POLITECNICO DI TORINO

INGEGNERIA ENERGETICA E NUCLEARE

A.A. 2023/2024

Sessione di Laurea Luglio 2024

Long-term energy system modelling: the impact of different time-series clustering algorithms

Supervisors

Prof. Giuliana MATTIAZZO

Dott. Paolo MAROCCO

Dott. Riccardo NOVO

Ing. Caterina CARÁ

Candidate

Tanooj JAGADEEP

Contents

FIGURES 3
TABLES 4
ABSTRACT 5
ACKNOWLEDGEMENT 6
1. INTRODUCTION 7
1.1 European Energy Scenarios and Initiatives 7
1.2 Energy and Scenario Modelling 7
1.3 Thesis Roadmap 9
2. THE ENERGY SYSTEM OF FAVIGNANA 9
2.1 Island of Favignana 9
2.2 Environmental regulations 10
2.3 Energy System of Favignana 11
3. ENERGY MODELLING FRAMEWORK 14
3.1 Osemosys Framework 14
3.2 Clustering algorithms 16
4. MODEL DEFINITION 20
4.1 Reference Energy System 20
4.2 Sets 21
4.3 Parameters 23
4.4 Scenario definition 27
5. MODEL AND SCENARIO DEVELOPMENT 27
6. RESULTS 37
6.1 Computational Time 38
6.2 Net Present Cost 42
6.3 Relative error in total system cost 46
6.4 Relative error in sizing of components 65
7. CONCLUSION 80
REFERENCES 82

FIGURES

Figure 1 Island of Favignana 10
Figure 2 Electricity demand for Favignana... 12
Figure 3 Solar Radiation for rooftop mounted PV.. 13
Figure 4 Solar Radiation for ground mounted PV.. 13
Figure 5 Suitable wind farm area for Favignana.. 14
Figure 6 OSeMOSYS workflow 16
Figure 7 Kmeans representation 18
Figure 8 Difference between Kmeans and Kmedoids 19
Figure 9 Reference Energy System of Favignana 20
Figure 10 Time Period representation with time brackets........................... 22
Figure 11 Time Series Configuration settings.. 28
Figure 12 Scenario Setting Configuration Settings 29
Figure 13 Demand Evolution Configuration Setting 30
Figure 14 Sets sheet of Favignana... 31
Figure 15 Parameters sheet of Favignana... 32
Figure 16 Structure of the code.. 32
Figure 17 Python terminal for Main_batch.py while simulation 34
Figure 18 Simulation Flowchart 35
Figure 19 json to excel code. 36
Figure 20 Layout of POWER BI.. 36
Figure 21 Flowchart of graphical representation.. 37
Figure 22 Computational Time-Kmeans 39
Figure 23 Computational Time-Kmedians 40
Figure 24 Computational Time-Kmedoids 41
Figure 25 Net Present Cost-kmeans.. 43
Figure 26 Net Present Cost-kmedians... 44
Figure 27 Net Present Cost-kmedoids 45
Figure 28 Relative Error in total system cost kmeans................................. 46
Figure 29 Relative Error in total system cost kmedians 47
Figure 30 Relative Error in total system kmedoids 48
Figure 31 Sizing of PV Kmeans 50
Figure 32 Sizing of PV Kmedians 50
Figure 33 Sizing of PV Kmedoids.. 51
Figure 34 Figure 32 Sizing of Wind Kmeans... 52
Figure 35 Sizing of Wind Kmedians 53
Figure 36 Sizing of Wind Kmedoids.. 53
Figure 37 Sizing of Battery Technology Kmeans 55
Figure 38 Sizing of Battery Technology Kmedians 55
Figure 39 Sizing of Battery Technology Kmedoids 56
Figure 40 Sizing of Battery Storage Kmeans 57
Figure 41 Sizing of Battery Storage Kmedians 57
Figure 42 Sizing of Battery Storage Kmedoids 58
Figure 43 Sizing of Fuel Cell Kmeans 59

Figure 44 Sizing of Fuel Cell Kmedians 60
Figure 45 Sizing of Fuel Cell Kmedoids 60
Figure 46 Sizing of Electrolyser Kmeans.. 62
Figure 47 Sizing of Electrolyser Kmedians 62
Figure 48 Sizing of Electrolyser Kmedoids... 63
Figure 49 Sizing of Hydrogen Storage Kmeans... 64
Figure 50 Sizing of Hydrogen Storage Kmedians 65
Figure 51 Sizing of Hydrogen Storage Kmedoids.. 65
Figure 52 Error in sizing of Photovoltaics Kmeans... 67
Figure 53 Error in sizing of Photovoltaics Kmedians 67
Figure 54 Error in sizing of Photovoltaics Kmedoids...................................... 68
Figure 55 Error in sizing Wind Power Kmeans 69
Figure 56 Error in sizing Wind Power Kmedians 69
Figure 57 Error in sizing Wind Power Kmedoids... 70
Figure 58 Error in sizing Battery Technology Kmeans 71
Figure 59 Error in sizing Battery Technology Kmedians 71
Figure 60 Error in sizing Battery Technology Kmedoids 72
Figure 61 Error in sizing Battery Storage Kmeans 73
Figure 62 Error in sizing Battery Storage Kmedians 73
Figure 63 Error in sizing Battery Storage Kmedoids 74
Figure 64 Error in sizing Hydrogen Fuel Cell Kmeans 75
Figure 65 Error in sizing Hydrogen Fuel Cell Kmedians 75
Figure 66 Error in sizing Hydrogen Fuel Cell Kmedoids 76
Figure 67 Error in sizing Hydrogen Electrolyser Kmeans 77
Figure 68 Error in sizing Hydrogen Electrolyser Kmedians 77
Figure 69 Error in sizing Hydrogen Electrolyser Kmedoids............................. 78
Figure 70 Error in sizing Hydrogen Storage Kmeans...................................... 79
Figure 71 Error in sizing Hydrogen Storage Kmedians.................................... 79
Figure 72 Error in sizing Hydrogen Storage Kmedoids.................................... 80

TABLES

Table 1 Time Slices 22
Table 2 Availability Factor 24
Table 3 Operational Life... 25
Table 4 Costs 25
Table 5 Scenarios 27

ABSTRACT

Climate change poses severe challenges to human and environmental the environment around them. Therefore, it is necessary to act in the energy domain especially as a significant source for climate change is greenhouse gas emissions. Effective policy measures are imperative for addressing this critical issue. A key strategy is the transition from conventional fossil fuels to renewable energy sources. This strategy involves a fundamental shift from carbon-based electricity to a diversified energy portfolio predominantly constituted of renewables, complemented by a minor share of gas.

The thesis at hand aims to conduct extensive energy system modelling for the island of Favignana over three different scenarios, identifying the most efficient clustering algorithm to reduce computational demands while maintaining precise outcomes. Utilizing an advanced iteration of the Open-Source energy Modelling System (OSeMOSYS), the energy systems are constructed through the clustering method applied to time-series data. This approach uses representative days (RDs) for various years, considering critical attributes on both demand and supply fronts. These RDs are to be clustered using diverse algorithms, with the objective of pinpointing the optimal one.

The scenarios developed characterize distinct operational conditions of the island's energy systems, which are sole utilization of photovoltaic systems, sole dependence on wind energy, and a mixture of both, alongside the incorporation of various storage technologies such as lithium-ion batteries and hydrogen storage. The findings underscore that an aggressive decarbonization strategy is not only viable but also advantageous, and that different clustering algorithms exhibit varying degrees of suitability across different scenarios.

ACKNOWLEDGEMENT

लोकाः समस्ताः सुखिनो भवन्तु।।

“Che tutti gli esseri in tutti i mondi siano felici e in pace”

I would like to thank my supervisor Giuliana Mattiazzo who gave me the opportunity to undertake this work, which has allowed me to deepen my knowledge and learn new concepts.

Then I would also like to thank Paolo Marocco, Riccardo Novo and Caterina Cara who guided me with their expertise and improved my skills during this period.

Finally, I would like to express my gratitude to my family, friends, and the universe for their love and support, which have played a crucial role in my success thus far.

1. INTRODUCTION

1.1 European Energy Scenarios and Initiatives

As of 2023, the energy scenario in Europe reflects a significant shift towards renewable energy, alongside a diversified energy mix. Oil and petroleum products remain the largest contributors to the EU's gross available energy at 34.5%, followed by natural gas (23.7%) and solid fossil fuels (10.2%). Collectively, fossil fuels constitute 68.4% of the EU's energy production. However, the share of renewable energy has notably increased, reaching 19% of the EU's energy mix in 2021, underscoring the region's commitment to sustainable energy [1][2]. In terms of electricity generation, renewables accounted for 39% in 2022, with wind and solar surpassing fossil fuels for the first time in EU electricity generation during May of that year. This progress is aligned with the EU's strategic initiatives such as the European Green Deal and its efforts in energy diversification and climate action [3]. The adoption of renewable energy varies across EU member states, with Sweden, Finland, and Denmark leading in renewable energy shares, whereas other countries like Malta and Belgium have lower shares. The final energy consumption in the EU in 2021 was marked by a decrease in the use of solid fossil fuels and an increase in renewable energy sources, reaching 11.8% of the total energy consumption [2].

The European Union (EU) has consolidated its commitment to decarbonization and renewable energy through several key policies. The 2030 Climate & Energy Framework, adopted in 2014, targets a 40% reduction in greenhouse gas emissions by 2030, at least a 27% share of renewable energy, and a 27% increase in energy efficiency [4]. The European Green Deal, launched in 2019, aims to make the EU climate-neutral by 2050, involving a comprehensive range of policy initiatives [5]. The greenhouse gas emissions are targeted by the EU policies to be reduced to 55% according to the Fit for 55 Package by 2030[6]. The Renewable Energy Directive (RED II), revised in 2018, sets binding renewable energy targets, aiming for at least 32% renewable energy by 2030 [7]. Additionally, the Energy Efficiency Directive (EED), updated in 2018, seeks a 32.5% improvement in energy efficiency by 2030 [8]. These frameworks collectively signify the EU's strategy towards a sustainable and climate-resilient future.

1.2 Energy and Scenario Modelling

Energy and scenario modelling have become essential tools in shaping our understanding and responses to the energy challenges of the future. One of the prominent tools in this field is the Integrated Framework of the International Energy Agency's (IEA) Global Energy and Climate Model (GEC Model). This model is a principal tool for generating detailed, sector-by-sector, and region-by-region long-term scenarios. It incorporates both the World Energy Model (WEM), a large-scale simulation model designed to replicate how energy market's function, and the Energy Technology Perspectives (ETP) model, a technology-rich bottom-up model. This hybrid approach, combining the strengths of both models, is used in the IEA's comprehensive studies, including their Net Zero Roadmap and the World Energy Outlook series [9]. The World Energy Outlook 2023, for instance, explores three fully updated scenarios that provide a framework for examining the implications of various policy choices, investment, and technology trends. These scenarios include the Stated Policies Scenario, the Announced Pledges Scenario, and the Net Zero Emissions by 2050 Scenario. Each scenario considers different assumptions about global economic and population growth, energy, carbon and mineral prices, and potential for volatility [10]. Apart from traditional modelling frameworks, machine learning is increasingly being recognized for its potential in energy projections. With today's advancement in machine learning information and data from already occurred scenarios derived from scenario tools can be used to predict and make projections for the future. It's particularly effective in handling a variety of drivers that shape the evolution of energy systems, providing more refined 'what if' analyses and projections, especially over the near term. This approach allows for a more nuanced understanding of the real-world phenomena that long-term scenario modelling frameworks might overlook [11].

There are two main approaches to energy modelling, top-down and bottom-up. Energy modelling is an essential tool for comprehending and controlling energy systems. Top-down models provide a macroeconomic viewpoint by concentrating on how more general economic issues affect energy supply and demand. They are usually based on economic concepts. Bottom-up models, on the other hand, place more emphasis on technical details and provide specifics about each energy technology and facility, such as operating costs and efficiency curves. These models better reflect the technical and physical characteristics of energy systems. Although integrating these two methods into hybrid models presents difficulties, it also provides more thorough insights. Prominent models in this domain, each with a distinct focus and methodology, include EnergyPLAN, TIMES and OSeMOSYS [12]. Energy planning tools commonly employ linear programming (LP) or mixed integer linear programming (MILP) methods for resolution. Models like TIMES and OSeMOSYS utilize an approach that spans multiple years, a strategy that effectively addresses limitations inherent in single-year models [13][14]. Single-year frameworks often fail to capture key dynamics such as rising energy demand over time, the evolving behaviour of renewable energy sources (RESs), and technological cost variations. In contrast, multi-year capacity expansion models require an approximation of time series to manage computational demands. These models typically divide each year into distinct segments based on seasonal, daily, and hourly variations [15]. Increasing the number of time slices improves the accuracy of the time series representation, but at the expense of a more complex problem. The correct representation of time series is important for scenarios with high-RES penetration and energy storage involvement [16]. Use of representative days (RDs) are beneficial while considering a multi-year modelling approach as they reduce the number of time slices and computational time and still give good results [17]. The interconnection of RDs is necessary to accurately model multi-energy systems with seasonal storage [18]. In order to couple the RDs different clustering algorithms can be used like K-means, K-medoids, K-medians, K-centres etc. K-medoids approximates the demand related costs the best while building energy systems when compared to the other clustering algorithms [19].

1.3 Thesis Roadmap

The current situation in Europe regarding the energy Renewable Energy Sources (RES) share, as well as the initiatives undertaken to augment them and advance towards decarbonization, are comprehensively discussed in Chapter 1. This chapter also provides a succinct introduction to energy and scenario modelling tools, elucidating their functionalities. Chapter 2 delves into the diverse parameters and objectives employed in developing the reference energy system for Favignana, emphasizing the island's existing energy system and its renewable energy potential. In Chapter 3, the framework of the energy modelling tool Osemosys and the functioning of the clustering algorithms is discussed. Chapter 4 discusses about the parameters considered in the conceptualization of the reference energy system model. Finally, Chapters 5 and 6 shows the development of the model and various scenarios and elaborates upon the results, respectively.

2. THE ENERGY SYSTEM OF FAVIGNANA

2.1 Island of Favignana

Favignana, the largest island within the Egadi archipelago in the Mediterranean Sea, lies off the west coast of Sicily, Italy, encompassing an area of 19.8 km². Its topography is predominantly flat, with the notable exception of a central hill that bisects the island from north to south. Favignana is entirely reliant on imported primary energy. The island's electrical energy requirements are met through a combination of diesel generators and renewable energy sources. Notably, Favignana receives an average global solar horizontal irradiation of 1,800 kWh/m²/year [20]. Additionally, the island experiences average wind speeds ranging from 7 to 10 m/s at a height of 100 meters [21]. However, due to its designation as a protected area, specific regulatory considerations apply. The surrounding waters are also designated as a marine protected area, further emphasizing the island's ecological significance.

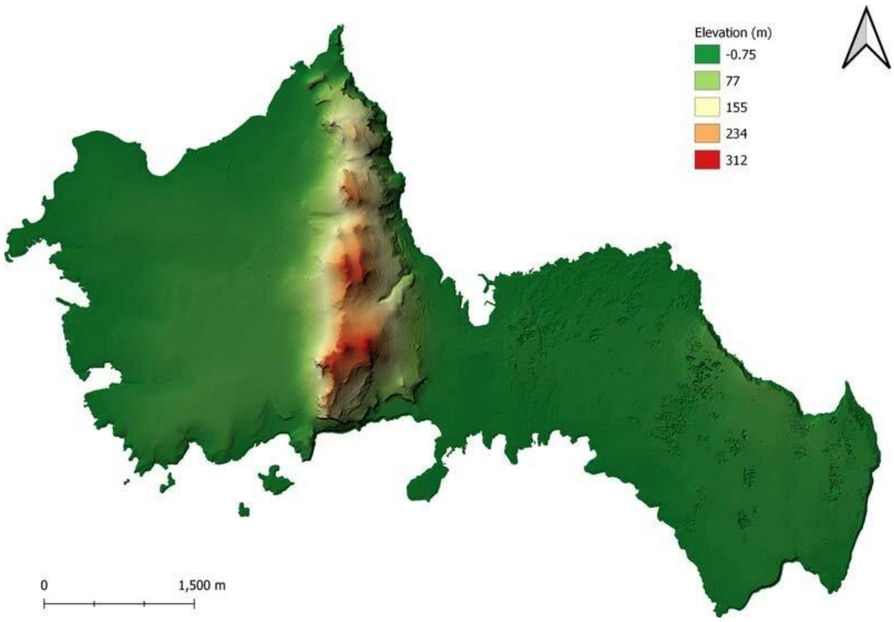


Figure 1 Island of Favignana

2.2 Environmental regulations

The Piano Energetico Ambientale della Regione Siciliana (PEARS 2030) is a comprehensive energy and environmental plan approved by the Sicilian Regional Government with Deliberation No. 67 on February 12, 2022. This plan is designed to address and direct both structural and infrastructural interventions in the energy field and serves as a reference framework for public and private entities undertaking energy initiatives. PEARS 2030 sets two main objectives: reducing energy consumption in end uses, particularly in the civil-agricultural sector and the transport sector (smart mobility) and increasing the share of renewable energy. The goal is to achieve a 68% incidence of renewable energy in the total regional energy production by 2030, compared to 33% in 2019. This implies tripling renewable energy production and halving non-renewable sources by 2030 [22]. As stipulated in the ordinance dated December 27, 1991 [23], the area was declared a special conservation zone. Subsequently, in alignment with the European Directive 92/43/ECC, it was designated as a Special Area of Conservation (SAC) [24]. Consequently, Italy is committed to protecting this area through appropriate conservation measures.

According to the Ministerial Decree of 20 October 2017 the Sicilian Region has identified the areas considered unsuitable for the installation of electricity generation systems using wind sources. The Decree categorizes wind power turbines based on their power output into three groups. Turbines with a nominal power below 20 kW are classified as EO1, those between 20 kW and 60 kW as EO2, and those with a nominal power exceeding 60 kW as EO3. In this study, the focus is on the EO3 category according to the CETA Favignana [25]. The Favignana island is affected by a decidedly extensive environmental protection regime, which takes the form of the establishment of various ZSC: ITA010004 "Isola di Favignana". Furthermore, on the Archipelago there is an area IBA (IT157) and an area ZPS (ITA010027 "Arcipelago delle Egadi – area marina e Terrestre"). Based on the environmental protection regulatory framework currently in force on the Archipelago, the installation of any class of wind turbine is prohibited. Furthermore, the stretch of sea affected by the presence of the Egadi Archipelago has been recognized as having environmental value through the establishment of the Marine Protected Area.

Through the Piano Nazionale Integrato per l'Energia e il Clima (PNIEC) of December 2019, Italy declared how an accelerated process of decarbonisation and electrification of consumption with renewable sources would be tested in some small non-interconnected islands, including Favignana. In this context, the Ministerial Decree of 02/14/2017 set specific objectives for each island to cover consumption with renewable sources. The decree aimed to promote the modernization of electricity networks to allow the use of more renewable sources and the implementation of pilot projects aimed at increasing the use of renewables, using storage systems. The Decree promotes the installation of RES plants by private individuals, remunerating the production and self-consumption of electricity with tariff incentives dedicated to the smaller non-interconnected islands. Certain regulations were ignored because the primary objective of decarbonization is to eliminate carbon emissions, a goal achievable only with sufficient renewable energy resources. Therefore, any existing regulation that may hinder this goal, whether now or in the foreseeable future, should be re-evaluated for potential modification. The analysis conducted for the for the island of favignana considers the fluctuating potential of renewable energy, factoring in both current and potentially more rigorous environmental regulations.

2.3 Energy System of Favignana

The island of Favignana is not connected to the national grid and therefore the electricity demands of Favignana is met by Società Elettrica di Favignana S.p.A. with a diesel power plant of 12 MW installed capacity [26] and photovoltaics with an installed capacity of 360.87kW as of June 2021.

The cumulative energy demand of Favignana for the year 2019 was 14.865GWh. This demand is divided into three categories according to the bill tables.

a) Residential Demand which accounted for 3.862GWh.
b) Non-Residential Demand accounted for 1.864GWh.
c) Activities and Services demand accounted for 9.139GWh.

There is clear spike in electricity demand during the summer season due to use of air conditioners and tourism as shown in Fig.2.

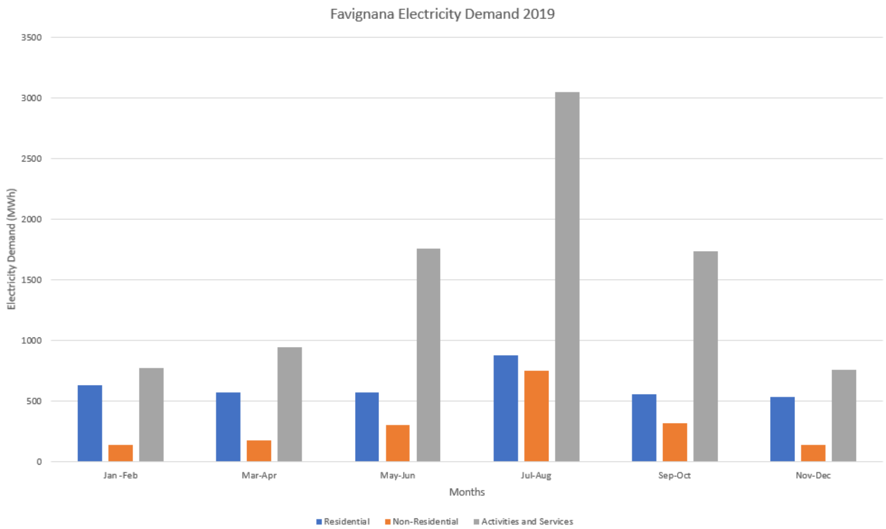


Figure 2 Electricity demand for Favignana

Potential for future photovoltaics and wind turbine installation

According to CETA Favignana [25], the suitable rooftop area available for photovoltaics is 303,734 m^2 which can accommodate total PV panel of 85,622 m2. The total maximum installed capacity is 16.76 MW with an annual energy production of 27.51GWh/year. And there is also mention of

Figure 3 Solar Radiation for rooftop mounted PV

three separate ground mounted photovoltaic plants which can be installed with annual energy production of 12-13GWh/year.

