Universidad Distrital Francisco Jose de Caldas

The system analysis of a Kaggle competition

Diego Angelo Ruano Vergara
20242020278

Arlo Nicolas Ocampo Gallego
20221020104

Tomás Cárdenas Benítez
20221020021

Sebastian David Trujillo Vargas
20242020217

THE COMPETITION

Competition's summary and goal

Before the analysis is made, it's essential to know what the competition embarks and what the goal of the it is. First of all the competition on Kaggle aims to build predictive models using the provided data for *load forecasting*. The goal of the model is to perform backcasting and forecasting of hourly loads (in kW) for a U.S. utility company across 20 zones. With that being said, participants must produce both zonal forecasts (20 time series) and a system-level forecast (sum of the 20 zonal series), for a total of 21 series.

Load Forecasting One of the key elements of why the competition its itself is the load forecasting. The load forecasting is a procedure based in predicting electricity demand to ensure sufficient energy supply for an efficient grid operation. This operation is really important for electricity usage because its importance comes from minimizing costs while ensuring sufficient electricity for the power grid; and it also ensures enough energy to the power grid to prevent failures and overloads. Another essential fact about load forecasting is the use of historical data (weather, location, date, hour, among others) in order to make the prediction, which is exactly what the competition wants.

Competition's Dataset

Another element needed in the competition comes in the datasets given to the participants. The competition gives six datasets in order for working on the model making, these being

1) "$Benchmark.csv$": Contains the hourly forecast results generated by a simple or naive model, providing a baseline performance score against which contestants' models are measured.
2) "$Load_history.csv$": Contains the historical hourly electricity load (in kW) for 20 individual zones and the overall system total (Zone 21). This data typically spans a period from 2004/1/1 to 2008/6/30.
3) "$submission_template.csv$": Provides the exact format needed for the final submission file, including the required column names and time indices for the forecasted and backcasted loads.
4) "$temperature_history.csv$": Contains the historical hourly temperature data recorded at 11 different weather stations, which are correlated with the electricity consumption zones.
5) "$test.csv$": This file defines the time steps for which forecasts and backcasts are required. This includes 8 weeks scattered throughout the history that need to be "backcasted" (filled in) and the final 1 week (2008/7/1 to 2008/7/7) that needs to be "forecasted."

6) "$weights.csv$": Contains the weighting scheme used to calculate the competition's evaluation metric, the Weighted Root Mean Square Error (WRMSE). The weights are assigned differently to various tasks (e.g., system-level forecasts have a much higher weight than zonal backcasts).

Competition's evaluation metric

Finally, knowing the goal and the datasets given to the participant, it should be know what type of evalutaion metric while be done by the competition. The metric is presented next:

- The forecasting accuracy will be evaluated by weighted root mean square error.
- The weights are assigned as following:
- Each hour of the 8 backcasted weeks at zonal level: 1;
- Each hour of the 8 backcasted weeks at system level: 20;
- Each hour of the 1 forecasted week at zonal level: 8;
- Each hour of the 1 forecasted week at system level: 160;
- Details of weight assignment are shown in data file: weights.csv.

THE SYSTEM ANALYSIS

Elements

1) **Wind Farms (7 units):** The physical sources of power generation, anonymized and normalized in the dataset.
2) **Historical Data:** Hourly power measurements over three years, providing training and validation material.
3) **Meteorological Forecasts:** Exogenous variables representing the main drivers of wind power variability.
4) **Forecast Horizon:** Continuous predictions for lead times of 1 to 48 hours ahead.
5) **Predictive Models (Regression Approaches):** Statistical, machine learning, or hybrid models that map meteorological inputs and past observations to continuous forecasts.
6) **Evaluation Metric:** Accuracy of continuous forecasts, measured through regression error metrics such as RMSE or MAE, with persistence forecasts as the baseline benchmark.
7) **Stakeholders:**
   - Grid operators, who require reliable forecasts for balancing supply and demand.
   - Energy traders, who depend on accurate regression forecasts for bidding strategies.
   - Wind farm operators, who benefit from predictive maintenance and scheduling.
   - Researchers and competition participants, who benchmark algorithms and forecasting methods.

