Workshop 1 - Systems Analysis: Global Energy Forecasting Competition 2012 - Load Forecasting

David Santiago Tellez Melo – 20242020107
Ana Karina Roa Mora – 20232020118
Daniela Bustamante Guerra – 20241020131
Andrés Felipe correa Méndez – 20221020141

27 de septiembre de 2025

1. Competition Overview

1.1. Goal

The *Global Energy Forecasting Competition 2012 — Load Forecasting Track (GEFCom2012)* was designed to foster innovation in short-term electricity demand forecasting. The primary goal was to predict hourly system load for a U.S. utility across multiple zones, while maintaining hierarchical consistency between individual zones and the aggregated system load. Participants were asked to carry out both *backcasting* (reconstructing missing historical values) and *forecasting* (predicting future demand) under conditions that mimic operational challenges.

Forecasting electricity load is of strategic importance to utilities and system operators, as accurate forecasts enable better scheduling of generation, trading, and maintenance. The competition encouraged the application of statistical learning, time series analysis, and ensemble methods to improve forecast accuracy over standard benchmarks.

1.2. Dataset Structure

The dataset provided by the organizers contained several components central to hierarchical load forecasting:

- **Load History:** Hourly demand data for 20 individual zones, with an additional zone (*Zone 21*) representing the total system load as the sum of all zones. This structure explicitly enforced hierarchical relationships in the system.
- **Temperature History:** Hourly temperature records from multiple weather stations geographically associated with the load zones. The relationship between load and temperature is nonlinear, making this input critical for predictive modeling.
- **Holiday List:** Calendar information identifying U.S. public holidays, which significantly influence electricity consumption patterns.
- **Benchmark Forecasts:** A baseline set of forecasts supplied as a reference point against which participants could measure improvements.
- **Solution Files:** True load and temperature values for the evaluation period. These were hidden during the competition, with only subsets revealed for public leaderboard scoring.

The training set consisted of historical records, while test sets contained missing values for both load and temperature that participants were required to forecast.

1.3. Constraints

Several explicit and implicit constraints shaped the competition:

- **Hierarchical Consistency:** Forecasts for the 20 individual zones needed to be coherent with the aggregated system load. Inconsistencies in aggregation could reduce performance.

- **Incomplete Information:** Temperature data for the forecasting horizon was not fully available. Participants were required to model or approximate temperature values, introducing additional uncertainty.

- **Nonlinear Load-Weather Relationship:** The effect of temperature on load demand is nonlinear and exhibits threshold effects, such as sharp increases during extreme heat or cold. Capturing these nonlinearities was essential for competitive performance.

- **Evaluation Procedure:** Kaggle implemented a public/private leaderboard split, with scores based on randomly sampled subsets of the test set. This setup made exact replication of rankings outside the platform impossible.

- **Metrics:** Accuracy was assessed primarily using error measures such as Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), with improvements reported relative to the benchmark forecasts.

- **Submission Rules:** Kaggle imposed submission limits, requiring participants to balance exploration and optimization. Final rankings were determined solely by the private leaderboard results.

1.4. Key Takeaways

The competition demonstrated the inherent complexity of hierarchical load forecasting, the challenges of dealing with incomplete exogenous data (temperature), and the importance of domain-specific factors such as holidays. It also established GEFCom2012 as a benchmark dataset for energy forecasting research, showcasing the effectiveness of ensemble models, feature engineering, and hierarchical reconciliation techniques in large-scale predictive tasks.

2. Systems Analysis Report

2.1. Systemic Analysis

From a systems perspective, the GEFCom2012 Load Forecasting competition can be described as a set of interconnected elements that together form a forecasting system. These elements include the data provided, the methods used by participants, the expected results, and the evaluation rules defined by the organizers.

- **Inputs:** The system starts with the information provided to the competitors:
    - Historical electricity load for 20 zones and an additional aggregated zone.
    - Hourly temperature data from several weather stations.
    - A list of holidays, which often change normal patterns of electricity demand.
- **Processes:** Participants had to transform and analyze the inputs in order to create useful predictions:
    - Preparing and cleaning the data, for example dealing with missing values.
    - Creating models that connect past load and temperature with future demand.
    - Adjusting the zone forecasts so that they remain consistent with the total system load.
- **Outputs:** The main outputs of the system were:
    - Forecasted electricity demand for each zone and for the total system.
    - Reconstructed backcasted values for parts of the historical data that were incomplete.
- **Evaluation:** The quality of the forecasts was judged using error measures such as Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Kaggle's platform used a public leaderboard based on part of the test set, and a private leaderboard with hidden data to determine the final ranking.