Figure 4 Solar Radiation for ground mounted PV

The CETA also mentions a suitable area located in the western part of the island adopting the relaxation of IBA criteria for wind energy potential.

Figure 5 Suitable wind farm area for Favignana

A single Vestas V90 model with a nominal power of 3MW was considered which has an annual energy production of 8GWh/year.

Possibilities for energy storage

Two kinds of energy storage Hydrogen Storage and Lithium-Ion Battery storage are considered in this energy system.

3. ENERGY MODELLING FRAMEWORK

3.1 Osemosys Framework

The energy modelling is carried out by the software Osemosys. The Open-Source Energy Modelling System, is a tool designed for analysing energy strategy development at various levels, including local, national, and multi-regional. It operates as a deterministic, linear optimization framework for long-term modelling, focusing on optimizing the net present value cost within an energy system [27]. It is a deterministic, linear optimization, long-term modelling framework. It is also possible to apply mixed-integer linear programming for specific functions. Multiple complex iterations are needed to get to the final solution for such problems and to do these specific solvers are used. As for scenario modelling frameworks, there are several available solvers. In the simulations presented here, the primary solver utilized was IBM ILOG CPLEX©, known for its implementation of both primal and dual simplex optimization algorithms. The time frame for these simulations varies, with a time step resolution that's determined by the specific domain of application and can cover a period ranging from several years to decades. It's worth noting that the finer the time step or the longer the period analysed, the greater the computational resources required [28][29].

The aim of this analysis is to develop a robust and realistic model of the energy system, enabling accurate estimations of the system's behaviour in response to applied constraints and the introduction of new technologies.

The OSeMOSYS framework is structured around sets, parameters, and variables. Sets define the model's physical structure, including aspects like time frames (YEAR), technologies (TECHNOLOGY), time slices (TIMESLICE), fuels (FUEL), and regions (REGION). Parameters are user-defined numerical inputs that vary across scenarios, including global parameters (like Discount Rate), technology costs, storage details, and capacity constraints. Variables are the computed outputs, encompassing aspects like demands, storage, capacity, activity levels, and costs. This structure enables flexible modelling of energy systems under various scenarios and conditions.

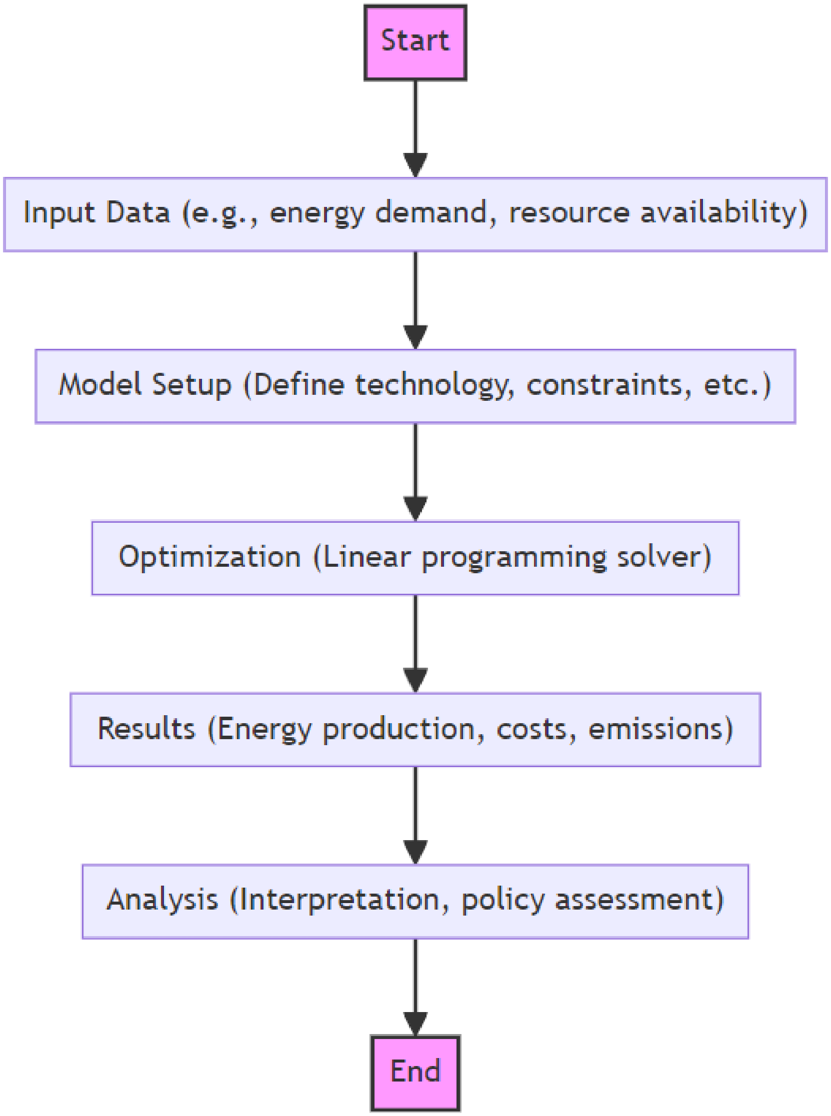


Figure 6 OSeMOSYS workflow

In the Python-based version of OSeMOSYS, the inputs for both SETS and PARAMETERS are organized within an Excel file. This approach facilitates the input management for various components of the model, enabling efficient and structured data handling for energy system analysis. As we are using the representative days approach, clustering algorithms are used to compute blocks in parameters which are dependent on time series.

3.2 Clustering algorithms

Clustering algorithms are a type of unsupervised learning technique used to group sets of objects based on their similarities and distinct features. Unlike supervised learning where the data is labelled, clustering algorithms organize data into clusters without pre-labelled responses. These algorithms are particularly useful for exploring data structure, identifying patterns, and extracting meaningful information from large datasets. Clustering algorithms offer several key benefits, particularly when dealing with large datasets.

- **Pattern Recognition:** Helps in identifying patterns and structures in large datasets that are not immediately apparent.
- **Data Simplification:** Clustering simplifies large datasets by grouping similar items, making it easier to analyse and interpret the data.
- **Anomaly Detection:** It can be used to detect anomalies or outliers in the data, which are data points that do not fit well into any cluster.
- **Feature Learning:** Clustering can assist in feature learning, where new features are derived based on the clustered groups, which can be useful for other machine learning tasks.
- **Scalability:** Many clustering algorithms are scalable to large datasets, although this can depend on the specific algorithm and its implementation.

The different clustering algorithms used in this work to compute the models are K-means, K-medians and K-medoids.

K-means

1. **Initialization**: The algorithm starts by randomly selecting 'k' points as the initial centroids, where 'k' is the number of clusters you want to identify.
2. **Assignment Step**: Each data point in the dataset is assigned to the nearest centroid, and thus clusters are formed. The distance is usually calculated using Euclidean distance, but other distance measures can also be used.
3. **Update Step**: Once all data points are assigned to clusters, the centroids of these clusters are recalculated. This is typically done by taking the mean of all points in each cluster.
4. **Iteration**: Steps 2 and 3 are repeated iteratively until the centroids no longer change significantly, which suggests that the clusters are as good as they can be given the current dataset.
5. **Convergence**: The algorithm stops when the centroids have stabilized, and the final clusters are defined.

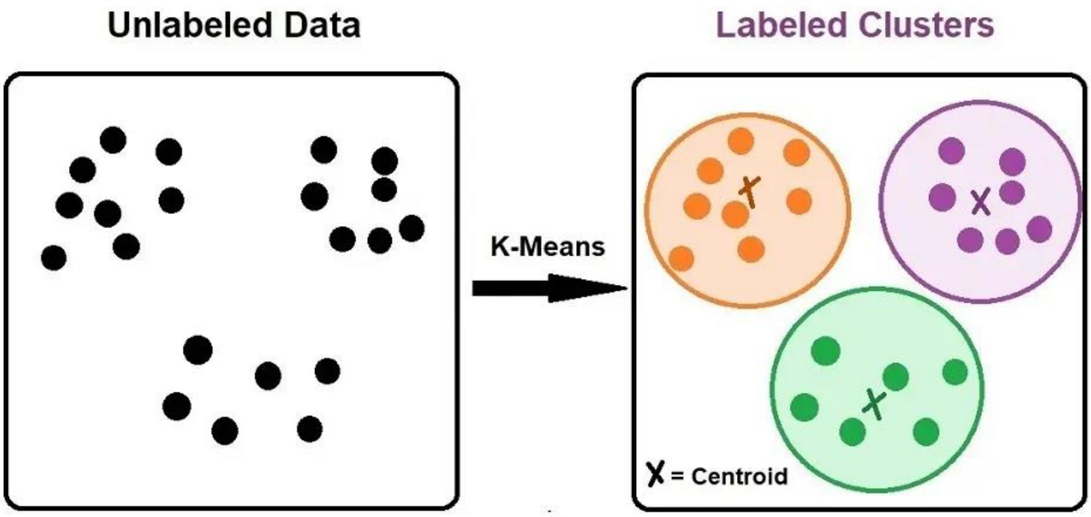


Figure 7 Kmeans representation

K-medians

1. **Initialization**: Similar to K-means, K-medians begins by selecting 'k' initial points as the centroids of the clusters. These points can be chosen randomly or through a more sophisticated method.
2. **Assignment Step**: Each data point in the dataset is assigned to the cluster whose median is closest to it. The distance is usually calculated using a metric like the Euclidean or Manhattan distance, though the latter is more common in K-medians due to its alignment with the median concept.
3. **Update Step**: The centroid of each cluster is updated instantly after the points are assigned to the clusters. In K-medians, this update is done by calculating the median of the data points in each cluster, rather than the mean (as in K-means). This involves finding the middle value (or the average of the two middle values if there is an even number of points) for each dimension of the data points in the cluster.
4. **Iteration**: The assignment and update steps are repeated iteratively. In each iteration, data points may be reassigned to different clusters based on the updated medians, and then the medians are recalculated.
5. **Convergence**: The process continues until a stable state is reached where there are no (or minimal) changes in the cluster assignments or median positions between successive iterations.

K-medoids

1. **Initialization**: The process starts by selecting 'k' representative objects from the dataset as the initial medoids. These are typically chosen

randomly, although there are more sophisticated methods for initial selection.

2. **Assignment Step**: Each data point in the dataset is assigned to the nearest medoid. The distance between data points and medoids is typically calculated using a distance metric like Euclidean or Manhattan distance. This step is similar to the assignment step in K-means and K-medians.

3. **Update Step**: Unlike K-means, which recalculates centroids, or K-medians, which recalculates medians, K-medoids involves finding a new medoid for each cluster. This is done by selecting a point within the cluster that minimizes the total distance to all other points in that cluster. In simple terms, the most centrally located point within each cluster is chosen as the new medoid.

4. **Iteration**: The algorithm iteratively performs the assignment and update steps. In each iteration, data points may be reassigned to different clusters based on the newly determined medoids, and then the medoids are updated.

5. **Convergence**: The process continues until the medoids no longer change, indicating that the clusters are as stable as possible given the current dataset.

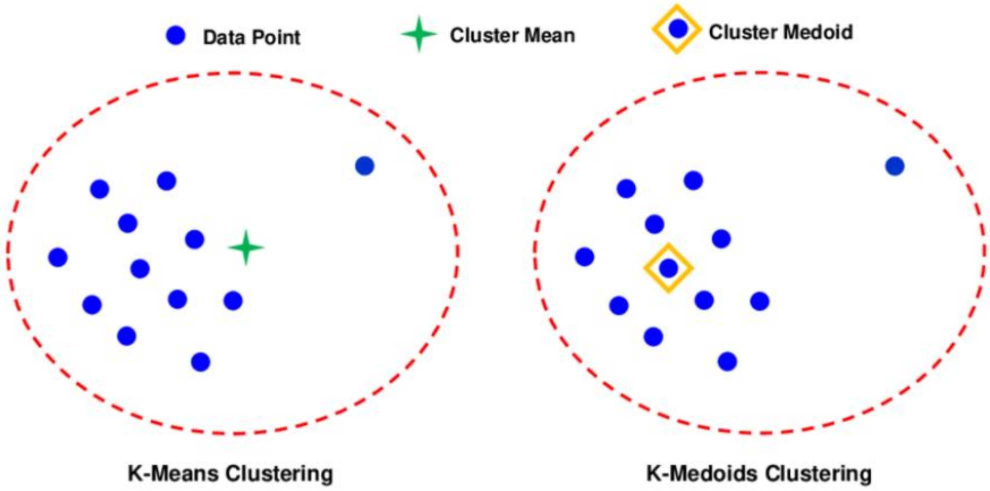


Figure 8 Difference between Kmeans and Kmedoids

Days with similar time series data for a year of wind and solar capacity factors and electrical load profiles are grouped together using clustering algorithms. These days are referred to as representative days RDs. A sensitivity analysis was made to determine the number of representative days that can be used for approximation of the final solution. In this model the comparison of different number of representative days was made with 144 days instead of 365 representative days due to lack of computational power.

4. MODEL DEFINITION

4.1 Reference Energy System

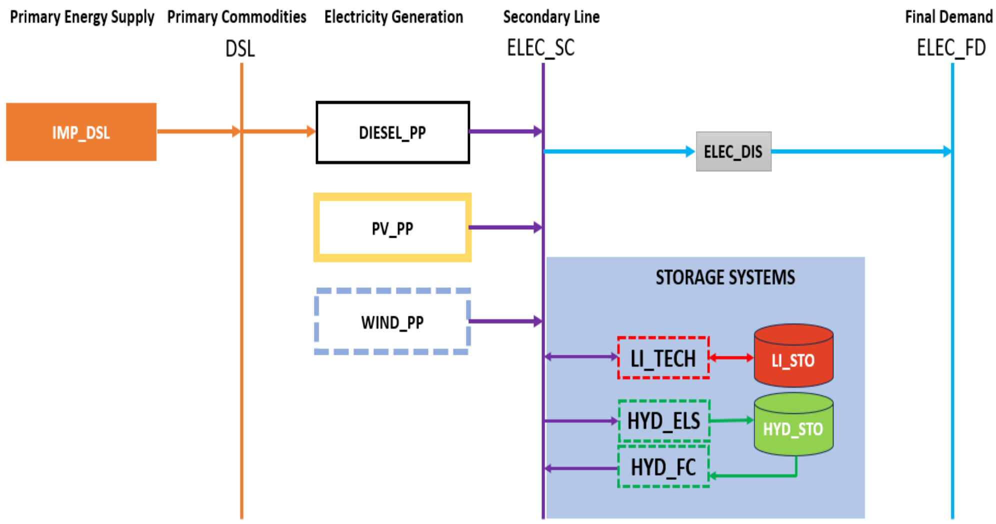


Figure 9 Reference Energy System of Favignana

Diesel Power Plant

The operational life of the diesel power plant is assumed to be 20 years [30]. The efficiency of the diesel power plant is assumed to be 45% [31]. The Input Activity ratio of the plant is taken as 2.85. The availability factor of the diesel power plant 0.91 [32].

Wind Turbine

Vestas V90 turbine is used in this model for the island of Favignana. The operational life of the wind turbine is set to be 20 years. The hourly capacity factor of the wind turbine when placed in Favignana is taken from the Renewables Ninja Website [33]. The availability factor for the PV is 0.95 [34].

Photovoltaic Power Plant

The operational life of the PV is 25 years [35]. The hourly capacity factor of the PV when placed in Favignana is taken from the Renewables Ninja

Website. The availability factor for the PV is 0.95 [36]. The solar panels chosen for the models are of monocrystalline type.

Lithium-ion Battery Storage

The operational life of Li-ion battery storage is 10 years [37] and the operational life of the Li-ion technology is 30 years [38]. The availability factor of the Li-ion technology is 0.92 [39]. With a round trip efficiency of 86% Li-ion batteries is an easy to install storage system. The input activity ratio of the Li-ion technology is 1.17 [40]

Hydrogen Storage System

The operational life of the hydrogen storage is 18 years, and the operational life of the electrolyser and fuel cell are 30 and 10 years respectively [41]. The availability factor of the hydrogen electrolyser and the fuel cell are both 0.95 [42]. The input activity ratio of the hydrogen electrolyser is 1.36 [43].

4.2 Sets

Region

It sets the regions to be modelled. In this model the region is defined as FAVIGNANA.

Emission

This section is used to contain all kinds of emissions that will be produced during the operation of different technologies. Carbon -dioxide is the only pollutant considered in this model for ease of the simulation.

Fuel

This section is used to incorporate various technologies associated with fuel like energy vectors, energy services or proxies entering or exiting technologies. In this model the Fuels are Diesel (DSL), Electricity second line (ELEC_SC) and Electricity Final Demand (ELEC_FD).

Time Period

It gives indication of how many parts the day is split into, and in which order these parts are sorted. In this model the number of time brackets considered are 5 and the length of the brackets are in the ratio 6;4;4;4;6. Therefore the timeperiod is number of time brackets multiplied by number of days in a year. In this case the timeperiod is equal to 1825.

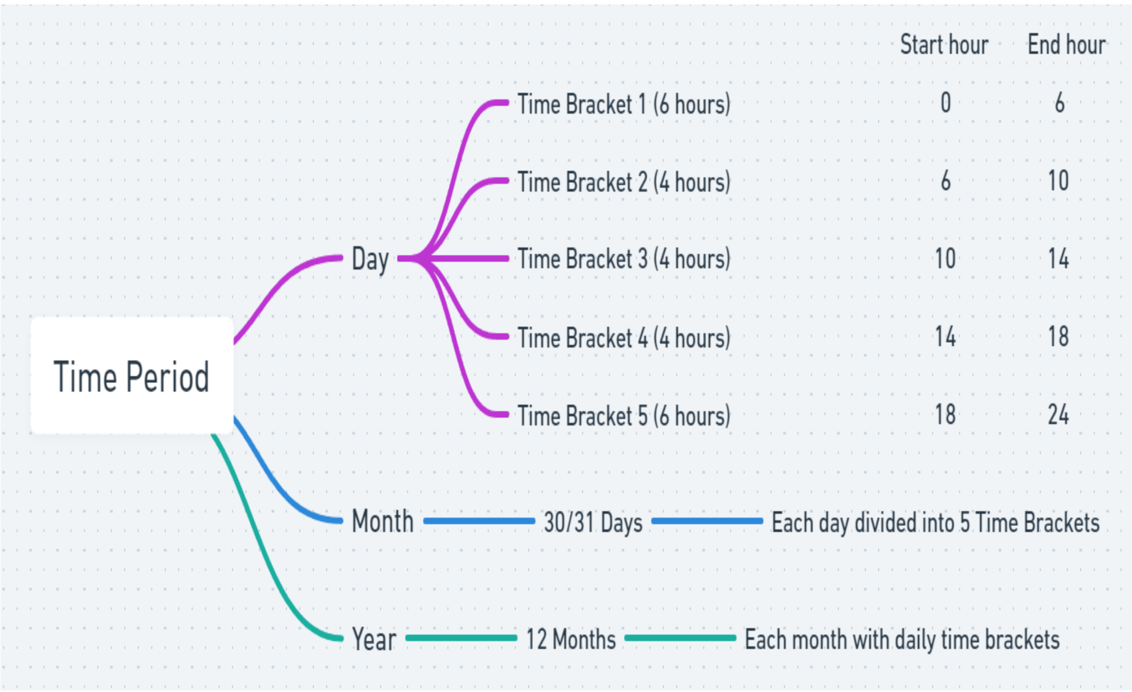


Figure 10 Time Period representation with time brackets.

Time Slice

It is used to represent the time split of each modelled year. The time slice is the product of the number of representative days and number of time brackets. In this model the different number of representative days are considered, therefore the time slice all varies.

No: of time brackets	No: of RDs	No: of time slices
5	6	30
5	12	60
5	24	120
5	36	180
5	48	240
5	72	360
5	144	720

Table 1 Time Slices

Mode of Operation

The number of modes with which a technology functions is defined by the mode of operation. Generally, two modes of operation are assigned to the storage system when the technology is connected. In this case for Li-ion technology there are two modes in which it works “charging” and “discharging”.

Storage

It includes storage facilities in the model. In this model the storages involved are Lithium-ion battery storage (LI_STO) and hydrogen storage (HYD_STO).

Technology

It includes all the subsystems of the energy system whose objective is to change a commodity from one form to another, usage of the commodity or supply of the commodity itself. In this model the different technologies are for import of fossil fuels, electricity generation, conversion, and distribution.

a) IMP_DSL which defines the import of diesel for the island of Favignana. There are no other fossil fuels used in the model of Favignana.

b) DIESEL_PP is used to refer the diesel power plant being used in Favignana.

c) PV_PP is used to refer the photovoltaic power plant setup.

d) WIND_PP is used for the wind turbine setup in Favignana.

e) LI_TECH is used to define the rectifiers and inverters used for the LI-ion battery technology.

f) HYD_ELS is used to refer to the hydrogen electrolyser used in the production of hydrogen.

g) HYD_FC is used for the fuel cells and inverters which are involved in the conversion of hydrogen to electricity.

h) ELEC_DIS is used to define the distribution line which is responsible to supply electricity from the generators to the users.