Relationships Between Elements

- Historical wind power data and meteorological forecasts feed into regression models as features and labels.
- Regression models generate continuous predictions of wind power generation for each farm and each forecast horizon.
- Evaluation metrics provide feedback by comparing predicted values with observed outcomes, guiding model refinement.
- The persistence model serves as a baseline, establishing a minimum performance threshold.
- Stakeholders utilize forecasts in operational, economic, and research contexts.

System dynamics

- **Feedback loops:** Forecasting errors highlight system limitations and inform iterative model improvement.
- **Error propagation:** Larger errors at longer horizons can significantly affect operational and economic outcomes.
- **Temporal dependencies:** The system must capture autocorrelations, seasonal patterns, and the impact of exogenous meteorological variables.

Constraints and Challenges

- Accurately modeling nonlinear relationships between wind conditions and power output.
- Addressing missing or noisy data in both training and operational contexts.
- Designing models that generalize across multiple wind farm sites.
- Maintaining accuracy across all 48 forecast horizons.
- Working under operational constraints, where only forecasts available at a given time can be used.

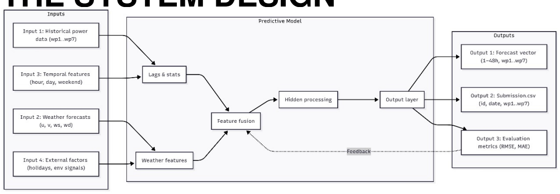
System Boundaries

- **Inside the system:** Wind farms, historical power data, meteorological forecasts, regression models, forecasts, evaluation metrics.
- **Outside the system:** Long-term energy policies, turbine mechanical failures, and broader market regulations that are not directly modeled.

Outcome

The systemic objective is to design regression models capable of producing accurate and reliable short-term forecasts of wind power generation at multiple sites. These models reduce uncertainty, enhance grid stability, support energy market efficiency, and contribute to the broader integration of renewable energy sources.

THE SYSTEM DESIGN



Explanation of the Model in the System

The predictive model constitutes the central component of the wind power forecasting system. Its role is to translate the information from historical generation and meteorological forecasts into accurate predictions of future wind power output. The model works by combining two key types of input data: past observations of wind power generation from the seven wind farms, and weather forecasts that include wind speed, wind direction, and wind components provided for up to 48 hours ahead. Before reaching the model, these raw inputs are preprocessed, aligned in time, and enriched with additional features such as lagged power values, rolling averages, or transformations of wind direction. This step ensures that the model can capture both temporal patterns and meteorological dynamics.

Model prediction

Once the inputs are prepared, the model is trained to predict wind power at horizons from 1 to 48 hours ahead. Two modeling strategies are possible. In the first, a specific model is trained for each horizon, which allows for fine-tuned optimization of short-term and long-term forecasts separately. In the second, a single multi-output model produces the full sequence of 48-hour forecasts at once, which better captures dependencies between consecutive horizons. Both approaches aim to minimize the Root Mean Square Error (RMSE), the official evaluation metric of the competition.

Model training

The model is trained on historical data from July 2009 to December 2010, and validated using a rolling-origin expanding window, which mimics real operational forecasting conditions. This validation strategy prevents data leakage and ensures that predictions are based only on information that would have been available at the time. Once deployed, the trained model is integrated into the forecasting pipeline: for each new forecast cycle, it receives the latest available observations and weather predictions, processes them into features, and outputs wind power forecasts for the next 48 hours.

Final output

Finally, the results are constrained to remain within the normalized range [0,1] and exported in the required format for evaluation. Beyond accuracy, the model also provides interpretability through feature importance analysis, allowing operators to understand which meteorological and temporal factors drive the predictions. In this way, the model acts as the bridge between raw data and actionable forecasts, ensuring both reliability and practical relevance for wind power integration into the energy system.

