System Analysis of the Kaggle GEFCom2012 Wind Forecasting Competition

Diego Angelo Ruano Vergara Tomás Cárdenas Benítez
Arlo Nicolas Ocampo Gallego
Sebastián David Trujillo Vargas

September 2025

Abstract

This report presents a complete system analysis of the Kaggle *Global Energy Forecasting Competition 2012 - Wind Forecasting*. The competition aims to predict short-term wind power generation based on historical and meteorological data. We describe the structure of the dataset, the evaluation metric, the systemic elements involved, and the model architecture that underlies forecasting processes. The analysis also discusses the system's sensitivity, complexity, and stochastic characteristics from a system science perspective.

1 The Competition

1.1 Competition's Summary and Goal

Before the analysis is made, it is essential to understand what the competition entails and its main objective. The *GEFCom2012 Wind Forecasting* competition on Kaggle aims to build predictive models that forecast wind power generation from historical measurements and meteorological data.

This competition is often confused with the *Load Forecasting* track of the same year, but this work focuses exclusively on the **Wind Forecasting** subcompetition, which concerns wind power generation, not electricity demand.

Participants must forecast the normalized hourly wind power output for seven anonymized wind farms over several forecast horizons (from 1 to 48 hours ahead). The goal is to design accurate models that can predict future wind generation, improving upon the baseline persistence forecast.

Accurate wind power forecasting is critical for modern energy systems since wind is an intermittent resource. Reliable forecasts help grid operators maintain stability, assist energy traders in market operations, and support wind farm managers in maintenance and scheduling.

1.2 Competition's Dataset

The competition provides several datasets that contain both target variables (wind power output) and explanatory variables (weather forecasts). These are:

1. `train.csv`: Historical hourly wind power generation data (normalized) for seven wind farms, including meteorological forecasts such as wind speed components and wind direction.

2. `test.csv`: Contains timestamps and meteorological forecast variables for which participants must predict wind power output.

3. `benchmark.csv`: Results from a persistence model that assumes future power equals the most recent past observation, serving as a baseline reference.

4. `metadata.txt`: Describes the structure and meaning of variables.

Unlike the Load Forecasting dataset, the Wind Forecasting dataset does not include additional temperature or load-related files. All meteorological information is embedded in the CSV files and limited to wind-related variables (u, v, ws, wd). There are no holiday or external signal variables in the original data.

1.3 Competition's Evaluation Metric

The competition evaluates models using the **Root Mean Square Error (RMSE)**:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

where $\hat{y}_i$ is the predicted and y_i the observed wind power.

Only RMSE is applied; there is no weighting scheme (WRMSE). The persistence model provides the baseline to beat.

2 System Analysis

2.1 Elements

1. **Wind Farms (7 units):** Physical generation sources, anonymized and normalized in the dataset.

2. **Historical Data:** Hourly normalized power output records used for model training and validation.

3. **Meteorological Forecasts:** Exogenous predictors such as wind speed components (u, v), magnitude (ws), and direction (wd).

4. **Forecast Horizon:** Predictions 1–48 hours ahead for each wind farm.

5. **Predictive Models:** Machine learning or regression models mapping meteorological inputs and past outputs to forecast future wind power.

6. **Evaluation Metric:** RMSE to measure accuracy against true observed values.

7. **Stakeholders:**

   - Grid operators: maintain supply–demand balance.
   - Energy traders: optimize bidding and hedging.
   - Wind farm operators: schedule maintenance and generation.
   - Researchers: design and benchmark new forecasting algorithms.

2.2 Relationships Between Elements

- Historical power and meteorological data are inputs for training regression models.

- Models generate forecasts for multiple farms and horizons.

- RMSE evaluates model outputs against true values.

- The persistence model acts as a baseline.

- Stakeholders use forecasts for operational and market decisions.

2.3 System Dynamics

- **Feedback loops:** Forecasting errors drive iterative model improvement.
- **Error propagation:** Forecast uncertainty increases with horizon length.
- **Temporal dependencies:** Models must capture seasonality, autocorrelation, and meteorological dynamics.

2.4 Constraints and Challenges

- Modeling nonlinear wind–power relationships.
- Handling missing or noisy meteorological data.
- Maintaining generalization across multiple sites.
- Achieving accuracy over all 48 horizons.
- Preventing data leakage by respecting time order.