4.3 Parameters

The parameters are divided into different categories, and they are global parameters, Demands, Performance, Technology costs, Storage, Capacity constraints, Activity constraints, Reserve margin, RE generation target and Emissions.

Global parameters

a) The YearSplit is the length of a time slice when represented as a fraction of a year. It has been calculated by dividing the length of each time slice in hours by the total number of hours in a year.

b) The DaySplit is the length of a single time bracket of a day as a fraction of the year. It is calculated by dividing the length of the time bracket by the number of hours in a year.

c) The DiscountRate for the region is set as 4% for the region of Favignana [44]. The depreciation method is set to sinking fund depreciation.

Demands

The energy demand data seen in section 2.3 is used for this parameter. The specific demand profile and the accumulated annual demand are calculated with the time series data that we have for a year for the specific annual demand. The specific annual demand is considered to increase 1.5% per year as considering an increase in electrification and tourism.

Performance

a) The CapacityToActivityUnit is the factor for used for calculating the amount of energy that would be produced when one unit of capacity which is equal to 1GW is fully used in one year which is 24h*365=8760h/year. Therefore, the activity in one year will be equal to 8760GWh/year. Since we are considering all calculations GWh, therefore the conversion factor is 1. So, except for IMP_DSL and ELEC_DIS for all the other technologies the value for this parameter will be 8760 and for these two the values will be equal to 1.

b) The CapacityFactor will give insights of the capacity available for each time slice and it's expressed as a fraction of the total installed capacity. As mentioned earlier the capacity factor for PV and Wind are retrieved from renewable ninjas [33].

c) The AvailabilityFactor is the maximum time a technology can run in a year.

Technology	Availability Factor
DIESEL_PP	0.91
PV_PP	0.94
WIND_PP	0.95
LI_TECH	0.92
HYD_ELS	0.95
HYD_FC	0.95
ELEC_DIS	1.00
IMP_DSL	1.00

Table 2 Availability Factor

d) The OperationalLife is the useful lifetime of a technology expressed in years.

Technology	Operational life
DIESEL_PP	20
PV_PP	25
WIND_PP	20
LI_TECH	30
HYD_ELS	30

HYD_FC	10
ELEC_DIS	30
IMP_DSL	1

Table 3 Operational Life

e) The ResidualCapacity is the capacity available from before the modelling period. This value is calculated with the help of the operational life, installed year and the annual capacity to calculate the annual decrease in capacity.

f) The InputActivityRatio is the ratio of rate of use of a commodity by a technology, to the rate of activity.

g) The OutputActivityRatio is the ratio of rate of commodity output from a technology to the rate of activity.

Technology Costs

The technology costs are split into Capital cost, Fixed cost and Variable cost. The CapitalCost accounts for the capital investment cost of a technology, per unit of capacity. The FixedCost accounts for the Operational and Maintenance cost of a technology per unit of capacity and VariableCost accounts for the variable O&M cost with different modes of operation. The costs of PV panels and wind turbines are assumed to decrease [45] and for the rest of the technologies the costs are considered constant.

	Capital costs	Fixed Costs	Variable Costs
DIESEL IMPORT	0.00001 k€/MW	0.00001 k€/MW/y	1.77 k€/t
DIESEL PP	1023.5 k€/MW	30.705 k€/MW/y	0.019 k€/MWh
PV PP	598_{2023} -330_{2050} k€/MW	8.16 k€/MW/y	0 k€/MWh
WIND PP	1304_{2023}-1118_{2050} k€/MW	14.575 k€/MW/y	0.003 k€/MWh
LI TECH	1596 k€/MW	3.916 k€/MW/y	0.006 k€/MWh
HYDROGEN ELECTROLYZER	1691 k€/MW	12.91 k€/MW/y	0.0056 k€/MWh
HYDROGEN FUEL CELL	1234.4 k€/MW	11.92 k€/MW/y	0.00044 k€/MWh
ELEC_DIS	0.00001 k€/MW	0 k€/MW/y	k€/MWh

Table 4 Costs

Storage

In this model two storage systems are used, Hydrogen Storage and Lithium-ion battery storage. The Hydrogen storage is connected to the electrolyser during the charging phase and to the fuel cell during the discharging phase. The Lithium-ion storage uses only one technology which charges during mode of operation 1 and discharges during mode of operation 2. The state of charge is set to zero for the storages when they are installed. The lifetime of the hydrogen storage is considered as 18 years [46] and the lifetime of the Lithium-ion storage is considered as 10 years [47].

Capacity Constraints

The capacity constraints are parameters used to set limits to the renewable power potentials that can be installed in the region and the investments that can be made. In this model we are setting it to invest freely in the installation of technology and the installation of renewable energy technology. This is done because the primary aim of this project is to study how different clustering algorithms affect long term energy modelling.

Activities constraints

The activity constraints are similar to capacity constraint in function, it is used to limit the activity of renewable energy technology. In this model we don't need to limit the technologies activities.

Reserve Margin

The excess installed capacity with respect to the peak demand is called as reserve margin. For this model we are not considering any extra reserve margin, so the value has been set to 1.

Renewable generation target

The renewable energy technologies used in this model are PV panels and wind turbines and they are set to 0 and the Diesel power plant is set to 0. This is done because when emissions are considered it is better because when renewable generation target is applied the system is forced to satisfy the electricity demand, but this doesn't reduce the emissions of the technologies which is disconnected from the network.

Emissions

In our model we have considered CO2 gas as the only emission. The emissions of CO2 are linked with the diesel import. It's considered to be 3.15 tons of CO2 per ton of diesel imported [48]. When the limits are set, the system is forced to reduce the pollutants and tries to satisfy all the demands using CO2 free technology.

4.4 Scenario definition

Different scenarios are modelled to test the functionality of the different clustering algorithms. In this model we have modelled a total of 3 scenarios. The three different scenarios are namely only PV, only Wind and Hybrid. As the name suggests “only PV” scenario involves the use of PV panels alone as renewable energy source and in the “only Wind” scenario the wind turbines are the only source of renewable energy. In the “Hybrid” scenario both wind turbines and PV panels are used as renewable energy sources.

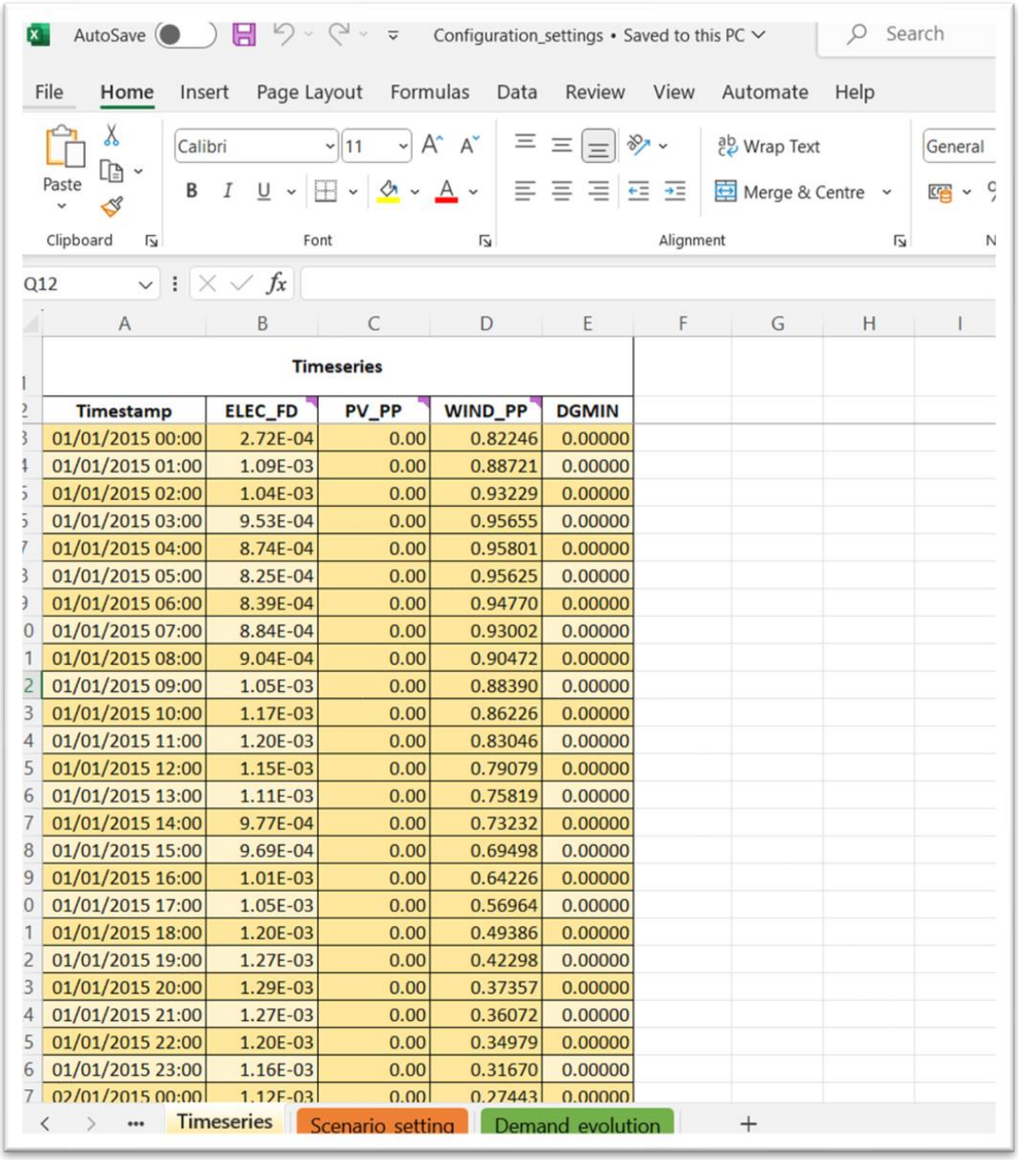
For each scenario the model is run using k-means, k-medoids and k-medians algorithms. As described earlier the model is run for different number of representative days which are 6,12,24,36,48,72 and 144. So, in the end a total of 63 different models were simulated to understand the working of the clustering algorithms.

	K-MEANS	K-MEDIANS	K-MEDOIDS
ONLY PV	6, 12, 24, 36, 48, 72, 144 REPRESENTATIVE DAYS		
ONLY WIND			
HYBRID			

Table 5 Scenarios

5. MODEL AND SCENARIO DEVELOPMENT

In this project for model development the osemosys software is used which is coded using the python language. The code is run using the spyder python software in the anaconda environment. The code is written in multiple sub-files which are then called upon by a main function. The code also takes inputs from excel files which contains the information of the sets and parameters as described in the previous chapter. The scenarios are developed with the help of the configurations sheet.



Timeseries				
Timestamp	ELEC_FD	PV_PP	WIND_PP	DGMIN
01/01/2015 00:00	2.72E-04	0.00	0.82246	0.00000
01/01/2015 01:00	1.09E-03	0.00	0.88721	0.00000
01/01/2015 02:00	1.04E-03	0.00	0.93229	0.00000
01/01/2015 03:00	9.53E-04	0.00	0.95655	0.00000
01/01/2015 04:00	8.74E-04	0.00	0.95801	0.00000
01/01/2015 05:00	8.25E-04	0.00	0.95625	0.00000
01/01/2015 06:00	8.39E-04	0.00	0.94770	0.00000
01/01/2015 07:00	8.84E-04	0.00	0.93002	0.00000
01/01/2015 08:00	9.04E-04	0.00	0.90472	0.00000
01/01/2015 09:00	1.05E-03	0.00	0.88390	0.00000
01/01/2015 10:00	1.17E-03	0.00	0.86226	0.00000
01/01/2015 11:00	1.20E-03	0.00	0.83046	0.00000
01/01/2015 12:00	1.15E-03	0.00	0.79079	0.00000
01/01/2015 13:00	1.11E-03	0.00	0.75819	0.00000
01/01/2015 14:00	9.77E-04	0.00	0.73232	0.00000
01/01/2015 15:00	9.69E-04	0.00	0.69498	0.00000
01/01/2015 16:00	1.01E-03	0.00	0.64226	0.00000
01/01/2015 17:00	1.05E-03	0.00	0.56964	0.00000
01/01/2015 18:00	1.20E-03	0.00	0.49386	0.00000
01/01/2015 19:00	1.27E-03	0.00	0.42298	0.00000
01/01/2015 20:00	1.29E-03	0.00	0.37357	0.00000
01/01/2015 21:00	1.27E-03	0.00	0.36072	0.00000
01/01/2015 22:00	1.20E-03	0.00	0.34979	0.00000
01/01/2015 23:00	1.16E-03	0.00	0.31670	0.00000
02/01/2015 00:00	1.12E-03	0.00	0.27443	0.00000

Figure 11 Time Series Configuration settings

Start_year	End_year
2023	2050

Short_name	Clustering_alg	Model_templa	n_days	n_bracket	length_bracke	n_timeslic	n_timep	n_reprda
FAVIGNANA_6D_5B_kmedians_2023_2050	kmedians	FAVIGNANA	6	5	6;4;4;4;6	30	1825	6
FAVIGNANA_12D_5B_kmedians_2023_2050	kmedians	FAVIGNANA	12	5	6;4;4;4;6	60	1825	12
FAVIGNANA_24D_5B_kmedians_2023_2050	kmedians	FAVIGNANA	24	5	6;4;4;4;6	120	1825	24
FAVIGNANA_36D_5B_kmedians_2023_2050	kmedians	FAVIGNANA	36	5	6;4;4;4;6	180	1825	36
FAVIGNANA_48D_5B_kmedians_2023_2050	kmedians	FAVIGNANA	48	5	6;4;4;4;6	240	1825	48
FAVIGNANA_72D_5B_kmedians_2023_2050	kmedians	FAVIGNANA	72	5	6;4;4;4;6	360	1825	72
FAVIGNANA_144D_5B_kmedians_2023_2050	kmedians	FAVIGNANA	144	5	6;4;4;4;6	720	1825	144

Timeseries | Scenario_setting | Demand_evolution

Figure 12 Scenario Setting Configuration Settings

Commodity	Year	Value
ELEC_FD	2023	15.469
ELEC_FD	2024	15.624
ELEC_FD	2025	15.780
ELEC_FD	2026	15.938
ELEC_FD	2027	16.097
ELEC_FD	2028	16.258
ELEC_FD	2029	16.421
ELEC_FD	2030	16.585
ELEC_FD	2031	16.751
ELEC_FD	2032	16.918
ELEC_FD	2033	17.088
ELEC_FD	2034	17.259
ELEC_FD	2035	17.431
ELEC_FD	2036	17.605
ELEC_FD	2037	17.782
ELEC_FD	2038	17.959
ELEC_FD	2039	18.139
ELEC_FD	2040	18.320
ELEC_FD	2041	18.504
ELEC_FD	2042	18.689
ELEC_FD	2043	18.875
ELEC_FD	2044	19.064
ELEC_FD	2045	19.255
ELEC_FD	2046	19.447
ELEC_FD	2047	19.642
ELEC_FD	2048	19.838
ELEC_FD	2049	20.037

Timeseries | Scenario_setting | Demand_evolution

Figure 13 Demand Evolution Configuration Setting

In figure 11, 12 and 13 we can see the different sheets which are present in the configuration settings excel file. In the first sheet we have the Timeseries data for the island of Favignana. The timeseries is recorded for every hour of every day of the month for a whole year. The timeseries data includes the electricity final demand of Favignana and the timeseries data for the capacity factor of solar and wind energy retrieved from renewable ninja are also part of this first sheet seen in figure 11.

The time series sheet helps switching between scenarios from ONLY PV, ONLY WIND and HYBRID. This is done by bringing changes to the capacity factors of solar and wind energy. For ONLY PV scenario the capacity factors of wind are set to zero and the PV intact. For ONLY WIND scenario the capacity factors of PV are set to zero and the wind intact. In the HYBRID scenario both the capacity factors are kept intact.

In the next sheet shown in figure 12, the different parameters that are used to set the model are seen. Under the column clustering algorithm, it is possible to switch between the different clustering algorithms. Then there are options where the number of representative days and time brackets can be specified with intervals.

In the next sheet shown in figure 13, the final demand for the island of Favignana is given as an input. In this model we have used the values from the year 2019. These values are expressed in GW.

Figures 14 and 15 are the representation of the sheets that were used for sets and parameters for the island of Favignana respectively. Both these sheets were inputted with data which has been described in the previous chapters got after thorough research.

REGION	EMISSION	FUEL	TIMEPERIOD	TIMESLICE	MODE_OF_OPERATION	STORAGE	TECHNOLOGY	YEAR
FAVIGNANA	CO2	DSL			1	LI_STO	IMP_DSL	2023
		ELEC_SC			2	HYD_STO	DIESEL_PP	2024
		ELEC_FD					PV_PP	2025
							WIND_PP	2026
							LI_TECH	2027
							HYD_ELS	2028
							HYD_FC	2029
							ELEC_DIS	2030
								2031
								2032
								2033
								2034
								2035
								2036
								2037
								2038
								2039
								2040
								2041
								2042
								2043
								2044
								2045
								2046
								2047
								2048
								2049
								2050

Sets

Figure 14 Sets sheet of Favignana

ull name	Short name	Default value	SET n.1	SET n.2	SET n.3
iscountRate	1_DiscRat	0.04	REGION		
radeRoute	2_TraRou	0	REGION	REGION	FUEL
epreciationMethod	3_DepMet	1	REGION		
imeSliceSplit	4_TimSlicSplit	0	TIMESLICE		
onversiontpl	5_Convtpl	0	TIMEPERIOD	TIMESLICE	
earSplit	6_YeaSplit	0	TIMESLICE		
pecifiedAnnualDemand	7_SpeAnnDem	0	REGION	FUEL	YEAR
pecifiedDemandProfile	8_SpeDemPro	0	REGION	FUEL	TIMESLICE
ccumulatedAnnualDemand	9_AccAnnDem	0	REGION	FUEL	YEAR
apacityToActivityUnit	10_CapToActUni	1	REGION	TECHNOLOGY	
apacityFactor	11_CapFac	0.95	REGION	TECHNOLOGY	TIMESLICE
vailabilityFactor	12_AvaFac	0.95	REGION	TECHNOLOGY	YEAR
perationalLife	13_OpeLif	1	REGION	TECHNOLOGY	
esidualCapacity	14_ResCap	0	REGION	TECHNOLOGY	YEAR
nputActivityRatio	15_InpActRat	0	REGION	TECHNOLOGY	FUEL
utputActivityRatio	16_OutActRat	0	REGION	TECHNOLOGY	FUEL
apitalCost	17_CapCos	0.00001	REGION	TECHNOLOGY	YEAR
ariableCost	18_VarCos	0.00001	REGION	TECHNOLOGY	MODE_OF_OPERATION
ixedCost	19_FixCos	0	REGION	TECHNOLOGY	YEAR
echnologyToStorage	20_TecToSto	0	REGION	TECHNOLOGY	STORAGE
echnologyFromStorage	21_TecFroSto	0	REGION	TECHNOLOGY	STORAGE
torageLevelStart	22_StoLevSta	0	REGION	STORAGE	
torageMaxChargeRate	23_StoMaxChaRat	99999999	REGION	STORAGE	
torageMaxDischargeRate	24_StoMaxDisRat	99999999	REGION	STORAGE	
linStorageCharge	25_MinStoCha	0	REGION	STORAGE	YEAR
perationalLifeStorage	26_OpeLifSto	0	REGION	STORAGE	
apitalCostStorage	27_CapCosSto	0	REGION	STORAGE	YEAR

Index | 1_DiscRat | 2_TraRou | 3_DepMet | 4_TimSlicSplit | 5_Convtpl | 6_YeaSplit | 7_SpeAnnDem | 8_SpeDemPro | 11_CapFac | 9_AccAnnDem | 10_

Figure 15 Parameters sheet of Favignana

For long-term energy modelling using python, different programs are written for executing the model. The structure of the code is given below in figure 16.