2.2. Complexity & Sensitivity

The forecasting system in the competition presented multiple layers of complexity due to the variety of data sources, the temporal dimension, and the hierarchical nature of the predictions. Each zone had its own consumption behavior, but at the same time all zones were interdependent through the requirement that their values must sum to the total system demand.

Several constraints influenced the modeling process. The presence of missing or incomplete data required careful handling to avoid bias. Seasonal patterns, such as increased load during winter and summer, introduced variability that had to be captured by the models. Holidays and special events also created abrupt changes that were difficult to predict using traditional linear models.

The sensitivity of the system was evident in the way small changes in input variables could significantly affect the forecasts. A slight deviation in temperature predictions could lead to large errors in estimated demand, especially during peak hours. The aggregation constraint also amplified sensitivity, since errors in individual zones accumulated in the total system forecast.

Different modeling choices further increased variability. The selection of features, the tuning of hyperparameters, and the choice of algorithms all produced differences in accuracy. The evaluation metrics, particularly RMSE and MAPE, highlighted these sensitivities because they penalized larger deviations and made models less robust to extreme cases.

2.3. Chaos and Randomness

The forecasting system exhibited characteristics that can be associated with chaotic and random behavior. Electricity demand is not only influenced by predictable seasonal and daily cycles but also by irregular human activities and external events. Sudden changes in consumer behavior, such as those caused by extreme weather conditions or unexpected social events, introduced nonlinear patterns into the data.

Temperature effects demonstrated feedback-like behavior. A small increase in temperature during hot seasons could trigger widespread use of cooling systems, which in turn caused a sharp rise in electricity consumption. Similarly, in cold seasons, minor temperature drops could lead to increased heating demand, creating nonlinear responses in the system.

Noise in the weather data added additional randomness. Forecast errors in temperature propagated directly into demand forecasts, amplifying uncertainty. The requirement to maintain consistency between individual zone forecasts and the aggregated total further complicated this dynamic, as small random fluctuations at the zone level could accumulate and cause disproportionate effects in the overall system load.

The competition setting itself introduced another layer of unpredictability. Since participants developed models independently, unforeseen modeling interactions emerged. Different approaches produced varying sensitivities to outliers, nonlinearities, and hidden structures in the dataset, reflecting the inherent complexity and partially chaotic nature of the forecasting task.

2.4. Conclusion

The analysis of the Global Energy Forecasting Competition 2012 Load Forecasting highlights the main strengths and weaknesses of the system under study. A clear strength was the availability of multiple data sources, including historical electricity load, weather information, and calendar events, which provided participants with a rich context to build forecasting models. The hierarchical structure of the dataset and the well-defined evaluation metrics supported the development of systematic approaches for electricity demand prediction.

The system also revealed weaknesses that limited predictability. The presence of missing and incomplete data created challenges in preprocessing. Strong sensitivity to temperature fluctuations made the system vulnerable to small changes in input conditions, particularly during peak demand seasons. The aggregation constraint across zones increased the risk of error accumulation, reducing robustness. Chaotic influences from irregular human activities, unexpected events, and nonlinear feedback processes further complicated modeling efforts.

The competition structure emphasized both the opportunities and difficulties inherent in forecasting complex energy systems, showing the balance between available information, methodological choices, and the unpredictable dynamics of real-world electricity demand.

