the Farsees family will be trained on the optimizer output so that it is scored on the optimiser's result rather than on forecast error alone.

● 00 FARSEES 2.0 BENCHMARKS

Farsees 2.0 against Google's and Amazon's foundation models run zero-shot, the best public deep network and the naive baseline, on the same held-out data.

FARSEES 2.0 BENCHMARKS · MEAN ABSOLUTE ERROR ON HELD-OUT DATA · LOWER IS BETTER

Task METRIC · WINDOW	Farsees 2.0 AUGUST 2026 RECORD	Google TimesFM 3 ZERO-SHOT	amazon Chronos-2 ZERO-SHOT	Best public deep network SAME DATA · 9 TRIED	Naive baseline PER TASK
Day-ahead price MGP, Italy · €/MWh	**13.51**	25.79	23.14	16.19	19.14
National index price PUN · €/MWh	**13.49**	27.12	24.57	16.03	18.68
Intraday price MI-A2 · €/MWh	**7.53**	8.32	11.11	8.53	9.55
Intraday price MI-A3 · €/MWh	**12.40**	22.40	23.90	15.07	15.90
Continuous intraday price order book · €/MWh	11.88	**7.68**	8.42	14.66	14.87
Balancing price, mFRR up €/MWh	**32.51**	35.10	34.65	46.27‡	50.11
Balancing price, mFRR down €/MWh	**19.99**	22.39	20.91	20.02‡	33.26
Balancing price, aFRR down preliminary · €/MWh	11.53	**8.12**	8.29	–	15.34
Site consumption, day-ahead one industrial site · kW	**35.55**	39.09	35.56	–	–
Site generation, day-ahead same site · kW	16.05	16.28	**15.51**	–	–

Bold marks the lowest error in each row. Farsees 2.0 is lowest on seven of the ten tasks; TimesFM 3 leads on the continuous intraday product and aFRR down, Chronos-2 on site generation.

Foundation models are run zero-shot: no training on Solship's data. **Best public deep network** is the best of nine architectures trained on the same feature panel; ‡ two tabular architectures on the balancing products.

Naive baselines are the prior-day price (MGP, PUN), the day-ahead price for the same hour (intraday products) and conditional climatology (balancing). A dash means the reference was not evaluated on that task in the record.

Figure 1. Mean absolute error on held-out data, lower is better; bold marks the lowest error in each row. Price tasks: 15-minute Italian products, each on its own held-out window, Diebold-Mariano against each product's naive baseline (prior-day price, day-ahead anchor or conditional climatology) at the 5% level. Site tasks: one industrial site, May 2026, 35-hour forecast issued daily at 13:00. TimesFM 3 and Chronos-2 are run zero-shot on every task; the best public deep network is the best of nine architectures trained on the same feature panel. Public models are named because a reader can reproduce the comparison; the model inside Farsees 2.0 is not.

● 00 THE ARGUMENT, IN ONE PAGE

01

Revenue from a behind-the-meter battery depends on three forecast quantities, one of which is a set of prices rather than a single value.

The three quantities are the market price, the site's consumption and the output of its photovoltaic array. The market price is not a single series: in the Italian case used throughout this document, a battery's energy for a given day can be settled across six market gates. The installed base of behind-the-meter systems is large, yet globally only around 1% of behind-the-meter batteries currently participate in these markets.

02

Farsees 2.0 combines its own architecture with zero-shot models, each retained on the basis of measured performance

Solship's own architecture handles event detection, horizon structure and calibration, and includes a deep learning network trained on engineered grid features. It is combined with zero-shot forecasting models, which require no site history, to address consumption and photovoltaic generation at sites where historical data is frequently unavailable.

03

Performance is evaluated against named baselines on held-out windows, with a test of statistical significance

Farsees 2.0 outperforms its baselines by 20% to 40%. Baselines are mostly the strongest references identified after published models were scored on the same data. On a live site, consumption and generation error fell by 42% and 18% respectively, relative to the previous generation of the engine.

04

Every forecast is issued as a distribution; the decision layer determines how it is used

Each signal is forecast as ninety-nine calibrated quantiles, at fifteen-minute resolution, up to thirty-seven hours ahead. The forecast establishes a lower bound on attainable performance; realised revenue is determined by how the dispatch plan is constructed and executed. **That subject is addressed in Volume 02.**

● 00 HOW TO READ THIS DOCUMENT

FOUR PARAMETERS	**Every accuracy figure depends on these** Granularity, denominator, horizon and window. An error measured at fifteen-minute resolution is larger than the same error aggregated to a month; an error normalized by peak load appears smaller than one normalized by mean load; a one-hour-ahead forecast and a day-ahead forecast are distinct problems; a single favorable month is not representative of six. Every figure in this document states all four parameters and is scored on held-out data.
TWO METRICS	**Accuracy and capture are reported separately** Accuracy denotes forecast error against a stated denominator. Capture denotes the share of a perfect-foresight profit ceiling that a dispatch plan actually realized. This volume reports the former; **revenue is treated in Volume 02.**
WHAT IS DISCLOSED	**The class of every model and the protocol behind every result** Specific architectures, their hyperparameters and the manner of their composition are not disclosed. Public reference models scored on Solship's data are identified by name. Italy is used as the reference case throughout, so that results are comparable on a common basis.
BASELINES	**Every reduction is stated relative to a named baseline** The baseline is the prior-day price, the day-ahead anchor or a conditional climatology, as stated in each case. Reductions are comparable only when measured against the same baseline.

CHAPTER 01

INTRODUCTION TO BEHIND-THE-METER ASSETS

01 THE INSTALLED BASE

Behind-the-meter assets are photovoltaic and battery systems installed on the customer's side of the meter, typically at residential and commercial sites where the owner's primary purpose is self-consumption. These systems are widely deployed, yet they rarely offer their spare energy to the electricity markets. Commercial energy trading concentrates on utility-scale batteries, and behind-the-meter assets are left aside. Utility-scale storage is the more straightforward trading case: a behind-the-meter system adds technical complications that a utility-scale plant does not face, chiefly the absence of historical data on site consumption and generation, and the need to forecast the site's own load as a further layer of uncertainty. The scale of the behind-the-meter fleet is nonetheless large. More than 187 GWh of residential battery storage is installed worldwide, before counting commercial and industrial systems [24], the world's battery fleet grew by two thirds in 2025 alone [23], and in some markets the consumer side exceeds utility-scale outright: in Italy, 61% of installed storage energy is behind the meter [1]. Where participation in the electricity markets has been counted, it is around 1% of these systems.