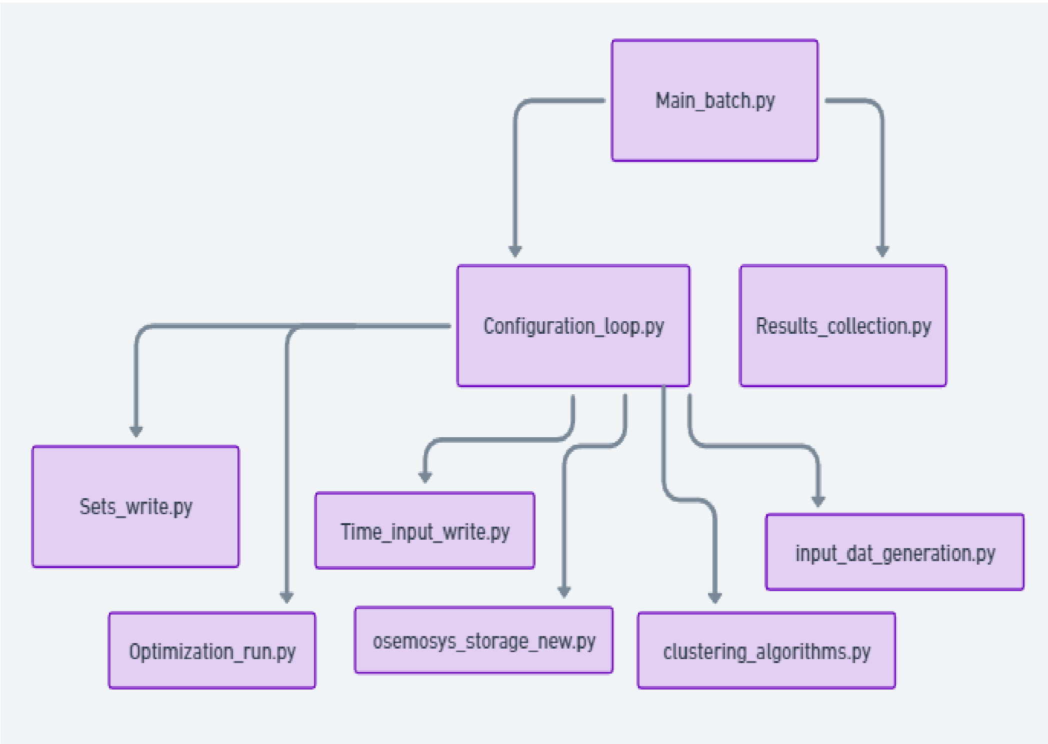


Figure 16 Structure of the code

Each of these python files has a function. Below the description of each of these files are given.

Main_batch.py

When this program is initiated, it first accesses the 'Configurations' Excel file. This file holds all the details about the models, as mentioned in the previous section like the sets, parameters and the timeseries and scenario settings. Subsequently, the tool sets up the entire simulation environment. This includes establishing the directories and importing additional functions.

Configuration_loop.py

The chosen model comprises various scenarios, each with its unique configurations. These settings are processed in a configuration loop using a for loop, which then initiates the simulation for each scenario. During this phase, a time series file is either generated or accessed if it pre-exists. After accumulating all the data for each scenario, the software proceeds to produce the sets file using 'sets_write.py'. It also creates an Excel file for the parameters using 'time_input_write.py' and then runs the scenario simulation through 'optimization_run.py'. 'input_dat_generation.py' carries out the operation of creating the input data for all the years of modelled. Upon completing a simulation, the program moves on to simulate the next scenario. The timeseries if its more than 8760 values because of more values because of less duration intervals it will be resized to 8760 values.

Time_input_write.py

This code creates the parameters specific to each scenario based on the conditions outlined in the configuration file. It utilizes a template parameters file that includes all essential parameters for the model's convergence, initially set for just one year. The 'time_input_write' script accesses this file and alters it by incorporating customizations specified in the configuration file, if any are present. Alternatively, it replicates the first year's data for the subsequent years. This process results in a fully formed parameter file, which is then saved for future use.

Clustering_algorithms.py

This code functions to switch between clustering algorithms that will be used while running the optimization program. Upon called by the code the clustering algorithm functions.

Optimization_run.py

Following the creation of the sets and parameters files, the subsequent step involves the 'optimization_run' code producing a .dat file. This file holds the necessary information for OSeMOSYS to carry out the simulation. Once this is done, the simulation commences, leading to the generation of an output JSON file that encapsulates all the results.

Results_collection.py

Once the model's various scenarios have been fully simulated, the 'results_collection' script cycles through all the results folders, extracting data from the JSON files. Each scenario of the model is stored in a dictionary, with a separate sub-dictionary dedicated to each scenario. Within each sub-dictionary, there are further subdivisions corresponding to the number of output variables. After completing this collection process, the script consolidates the results into a numpy file, which is then formatted for compatibility with the visualization tool.

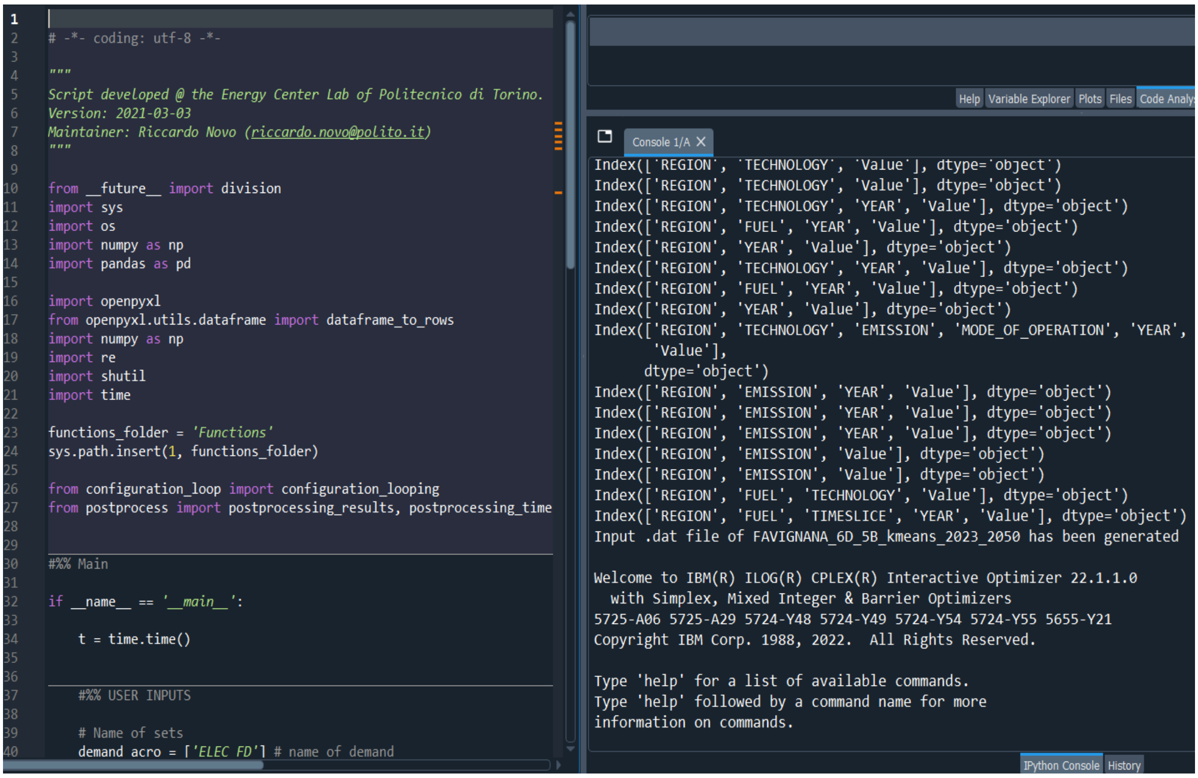


Figure 17 Python terminal for Main_batch.py while simulation

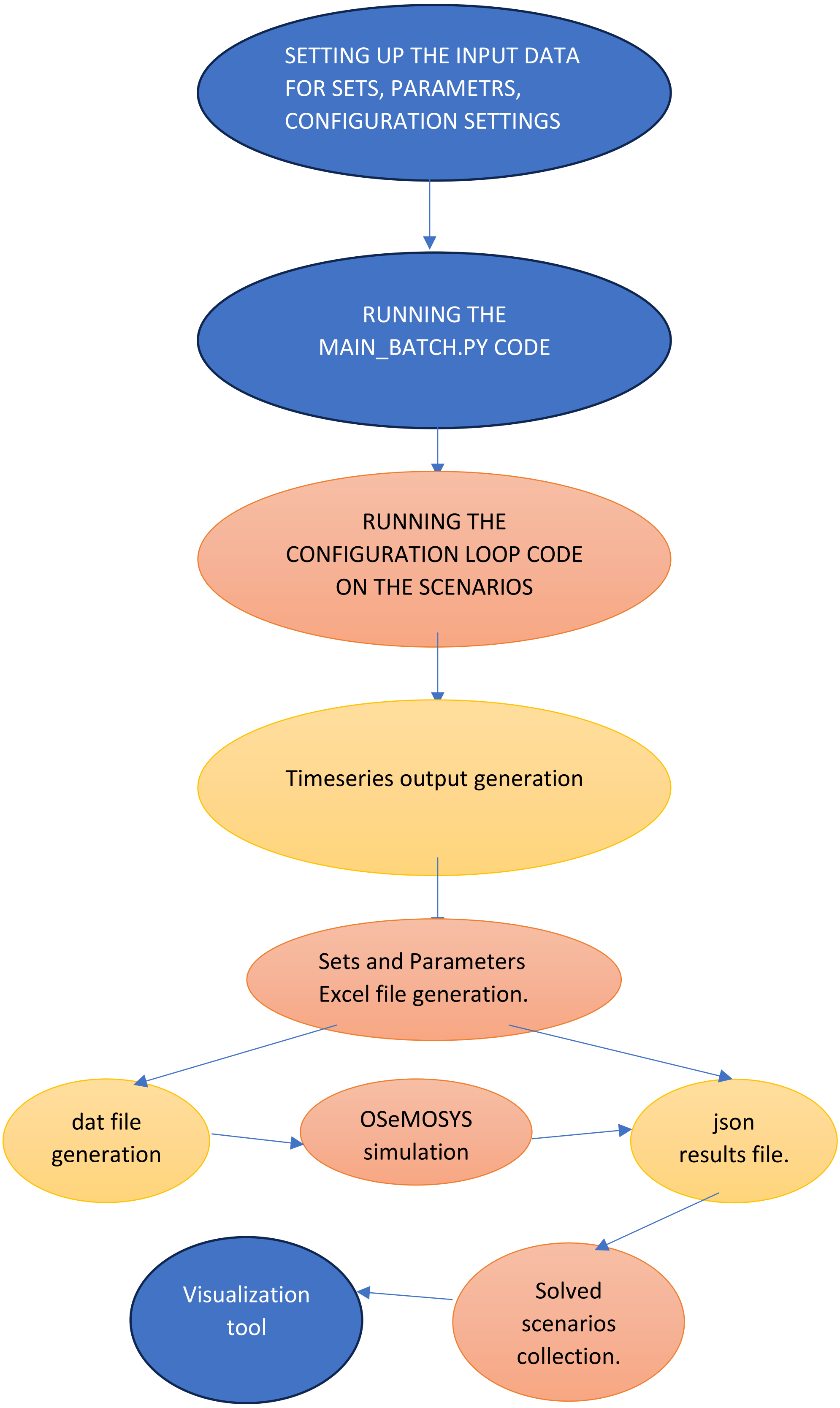


Figure 18 Simulation Flowchart

Visualisation

For visualisation of the results a code was written on python which helped in converting the json file into excel files with the important data needed.

```python
import json
import pandas as pd

file_path = 'C:/Users/ASUS/Desktop/FINAL FOLDER/OSeMOSYS_autoclust_20231022/Configurations/FAVIGNANA_144D_5B_kmedians_2023_
with open(file_path, 'r') as file:
    data = json.load(file)

    results = []

    for variable_name, variable_value in data["Solution"][1]["Variable"].items():
        # Check if the variable name contains any digit between 2023 and 2050
        year_present = any(str(year) in variable_name for year in range(2023, 2051))

        if year_present:
            value = variable_value['Value']
            year = None

            for year in range(2023, 2051):
                if str(year) in variable_name:
                    year = year
                    break
        else:
            value = variable_value['Value']
            year = None
# Replace 'FAVIGNANA' with an empty string in the variable name
        variable_name = variable_name.replace('FAVIGNANA,', '')
# Replace 'years' with an empty string in the variable name
        variable_name = variable_name.replace(',2023', '')
        variable_name = variable_name.replace(',2024', '')
```

Figure 19 json to excel code.

The resulting excel file was then imported into Power BI where the results were visualised in the form of graphs [49]. In figure 20 the layout of the Power BI software is shown.

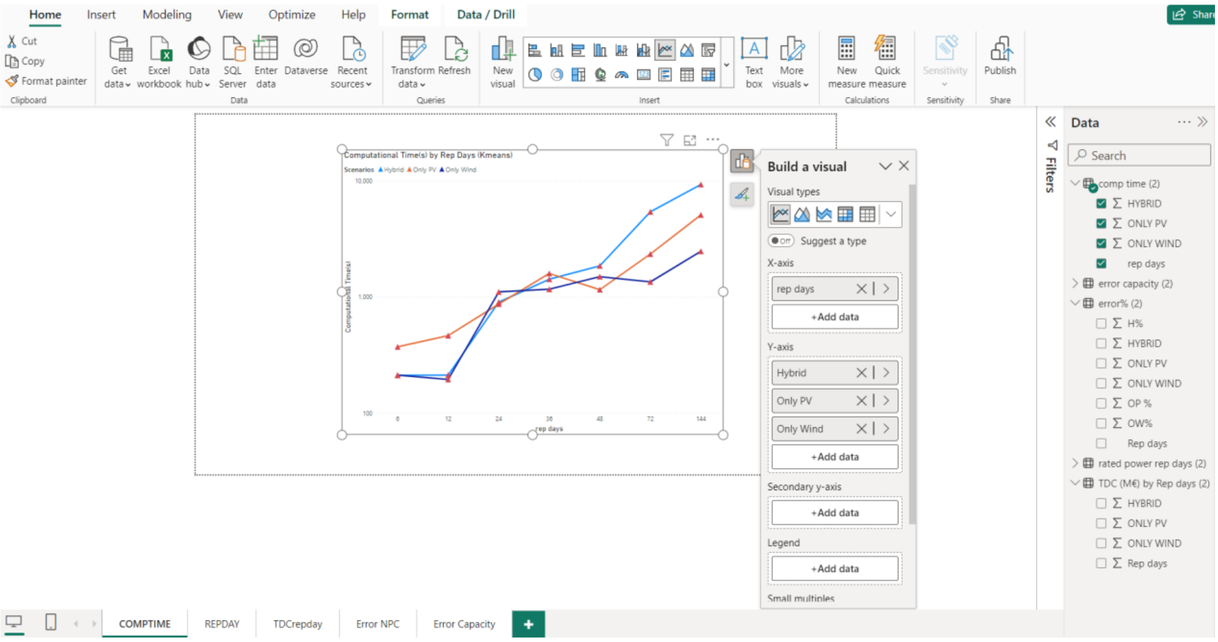


Figure 20 Layout of POWER BI

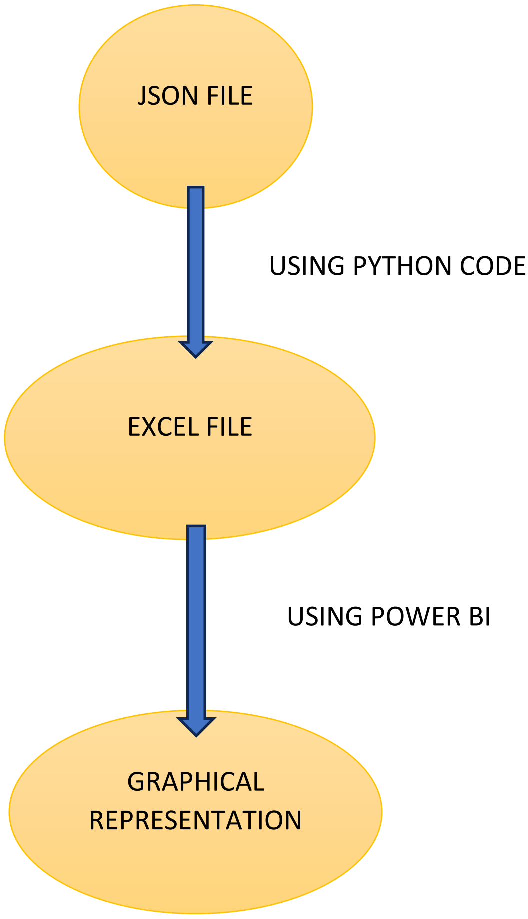


Figure 21 Flowchart of graphical representation

6. RESULTS

In this pivotal section of our thesis, we delve into the findings of our comprehensive energy system modelling for Favignana Island. Utilizing an enhanced version of the OSeMOSYS tool, we explored multiple scenarios with a focus on identifying the most suitable time-series clustering algorithm that effectively balances computational efficiency and result accuracy. Our scenarios, diverse in their approach, include exclusive reliance on photovoltaics, wind energy, and a hybrid combination of both, augmented with various storage technologies like Li-ion batteries and hydrogen storage.

The results derived from these scenarios provide crucial insights into the feasibility and desirability of ambitious decarbonization strategies. Notably, they highlight the differential suitability of various clustering algorithms in distinct scenarios, offering a nuanced understanding of their applicability in energy system modelling. The implications of these findings are significant, offering a path forward in energy planning that is both efficient and effective, particularly in contexts similar to Favignana Island.

In the following sections, we present a detailed analysis of these results, offering a clear, structured, and critical assessment of each scenario and the performance of the clustering algorithms employed.

We have thoroughly discussed the results obtained from our simulation runs, where we meticulously analyzed the performance of three clustering algorithms - KMEANS, KMEDOIDS, and KMEDIANS - across different energy scenarios like ONLY PV, ONLY WIND, and HYBRID. The study's crux lies in evaluating these algorithms based on various parameters such as computational time, net present cost, sizing of components, and relative errors in total system cost and sizing. The primary objectives of this investigation are twofold: firstly, to identify a clustering algorithm that significantly reduces computational time, thereby curtailing the costs associated with simulations; and secondly, to pinpoint the algorithm that yields the most accurate results. This dual focus on efficiency and precision in algorithm selection is pivotal for advancing the reliability and cost-effectiveness of renewable energy system modelling.

6.1 Computational Time

K-means Analysis

The variation of the computational time with respect to the representative days for different scenarios while using the k-means algorithm can be visualised in figure 22.

The Hybrid scenario shows a consistent increase in computational time as the number of representative days increases, indicating a direct relationship between data volume and processing time.

The Only PV scenario exhibits a non-linear pattern, with a significant peak at 36 rep days, suggesting that the nature of the PV data might have complexities that impact computational efficiency differently at various stages.

The Only Wind scenario like Hybrid, it displays a generally increasing trend, but with a distinct peak at 24 rep days, which could be due to specific data characteristics at that point.

The Hybrid scenario shows a considerable increase in computational time, most significantly from 24 to 36 representative days with an increase of 57.60%. In the Only PV scenario, a substantial peak is observed from 24 to 36 rep days, with an 83.39% increase, followed by a notable decrease of 27.48% from 36 to 48 rep days. The Only Wind scenario presents its most dramatic increase between 12 to 24 rep days, a striking 466.51%, indicating a significant computational demand during this period.

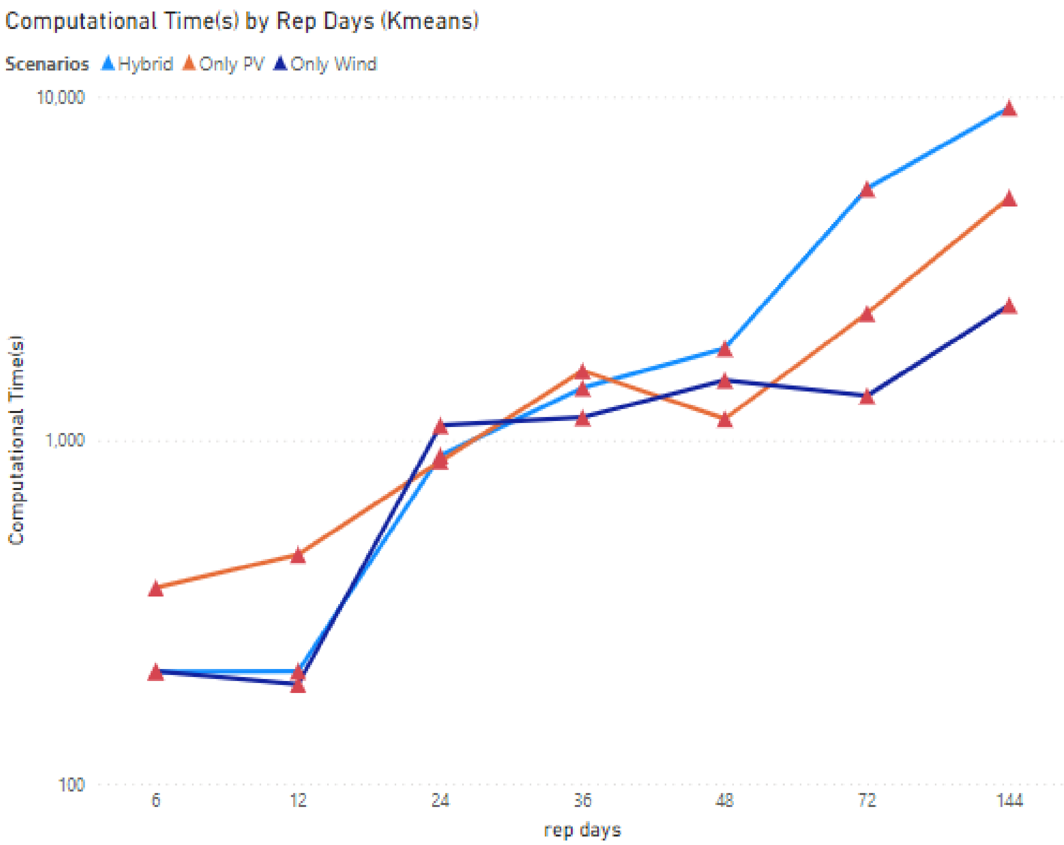


Figure 22 Computational Time-Kmeans

K-medians Analysis

The variation of the computational time with respect to the representative days for different scenarios while using the k-medians algorithm can be visualised in figure 23.

The Hybrid scenario has a fluctuating pattern, with a notable peak at 48 rep days, indicating a higher computational demand for larger datasets.

