Forecasting Spain Daily Electricity Generation (REE) January 2026

- Capstone Forecasting Project
- Time Series Analysis
- Top-7 generation technologies | Daily forecasts + uncertainty
- Team: Spanish Forecasters

Team & Contributions.

- Enrique Ruiz:
  - Data preparation (CSV → clean panel dataset)
  - Forecasting pipeline implementation (train/val/test split)
  - Evaluation and model selection (metrics + plots)
  - Problem description (what we want to predict and why it matters)
- Mireia Montoya:
  - Models (Baseline, AutoARIMA, MLForecast, NeuralForecast)
  - Final forecasts export (Jan 2026 CSV/Parquet)
  - Data exporation and insights
  - Results interpretation

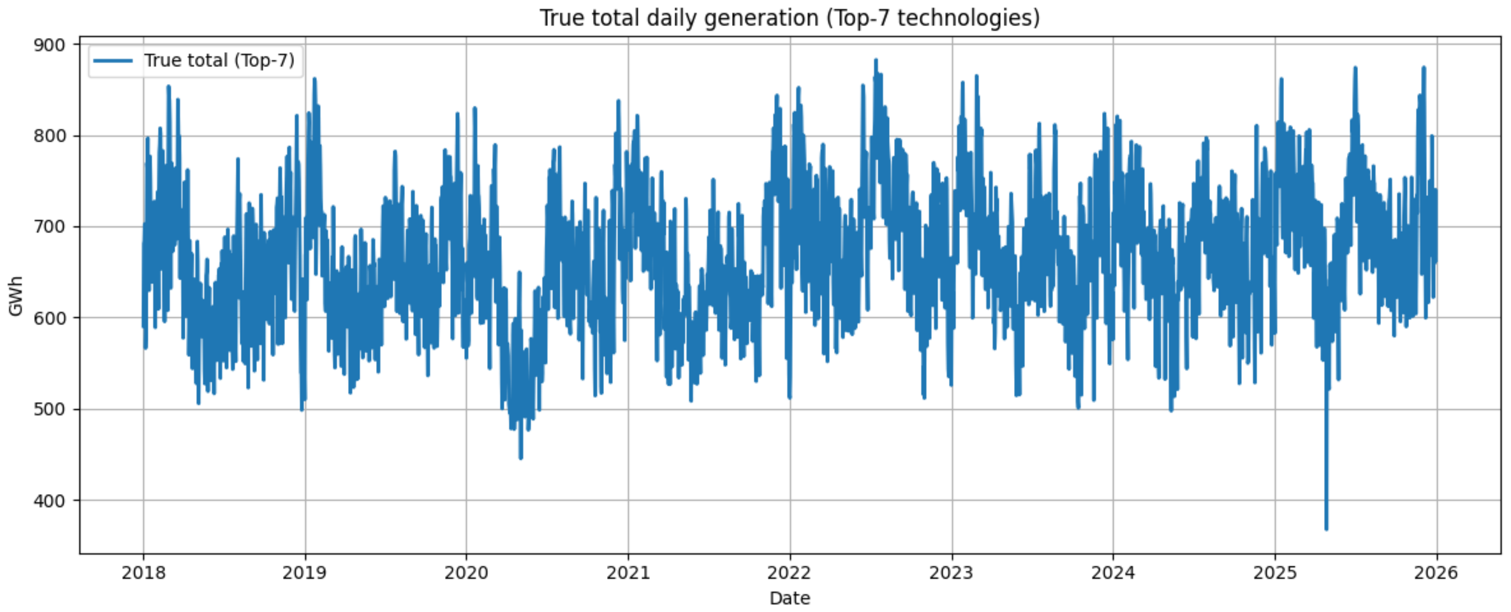
Business Problem

Problem:

- Electricity generation in Spain is highly variable due to weather conditions, demand fluctuations, and seasonal patterns. This variability introduces uncertainty in grid operation, generation planning, and energy market decisions.

Goal :

- Forecast daily electricity generation by production technology for January 2026 to support better planning decisions, reduce operational risk, and minimize economic losses caused by unexpected changes in generation.