THE INSTALLED BASE · BEHIND THE METER

A large fleet already sits behind the meter. Around 1% of it trades.

WORLD · RESIDENTIAL · END 2025

187 GWh

of residential battery storage installed worldwide, before counting commercial and industrial systems; 35 GWh shipped in 2025 alone [24]

ITALY · 30 JUNE 2025

61 %

of Italy's installed storage sits behind the meter: **10,011 MWh** in **815,092** residential and commercial systems, against 6,400 MWh in 69 utility-scale plants [1]

MARKET PARTICIPATION

~1 %

of these systems take part in the electricity markets, in the countries where participation has been counted [1, 2, 3]

GROWTH · 2025

+66 %

growth of the world's battery storage fleet in a single year, from 182 GW to 302 GW; the fastest-growing power technology [23]

Figure 2. The installed base behind the meter. Worldwide residential storage at end 2025 from [24]; Italy at 30 June 2025 from ANIE Rinnovabili on Terna Gaudì data [1], where utility-scale is Solship's sum of the two utility classes in the source; participation counts from [2] for France and from [1] and [3] for Italy; fleet growth in 2025 from the Energy Institute Statistical Review [23].

● 02 WHY FORECASTING IS THE TECHNICAL BARRIER

A bid for tomorrow commits a site to a position that depends on what it will consume, generate and be paid. For a utility-scale battery only the price has to be forecast, and price forecasting is a more mature discipline with open benchmarks. A behind-the-meter system carries two further forecasts that a utility-scale plant never needs: the site's own consumption and its own photovoltaic generation for small single units. Both are far less studied, because the field has concentrated on the power system as a whole rather than on the building behind the meter.

+25.2 pp of MAPE more when forecasting a single site rather than a whole power system, across 421 published models [4]	**43–74**% of the error disappears when individual building profiles are aggregated, 15-minute data [7]	**79.3**% normalised error of the best pretrained model on residential buildings; persistence does better, at 77.9% [6]

The gap is compounded by data. A newly installed system, which is where the fleet is growing fastest, has no meter history at all, and even an established site rarely offers more than a few months of clean readings; the consumption forecast therefore has to work zero-shot, from the first day, before any site-specific model can be trained. Getting the forecast wrong raises a microgrid's operating cost by about 10% [11]. A review calls load forecasting for buildings "often overlooked by the power engineering community" [8], only 44 of 221 low-voltage forecasting papers carry a probabilistic element [5], and day-ahead forecasting at low aggregation "remains underexplored" [9].

● 03 WHY SOLSHIP ADDRESSES IT NOW

01 **Pretrained AI models changed what a new site costs** Pretrained time-series foundation models forecast single-site load zero-shot [12, 13, 14, 15], comparably to transformers trained on household data [16]. Farsees 2.0 uses zero shot forecasting as the base.	02 **The installed base is large and under-served** Around 1% of behind-the-meter systems take part in the markets, where participation has been counted. Farsees 2.0 forecasts a site's generation from its first morning and its consumption from its first week, without a person tuning a model per site.

CHAPTER 02

THE FULL SOLSHIP ENGINE

04 EVERY BID IS PLACED BEFORE THE QUANTITIES ARE KNOWN

Three groups of inputs enter the engine. Two forecasters consume them, and both deliver the same output contract to the decision layer.

SIGNAL 01	SIGNAL 02	SIGNAL 03
Price	**Consumption**	**Generation**
One day's energy clears through a day-ahead auction, several intraday auctions, a continuous market and several balancing products, each with its own gate. A forecast that arrives after the gate cannot change the position, so every price forecast is issued and frozen before its gate.	Driven by production schedules, shift patterns, machine cycles and occupancy, which leave a strong signature in the meter data but publish no feed of their own. A pretrained model carries the shape, and Solship's own event layer carries the spikes.	The most tractable of the three. Sun geometry and module behaviour are computed from a physical model before the site has produced a kilowatt-hour; the learning is aimed only at that model's own error.

