Energy Storage System Modeling Using Time Series Clustering for Long-term Planning

Md. Abu Bakkar (0422062156)

*Abstract*— Accurate representation of variability in variable renewable energy sources (VRES) and storage capacity is crucial for studying future energy scenarios with high shares of VRES. However, traditional long-term energy system models often rely on coarse time series to reduce computational effort, leading to errors in the sizing of VRES plants and storage systems. This study proposes a novel approach to address these limitations, enabling precise long-term planning for high-VRES decarbonization pathways. The method leverages clustering techniques to process time series while maintaining the ability to model inter-day and intra-day energy storage. The approach incorporates interconnected, clustered representative days. The result showed that an ESS with a 2000 MW capacity and charge/discharge rates of 800.0 MWh/day effectively reduced unmet demand to 3,203.008 GWh, balancing the annual supply of 55,204.163 GWh and demand of 56,292.449 GWh. This novel approach offers a computationally efficient and accurate solution for planning storage energy systems with high shares of VRES.

Keywords— energy modeling, machine learning, clustering, intermittency of renewable, storage system

# Literature Review

Time series clustering has demonstrated significant potential in improving the modeling accuracy of energy storage systems and the variability of variable renewable energy sources (VRES). For instance, Gabrielli et al. [2] developed a mixed-integer linear programming (MILP) approach that utilizes clustering and coupling of representative days to enhance the modeling of high-capacity storage components. This method enables year-long simulations with hourly resolution while reducing the number of binary variables, thus decreasing the computational complexity of the optimization problem. The approach effectively captures the operational behavior of storage systems by preserving hourly dynamics while simplifying long-term computations. Kotzur et al. [3] introduced a two-layer state methodology by linking clustered representative days across a year. Their work demonstrated the advantages of this approach, particularly for optimizing seasonal storage systems. By connecting clustered days, the methodology ensures the accurate representation of storage dynamics over seasonal variations, improving the model's ability to evaluate long-term energy storage requirements and their impact on system costs and reliability. Building on this, Welder et al. [4] carried out a spatiotemporal optimization of energy systems for power-to-hydrogen applications, considering a one-year time horizon. Their approach optimized storage system deployment and operation to balance renewable generation and hydrogen demand across both spatial and temporal scales. Similarly, Limpens et al. [5] evaluated various clustering methods within the EnergyScope TD (Typical Days) framework. However, their model was limited to a single target year, offering insights into storage systems but lacking the multi-year perspective necessary for long-term planning. On a larger temporal scale, Nahmmacher et al. [6] applied a hierarchical clustering algorithm to the LIMES (Long-term Investment Model for the Electricity Sector)-EU model [7]. Although their work provided insights into VRES variability and storage system optimization, it did not account for inter-period storage, such as energy storage systems operating across consecutive clustered days or groups of days. This limitation restricts the ability to accurately model long-term storage interactions, particularly for systems requiring multi-period storage solutions like seasonal reservoirs.

Optimizing storage systems in these studies typically involves reducing system costs by balancing energy supply and demand efficiently. Clustering methods simplify the representation of temporal variability while retaining critical characteristics of storage operations, such as charge/discharge cycles, round-trip efficiency, and state-of-charge constraints. By reducing computational complexity, these approaches enable a detailed analysis of storage technology deployment, sizing, and utilization while maintaining feasible optimization timelines.

Despite these advancements, time series clustering is not yet widely applied in multi-year capacity expansion models. Furthermore, to the best of the authors’ knowledge, no research explicitly addresses time series clustering while incorporating inter-period storage in long-term energy system models.

To bridge this gap, the present study introduces a novel methodology integrating interconnected clustered representative days (RDs) into an open-source energy system modeling tool. This framework enhances the representation of VRES variability and simulates storage technologies with greater accuracy. The methodology optimizes storage systems by preserving inter-day and intra-day dynamics, ensuring accurate sizing and operation. The approach is applied to a reference case study involving energy storage systems under high-RES penetration levels. Comparative analysis with traditional methods highlights the advantages of interconnected clustered RDs in identifying optimal decarbonization pathways and improving storage system efficiency.