Value and Decisions

- Plan reserves and backup generation
- Schedule maintenance on safer days
- Reduce balancing costs: avoid over/under planning
- Intervals help decisions: plan for worst-case and best-case scenarios
- **Impact if no forecast /bad forecast:**
- **No forecast** → more last-minute actions, higher balancing cost and risk
- **Bad forecast** → wrong reserve planning

DATASET INFO

Item	Value
Source	Spanish Electic Network (REE)
Frequency	Daily
Target	Electricity Generation (GWh)
Date range	01/01/2018 – 31/12/2025

DATASET INFO

Advert!!!:

As the dataset is from Spain, the features are written in Spanish. Translation:

- **Carbón:** coal
- **Ciclo combinado:** combined cycle
- **Cogeneración:** cogeneration
- **Eólica:** Eolic
- **Hidráulica:** Hidraulic
- **Nuclear:** nuclear
- **Solear fotovoltaica:** Solar Photovoltaics

Solution Overview

1. **Business goal:** forecast daily electricity generation (Top-7 technologies).
2. **Data preparation:** clean panel data, full daily grid, missing values handled
3. **Feature engineering:** lags, calendar features and rolling statistics.

4. **Modelling:** baselines, statistical (ARIMA), ML and neural models.

5. **Evaluation:** time-based Train/Val/Test split (last 31 + 31 days) with error metrics.

6. **Output:** January 2026 forecast with uncertainty intervals and saved results.

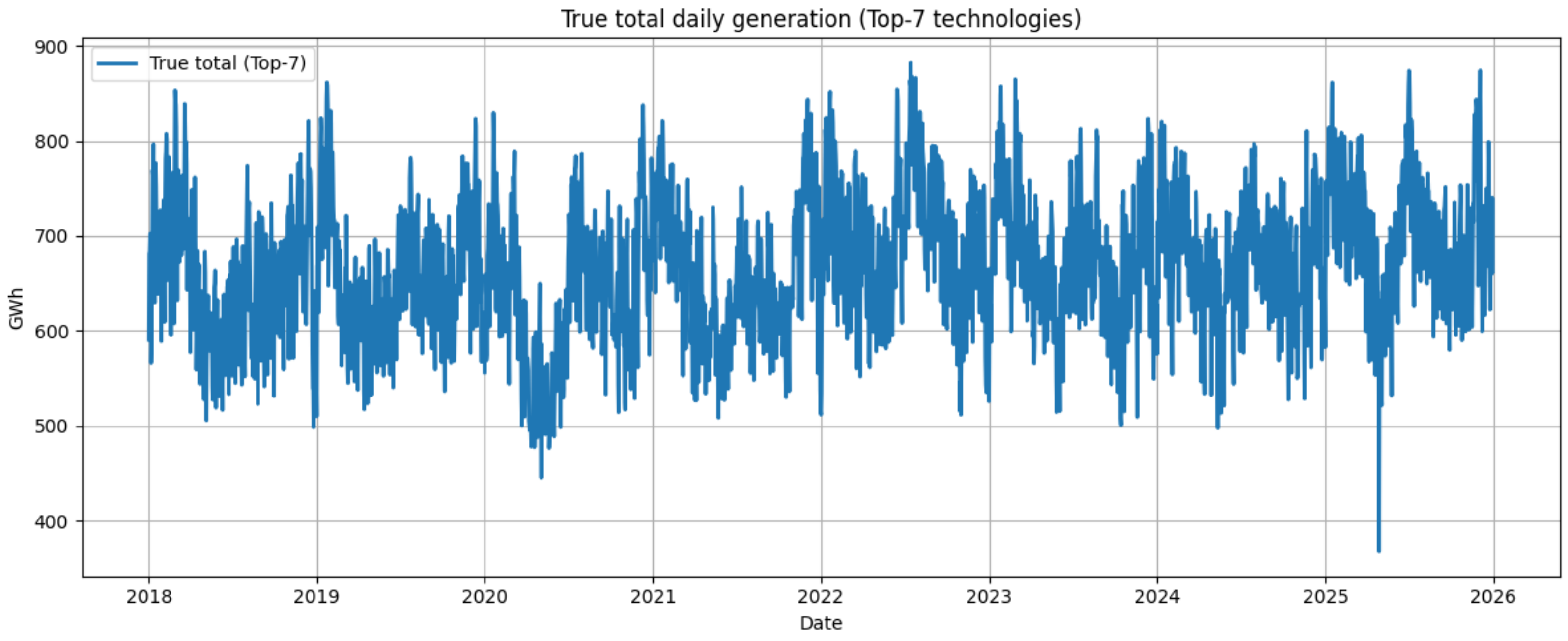
Data Engineering.