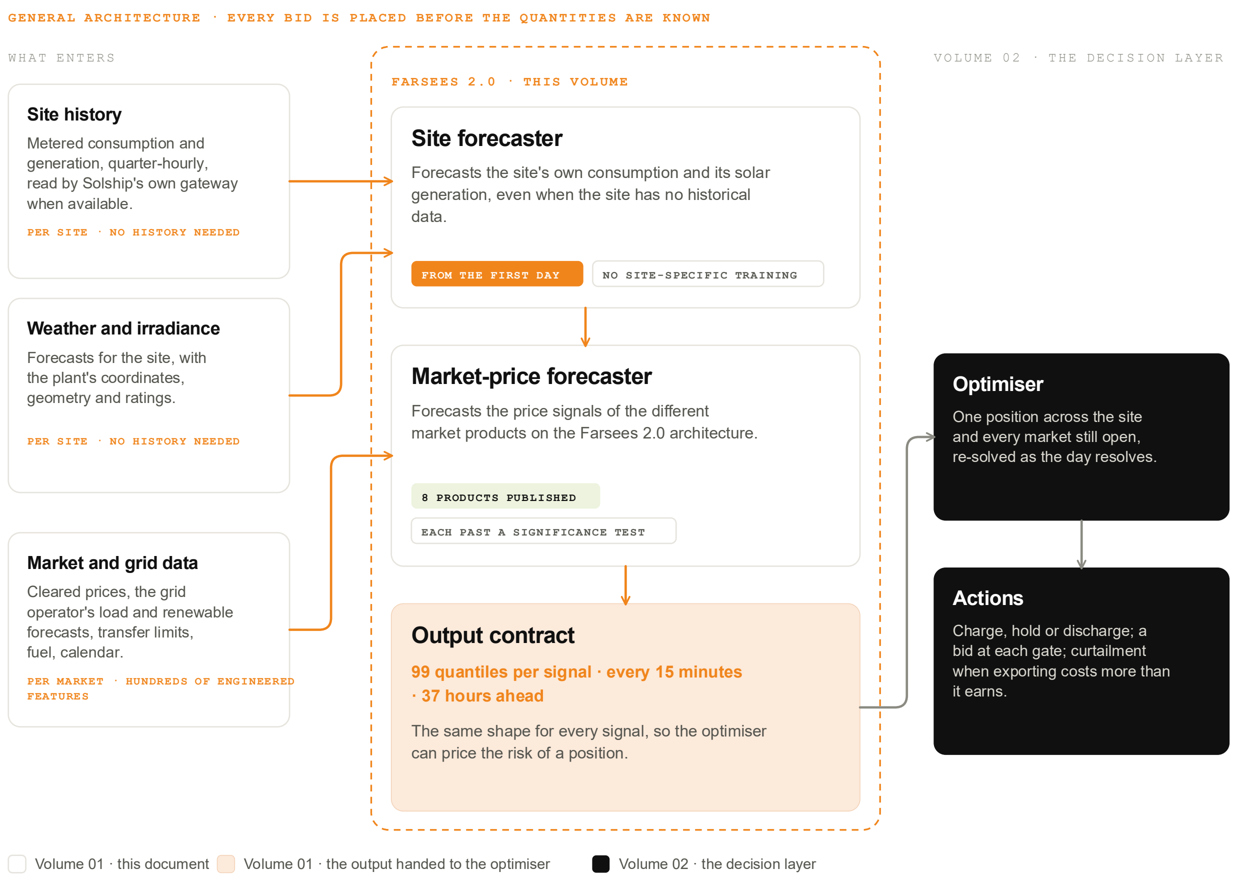

Figure 3. General architecture. What enters on the left; the two forecasters that make up Farsees 2.0 and the output they hand to the optimizer in the center; the decision layer, covered in Volume 02, on the right.

CHAPTER 03

INTRODUCING FARSEES 2.0

05 WHAT FARSEES 2.0 IS MADE OF

Farsees 2.0 is the flagship forecaster of Solship. It produces market prices, site consumption and site generation as calibrated distributions, and it is the model in production on Solship's Italian fleet as of August 2026

COMPOSITION · WHAT FARSEES 2.0 IS MADE OF

Solship's own architecture, wrapped around the models it is measured against.

01

Solship's own architecture

Designed and trained by Solship. It runs beside the base models and decides what reaches the optimiser. Serves consumption, generation and prices.

CONSUMPTION · GENERATION · PRICES

Event layer
Spike classifiers, one per window of the day, and regressors that size the spike; the sized spike shifts the whole distribution.
CONSUMPTION

Horizon blend
Raw and de-noised history fed to the base model; near steps lean raw, far steps lean smoothed.
CONSUMPTION

Calibration
Bias recentred on the last observed hours; interval width rescaled on the model's own recent errors.
ALL SIGNALS

Executed-hour layer
Regenerates the forecast at each rolling re-solve from the latest measured values for the hour being dispatched.
SITE SIGNALS

Own deep network
A deep-learning model of Solship's design, carried as a research track beside the production price models.
PRICES · RESEARCH

02
Zero-shot forecasting
Forecasts a site with no history of its own.

03
Deep-learning capabilities
Learned models for prices and site signals.

04
Physical model
Sun, array and inverter physics for generation.

One output contract for every signal
99 calibrated quantiles per signal, every 15 minutes, 37 hours ahead.
CONSUMPTION · GENERATION · PRICES

Figure 4. Composition. Solship's own architecture comes first, with its components across the top; the three capabilities it wraps are named below. Chips name the signal each component serves.

0 days	**136**	**23**	**99**
of site history before the first generation forecast	engineered grid features behind the market-price forecaster, each block kept or dropped by measurement	reference models scored on the same data.	quantiles per signal per quarter-hour, the one output contract all four components feed

06 HOW THE MODEL EVOLVED

Two generations are in the record, and the third is defined by what the second measured.

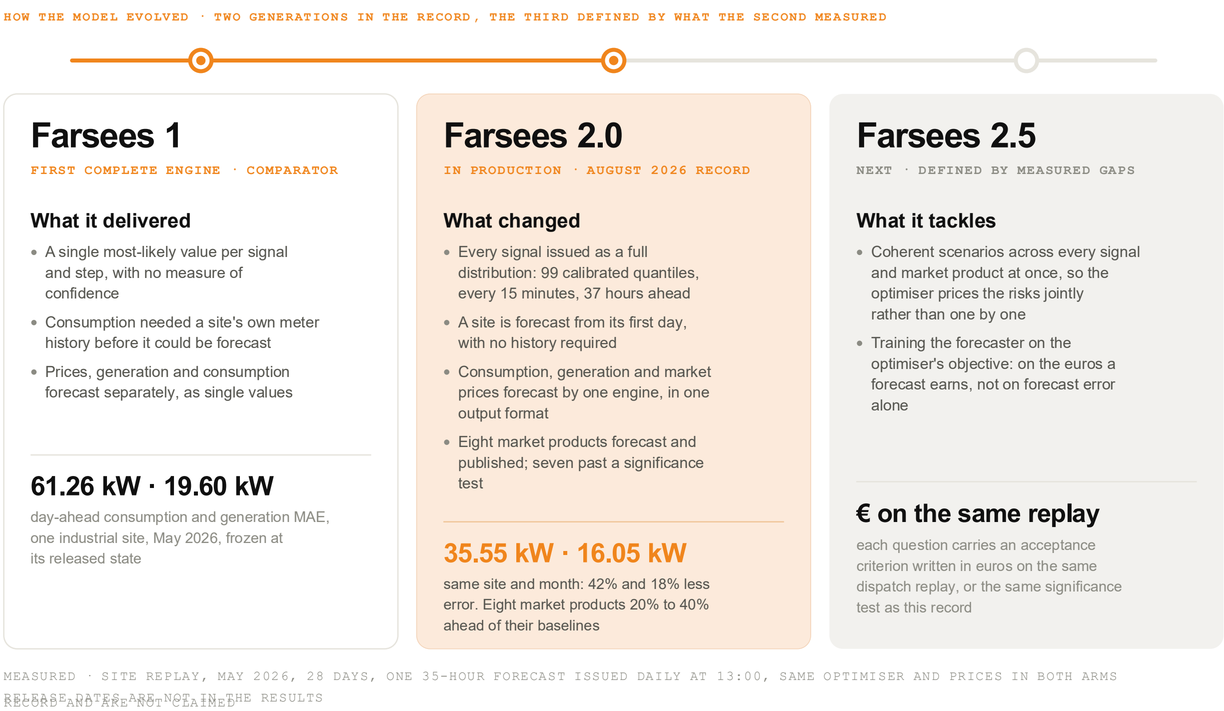

Figure 5. The Farsees generations. What each generation delivers to the user, not how it is built; Farsees 2.5 carries the open questions of section 13, each with its acceptance criterion. The record does not label results by generation, and release dates are not claimed.