# Methodology

This methodology introduces a novel approach for modeling time-series data, focusing on the accurate representation of variable renewable energy sources (VRES) and energy storage systems over long-term planning horizons. The key innovation lies in the clustering of the original time series into representative days, which are then used to simulate the energy system behavior. In addition, the state-of-charge (SOC) of energy storage systems is dynamically calculated based on supply and demand profiles, ensuring the accurate functioning of storage systems across different time intervals.

A. Chronological Representative Days (RDs)

In this study, we address the limitations of traditional time-series representations in long-term energy system modeling by introducing a clustering-based approach. The idea is to cluster days with similar characteristics of energy demand profiles into representative days, which are then used to model the entire year. This process allows us to reduce the complexity of modeling while retaining the critical temporal variability required for accurate energy system optimization.

### Clustering Method: K-means clustering is employed to group days based on the electrical load profile. By partitioning the year into clusters of similar days, we ensure that each representative day mirrors the temporal dynamics of the original days it represents.

### Periods: Each representative day consists of a sequence of time intervals (e.g., hourly). These time intervals are referred to as timeslices. The timeslices of representative days preserve the key features of the original time series and are used in the model to simulate the evolution of energy supply and demand.

### Chronological Sequence: In the revised framework, the chronological order of representative days is maintained over the entire year. This ensures that the temporal relationships between days, seasons, and demand-supply fluctuations are accurately represented. For instance, after clustering, the model assigns each timeslice of the year to a specific representative day, which helps track energy generation and consumption patterns across periods.

B. Optimization of Energy Storage System for Supply-Demand Balancing

This study aims to optimize the configuration of an energy storage system (ESS) to minimize unmet demand, thereby ensuring a stable supply of energy from renewable sources. The optimization process involves simulating different configurations of battery capacities, charge rates, and discharge rates over a year, while taking into account the fluctuations in energy supply and demand. The proposed methodology is divided into several key steps: defining system parameters, calculating the charge/discharge cycle, optimizing the configuration, and evaluating the system's performance.

### 1) System Parameters and Assumptions: The simulation models a time horizon of one year, divided into daily time steps, with predefined system parameters and assumptions. The initial State of Charge (SOC) starts at a specified value and is constrained between a minimum and maximum limit, where the maximum represents the energy storage system's (ESS) capacity. Multiple energy storage capacities are tested to evaluate performance under different configurations. Charging and discharging rates are simulated at varying levels, scaled proportionally to a base rate to reflect different system capabilities. The model assumes fixed charging and discharging efficiencies to account for energy losses during storage and retrieval processes, ensuring realistic system behavior. These parameters collectively provide a comprehensive framework to analyze the ESS performance over the simulation period.

*2) Energy Supply and Demand:* At each time step (i.e., each day), the system receives an energy supply from renewable sources (and faces an energy demand from the consumers. The balance between supply and demand determines whether energy is stored or retrieved from the ESS:

Surplus Energy (Supply > Demand): When the supply exceeds demand, the surplus energy is stored in the ESS. The amount of energy stored is limited by the available surplus and the charge rate of the system.

Deficit Energy (Supply < Demand): When the supply falls short of demand, the ESS discharges energy to meet the deficit. The amount of energy discharged is limited by the available storage and the discharge rate of the system*.*

### Charge/Discharge Dynamics: The energy flow between the ESS and the grid (or consumers) is modeled using the following logic:

### If there is surplus energy (i.e., net energy > 0), the battery charges. The effective charge is calculated based on the charging efficiency and is capped by the maximum charging rate.

### If there is a deficit energy (i.e., net energy < 0), the battery discharges. The effective discharge is adjusted for the discharging efficiency and is capped by the maximum discharge rate.