SYSTEM'S SENSIBILITY AND COMPLEXITY

Sensibility

The Kaggle load forecasting competition exhibits a high degree of sensitivity to its input parameters and modeling assumptions. Electricity demand is heavily driven by temperature fluctuations; therefore, even marginal changes in the temperature dataset may generate substantial deviations in predicted load curves. This implies that the forecasting system is strongly conditioned by external environmental variables that cannot be directly controlled. In addition, the historical load data introduces another layer of sensitivity: missing values, measurement errors, or temporal misalignments can propagate throughout the model training process, amplifying forecasting errors. The competition's evaluation metric, the Weighted Root Mean Square Error (WRMSE), accentuates this phenomenon. Because the system-level forecasts carry disproportionately high weights, even minor inaccuracies at the aggregate level may outweigh otherwise strong zonal predictions, penalizing participants severely. Sensitivity analysis can therefore be approached by systematically perturbing input variables, conducting stress tests with extreme weather scenarios, or applying Monte Carlo simulations to quantify how uncertainty in the inputs propagates into the output forecasts. This systemic vulnerability underscores the necessity of robust preprocessing, feature engineering, and model regularization techniques to stabilize predictions against small perturbations.

Complexity

Beyond sensitivity, the forecasting task embodies characteristics of complexity. The system involves the interaction of multiple subsystems: 20 regional zones, diverse weather stations, and heterogeneous temporal patterns that are not linearly correlated. Demand patterns are influenced simultaneously by climatic conditions, socio-economic behaviors, and unexpected events, creating nonlinear interdependencies that complicate prediction. These interdependencies can manifest as emergent phenomena, where the behavior of the aggregated system cannot be trivially inferred from the behavior of its parts. Moreover, the weighting scheme in the evaluation metric introduces conflicting objectives: optimizing for zonal accuracy may degrade system-level accuracy and vice versa, producing inherent trade-offs. Such complexity suggests that linear or static models are insufficient; instead, hybrid approaches combining statistical learning, time-series decomposition, and machine learning are required to capture the multi-scale dynamics of the system. Ultimately, the competition exemplifies a high-dimensional, tightly coupled system in which uncertainty, interdependence, and trade-offs must be explicitly managed.

SYSTEM'S CHAOS AND RANDOMNESS

The competition also highlights elements of chaos and randomness. From a systems theory perspective, the electricity demand process is partially chaotic: small perturbations in initial conditions, such as an unanticipated temperature spike or an unusual holiday, may trigger disproportionate deviations in consumption. These nonlinear effects limit the predictability horizon and emphasize the system's sensitivity to boundary conditions, resembling the behavior of chaotic dynamical systems. Randomness is further introduced by noise in sensor measurements, incomplete historical records, and exogenous shocks such as storms or blackouts. These factors inject stochastic variability that cannot be fully captured by deterministic models. Effective handling of chaos and randomness requires adopting probabilistic forecasting approaches, ensemble models, and uncertainty quantification techniques, which provide not only point forecasts but also confidence intervals. Recognizing the chaotic and stochastic nature of the system is thus essential to designing resilient forecasting models capable of adapting to unpredictable real-world conditions.

REFERENCES

1. "Model Benchmark," ScienceDirect Topics, Elsevier, accessed Sept. 25, 2025. Available: https://www.sciencedirect.com/topics/computer-science/model-benchmark#:~:text=A%20'Model%20Benchmark'%20refers%20to,assess%20their%20performance%20and%20reproducibility
2. Kaggle, "Global Energy Forecasting Competition 2012 - Load Forecasting," Kaggle Competitions. [Online]. Available: https://www.kaggle.com/competitions/global-energy-forecasting-2012-load-forecasting. [Accessed: Sep. 25, 2025].
3. T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *International Journal of Forecasting*, vol. 30, no. 2, pp. 357–363, Apr. 2014, doi: 10.1016/j.ijforecast.2013.07.001.