The only PV scenario varies significantly, peaking at 36 rep days, much like in the K-means analysis. This variation could be indicative of data anomalies or irregularities.

The Only Wind scenario shows a less consistent trend, with the highest time at 12 rep days, potentially highlighting the algorithm's sensitivity to wind data peculiarities.

For K-medians, the scenario with Only Wind shows an extremely high increase of 1113.65% from 6 to 12 rep days, suggesting a significant sensitivity of this algorithm to wind data. The Hybrid scenario also demonstrates considerable fluctuations, with a notable increase of 116.86% from 36 to 48 rep days. In the Only PV scenario, the change from 24 to 36 rep days stands out with a 122.24% increase, reflecting the algorithm's varying efficiency with different data sizes.

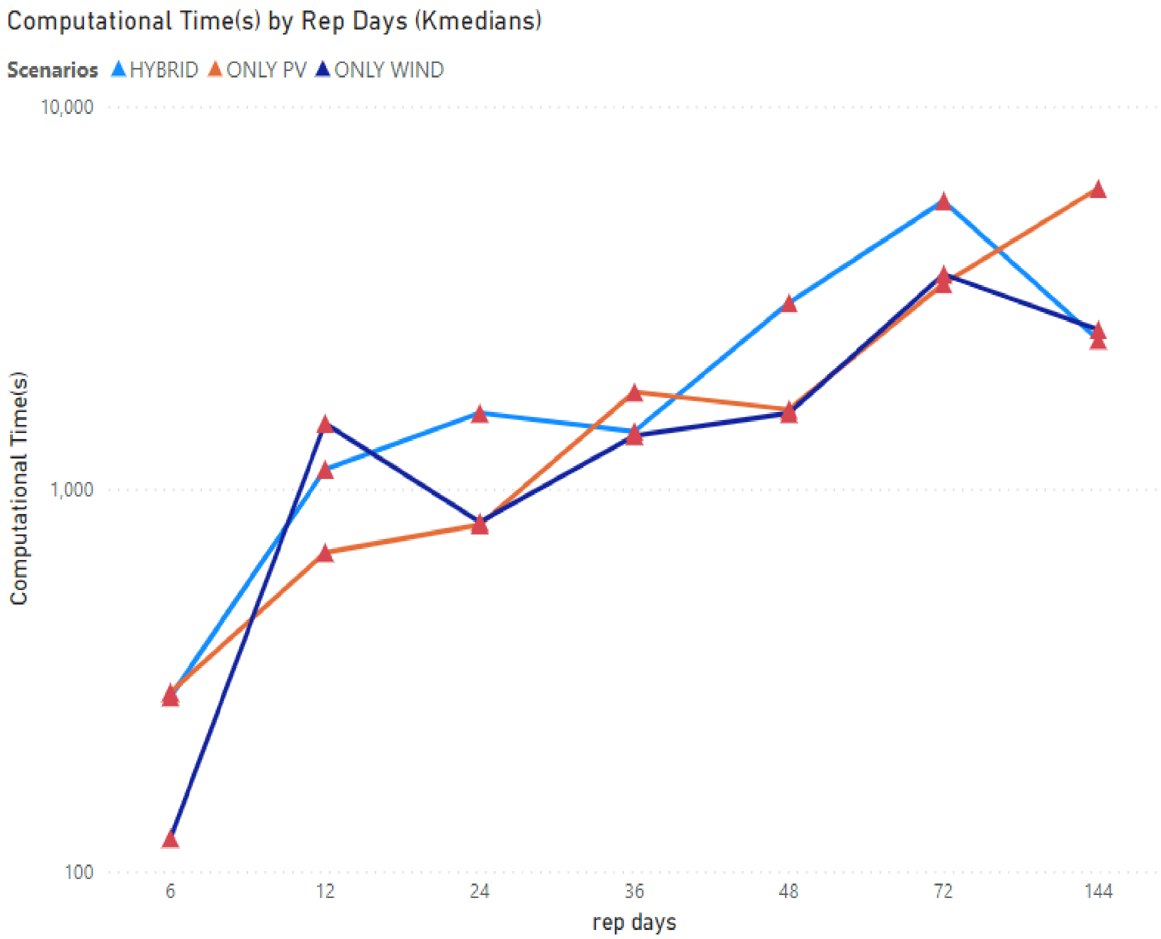


Figure 23 Computational Time-Kmedians

K-medoids Analysis

The variation of the computational time with respect to the representative days for different scenarios while using the k-medians algorithm can be visualised in figure 24.

The Hybrid scenario exhibits the most considerable variation in computational times, with an exceptionally high time at 36 rep days. This could be due to its more complex computation, particularly when handling diverse or outlier-heavy data.

The Only PV scenario shows a dramatic increase in computational time as the number of rep days grows, indicating that the PV data may contain numerous outliers or complex structures that significantly impact computational efficiency.

The Only Wind scenario follows a somewhat increasing trend but with less pronounced peaks compared to other scenarios, possibly due to the nature of wind data.

The Hybrid scenario records an exceptionally high increase of 511.04% from 24 to 36 rep days, followed by a sharp decrease of 63.94% from 36 to 48 rep days. The Only PV scenario shows an enormous increase of 2007.30% from 12 to 24 rep days, indicating a high computational demand for this data type in this period. Similarly, in the Only Wind scenario, there is a significant increase of 295.31% from 6 to 12 rep days.

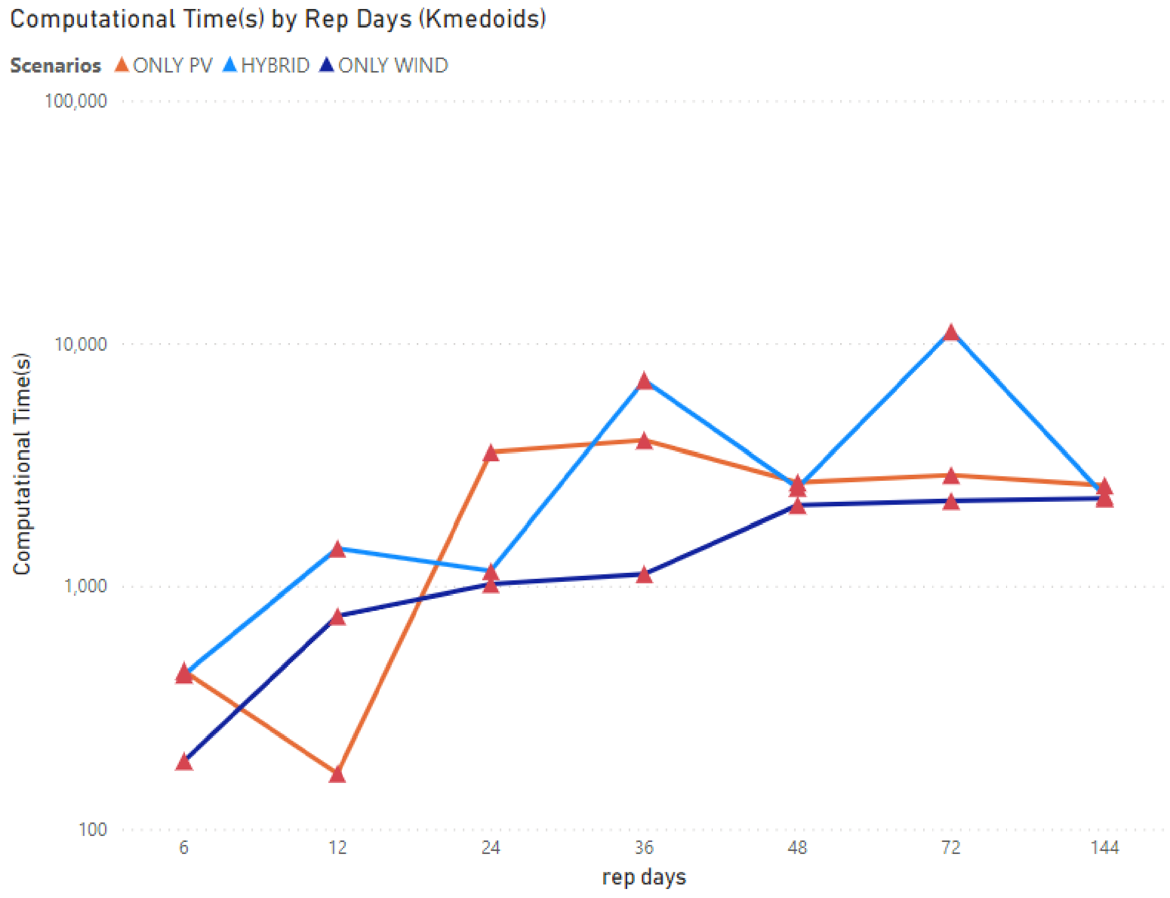


Figure 24 Computational Time-Kmedoids

Best Suited Algorithm

To determine the most suited clustering algorithm, we must consider both computational efficiency and the nature of the data:

Efficiency: K-means generally shows the most consistent and lowest computational times across scenarios, indicating its efficiency in handling various data types and volumes.

Robustness: K-medoids, while computationally intensive, is more robust to outliers and could be preferred if the data has many anomalies or non-standard distributions.

Data Nature and Complexity: K-medians might be a choice when the dataset has significant non-normal distributions, especially in the PV data.

Given the trade-off between computational efficiency and robustness to outliers, K-means emerges as the most suited for scenarios where computational efficiency is a priority and the data is relatively clean with few outliers. However, in cases where data robustness is crucial, and the presence of outliers is significant, K-medoids, despite its higher computational demand, would be a preferable choice. K-medians could be considered for specific datasets with non-normal distributions, particularly in the PV scenario.

6.2 Net Present Cost

K-means Analysis

The variation of the net present cost with respect to the representative days for different scenarios while using the k-means algorithm can be visualised in figure 25.

The Hybrid scenario shows a steady increase in NPC across representative days. Notably, there's a 15.69% increase from 48 to 72 days, suggesting a higher cost implication as the number of days increases. This trend could indicate the increasing complexity of data management and processing in hybrid scenarios.

The Only PV scenario demonstrates an upward trend in NPC, with a significant jump of 11.32% between 24 to 36 days. This indicates that PV data might become progressively complex or voluminous, impacting the cost.

The Only Wind scenario is consistent with the other scenarios, there's an increase over time. A notable peak of 13.78% from 72 to 144 days implies that long-term wind data processing could become increasingly costly, possibly due to variability in wind patterns over extended periods.

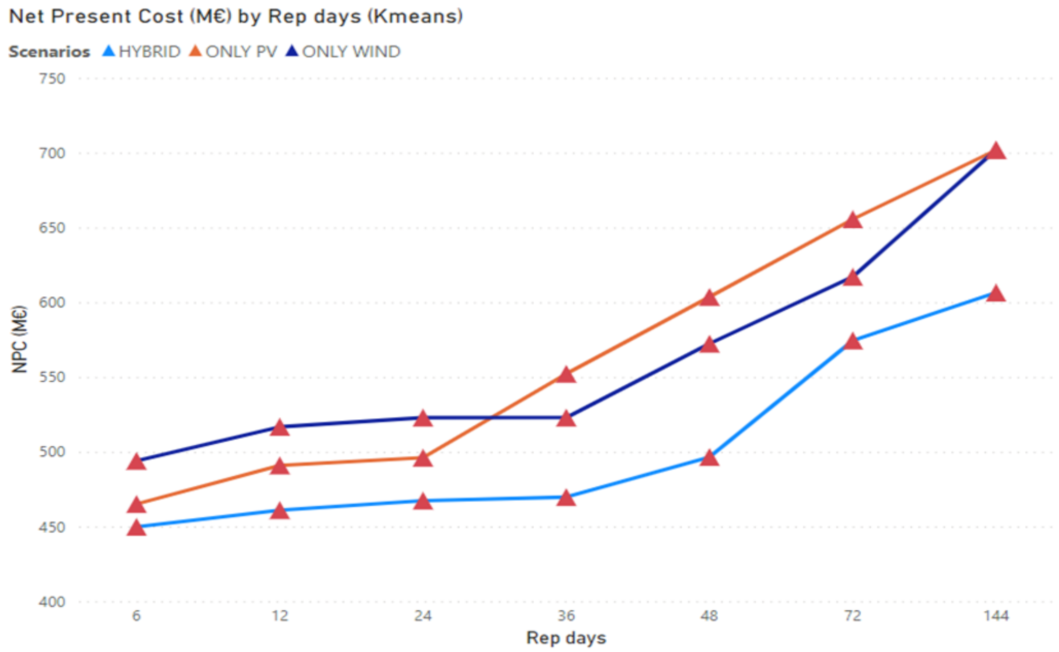


Figure 25 Net Present Cost-kmeans

K-medians Analysis

The variation of the net present cost with respect to the representative days for different scenarios while using the k-medians algorithm can be visualised in figure 26.

The only Hybrid scenario exhibits fluctuations with a significant rise of 20.44% from 36 to 48 days. This could reflect the sensitivity of the K-medians algorithm to the data distribution in hybrid energy scenarios, especially when median values shift notably with new data.

The Only PV scenario shows dramatic increases, especially a 49.68% rise from 6 to 12 days. This might suggest that median values in the PV data significantly differ even with small increases in representative days, indicating a diverse or non-uniform dataset.

The Only Wind scenario shows varying results, with a peak increase of 27.82% from 72 to 144 days. This could mean that the wind data's median values are significantly affected over longer periods, possibly due to seasonal or other environmental factors.

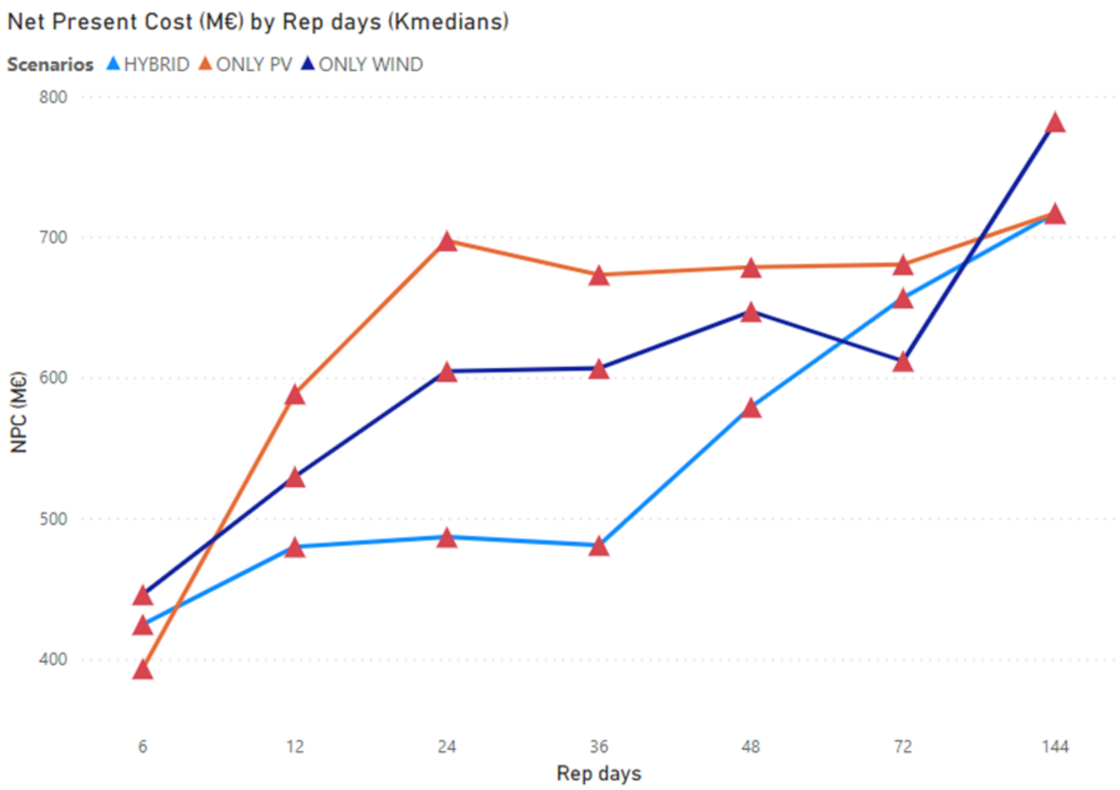


Figure 26 Net Present Cost-kmedians

K-medoids Analysis

The variation of the net present cost with respect to the representative days for different scenarios while using the k-medoids algorithm can be visualised in figure 27.

The only Hybrid scenario exhibits the most significant variability among the algorithms, with an enormous increase of 511.04% from 24 to 36 days. Such a dramatic rise suggests that the selection of actual data points as cluster centres in hybrid scenarios can lead to substantial cost differences as the dataset grows.

The Only PV scenario demonstrates an extreme increase of 2007.30% from 12 to 24 days, indicating a high sensitivity to changes in data representation. This could be due to the presence of significant outliers or highly diverse data in the PV scenario.

The Only Wind scenario shows considerable fluctuations, with a 295.31% increase from 6 to 12 days, again indicating the algorithm's sensitivity to the dataset's characteristics.

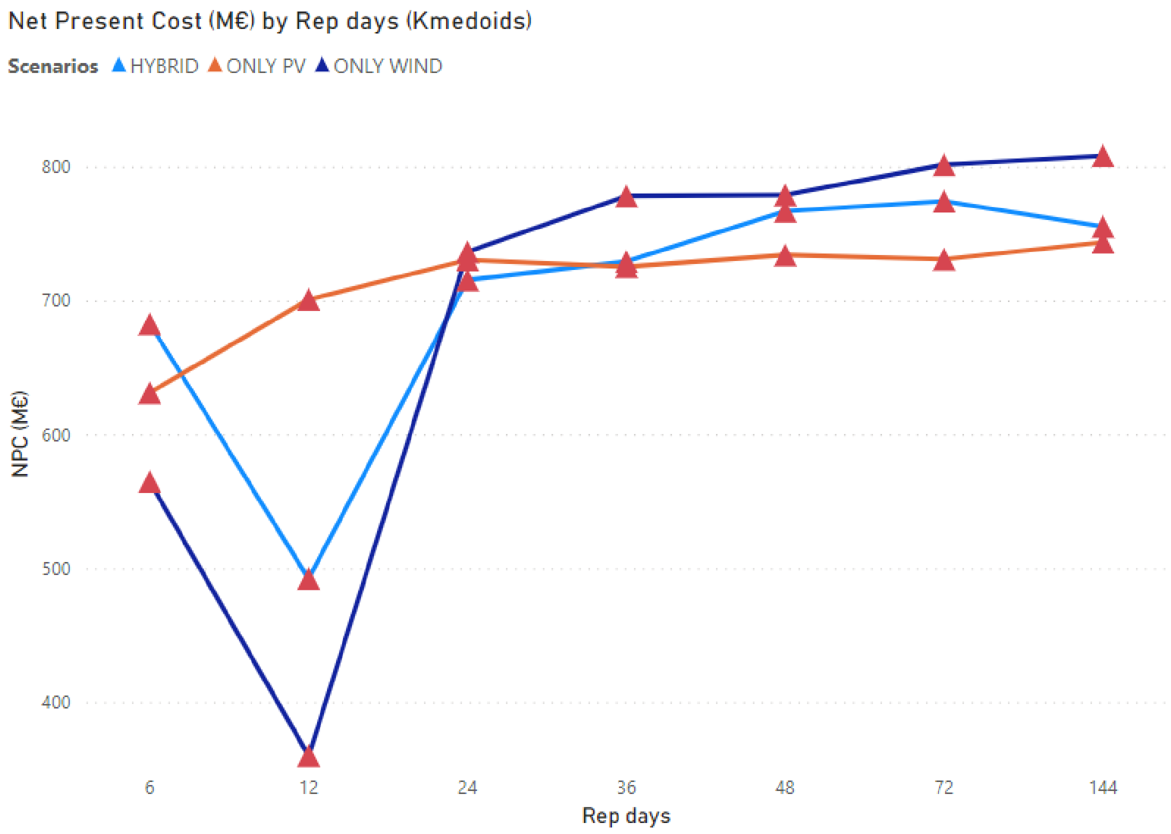


Figure 27 Net Present Cost-kmedoids

Best Suited Algorithm

K-means: Exhibits a more stable and gradual increase in Net Present Cost, suggesting its suitability for scenarios where a steady prediction of costs is crucial. Its less changes make it a reliable choice for consistent data patterns.

K-medians: Shows more variability than K-means but less than K-medoids. It could be suitable for scenarios where median values provide a more accurate cost prediction, especially in datasets with outliers.

K-medoids: While this algorithm is robust to outliers, its high variability in NPC predictions makes it less suitable for stable cost estimation. However, it could be preferred in highly irregular or outlier-prone datasets where the robustness to extreme values is critical.

In conclusion, K-means seems to offer the most stable NPC predictions, making it suitable for scenarios requiring consistent and gradual changes in cost estimations. K-medians could be used in situations where median values offer a more accurate prediction, and K-medoids is best suited for highly irregular datasets where outlier robustness is essential.

6.3 Relative error in total system cost

K-means Analysis

The variation of the relative error in total system cost with respect to the representative days for different scenarios while using the k-means algorithm can be visualised in figure 28.

Beginning with k-means, the HYBRID method stands out with its consistent error reductions across repetition days. At 6 days, it exhibits a -25.83% error, which then steadily declines to -23.99% (12 days), -22.94% (24 days), -22.55% (36 days), and finally -18.13% (48 days). This demonstrates smooth refinement of cost modelling precision, with accuracy improvements of 7.2%, 3.7%, and 20.1% respectively. By 72 days, accuracy is even higher at -5.28%, representing massive gains of over 370% from the original error.

