Technical Specification

Avangrid Hackathon 2025 Team

Contents

1 Scope **2**

2 Conceptual Overview **2**

3 Data Model and Notation **2**
3.1 Indices and Entities 2
3.2 Core Variables 2
3.3 Derived Quantities 3

4 Dimensional Analysis **3**

5 Data Ingestion and Normalization (`parse.py`) **3**
5.1 Functionality 3
5.2 Purpose 3

6 Descriptive Analytics and Forward Simulation (`analize.py`) **3**
6.1 Summary Statistics 3
6.2 Forward Curve Extraction 4
6.3 Monte Carlo Simulation 4

7 Risk-Adjusted Valuation (`risk.py`) **4**
7.1 Monthly Generation 4
7.2 Revenue and Price 4
7.3 Risk Premium 4

8 Model Validation and Interpretation **5**
8.1 Statistical Validation 5
8.2 Economic Interpretation 5

9 Glossary **5**

10 Assumptions and Future Work **5**
10.1 Assumptions 5

11 Validation Checklist **5**

12 Module Summary **6**

1 Scope

This document rigorously defines the mathematical, statistical, and computational logic behind the following Python modules:

- `data/parse.py` — ingestion and normalization of Excel data into clean CSVs.
- `analysis/analize.py` — descriptive analytics and forward price simulation.
- `analysis/risk.py` — risk-adjusted fixed price computation for merchant renewable assets.

The framework supports transparent and reproducible valuation of merchant renewable energy assets across multiple U.S. markets (ERCOT, MISO, CAISO) for the 2026–2030 horizon.

2 Conceptual Overview

The model blends data engineering, probabilistic simulation, and risk-based financial valuation.

- All forecasts are anchored on observable forward curves.
- Risk is quantified through percentile-based price distributions, analogous to Value at Risk (VaR).
- Historical generation patterns are repeated to capture realistic seasonal volume profiles.

This design ensures transparency and direct traceability from data to decision.

3 Data Model and Notation

3.1 Indices and Entities

- t: Hour index
- m: Month index
- $p \in \{\text{peak}, \text{off\_peak}\}$: Time-of-use period
- a: Asset identifier (ERCOT, MISO, CAISO)

3.2 Core Variables

$$G_t = \text{Generation (MWh)}, \quad P_t = \text{Price (\$/MWh)}, \quad F_{m,p} = \text{Forward price (\$/MWh)}.$$

3.3 Derived Quantities

$$\begin{aligned}\text{Basis}_t^X &= P_{\text{hub},t}^X - P_{\text{bus},t}^X, \quad X \in \{\text{RT, DA}\}\\ G_{m,p} &= \sum_{t\in(m,p)} G_t\end{aligned}$$

4 Dimensional Analysis

All derived quantities are dimensionally consistent:

$$[P] = \frac{\$}{\text{MWh}}, \quad [G] = \text{MWh}, \quad [R] = \$.$$

Thus $R = P \times G$ yields revenue in dollars, and R/G has units of \$/MWh.

5 Data Ingestion and Normalization (`parse.py`)

5.1 Functionality

a) Reads all Excel sheets.

b) Standardizes headers (lowercase, snake_case).

c) Cleans monetary strings (\$, parentheses) → float.

d) Builds timestamps:

$$\text{timestamp} = \text{datetime(date)} + (\text{HE} - 1)\text{h}.$$

e) Outputs one cleaned CSV per sheet to `data/clean_data/`.

5.2 Purpose

Ensures a consistent schema: `timestamp`, `gen`, `rt_hub`, `da_hub`, `peak_date`, `peak`, `off_peak`, etc.

6 Descriptive Analytics and Forward Simulation (`analize.py`)

6.1 Summary Statistics

For each asset:

$$\begin{aligned}\sigma_P &= \sqrt{\frac{1}{N-1}\sum_t (P_t - \bar{P})^2},\\ \rho(G, P) &= \frac{\text{Cov}(G, P)}{\sigma_G \sigma_P}.\end{aligned}$$

Computed metrics include:

- Capacity factor
- Price volatility
- Correlation of generation vs. price
- Negative price frequency
- Hub–busbar basis mean and variance

6.2 Forward Curve Extraction

From columns `peak_date | peak | off_peak`, create long-form $(m, p, F_{m,p})$ dataset for 2026–2030.

6.3 Monte Carlo Simulation

Monthly stochastic price draws:

$$P^{(i)}_{m,p} = F_{m,p} + \epsilon^{(i)}_{m,p}, \quad \epsilon^{(i)}_{m,p} \sim \mathcal{N}(0, \sigma_{\text{hub}}\kappa_p),$$

with $\kappa_{\text{peak}} = 1.0$, $\kappa_{\text{off}} = 0.7$. Percentiles:

$$\{P_{25}, P_{50}, P_{75}, P_{90}\} = \text{quantiles of simulated draws.}$$

7 Risk-Adjusted Valuation (`risk.py`)

7.1 Monthly Generation

Aggregate generation:

$$G_{m,p} = \sum_{t\in(m,p)} G_t.$$

Replicate across simulated years:

$$G_{m_y,p} = G_{m_{base},p}.$$

7.2 Revenue and Price

For each percentile $q \in \{25, 50, 75\}$:

$$\begin{aligned} R^{(q)}_{m,p} &= P^{(q)}_{m,p} \cdot G_{m,p}, \\ R^{(q)}_{\text{tot}} &= \sum_{m,p} R^{(q)}_{m,p}, \\ \bar{P}^{(q)} &= \frac{R^{(q)}_{\text{tot}}}{\sum_{m,p} G_{m,p}}. \end{aligned}$$

7.3 Risk Premium

$$\text{Risk Premium} = \bar{P}^{(75)} - \bar{P}^{(25)}.$$

This represents the compensation required to hedge against price uncertainty.

8 Model Validation and Interpretation

8.1 Statistical Validation

- Simulated mean $\approx$ historical mean $\pm 10\%$.
- $P_{25} < P_{50} < P_{75}$ ordering verified.
- Correlation $\rho(G, P)$ consistent with market expectations.

8.2 Economic Interpretation

$\bar{P}^{(75)}$: Price ensuring 75% probability of outperforming merchant exposure.

$$\text{Risk Premium} = P_{75} - P_{25}\text{: compensation for volatility.}$$

9 Glossary

Term	Meaning	Unit
PPA	Power Purchase Agreement	—
RT	Real-Time	—
DA	Day-Ahead	—
ISO	Independent System Operator	—
Hub	Regional market node	—
Busbar	Generator's physical node	—
P/OP	Peak/Off-Peak flag	—
REC	Renewable Energy Credit	—

10 Assumptions and Future Work

10.1 Assumptions

- Constant generation pattern (2026–2030).
- Gaussian price innovations.
- No degradation, curtailment, or O&M costs.
- Off-peak volatility scaled to $0.7\times$ peak.

11 Validation Checklist

- Clean CSV schema validated.
- Forward coverage spans 2026–2030.
- Non-zero monthly MWh confirmed.
- Percentile ordering correct.

12 Module Summary

Module	Purpose	Output
`parse.py`	Data normalization	`data/clean_data/*.csv`
`analize.py`	Forward simulation	`analysis_outputs/*_forward_sim.csv`
`risk.py`	Risk-adjusted valuation	`risk_outputs/risk_adjusted_summary.csv`