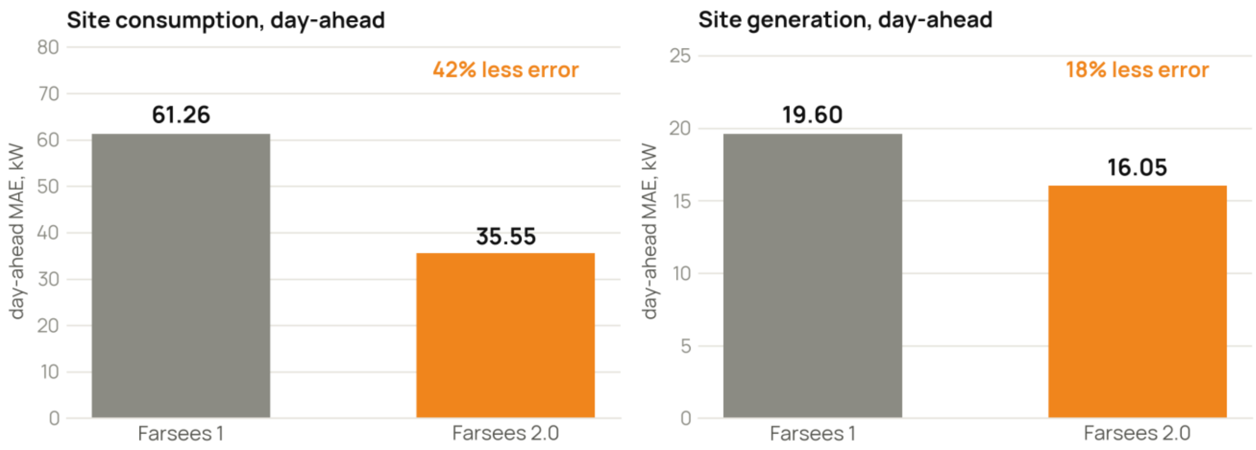

Figure 6. Error by generation on the site signals. Day-ahead MAE in kW, one industrial site, 28 days in May 2026, one 35-hour forecast issued daily at 13:00, through the same optimiser on the same live prices. The only difference between the two arms is the forecaster.

CHAPTER 04

MARKET PRICES

07 APPLYING FARSEES 2.0 ON THE ITALIAN MARKET

Italy prices one day's energy several times before delivery: a day-ahead auction, three intraday auctions, a continuous market and the balancing products [17, 18]. Each has a gate, and a forecast that improves after the gate is worth little, so every product is forecast and frozen before its gate.

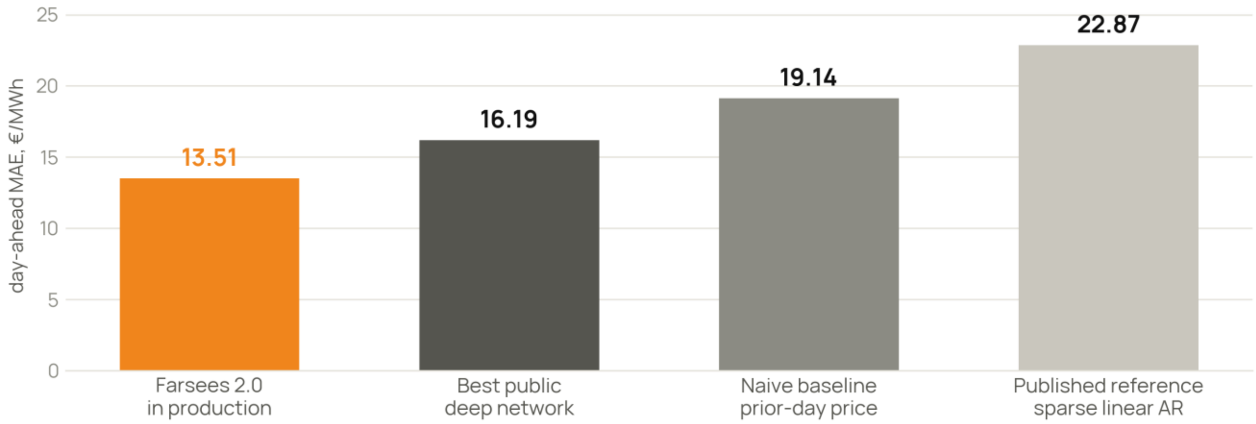

Figure 7. Day-ahead price forecasting on the Italian MGP. MAE in €/MWh of Farsees 2.0 against the best of deep-learning architectures trained on the same data, the naive prior-day price, and the field's published reference model, a sparse linear autoregression with exogenous variables [10], reimplemented on the same data. 89 held-out days, 15-minute products, forecast issued at 11:00 the day before.