From raw REE files to a clean daily panel dataset

1. Input: multiple yearly CSV exports (wide format)
2. We convert to long format: (unique_id, ds, y)
3. We clean dates and numeric values (missing values)
4. We remove “Generación total” and keep Top-7 technologies (by mean generation)
5. We build a complete daily calendar for each technology
6. We fill missing days (forward/backward fill) and the negatives values to 0
7. Output: one clean panel file (parquet)

Insights from the data

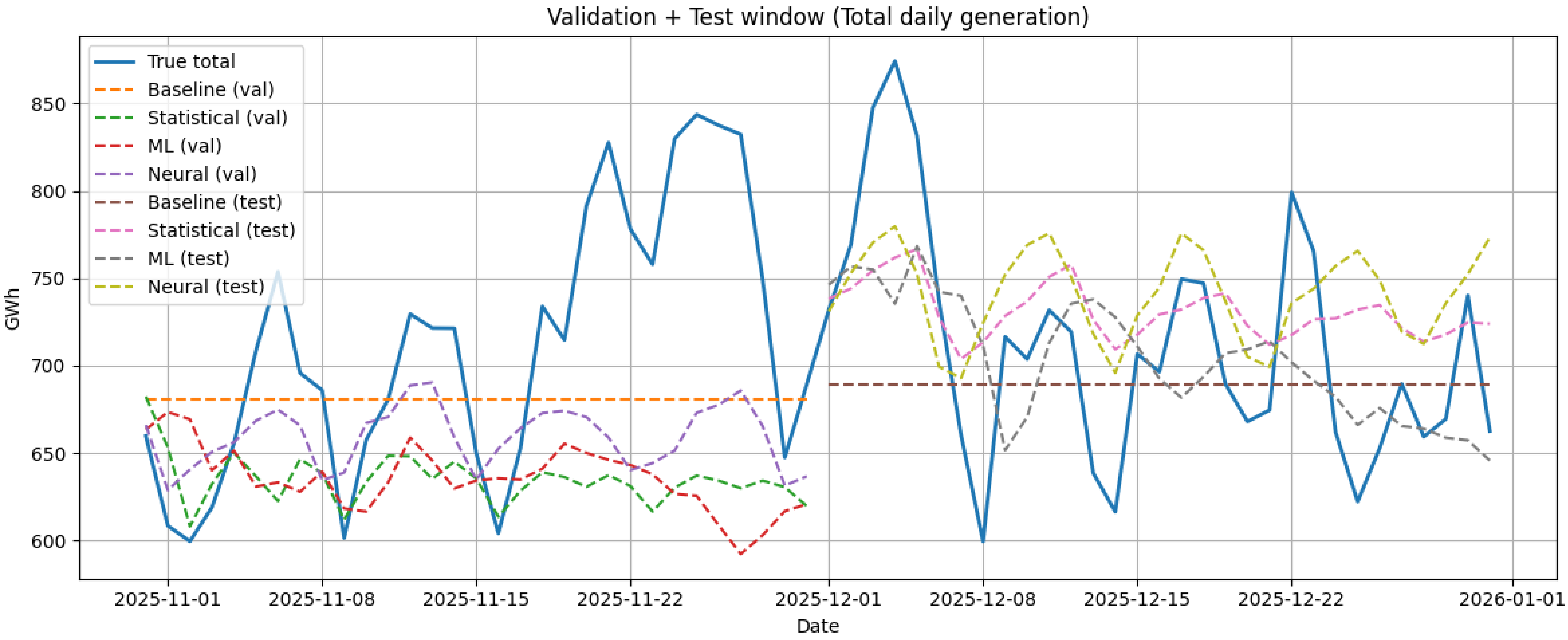
- **High variability:** total generation changes a lot day to day
- **Seasonal changes across years:** the level and volatility are different depending on the time of the year
- **Important events:** covid, blackout
- **Why it matters:** our models must handle volatility + rare shocks → we report prediction intervals