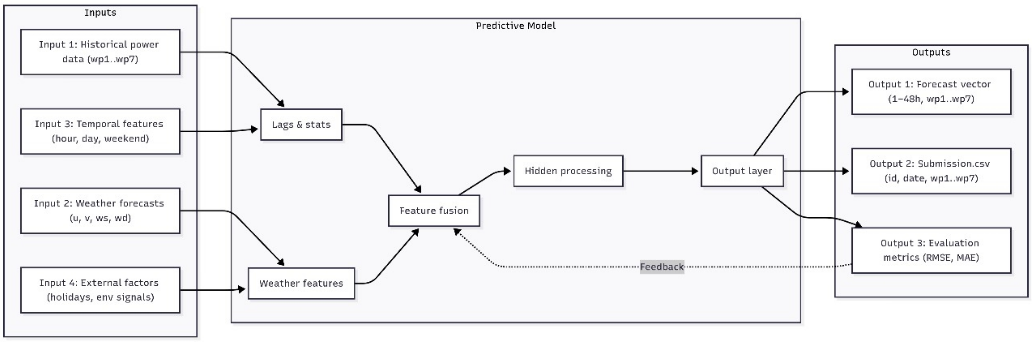
2.5 System Boundaries

- **Inside:** Wind farms, historical data, meteorological forecasts, predictive model, forecasts, RMSE.
- **Outside:** Energy policies, turbine failures, or market prices.

2.6 Outcome

The objective is to produce accurate short-term forecasts of wind power generation at multiple sites, reducing uncertainty, improving grid stability, and enabling better renewable energy integration.

3 System Design



Note on System Diagram: Inputs 1–3 (historical data, weather forecasts, and temporal features) correctly represent the dataset. Input 4 (external factors) is optional and not present in the real competition data. The evaluation output should only include RMSE as the metric.

3.1 Explanation of the Model in the System

The predictive model constitutes the central component of the system, transforming historical and meteorological data into forecasts. Before modeling, inputs are aligned and preprocessed—lag features, rolling means, and derived variables such as wind direction trigonometric components are added. This process captures temporal dependencies and meteorological variability.

3.2 Model Prediction

The model predicts wind power for 1–48 hour horizons. Two strategies exist: (1) one model per horizon, or (2) a single multi-output model that captures horizon dependencies. Both minimize RMSE against the persistence baseline.

3.3 Model Training

Training uses data from mid-2009 to 2010, validated by a rolling-origin approach that simulates real-time forecasting without data leakage. Once trained, the model processes new weather inputs each cycle and outputs forecasts for all 48 future hours.

4 System Sensitivity and Complexity

4.1 *Sensibility*

The Kaggle *Wind Forecasting* competition exhibits a high degree of sensitivity to its input parameters and modeling assumptions. Wind power output depends nonlinearly on meteorological conditions—particularly wind speed and direction. Even small deviations in forecasted wind speed near the turbine's operating thresholds can lead to large changes in predicted power. This sensitivity is amplified by temporal correlations and spatial variability among the seven wind farms.

Uncertainty in meteorological forecasts, missing values, or temporal misalignments may propagate through the model training process, resulting in significant forecast errors. Sensitivity analysis can be performed by perturbing meteorological inputs (e.g., $\pm 10\%$ variation in wind speed) or by conducting Monte Carlo simulations to observe how uncertainty in inputs affects the final RMSE. This highlights the necessity of robust preprocessing, feature engineering, and model regularization to stabilize predictions against small perturbations.

4.2 *Complexity*

Beyond sensitivity, the wind forecasting problem exhibits high system complexity. The dataset integrates multiple interdependent subsystems—seven wind farms, each influenced by meteorological variables that are themselves dynamically correlated in space and time. Wind behavior is nonlinear and chaotic, producing emergent patterns that cannot be directly inferred from single-site observations.

This complexity requires models capable of learning multi-scale temporal dynamics and nonlinear interactions. Linear regression models are insufficient; hybrid approaches combining statistical time-series techniques and machine learning (e.g., gradient boosting, neural networks, or ensemble models) are more appropriate for capturing the system's multi-dimensional dependencies.

5 SYSTEM'S CHAOS AND RANDOMNESS

Wind power generation inherently exhibits chaotic and stochastic behavior. Small perturbations in initial meteorological conditions—such as slight shifts in wind direction or localized gusts—can cause disproportionately large deviations in turbine output. This nonlinearity limits the effective forecasting horizon and aligns the system's behavior with that of chaotic dynamical systems.

Randomness arises from noise in meteorological forecasts, measurement uncertainty, and external environmental disturbances (e.g., storms). Probabilistic and ensemble-based forecasting methods can help quantify uncertainty by providing prediction intervals instead of point estimates. Recognizing the chaotic nature of wind dynamics is essential for developing resilient and adaptive forecasting models capable of handling real-world variability.

References

1. Kaggle, "Global Energy Forecasting Competition 2012 - Wind Forecasting," Kaggle Competitions. [Online]. Available: https://www.kaggle.com/competitions/GEF2012-wind-forecasting. [Accessed: Sep. 25, 2025].

2. Hong, T., Pinson, P., & Fan, S. (2014). *Global Energy Forecasting Competition 2012.* International Journal of Forecasting, 30(2), 357–363. https://doi.org/10.1016/j.ijforecast.2013.07.001