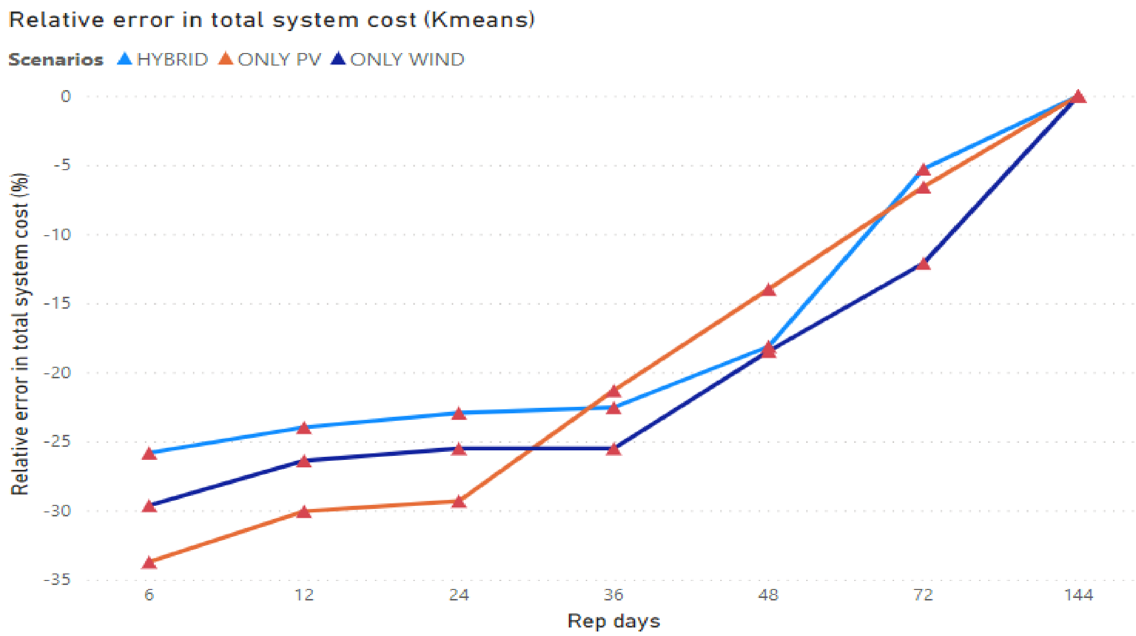


Figure 28 Relative Error in total system cost kmeans

K-medians Analysis

The variation of the relative error in total system cost with respect to the representative days for different scenarios while using the k-medians algorithm can be visualised in figure 29.

As for k-medians, errors remain elevated across all categories. At 6 days, HYBRID sits at -40.78%, ONLY PV at -45.15%, and ONLY WIND at -42.99%, significantly higher than other algorithms. Subsequent gains also lag, with HYBRID only reaching -19.23% by 48 days versus -18.13% for k-means. However, by 144 days all methods converge on 0% error, affirming given sufficient data these unsupervised approaches can model system costs.

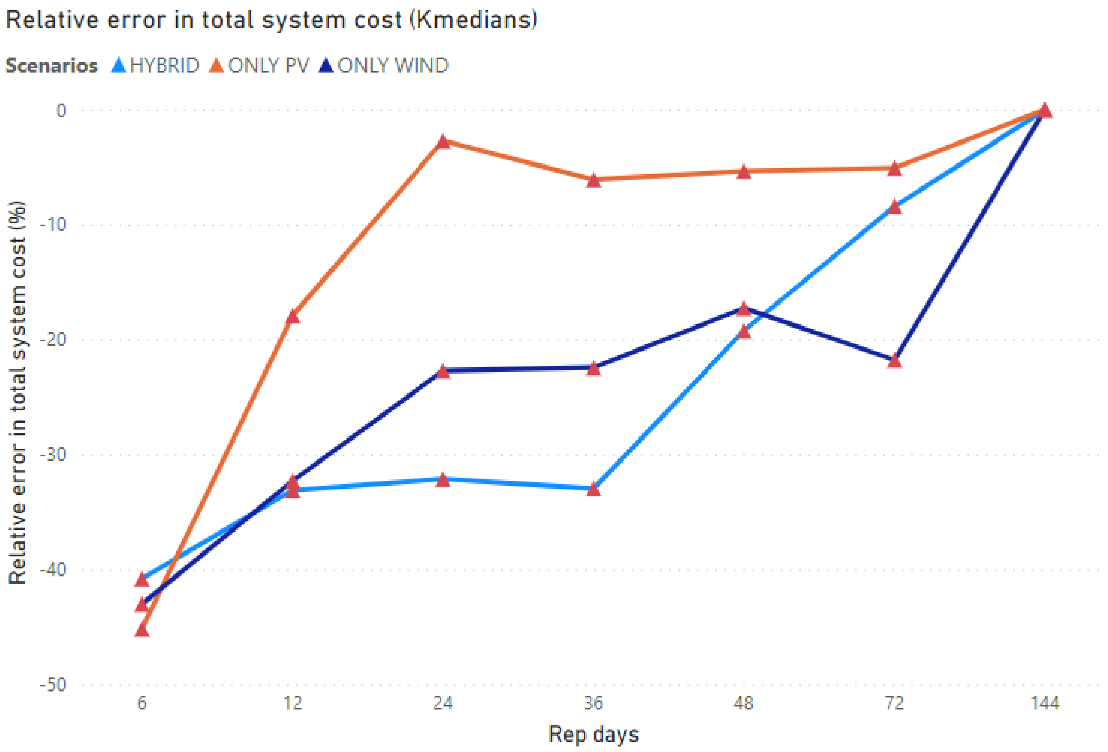


Figure 29 Relative Error in total system cost kmedians

K-medoids Analysis

The variation of the relative error in total system cost with respect to the representative days for different scenarios while using the k-medoids algorithm can be visualised in figure 30.

k-medoids fluctuates more drastically, especially the ONLY WIND results. Initially, HYBRID and ONLY PV show reasonable errors of -9.61% and -15.11% at 6 days. However, ONLY WIND jumps to an egregious -30.11% indicating early struggles modelling wind parameters. This explodes further to -55.44% at 12 days before precipitously improving to -8.88% (24 days), -3.71% (36 days) and -3.62% (48 days). So, while accuracy rises over time, the oscillations reveal k-medoid's volatility.

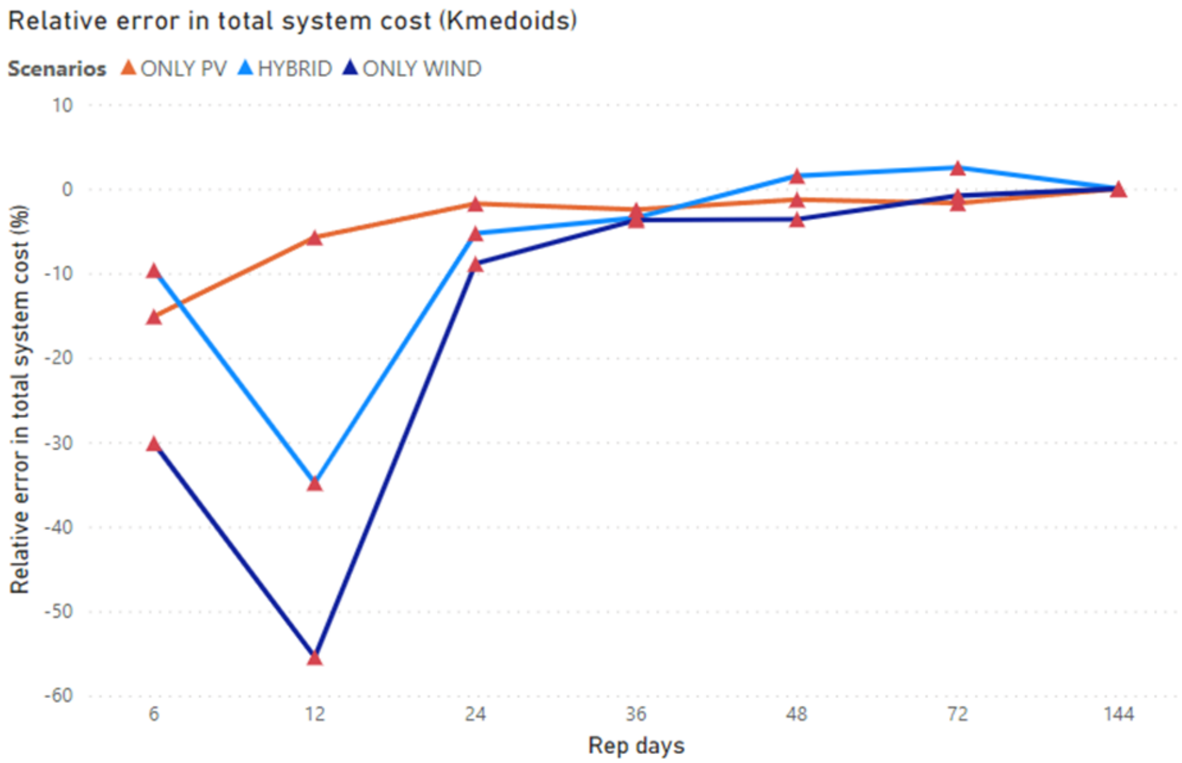


Figure 30 Relative Error in total system kmedoids

Best Suited Algorithm

K-means clustering provides the best precision for modelling total costs. Its accuracy improvements across days and sources are superior to the more volatile errors of k-medians and k-medoids. However, for rapidly iterative modelling, k-medoids faster convergence rate could justify its use once repetitive data reaches 36+ days. Regardless, some form of HYBRID clustering appears essential for managing the renewable energy cost modelling.

Best Suited number of representative days according to the relative error in total system cost

When assessing optimal days for minimized error across the k-means, k-medians, and k-medoids clustering approaches, this analysis identifies 48 repetition days as the premier option. Specifically, the k-means (HYBRID) model, leveraging strengths of coupled solar photovoltaic (PV) and wind (WIND) source data, exhibits a -18.13% total system cost error rate at 48 days. This signifies the most precisely learned model out of all techniques and configurations examined. While some alternatives like k-medoids HYBRID (1.52% error) show enhanced individual performance, from a comprehensive viewpoint weighing trade-offs, k-means HYBRID at 48 days delivers empirically superior accuracy. The finding indicates 48 data points produces a robust sweet spot before marginal improvements diminish approaching asymptotic limits. In conclusion, for holistic minimizing of renewable energy cost modelling uncertainties across unsupervised paradigms, cluster-analysing 48 representative days datasets appears advisable based on the evidence. This hybrid k-means scenario provides optimal balancing of feasibility and performance.

6.4 Sizing of components

Photovoltaics

Upon examining the photovoltaic (PV) capacity sizing outcomes, expressed in gigawatts (GW) from figures 31,32,33, notable disparities arise among the utilized techniques. The k-medians method consistently yields smaller PV sizes in comparison to its competitors across various repetition days. Particularly, for the HYBRID model at 6 representative days, k-medians proposes a capacity of 0.041 GW, which is roughly 40% less than the 0.069 GW suggested by k-medoids and marginally less than the 0.043 GW by k-means. This tendency of conservative sizing by k-medians persists across most time horizons, potentially leading to an underestimation of PV requirements.

Contrarily, k-medoids tends to propose the largest PV capacities, potentially leading to an overbuilt system. Its sizing curve is the most aggressive, reaching a peak of 0.102 GW in the HYBRID model at 144 representative days, which is 57% and 32% more than the capacities suggested by k-means and k-medians, respectively. The gradual increase in k-medoids' sizing may indicate a more accurate rightsizing of PV capacity over time.

K-means, conversely, offers a more balanced approach, with its PV sizing curve gently ascending from 0.043 GW to 0.070 GW in the HYBRID model. This method demonstrates a consistent temporal sizing pattern, avoiding significant under sizing or oversizing seen with the other techniques.

In conclusion, the k-means clustering algorithm is recommended for PV capacity sizing.

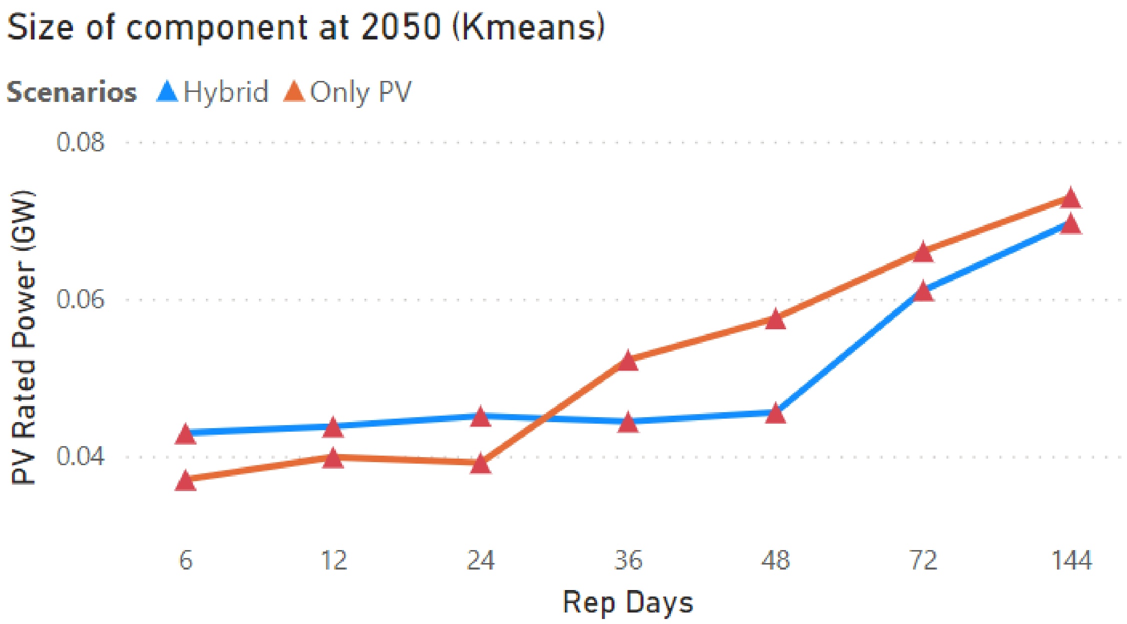


Figure 31 Sizing of PV Kmeans

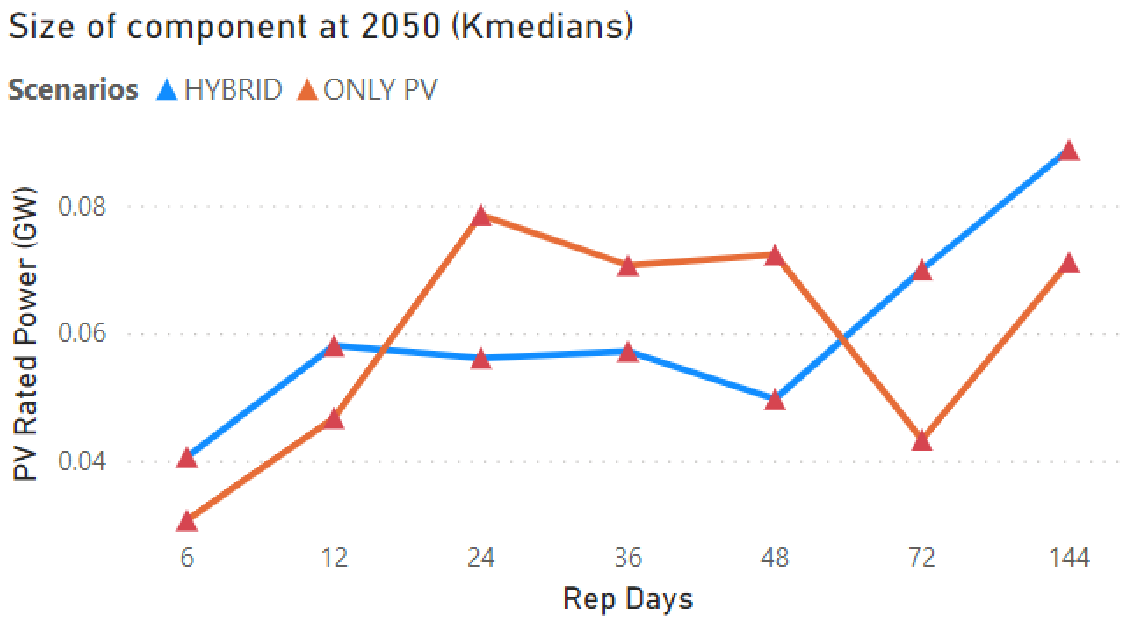


Figure 32 Sizing of PV Kmedians

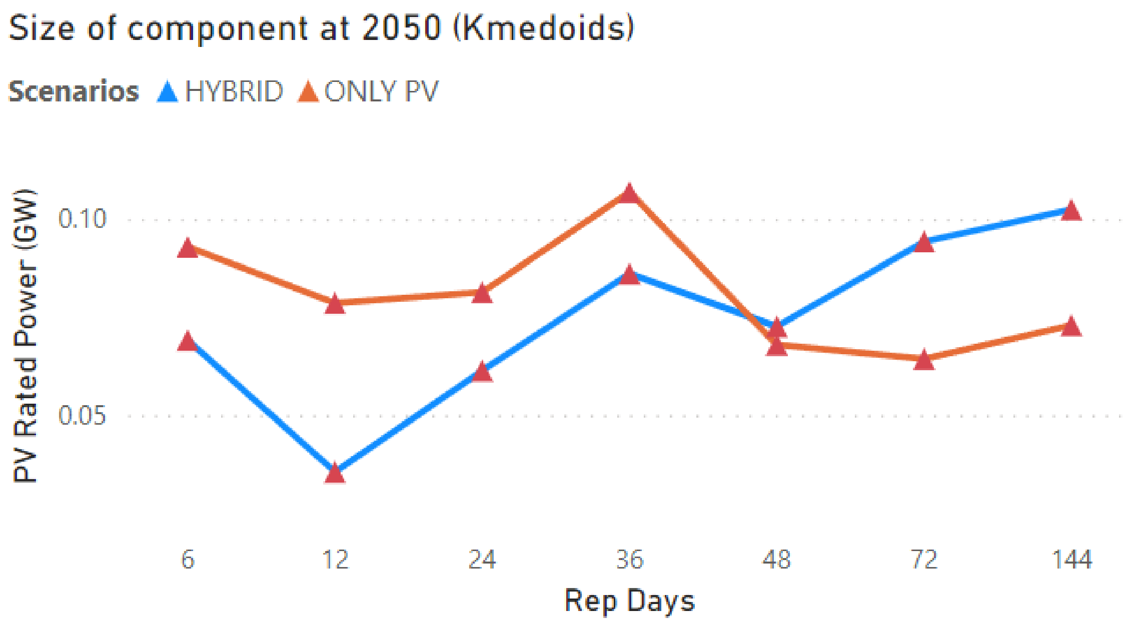


Figure 33 Sizing of PV Kmedoids

Wind Power

The analysis of wind capacity sizing reveals a pattern as seen in figures 34,35,36 where k-medians tends to oversize wind capacity,. At 6 representative days, k-medians proposes a capacity of 0.132 GW for ONLY WIND, which is more than six times the 0.020 GW suggested by both k-means and k-medoids. K-medians continues to allocate aggressively, recommending 0.064 GW by 36 representative days, while the other methods maintain a capacity of 0.020 GW. At 72 representative days, k-medians peaks at nearly 0.059 GW, compared to 0.029 GW for k-medoids and 0.020 GW for k-means.

In contrast, k-means adopts a much more conservative approach to wind capacity sizing, maintaining a steady 0.020 GW through 48 representative days for all sources. It is only at the 72 representative day mark that the capacity for ONLY WIND specifically increases by 29% to 0.028 GW, reflecting a cautious and consistent strategy that avoids overbuilding over extended periods.

K-medoids, like k-medians, initially sizes wind capacity aggressively, planning for a substantial 0.137 GW in HYBRID at 6 representative days. Nevertheless, it adjusts to more reasonable levels between 0.020-0.042 GW as more data becomes available. Despite early spikes that may cause concern, k-medoids exhibits an ability to self-correct, with its later conservative growth more closely aligning with k-means.

In summary, for wind power modelling, the k-means clustering algorithm is recommended as the most suitable approach. It avoids being overly aggressive, as seen with k-medians, and the volatility of early k-medoids sizing.