### At each time step, the State of Charge (SOC) is updated based on the energy flow (charge or discharge) and the charging/discharging characteristics of the ESS.

### The SOC at time step t is calculated as follows:

### SOCt=SOCt-1+Effective Charge−Effective Discharge

### Where:

### Effective Charge = min (Surplus Energy × Charging Efficiency, Charge Rate)

### Effective Discharge= min (Deficit Energy × Discharging Efficiency, Discharge Rate)

### Unmet Demand Calculation: Unmet demand is calculated when the supply is insufficient to meet the demand at a given time step. The unmet demand for each day is given by:

### Unmet Demandt = max (Demandt −Supplyt, 0)

### The total unmet demand over the entire simulation period is the sum of daily unmet demands:

Total Unmet Demand =

*5) Optimization Process:* The simulation runs for each combination of battery capacity, charge rate, and discharge rate. For each configuration, the SOC is updated at each time step using the charge/discharge function defined earlier. After simulating the energy storage dynamics over the entire year (365 days), the total unmet demand is calculated for each configuration. The configuration that results in the minimum total unmet demand is considered the optimal configuration. The corresponding battery capacity, charge rate, and discharge rate are identified as the best-performing setup.

By clustering the original time series into representative days and maintaining their chronological order, the model captures the critical features of energy supply and demand variations throughout the year. The state-of-charge of storage systems is calculated dynamically, ensuring that the storage facilities are optimally utilized in response to the evolving energy balance over time. This novel approach improves the modeling of VRES variability and storage systems, offering more accurate long-term energy system planning for decarbonization pathways.

# Result and Discussion

The dataset, used here, provides hourly time-series data on electricity consumption and production over a span of more than five years. It includes detailed information on total electricity consumption and production, with production categorized by source: nuclear, wind, hydroelectric, oil and gas, coal, solar, and biomass. All values are recorded in megawatts (MW), with timestamps marking each hour. This extensive dataset is ideal for analyzing trends like seasonal variations, renewable energy contributions, and fossil fuel dependency. It also supports innovation in energy technologies, such as optimizing storage systems, and enables detailed statistical analyses and forecasting for improved energy production and consumption planning.

The clustering of representative days using the k-means method was carried out successfully with different cluster sizes (6, 12, 24, and 36). The clustering approach helped to reduce the complexity of daily profiles while retaining key features of energy supply and demand. To assess the quality of the clustering, we computed the Silhouette Score, which measures how well each data point fits within its assigned cluster. Higher silhouette scores indicate that the clustering is effective in capturing the inherent patterns of the data. The silhouette score increased with the number of clusters, indicating that finer granularity in the clustering process allowed for a better fit between representative days and the actual time series.

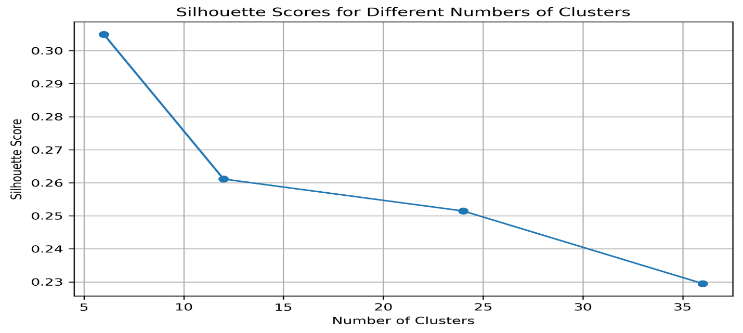


Fig. 1. Silhouette score for different numbers of clusters

The cluster centers represented the average daily profiles for each cluster. The representative days derived from these cluster centers closely matched the typical daily load. After evaluating the silhouette scores for various cluster sizes, it was found that using 6 clusters provided a good balance.