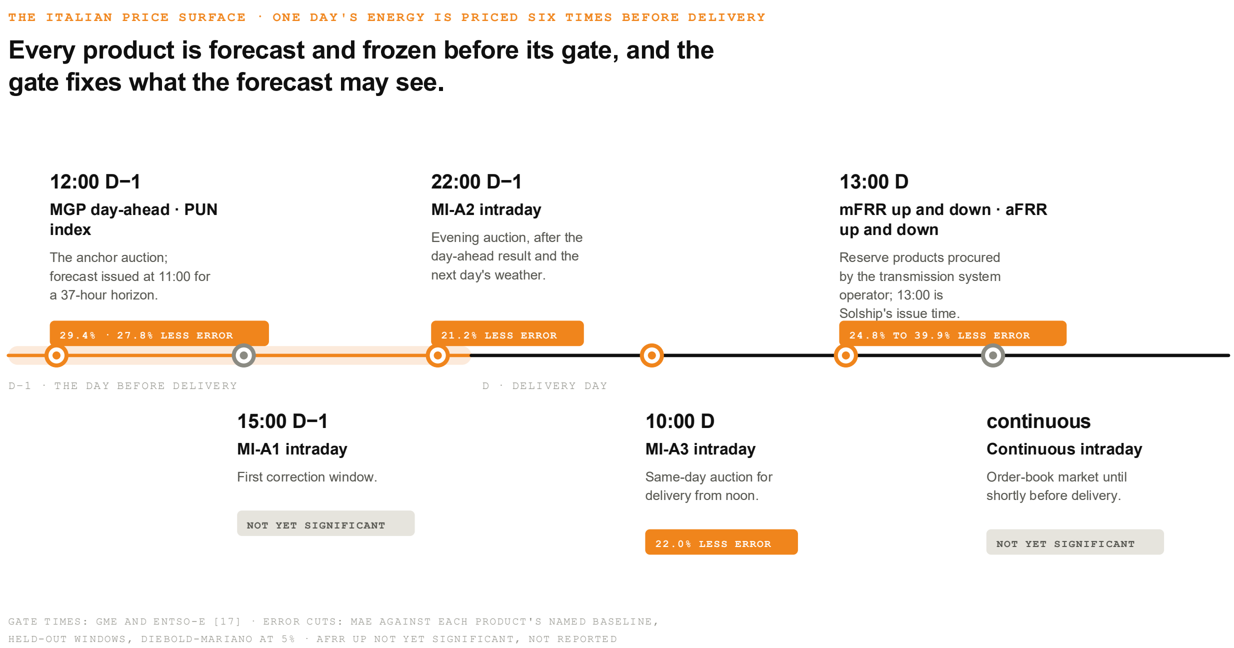

Figure 8. One day's energy is priced six times before delivery. Each gate with the product it closes and Farsees 2.0's error cut against that product's named baseline on its held-out window: the prior-day price for the day-ahead products, the day-ahead anchor for the intraday products, conditional climatology for the balancing products. Grey chips are products whose lead has not yet passed the significance test.

● 08 THE SCORECARD

Farsees performance score against the baselines

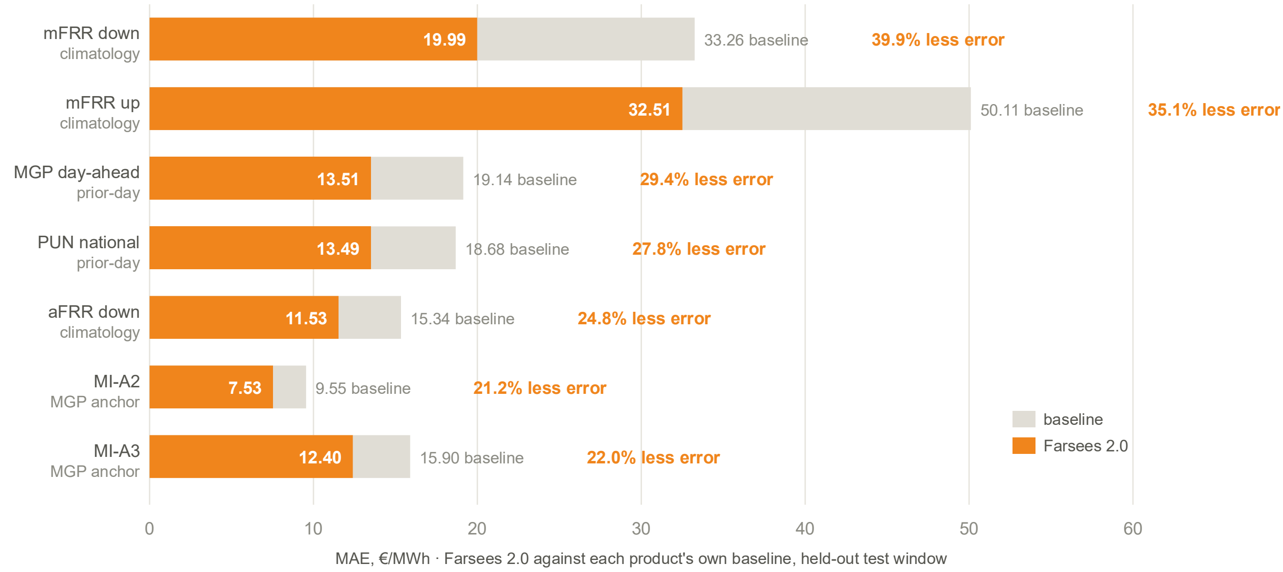

Figure 9. Mean absolute error, Farsees 2.0 against each product's own baseline, €/MWh at 15-minute resolution, each product on its own held-out window; results record as of August 2026. MI-A1, the continuous intraday product and aFRR up do not yet pass the test and are not reported.

What the baselines are

The prior-day price is yesterday's price for the same quarter-hour. The day-ahead anchor is the MGP price already cleared for the same period, known at every intraday gate. Conditional climatology is the average a balancing product has historically cleared at in this quarter-hour and season, conditioned on the calendar and nothing else. Each is the strongest reference a participant has without a model.

What a 29.4% reduction means

The average size of the Farsees 2.0 error is 29.4% smaller than the average size of the baseline's error, on the same days and quarter-hours, in euros per megawatt-hour. Where the baseline is out by €10, Farsees 2.0 is out by €7.06. It is a statement about error, comparable only across products that share a baseline.

WHAT PASSES

A product is published only if it beats its own baseline under a Diebold-Mariano test at the 5% level.

WHAT IS PRELIMINARY

aFRR down rests on sixteen scored days. The window is not yet long enough to call it settled.

WHAT IT IS NOT

These are error reductions. What a battery earns from a better price forecast depends on the optimizer that consumes it. **That is Volume 02.**

● 09 FARSEES 2.0 AGAINST THE FULL FIELD

The obvious criticism of any forecasting result is that the benchmark it beat was weak. So the whole field was scored on the day-ahead product, behind the same gate, on the same 89 days: deep-learning architectures, published methods, pretrained foundation models and the naive baselines; the foundation models are in Figure 1.