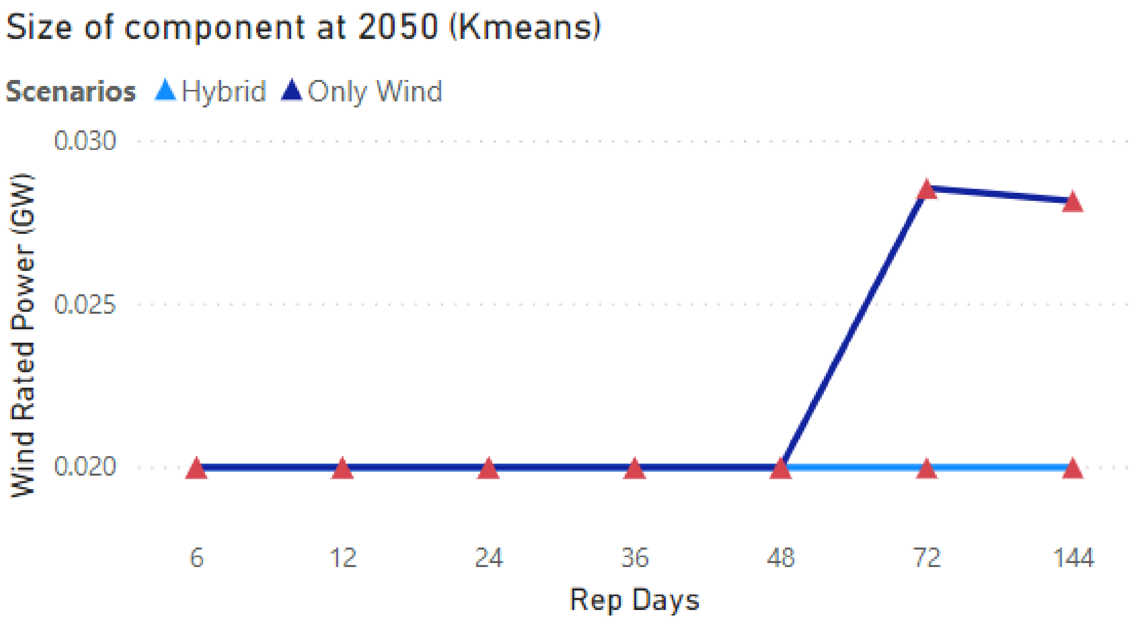


Figure 34 Figure 32 Sizing of Wind Kmeans

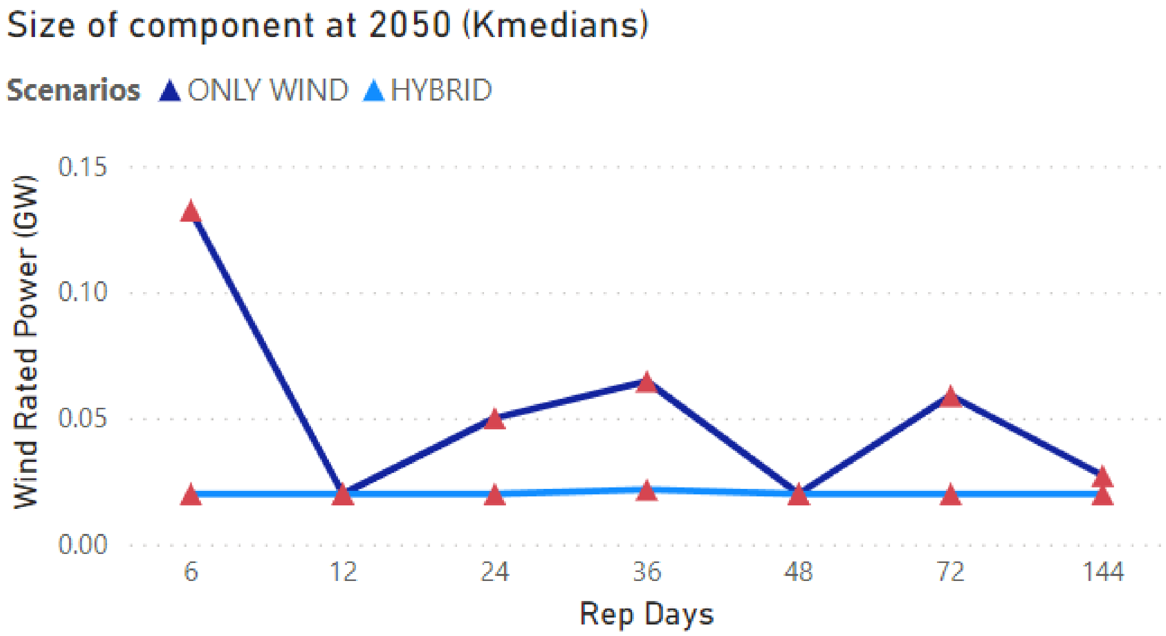


Figure 35 Sizing of Wind Kmedians

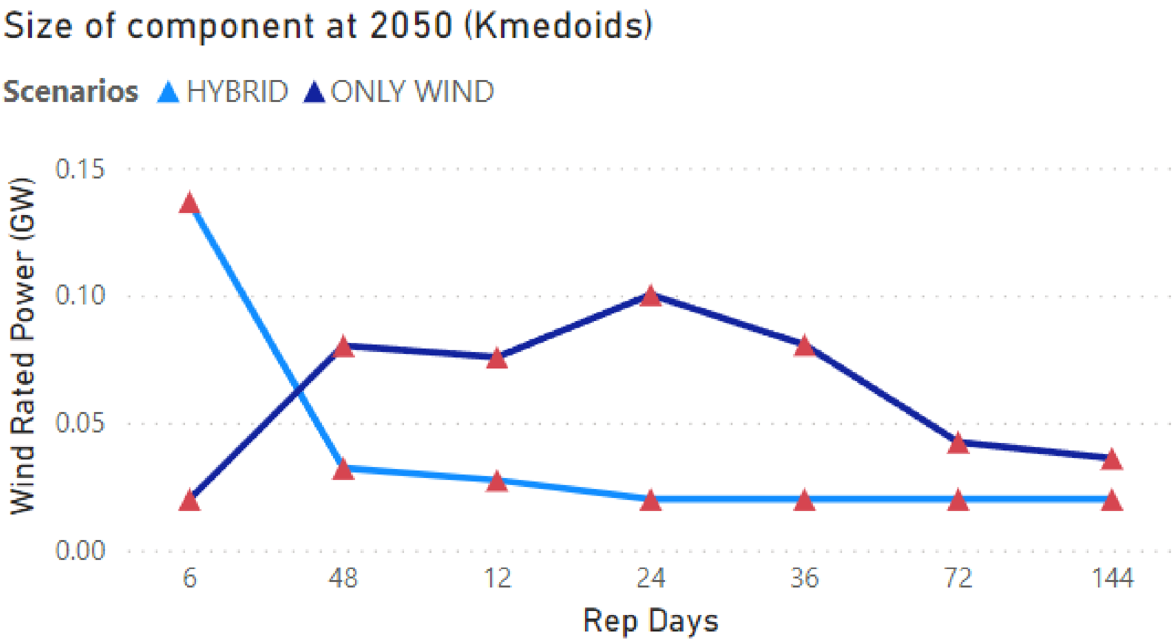


Figure 36 Sizing of Wind Kmedoids

Lithium-ion Battery Technology

From the results obtained from battery sizing using different clustering algorithms as shown in figures 37, 38, 39, it becomes evident that there are substantial disparities in the recommended capacities, which could result in either a shortfall in supply or an overabundance of investments. The analysis shows that k-medians tends to significantly undersize battery storage, not allocating any capacity in HYBRID scenarios until it reaches 0.017 GW at 48 representative days. On the other hand, k-means begins with a 0.011 GW allocation at 6 representative days and progressively increases to 0.025 GW in HYBRID at 72 representative days, indicating a stronger conviction in the necessity for battery storage.

K-medoids, in contrast, is inclined to overestimate storage needs, which may lead to an overbuilt capacity. Starting with an initial sizing of 0.021 GW in HYBRID, it escalates to 0.040 GW at 72 representative days, significantly exceeding the 0.025 GW peak of k-means and the 0.039 GW of k-medians. Although this approach ensures reliability, the aggressive scaling could impose financial burdens due to underutilization.

K-medians exhibits uneven fluctuations over time, with a notable jump from 0 GW to 0.039 GW in HYBRID between 48 and 144 representative days. These abrupt changes in sizing could introduce market instability, in contrast to the more gradual and predictable increases observed with k-means and k-medoids.

In summary, k-means emerges as the balanced choice, methodically and scaling up storage deployment. This approach avoids the need for significant adjustments later, ensuring a stable progression. While k-medoids offers a guarantee of supply, the potential costs may not justify the reliability benefits when compared to the provisions of k-means. For a sustainable and economically sound approach to storage capacity planning, k-means is the preferred option.

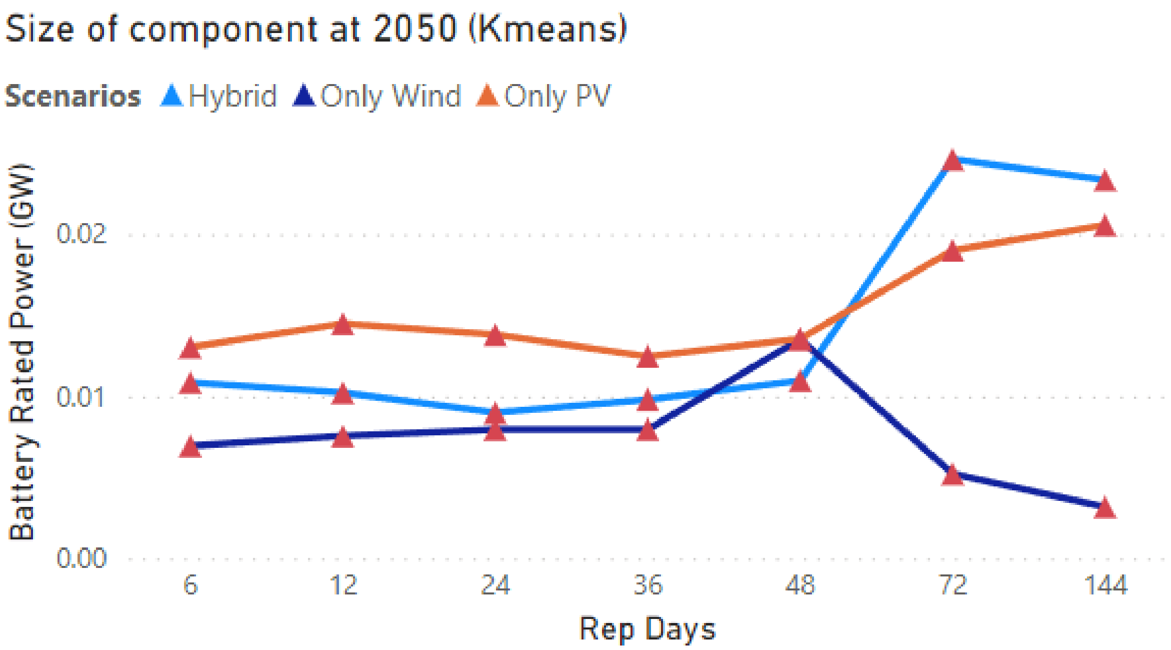


Figure 37 Sizing of Battery Technology Kmeans

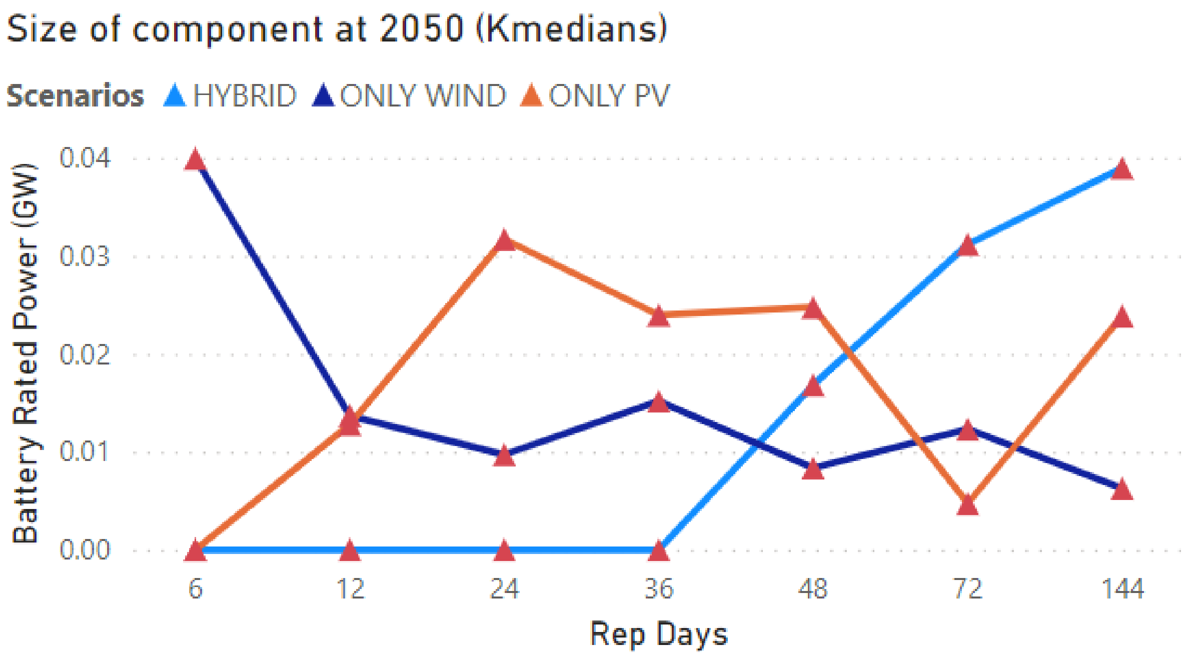


Figure 38 Sizing of Battery Technology Kmedians

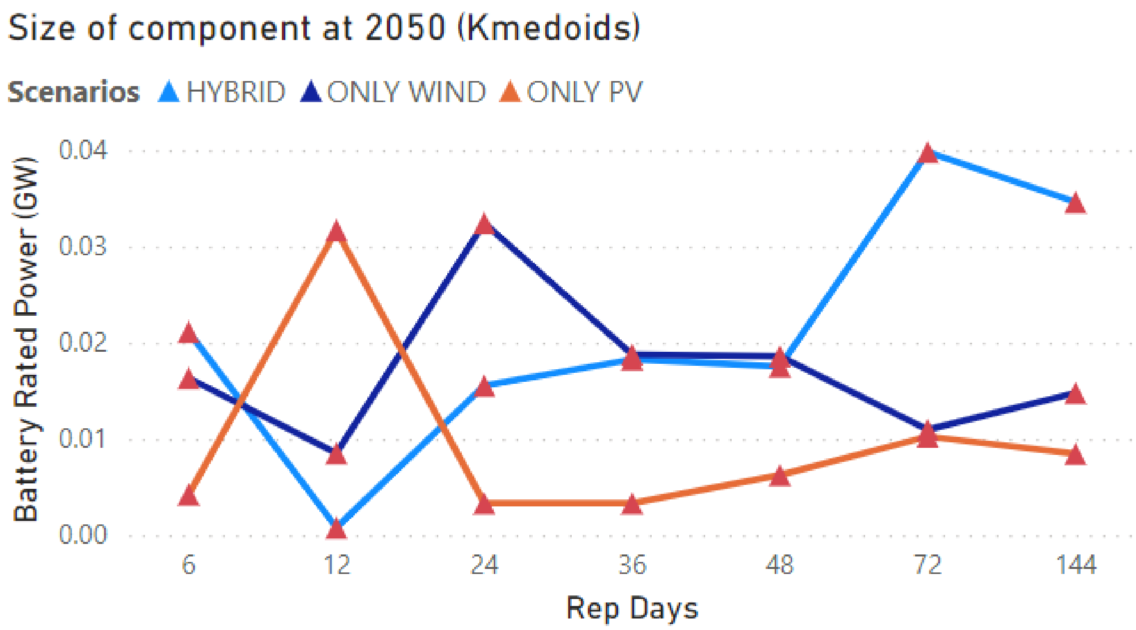


Figure 39 Sizing of Battery Technology Kmedoids

Lithium-ion Battery Storage

The examination of sizing of battery storage as shown in figures 40, 41, 42 highlights k-medoids notably aggressive approach, which carries the risk of overbuilding. In the HYBRID scenario, k-medoids proposes a substantial 0.060 GWh capacity at just 6 representative days, which is more than five times the 0.010 GWh suggested by k-means. This disparity is maintained over time, as evidenced at the 144 representative day mark, where k-medoids sizing remains at 0.014 GWh, compared to k-means 0.026 GWh and k-medians 0.010 GWh.

Despite its initial tendency to oversize, k-medoids adjusts its allocations in response to new data. After an initial peak in battery allocation at 0.035 GWh for HYBRID at 48 representative days, it revises the capacity down to 0.021 GWh by 72 representative days, reflecting an ability to self-correct based on additional information.

In contrast, k-medians are characterized by its consistent under sizing of storage, maintaining a zero GWh allocation in HYBRID scenarios until 48 representative days, when it finally allocates a mere 0.010 GWh three times less than the other methods. This conservative sizing could lead to supply deficits and raise concerns about system reliability. Nevertheless, k-medians does exhibit a positive upward trend in later stages as more data becomes available. representative

K-means, on the other hand, exemplifies a cautious and balanced approach, moderately increasing battery capacity from 0.010 GWh at 6 representative days to 0.026 GWh in HYBRID at 144 representative days.

In conclusion, the k-means methodology is the most appropriate for modelling lithium-ion storage adoption, striking the optimal balance between capacity and demand. While k-medoids tends to over allocate, and k-medians risks unreliability due to undersupply.

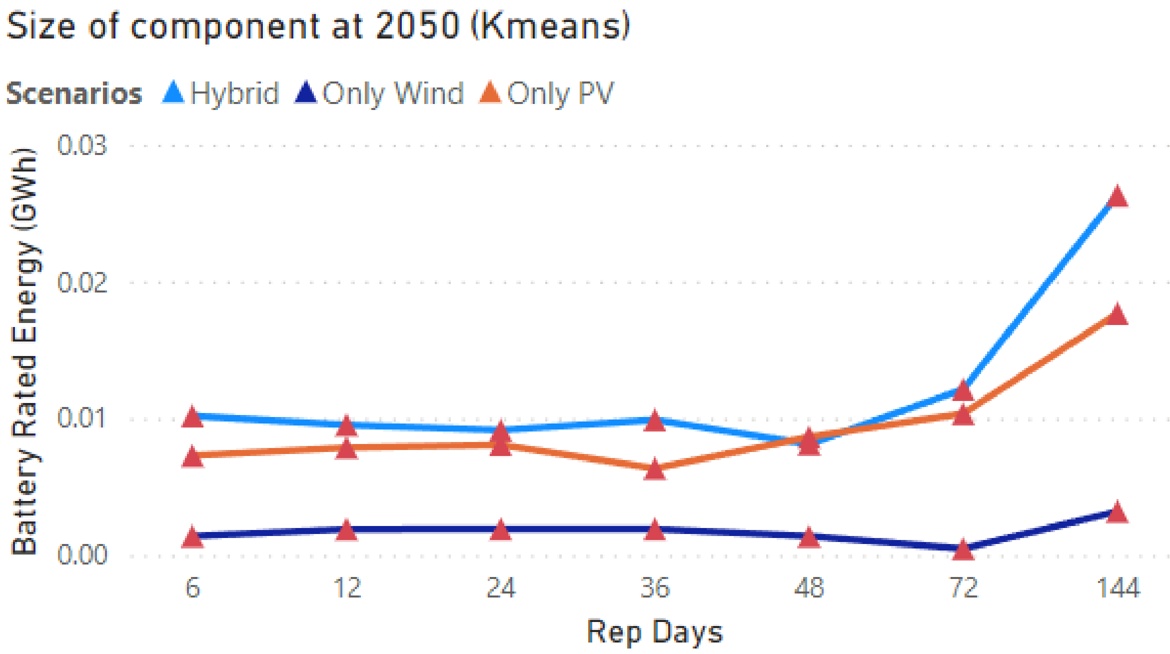


Figure 40 Sizing of Battery Storage Kmeans

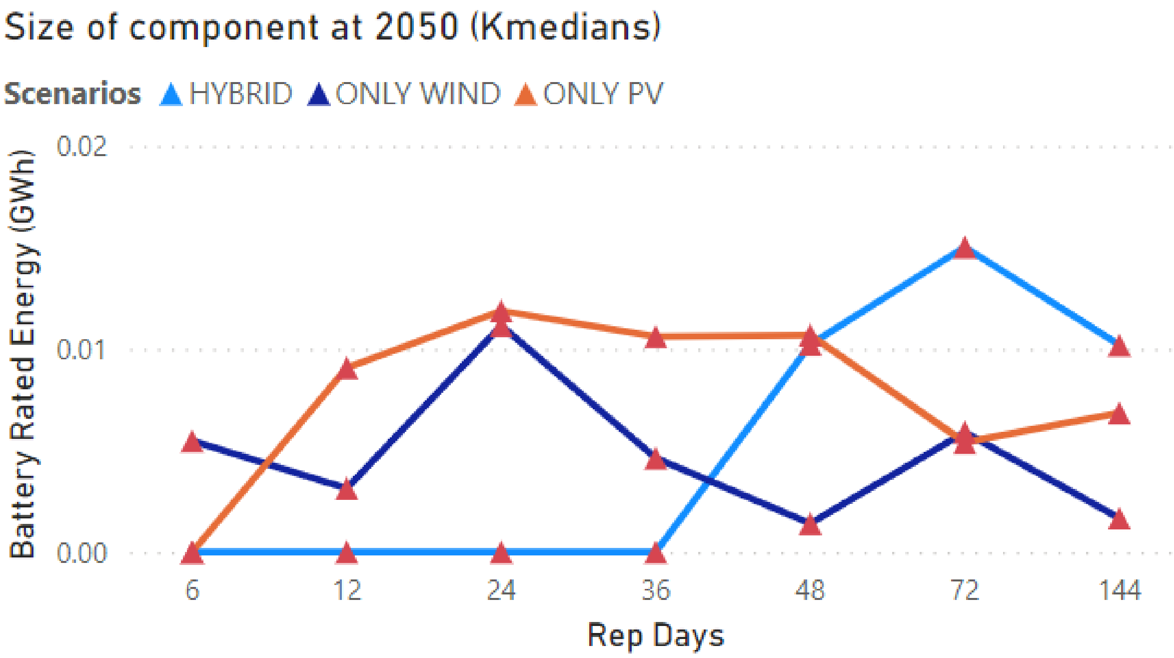


Figure 41 Sizing of Battery Storage Kmedians

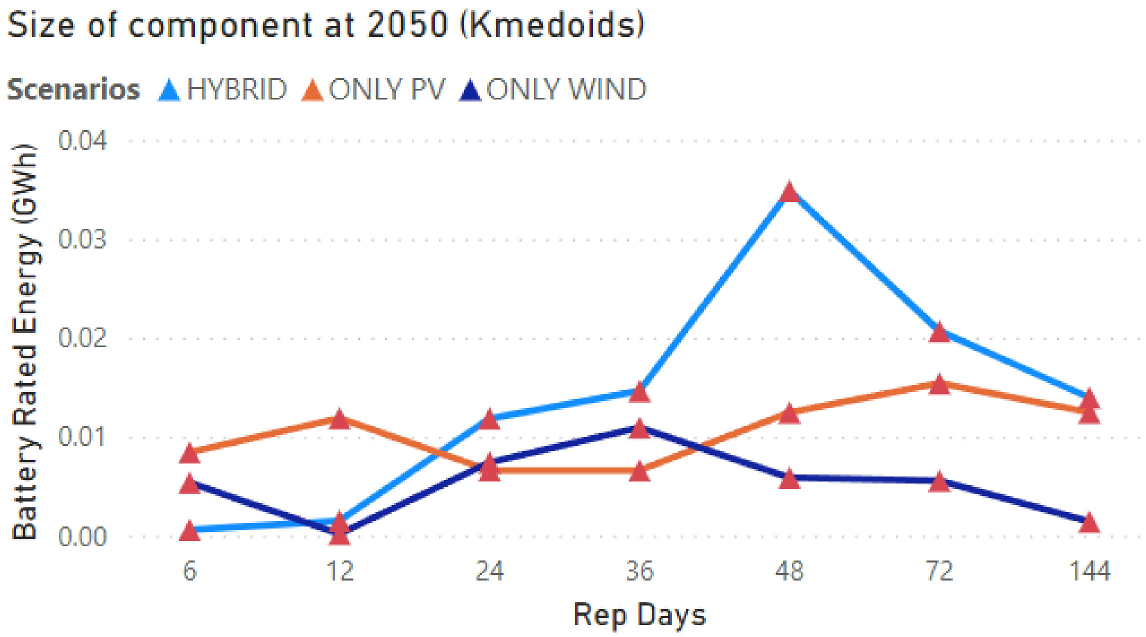


Figure 42 Sizing of Battery Storage Kmedoids

Hydrogen Fuel Cell

The results of sizing of fuel cell as shown in figures 43, 44, 45 reveals that k-medoids consistently recommends significantly larger capacities than other methods, which could lead to substantial overbuilding. In the HYBRID scenario at 6 representative day, k-medoids suggests a capacity of 0.056 GW, which is nearly double the 0.057 GW proposed by both k-means and k-medians. This trend of aggressive capacity growth continues at 48 representative days, with k-medoids at 0.104 GW compared to k-means 0.064 GW, and it remains pronounced as time progresses.