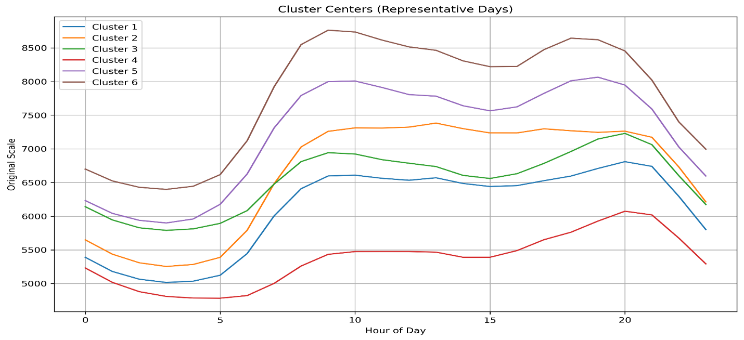


Fig. 2. Cluster center (representative days)

After clustering the days, we mapped each representative day to the model's time structure, consisting of timeslices and time periods. Each timeslice represents a fixed time interval (e.g., 6-hour blocks) and is mapped to one of the representative days identified in the clustering phase. This ensures that each timeslice accurately reflects the temporal variability of demand patterns. By clustering the days, we were able to assign each timeslice with a day profile that matches its energy generation and demand characteristics.The model maintained the chronological sequence of days, ensuring that energy demand and supply behaviors evolved throughout the year. Each representative day was mapped sequentially to preserve the natural progression of seasons and weather patterns, such as the transition from summer to winter. This helped to model long-term system behavior more accurately, as it accounted for shifts in renewable energy generation and demand in a time-ordered manner.

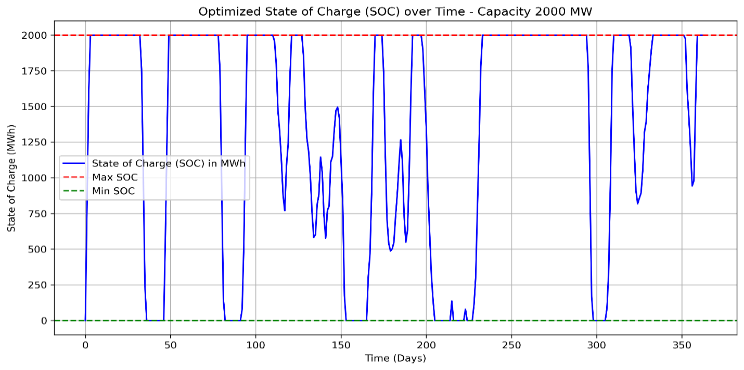


Fig. 3. Optimized state of charge over time

The state-of-charge (SOC) of the storage system was calculated over time, based on the energy imbalance (supply-demand) within each timeslice. The simulation was conducted with storage capacities, charge rates, and discharge rates defined for the energy storage system.

The SOC evolved over time as the system was charged or discharged based on the difference between supply and demand in each timeslice. When demand exceeded supply, the storage system discharged energy to meet the deficit, and when supply exceeded demand, energy was stored. The SOC was updated at the beginning of each timeslice, and the energy balance was tracked accordingly.