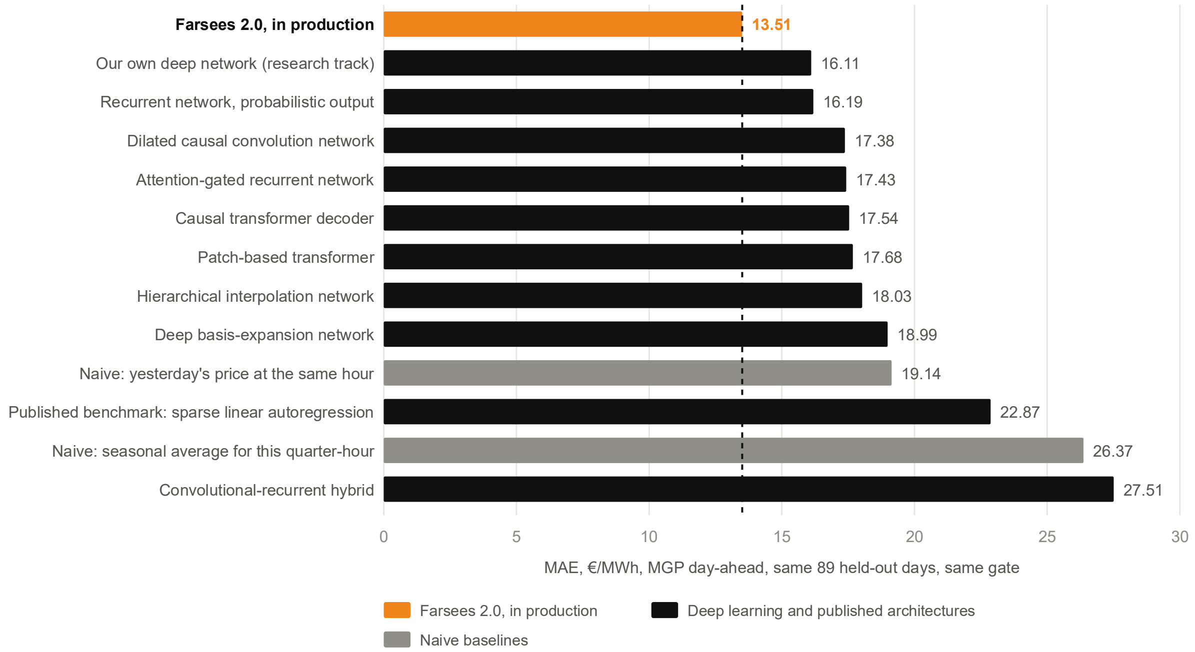

Figure 10. MGP day-ahead, 89 held-out days, MAE in €/MWh. Every row is a model that was trained and scored on this product, labelled by what it is rather than by its library name. TAll within 0.9 €/MWh of the production model, are omitted for clarity; two literature replications that diverged numerically are off-scale.

CHAPTER 05

SITE SIGNALS: CONSUMPTION AND GENERATION

ZERO-SHOT, SITE TO SITE

No site-specific training

The base model was never trained on a Solship site, so the same forecaster moves from one site to the next without retraining. What adapts to the site is learned fresh at each issue from its own history.

SPIKES

An event layer carries them

Classifiers return a spike probability per window of the day, regressors size it, and the sized spike shifts the whole distribution. They read the calendar and holidays, the site's recent history, its hour-and-weekday climatology, the season and the weather.

MEASURED

85.6% day-ahead accuracy on the hardest live site

Scored on several loads spanning three orders of magnitude, 30 to 31 forecast windows each, retrained per window: accuracy runs from 73.5% to 99.0% by site and layer, and the weakest result is reported beside the strongest (Figure 12).

● 11 GENERATION IS COMPUTED BEFORE IT IS LEARNED

A solar array's output is largely set by geometry and materials science settled long before the site existed. The physics is computed first, and the learned part predicts only the physics model's error.

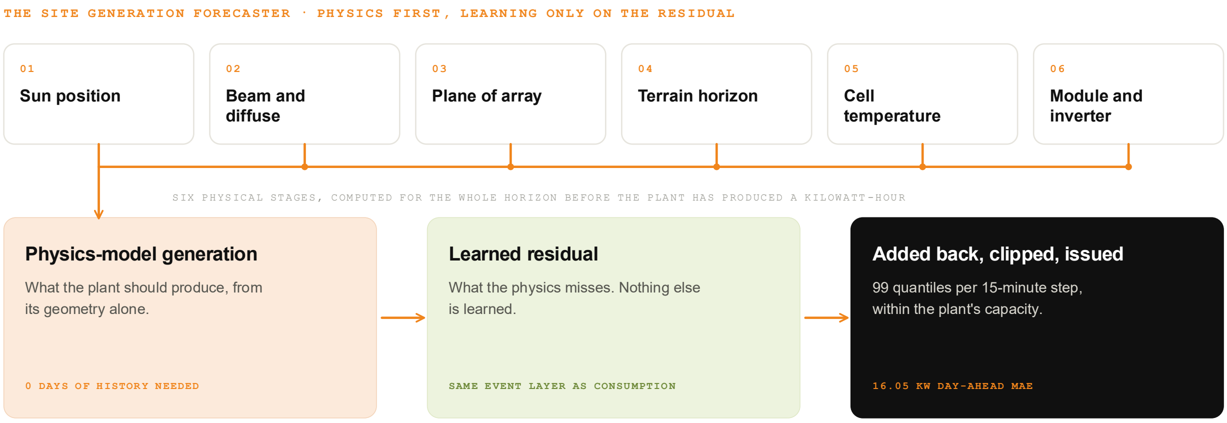

Figure 11. The site generation forecaster. Several physical stages computed for the whole horizon, a learned residual on top, and the clipped 99-quantile output. In the published record a physics-plus-learning hybrid cut MAE by 5.2% against an optimized physical chain and 10.4% against pure machine learning on 14 plants [19]; the choice of physical chain alone moves MAE by 13% [20].

18%	**0** days	**8.3**%	**6–700** kWp
less day-ahead generation error than Farsees 1, same site and window: 16.05 kW from 19.60 kW	of site history required before the first generation forecast	of the generation error removed by one verified instantaneous radiation input	range of plant ratings the physical chain serves, tilts 10° to 44°, azimuths 135° to 180°

● 12 FARSEES 2.0 PERFORMANCE ON CONSUMPTION FORECASTING

A forecaster that performs on a single site is not a product. Farsees 2.0 is therefore benchmarked on several consumption datasets whose peak loads span three orders of magnitude, from a 5.6 kW unit to a 449 kW feeder, and the weakest result is reported beside the strongest. Figure 12 shows two layers for each dataset. The day-ahead layer, which issues the whole 33-hour block once, reaches 73.5% to 93.0% accuracy; the executed-hour layer, which regenerates the forecast one hour ahead at each rolling re-solve, lifts every dataset to between 87.5% and 99.0%. The gain from the second layer is largest where the load is most volatile: the live residential site moves from 73.5% to 92.1%, while the live industrial site, already the steadiest a full day ahead at 85.6%, gains two points to 87.5%. Accuracy is normalized by each dataset's own peak load, so the figures are comparable within a row but not across rows.