Model Families Tested

- **Baselines:**
  - Naive (last value)
  - Moving average (7-days window)
  - Seasonal Naive (weekly)
- **Statistical:**
  - AutoARIMA (weekly seasonality)
- **Machine Learning:**
  - Random Forest
  - Gradient Boosting
  - **Features:** Lag features, calendar (day, month, weekday), and rolling & expanding statistics.
- **Neural Forecast:**
  - NBEATS
  - NLinear

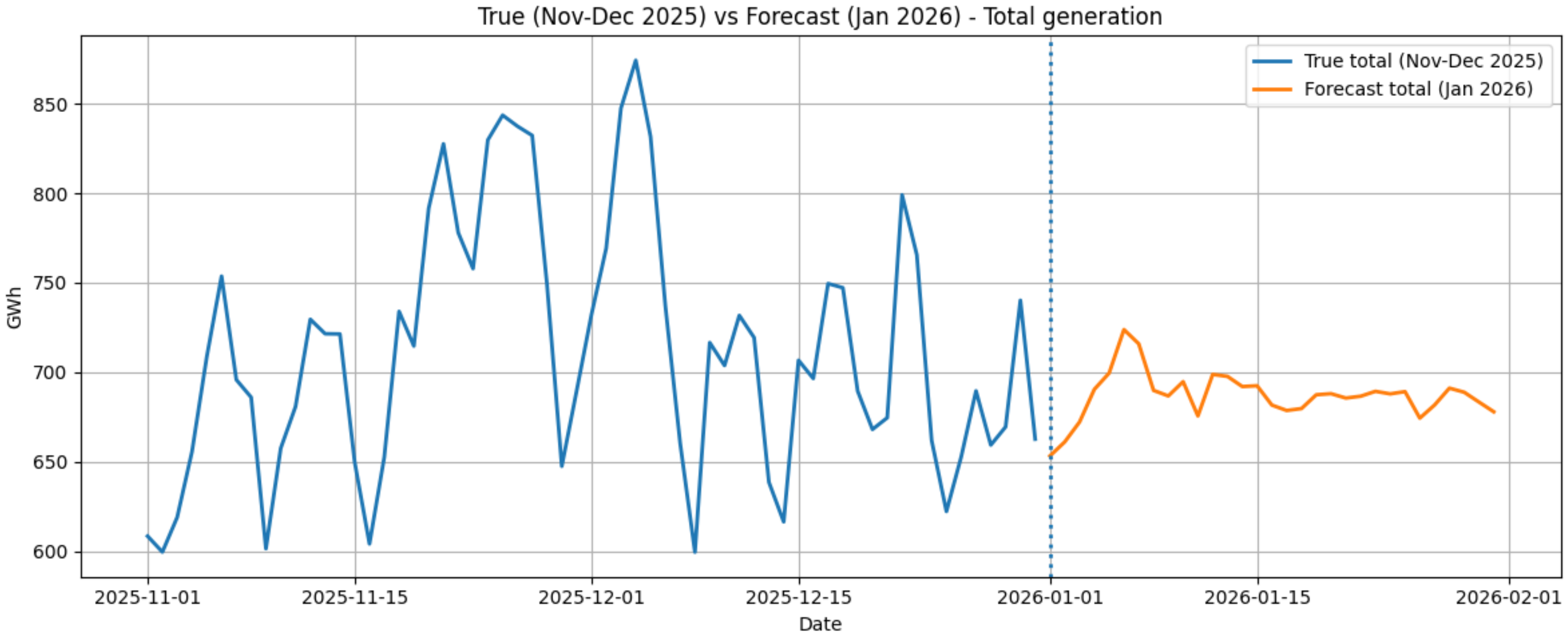
Evaluation



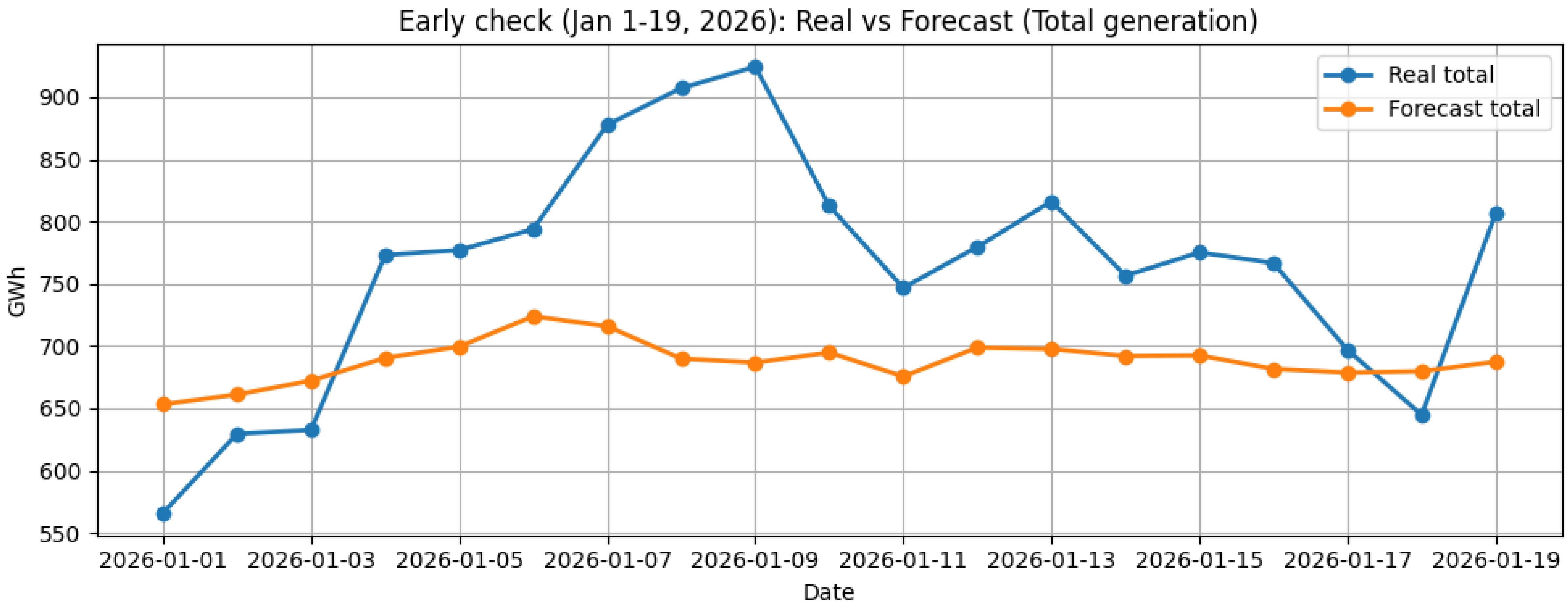
Validation + Test window (Total daily generation)
True total
Baseline (val)
Statistical (val)
ML (val)
Neural (val)
Baseline (test)
Statistical (test)
ML (test)
Neural (test)
GWh
850
800
750
700
650
600
2025-11-01
2025-11-08
2025-11-15
2025-11-22
2025-12-01
2025-12-08
2025-12-15
2025-12-22
2026-01-01
Date

January 2026 Forecst (Top-7 total)

- Daily generation for January 2026 forecast
- Orange line = point forecast (yhat)



Early check (Jan 1-19, 2026) Real vs Forecast



Early check (Jan 1-19, 2026) Real vs Forecast

- **Final model used for the forecast →RandomForestRegressor(lags)** selected as best overall by Validation MAE across 7 technologies)
- **Accuracy:** MAE = 94,64 GWh, MAPE = 11,91%
- **Interpretation:** the model captures the level, but it misses short-term spikes in the first three weeks of January.
- **Why this is expected**: we only used historical generation (lags+calendar) and did not include external drivers (weather, demand..)
- **Further improvements**
  - Add features like temperature, wind speed, holidays, demand)

- Even with imperfect accuracy, a forecast is useful for planning and risk management (better than “no forecast”)

THANK YOU FOR YOUR ATTENTION :)