**Japan Day-Ahead → Imbalance: Interim Report**

**Overview**

This document summarizes work to date on forecasting the Japan day-ahead to imbalance spread. It covers the current production baseline, preliminary results, and early hypotheses for why the signal works in Japan.

**Scope & period.** Data spans Japan’s regional pricing zones from **31-03-2022** to **31-12-2023**. For EDA and model testing, **31-09-2023** is used as the hard cutoff between train and test.

**Data available for modeling**

* **Prices:** Imbalance and spot prices, with flags describing calculation method (market vs. additional pricing mechanism).
* **Forecasts:** EC12 and EC00 for residual demand/load, solar, wind, temperature, and consumption.
* **Actuals:** Pumped-hydro storage and the **reserve-rate** variable used in the dual-price imbalance calculation.

*Note:* For **2024–2025**, JEPX/TSO/vendor datasets contain additional fields that should be incorporated in follow-on analysis.

**Production Baseline (for reference)**

The current production signal is a simple trend model: an **EWMA** with **10-day half-life**, computed **per 30-minute bucket**. Positions are taken from the sign/magnitude of this EWMA by bucket.

**Performance (test period)**

* [Figure: Cumulative PnL — Tokyo]
* [Figure: Cumulative PnL — Kansai]

**Why does the trend work (in Japan)?**

Performance is noticeably stronger in Japan than in Europe or the US. Two market features appear to explain much of the hourly shape of the spread.

**Mean spread by hour**

[Figure: Mean spread by hour]

**1) Reserve-rate threshold effects (17:00–20:00)**

The most pronounced pattern is a **negative spread** during **17:00–20:00**. This likely reflects **risk preferences around the reserve-rate mechanism**: participants bid up **spot** for later hours when the threshold is most likely to be hit, pulling spot above imbalance in those windows.

* [Figure: Frequency of reserve-rate threshold hit by hour]

**2) Pumped-hydro behavior & ramp constraints**

Pumped hydro accounts for roughly **~15%** of “production” in Tokyo (similar in other regions). Operators tend to **buy in the spot** when power is cheap early in the day and **sell later in the evening** when spot is stronger—reinforcing the hourly shape of the spread. In practice, **ramp constraints** may limit how quickly outage can be adjusted, strengthening the intra-day pattern observed.

* [Figure: Yearly mean of hydro outage by hour]

*Additional details and supporting plots are provided in the Appendix.*

want me to go on and reword **“Simple Improvements to the model”** (half-life sweep, floor handling, clipping/ACF) in the same style next?

**Simple Improvements to the Baseline**

**1) Half-life tuning**

We first tested whether the 10-day half-life in the baseline EWMA is close to optimal.  
Result: across a sweep of half-life values, performance peaks near the current setting; gains from further tuning are marginal.

* [Figure: EWMA half-life vs. R²/performance]

**Takeaway:** keep **HL = 10 days** as the default until other features materially change the optimum.

**2) Handling the spot floor**

Japan’s spot can clear at a hard floor of **0.01**. Short positions perform poorly on floor days because spot can’t move lower while imbalance can. This shows up clearly in conditional bin plots.

* [Figure: Bin plots — floor hit vs. not]

Two simple remedies:

1. **Masking:** avoid opening shorts during floor conditions. Simple, but it ignores sizing.
2. **Conditional averaging:** maintain **separate rolling means** of the spread for (a) floor-hit periods and (b) non-floor periods, and size/tilt positions accordingly.

**Takeaway:** conditional means preserve signal while reducing systematic floor bias.

**3) Spike memory / clipping**

Because EWMA averages **all** past values, large spikes (≈ 3σ events) bleed into subsequent days and bias the bucket-level signal. We observe elevated positive autocorrelation right after spike days that then decays back to baseline.

* [Figure: ACF around spike events — before vs. after]

Working remedy: maintain a **rolling mean of prior spike outcomes** and clip or dampen their influence in the EWMA update so the model **reacts quickly** on the day, then **forgets faster** afterward.

**Takeaway:** controlled clipping improves responsiveness without locking in spike bias.

**4) Combined tweaks (Floor + Clip)**

Applying the floor-aware conditioning **and** spike-aware clipping yields small but consistent gains versus the raw EWMA.

**Takeaway:** modest improvements across error metrics and PnL; serves as a stronger reference baseline before adding new features/hypotheses.

Yes—adding a short “bridge” block right after **Data available for modeling** will make the rest flow. Drop in these subsections (keep them tight—1–3 lines each) and then move into the baseline model.

Here’s a clean set you can paste in (use Heading 2/3 in Word):

1. **Data quality & preprocessing**
   * Coverage, missingness, outliers; resampling/alignment to 30-min.
2. **Leakage controls**
   * What was known at decision time; vendor lag rules; forecast cutoffs.
3. **Target definition**
   * Exact formula for day-ahead → imbalance spread; units; sign convention.
4. **Feature snapshot (current)**
   * Brief list by category (Prices, Forecasts EC00/EC12, Actuals: reserve-rate, pumped hydro).
   * Note: extended features for 2024–25 deferred to backlog.
5. **Train/validation/test split**
   * Calendar + hard cutoff (**31-09-2023** test start); region handling.
6. **Backtest protocol & metrics**
   * Walk-forward / bucketed evaluation; MAE/MSE/R², IC, hit-rate, PnL, t-stat, ESS.
7. **Costs & trading constraints**
   * Transaction costs, borrow/shortability, position limits, execution assumptions.
8. **Reproducibility notes**
   * Seeds, environment, data versions.