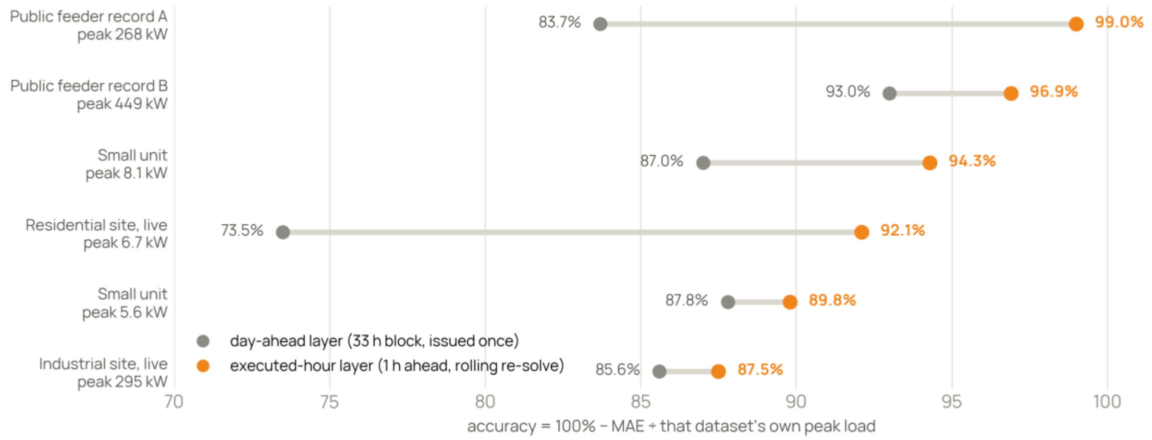

Figure 12. Accuracy as 100% minus MAE over each dataset's own peak load. Several datasets, 30 to 31 forecast windows each, a 33-hour window anchored at 15:00, retrained per window. Rows A and B are public feeder records on an archival year, never presented as live results. The day-ahead layer emits the whole block at once; the executed-hour layer regenerates at each rolling re-solve for the hour being dispatched. The number to quote is the 85.6% of our hardest live site a full day ahead, rather than the 99.0% of the easiest dataset.

CHAPTER 06

FARSEES 2.5

13 FARSEES 2.5: THE OPEN QUESTIONS IT WILL TACKLE

Two questions are left open by the measurements in this volume, and they define the next generation. The first is what the forecaster is trained on: everything here is scored against forecast error, yet the optimizer study shows that better error did not translate into more euros. The second is how uncertainty is represented: each signal is calibrated on its own, while the battery is exposed to all of them at once. Each question below states what Solship would accept as an answer, in euros on the same replay.

01 **Train the forecaster on the optimizer's objective.**

Everything in this volume is scored against mean absolute error, and the optimizer study is the strongest evidence that this is the wrong target: a month of replay moved a forecast-quality metric and moved the euros not at all. Farsees 2.5 trains the forecaster against the dispatch cost it incurs, through a smoothed version of the optimization program or a substitute that preserves the ranking of decisions [22, 21]. **Accept when:** a decision-trained model is ahead of an error-trained model in euros on the same replay, at equal or worse MAE.

02 **Coherent scenarios across every signal at once.**

For the site signals the quantiles are already stitched into coherent paths with a rank-shuffle construction. No construction yet makes the paths coherent across signals and market products, so the optimizer prices each risk separately while the battery is exposed to them jointly. **Accept when:** a jointly coherent construction is ahead of the current one in euros on the same replay.

14 PROTOCOL AND DEFINITIONS

Everything needed to reproduce or attack the numbers in this document.

ITEM	DEFINITION USED THROUGHOUT THIS DOCUMENT
MAE	Mean absolute error, in kW for site signals and €/MWh for prices, per forecast step over the whole test window, unweighted.
Accuracy, %	100% minus MAE divided by the dataset's own peak load over its scored window. Not comparable across datasets with different peaks.
Capture	The share of a perfect-foresight profit ceiling a plan actually earned when run through the optimiser. A money metric, reported in Volume 02.
Baselines	Prior-day persistence on the day-ahead products; the same-period day-ahead price on the intraday products; conditional climatology on the balancing products. Reductions are comparable only within a baseline.
Significance	Diebold-Mariano test against the product's own baseline at the 5% level on daily-aggregated losses. Products that do not pass are reported but not claimed.
Windows and record	89 to 90 held-out days for day-ahead and intraday products, 60 for mFRR, 16 for aFRR; training history from 2023; validation windows held out separately. Price figures are read from Solship's results record as of August 2026, which holds every candidate scored on each product alongside the champion.
Horizons	37 hours in production; the May 2026 site replay used a 35-hour forecast issued at 13:00, the six-dataset study a 33-hour block anchored at 15:00. Every lag is shifted by the horizon so that no feature can read inside its own window.
Site replay	One industrial site, May 2026, 28 days, one 35-hour forecast per day, executed through a mixed-integer optimiser against realised load and generation and settled at live market prices. Both arms identical except for the forecaster; the comparator, Farsees 1, frozen at its released state with two pre-existing defects repaired so that the replay would run.
Optimizer study	Same site and month, 15-minute replay through the full optimizer, several scenario stochastic program against a median forecast; each arm plans on its own forecast, executes onto realized load and generation, and settles at realized prices.
Reference methods	Ten published methods reimplemented on Solship's panels, splits and gates, including the field's canonical sparse-linear benchmark [10]. Several pretrained foundation models scored on the continuous intraday product against the same anchor. Public deep-learning architectures trained on market products on the same feature panel as the champion, August 2026.
Forecast layers	The day-ahead layer emits a 33-hour block in one shot; the executed-hour layer regenerates at each rolling re-solve and overrides the block for the hour being dispatched.
Anonymisation	Site names, suppliers and offtakers are withheld; capacities, windows and sample counts are exact.