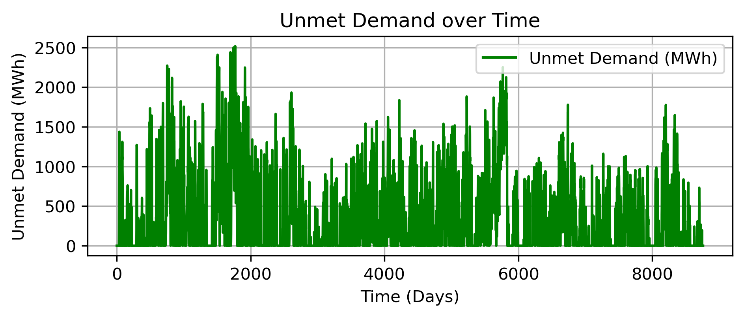


Fig. 4. Unmet demand over time

The analysis highlighted that larger storage capacities effectively reduce reliance on fossil fuels during high-demand periods, particularly in systems with significant renewable energy integration. Additionally, enhancing discharge rates during peak demand was shown to minimize the risk of unmet demand, ensuring optimal performance of ESS during periods of low renewable generation. The results also emphasized that high unmet demand indicates the need to increase storage capacity or improve charge/discharge rates. In some cases, demand-side management strategies, such as demand response programs, could complement ESS optimization by reducing peak demand and maximizing the utilization of stored energy.

The simulation evaluated the ESS performance over a one-year time horizon, during which the total annual energy supply was 55,204.163 GWh, while demand was 56,292.449 GWh. The optimal configuration identified for the ESS featured a storage capacity of 2000 MW, with charging and discharging rates of 800.0 MWh/day each. This setup reduced the total unmet demand to 3,203.008.19 GWh, demonstrating the ESS's capability to mitigate supply-demand imbalances within the defined system parameters.

# **Conclusion**

This study presents a novel approach for modeling renewable energy systems and storage behavior using a clustering-based methodology and dynamic SOC simulation. The k-means clustering of representative days successfully reduced the complexity of time-series data while retaining critical information on energy generation and demand patterns. The revised time framework mapped these representative days to timeperiods and timeslices, preserving chronological order and ensuring realistic simulations. The SOC calculations demonstrated the effective use of energy storage systems in balancing supply and demand, providing insights into system optimization and real-world applications. The simulation demonstrated that an ESS with a 2000 MW capacity and 800.0 MWh/day charge/discharge rates reduced unmet demand to 3,203.008 GWh, balancing annual supply (55,204.163 GWh) and demand (56,292.449 GWh). Future work could focus on refining clustering techniques and incorporating more detailed energy storage characteristics.

REFERENCES

1. R. Novo, P. Marocco, G. Giorgi, A. Lanzini, M. Santarelli, and G. Mattiazzo, “Planning the decarbonisation of energy systems: The importance of applying time series clustering to long-term models,” *Energy Conversion and Management: X*, vol. 15, p. 100274, Aug. 2022, doi: 10.1016/j.ecmx.2022.100274.
2. Gabrielli P, Gazzani M, Martelli E, Mazzotti M. Optimal design of multi-energy systems with seasonal storage. Appl. Energy 2018;219(October 2017):408–24. https://doi.org/10.1016/j.apenergy.2017.07.142.
3. Kotzur L, Markewitz P, Robinius M, Stolten D. Time series aggregation for energy system design: Modeling seasonal storage. Appl Energy 2018;213(January): 123–35. <https://doi.org/10.1016/j.apenergy.2018.01.023>.
4. Welder L, Ryberg DS, Kotzur L, Grube T, Robinius M, Stolten D. Spatio-temporal optimization of a future energy system for power-to-hydrogen applications in Germany. Energy 2018;158:1130–49. https://doi.org/10.1016/j. energy.2018.05.059.
5. Limpens G, Moret S, Jeanmart H, Mar´echal F. EnergyScope TD: A novel open source model for regional energy systems. Appl Energy 2019;255(July):113729. <https://doi.org/10.1016/j.apenergy.2019.113729>.
6. Nahmmacher P, Schmid E, Hirth L, Knopf B. Carpe diem: A novel approach to select representative days for long-term power system modeling. Energy 2016;112: 430–42. <https://doi.org/10.1016/j.energy.2016.06.081>.
7. P. Nahmmacher, E. Schmid, and B. Knopf, “Documentation of LIMES-EU - A longterm electricity system model for Europe,” 2014. [Online]. Available: https://www .pik-potsdam.de/en/institute/departments/transformation-pathways/models/lime s/DocumentationLIMESEU\_2014.pdf.