3. Visual Representation

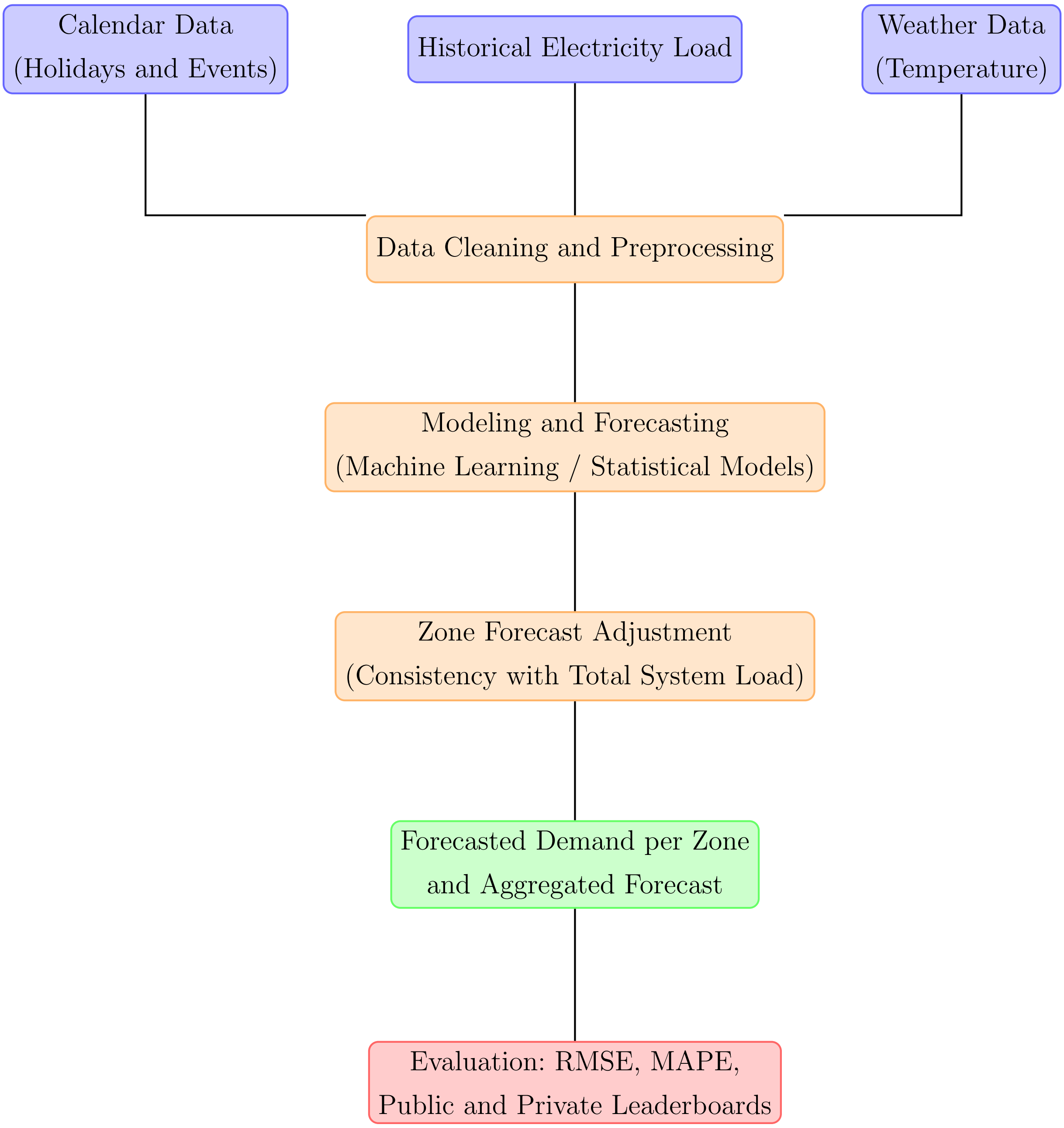

Figura 1: System architecture and data flow of the Global Energy Forecasting Competition 2012 Load Forecasting.

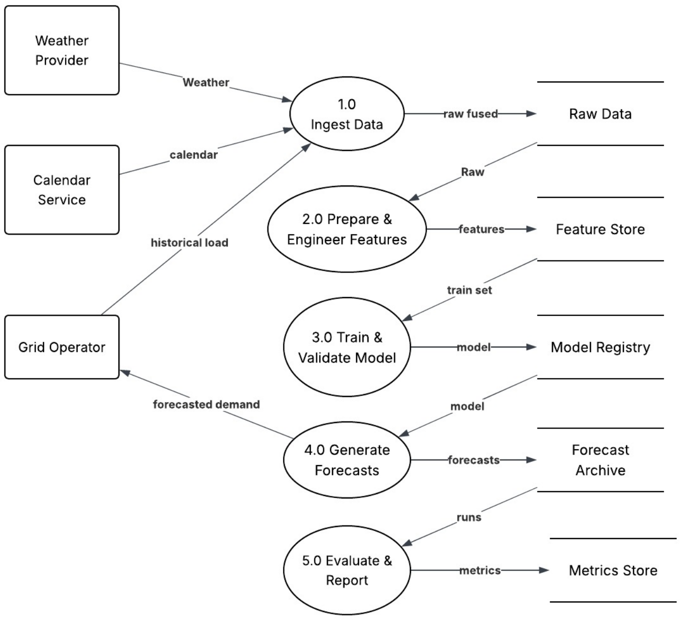
Weather Provider
Calendar Service
Grid Operator
Weather
calendar
historical load
forecasted demand
1.0 Ingest Data
2.0 Prepare & Engineer Features
3.0 Train & Validate Model
4.0 Generate Forecasts
5.0 Evaluate & Report
raw fused
Raw
features
train set
model
model
forecasts
runs
metrics
Raw Data
Feature Store
Model Registry
Forecast Archive
Metrics Store

Figura 2