TABLE A1 · SOLSHIP · VOLUME 01 OF 02, BEFORE THE GATE · VOLUME 02, AT THE GATE, COVERS THE DECISION LAYER

● 15 REFERENCES

[1] ANIE Rinnovabili (2025). Osservatorio Sistemi di Accumulo, data to Q2 2025, on Terna Gaudì data. https://anie.it/wp-content/uploads/2025/09/250911-Osservatorio-SdA.pdf

[2] Eurelectric (2025). Distributed Flexibility Report, November 2025, p. 24, Figure 21, citing Enedis data for October 2025 (NEBEF mechanism). https://www.eurelectric.org/wp-content/uploads/2025/11/Eurelectric-Distributed-Flexibility-Report.pdf

[3] Terna (2022). Progetti pilota UVAM, Lightbox, 1 June 2022; count as at December 2021. https://lightbox.terna.it/it/insight/progetti-pilota-uvam

[4] Hopf, K., Hartstang, H., Staake, T. (2023). Meta-regression analysis of errors in short-term electricity load forecasting. e-Energy '23 Companion. https://doi.org/10.1145/3599733.3600248

[5] Haben, S., Arora, S., Giasemidis, G., Voss, M., Vukadinović Greetham, D. (2021). Review of low voltage load forecasting: methods, applications, and recommendations. Applied Energy, 304, 117798. https://doi.org/10.1016/j.apenergy.2021.117798

[6] Emami, P., Sahu, A., Graf, P. (2023). BuildingsBench: a large-scale dataset of 900K buildings and benchmark for short-term load forecasting. NeurIPS 36, Datasets and Benchmarks Track. arXiv:2307.00142

[7] Groß, A., Lenders, A., Schwenker, F., Braun, D. A., Fischer, D. (2021). Comparison of short-term electrical load forecasting methods for different building types. Energy Informatics, 4 (Suppl. 3). https://doi.org/10.1186/s42162-021-00172-6

[8] Hong, T., Pinson, P., Wang, Y., Weron, R., Yang, D., Zareipour, H. (2020). Energy forecasting: a review and outlook. IEEE Open Access Journal of Power and Energy, 7, 376–388. https://doi.org/10.1109/OAJPE.2020.3029979

[9] Timmermans, J.-L., Henneaux, P. (2025). Comparative study of day-ahead residential load forecasting at low aggregation levels. IEEE PES ISGT Europe 2025. https://doi.org/10.1109/ISGTEurope64741.2025.11305544

[10] Lago, J., Marcjasz, G., De Schutter, B., Weron, R. (2021). Forecasting day-ahead electricity prices: a review of state-of-the-art algorithms, best practices and an open-access benchmark. Applied Energy, 293, 116983. https://doi.org/10.1016/j.apenergy.2021.116983

[11] Manzolini, G., et al. (2024). Impact of PV and EV forecasting in the operation of a microgrid. Forecasting, 6(3), 32. https://doi.org/10.3390/forecast6030032

[12] Das, A., Kong, W., Sen, R., Zhou, Y. (2024). A decoder-only foundation model for time-series forecasting. ICML 2024, PMLR 235. arXiv:2310.10688

[13] Ansari, A. F., et al. (2024). Chronos: learning the language of time series. Transactions on Machine Learning Research. arXiv:2403.07815

[14] Ansari, A. F., et al. (2025). Chronos-2: from univariate to universal forecasting. arXiv:2510.15821

[15] Ekambaram, V., et al. (2024). Tiny Time Mixers (TTMs): fast pre-trained models for enhanced zero/few-shot forecasting of multivariate time series. NeurIPS 37. arXiv:2401.03955

[16] Meyer, M., Zapata, D., Kaltenpoth, S., Müller, O. (2025). Benchmarking time series foundation models for short-term household electricity load forecasting. IEEE Access, 13, 218141–218153. arXiv:2410.09487

[17] GME (2026). Vademecum della Borsa Elettrica, 31 January 2026; ENTSO-E, intraday auctions (IDA) implementation page, gate closure times. https://www.mercatoelettrico.org/Portals/0/Documents/it-it/20260131VademecumBorsaElettrica.pdf · https://www.entsoe.eu/network_codes/cacm/implementation/ida/

[18] RSE, Ricerca sul Sistema Energetico (2024). Il Mercato per il Servizio di Dispacciamento e il Mercato del Bilanciamento, March 2024, on pay-as-bid remuneration of accepted offers. https://www.rse-web.it/wp-content/uploads/2024/03/05_MSD-MB.pdf

[19] Mayer, M. J. (2022). Benefits of physical and machine learning hybridization for photovoltaic power forecasting. Renewable and Sustainable Energy Reviews, 168, 112772. https://doi.org/10.1016/j.rser.2022.112772

[20] Mayer, M. J., Gróf, G. (2021). Extensive comparison of physical models for photovoltaic power forecasting. Applied Energy, 283, 116239. https://doi.org/10.1016/j.apenergy.2020.116239

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

[22] Mandi, J., et al. (2024). Decision-focused learning: foundations, state of the art, benchmark and future opportunities. Journal of Artificial Intelligence Research, 80, 1623–1701. arXiv:2307.13565

[23] Energy Institute (2026). Statistical Review of World Energy 2026, battery storage capacity by country and connection point (front-of-meter and behind-the-meter), as analysed in Rapier, R., "Battery Storage Grew 66% in 2025. China Now Controls Nearly Half", Forbes, 16 August 2026. https://www.forbes.com/sites/rrapier/2026/08/16/battery-storage-grew-66-in-2025-china-now-controls-nearly-half/

[24] Hoymiles Power Electronics and TÜV SÜD (2026). Global Residential Energy Storage White Paper — 35 GWh shipped in 2025, cumulative installed base above 187 GWh, compiled from Wood Mackenzie, BloombergNEF, InfoLink and EUPD Research data; as reported by ESS News, 27 May 2026. https://www.ess-news.com/2026/05/27/residential-storage-shipments-hit-35-gwh-in-2025-in-global-boom/

VOLUME 01 OF 02

01

BEFORE THE GATE.

Forecasting price, load and generation for behind-the-meter assets.

02

AT THE GATE.

Hedging battery decisions. Volume 02 measures what a battery earns once the forecasts become bids and charging decisions.

www.solship.ai · founders@solship.ai