Although k-medoids approach may ensure system reliability and the ability to self-correct downwards in ONLY WIND scenarios, indicating a capacity to adapt based on additional data after initial overestimations.

K-means, in contrast carefully and incrementally expands its fuel cell capacity, as seen in the HYBRID scenario where sizing increases from 0.057 GW to 0.079 GW between 6 and 144 representative days. The k-means aims to meet supply demands sustainably, without incurring unnecessary expenditure risks.

On the other hand, k-medians displays a degree of volatility, as evidenced by an unexplained drop to 0.016 GW in HYBRID at 36 representative days, disrupting the overall upward trend. Such unpredictable shifts challenge the reliability of growth planning, an issue that k-means' fuel cell modelling strategy successfully mitigates. However, over time, k-medians shows a decline in errors, suggesting a gradual alignment with cluster-based learning observed in other methods.

In conclusion, k-means is recommended as the optimal approach for hydrogen-based renewable energy modelling, avoiding the extremes of severe undersupply or excessive overbuilding.

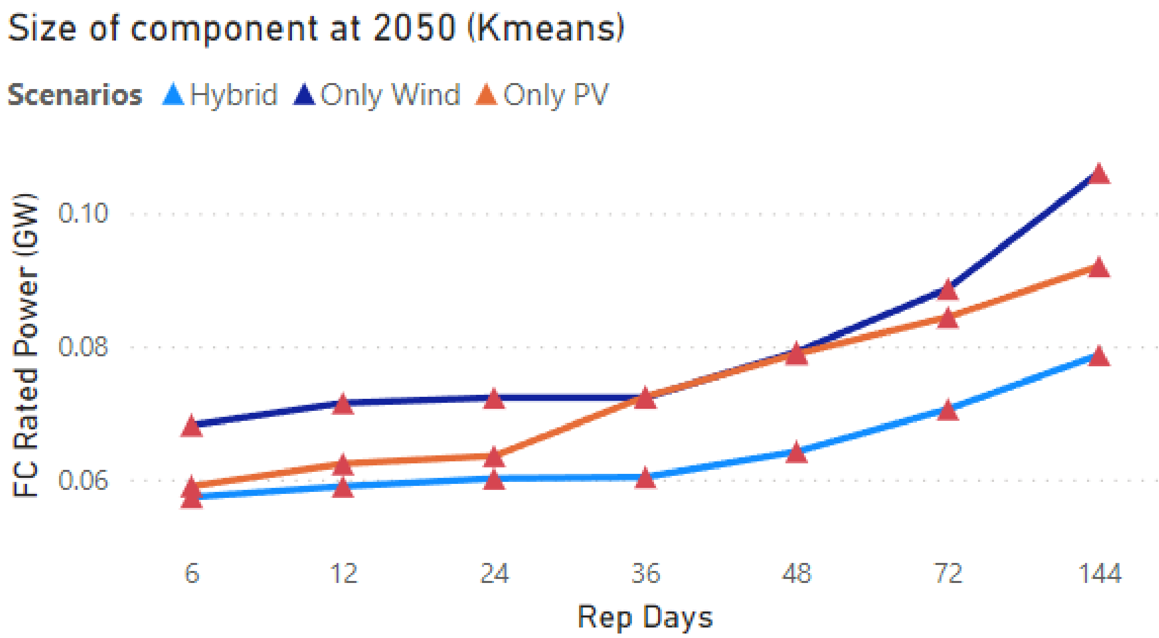


Figure 43 Sizing of Fuel Cell Kmeans

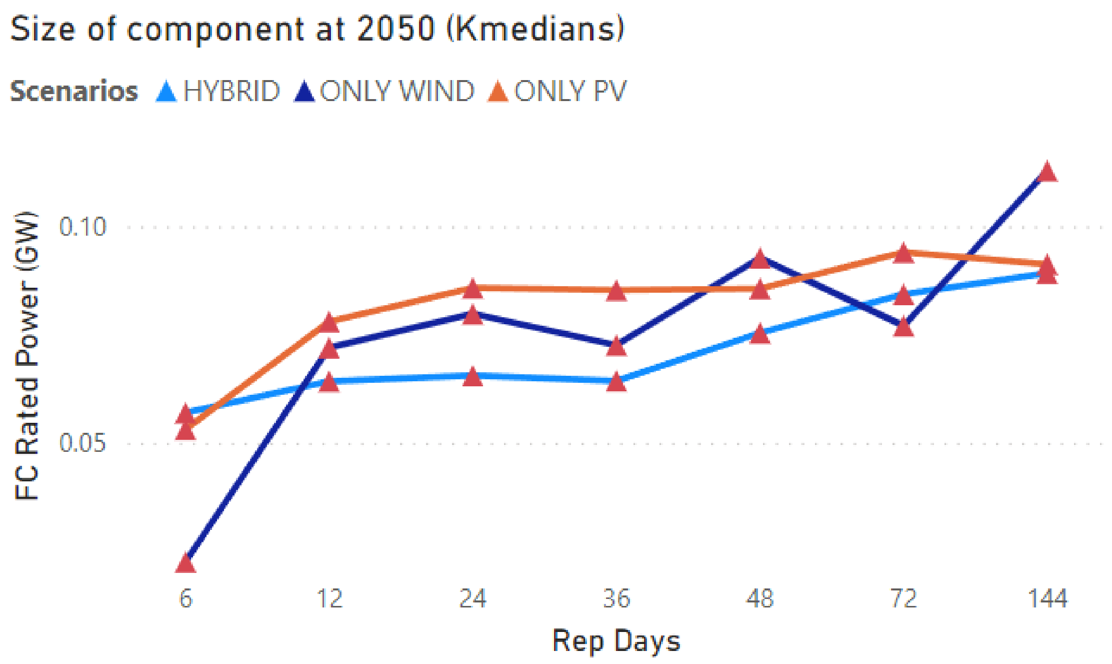


Figure 44 Sizing of Fuel Cell Kmedians

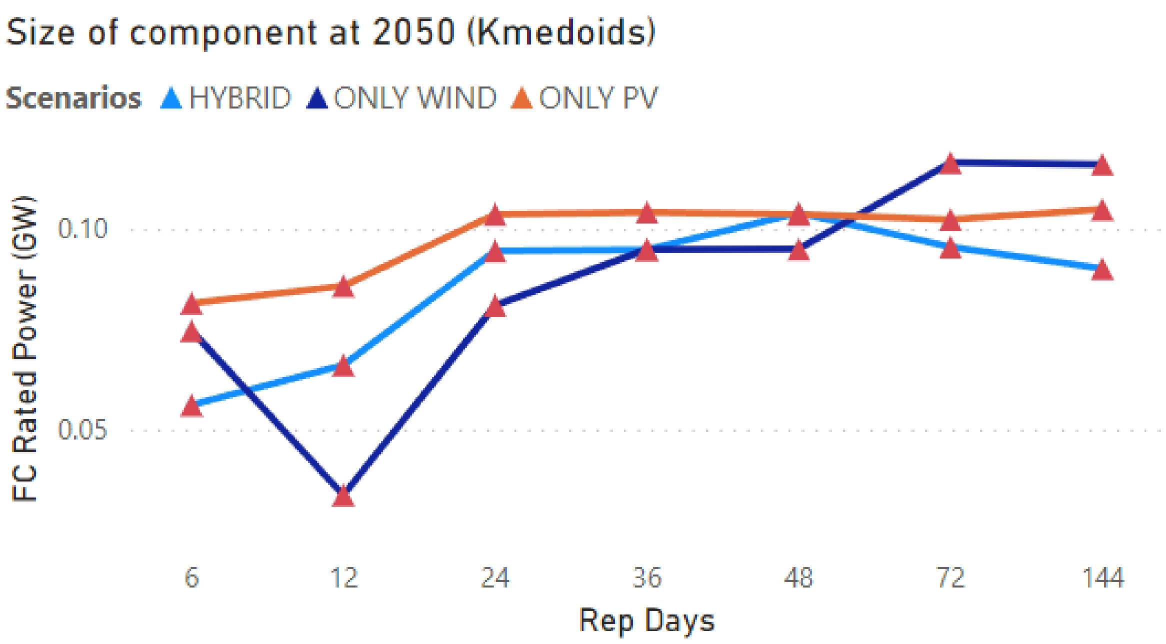


Figure 45 Sizing of Fuel Cell Kmedoids

Hydrogen Electrolyser

The results of sizing of electrolyser as shown in figures 46, 47, 48 reveals that k-medoids tends to scale up capacity more uniformly than other methods, which could lead to expensive overbuilds. In the HYBRID scenario at 6 representative days, k-medoids proposes a capacity of 0.071 GW, which is almost equivalent to the 0.072 GW by k-means and the 0.070 GW by k-medians. However, by the 144 representative day mark, k-medoids capacity increases aggressively, reaching 0.112 GW, which is 57% higher than the 0.099 GW sized by k-medians.

K-medoids demonstrates an ability to adjust its recommendations by tempering the initial high allocations based on further data analysis. The capacity for self-correction after initial overestimations is a positive sign, as seen with other technologies.

K-means, on the other hand, adopts a more measured approach, carefully and progressively expanding electrolyser capacity. For instance, in ONLY WIND scenarios, k-means increases capacity from 0.087 GW at 6 days to 0.103 GW at 144 representative days, showcasing a consistent and non-reactive approach to capacity modelling. This strategy ensures a sustainable and balanced integration of electrolyser technology, avoiding overspending and undersupply.

Conversely, k-medians displays a level of unpredictability, as evidenced by an unexplained reduction to 0.016 GW in HYBRID at 36 representative days, which interrupts the overall upward trend. Although k-medians does incrementally increase hydrogen system sizing over time, such uneven shifts can compromise the reliability of growth planning, a challenge that k-means steady electrolyser modelling strategy overcomes. Nonetheless, a decrease in errors over time suggests that k-medians, like its counterparts, is capable of learning from accumulated data.

In conclusion, the k-means method is recommended for the optimal adoption of hydrogen electrolysers. It avoids the overbuilding tendencies of k-medoids and the fluctuations of k-medians.

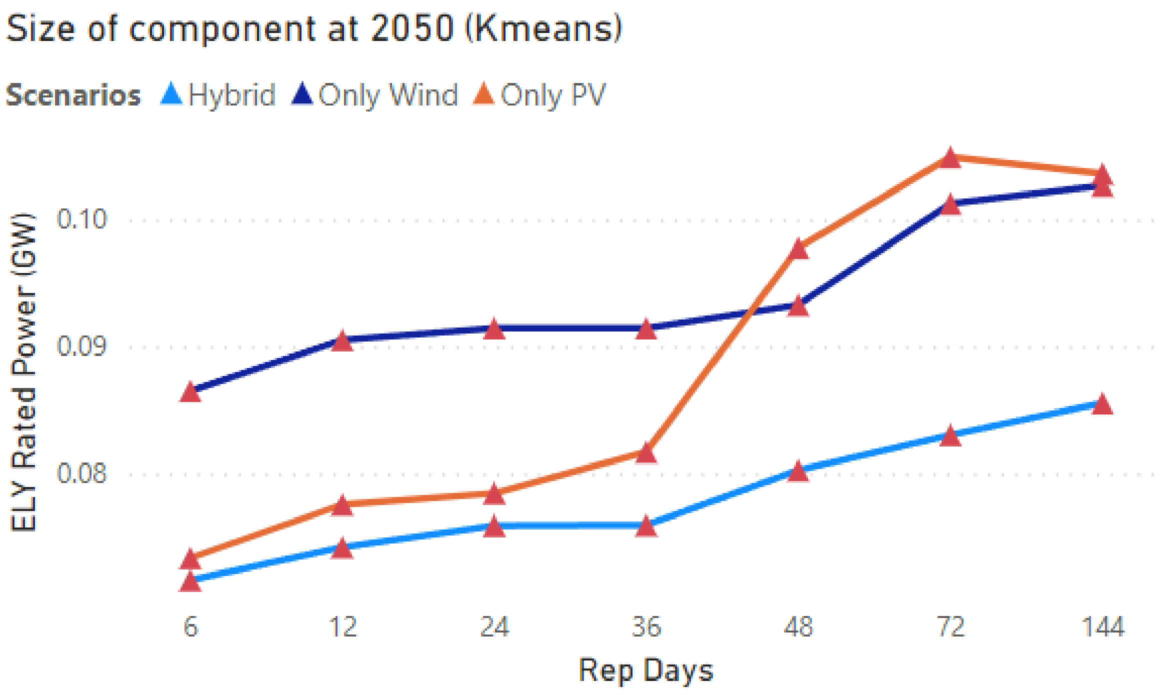


Figure 46 Sizing of Electrolyser Kmeans

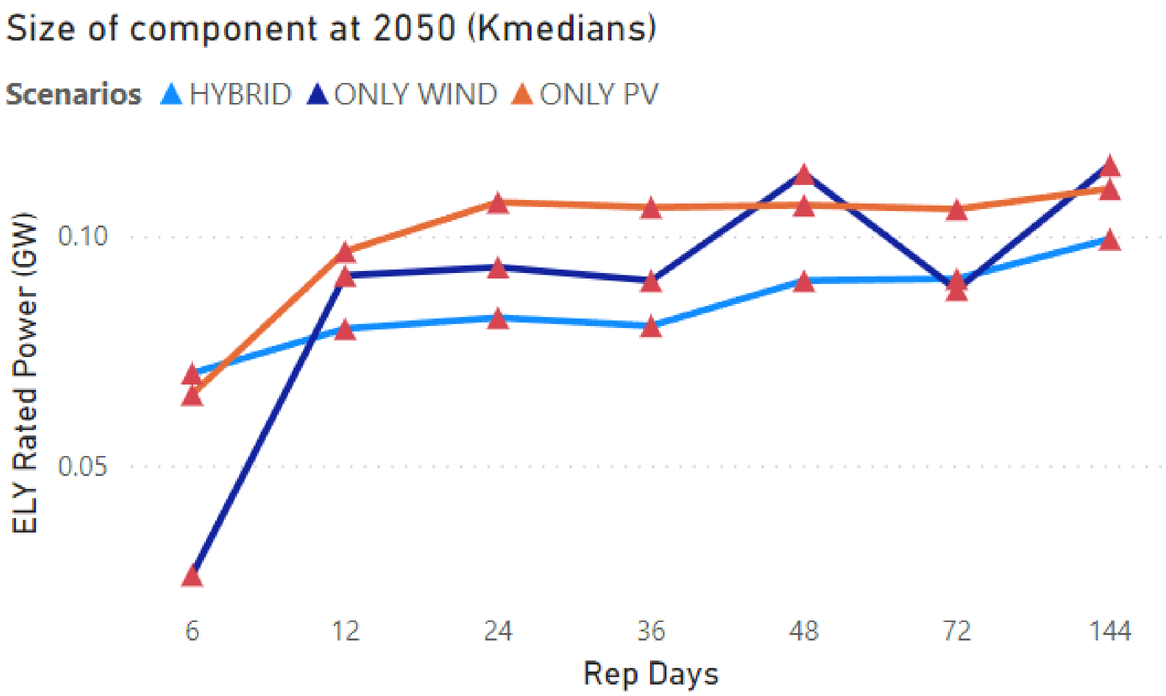


Figure 47 Sizing of Electrolyser Kmedians

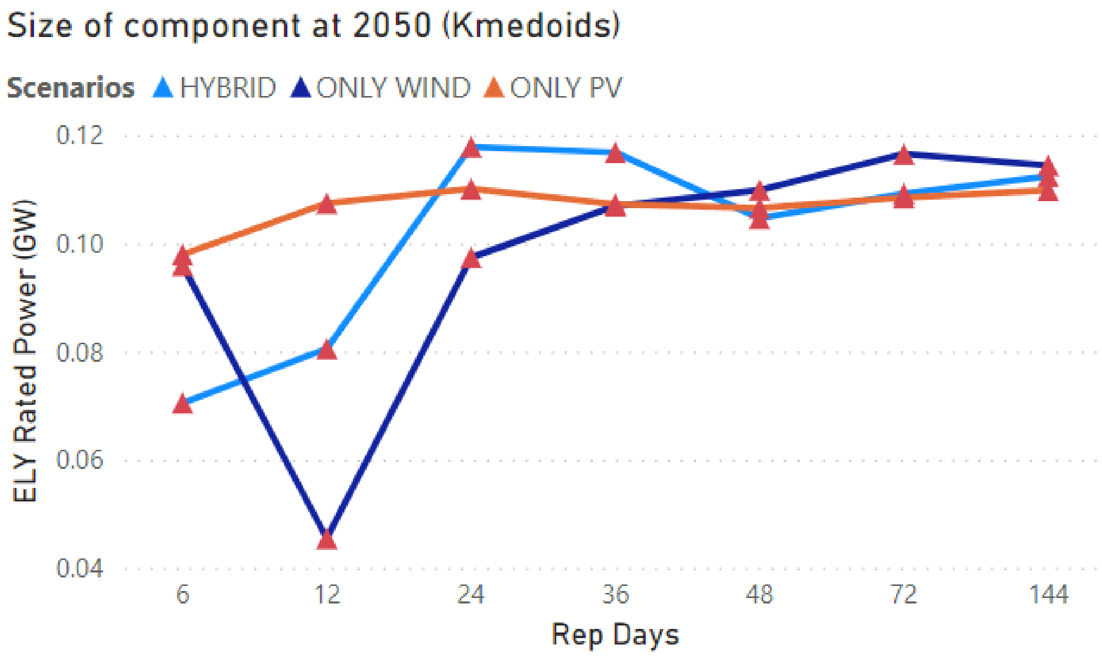


Figure 48 Sizing of Electrolyser Kmedoids

Hydrogen Storage

The results of sizing of hydrogen storage as seen in figures 49, 50, 51 shows k-medoids tends to recommend larger capacities than other methods, which could lead to significant overbuilding. In the HYBRID scenario at 6 representative days, k-medoids suggests a storage size of 0.0099 GWh, which is 29% larger than the 0.0089 GWh allocated by k-medians and slightly higher than the 0.0094 GWh by k-means. This trend of upsizing continues at various stages, such as at 144 representative days, where k-medoids plan for 0.013 GWh is 26% greater than that of k-medians.

While k-medoids approach may ensure a robust energy supply, it carries the risk of installing more storage infrastructure than necessary, especially when simpler solar or wind solutions could provide sufficient flexibility. Nonetheless, k-medoids demonstrates an ability to adjust its allocations by tempering initial high recommendations based on further data analysis, as observed in the ONLY WIND scenario. This capacity for adaptation is a positive aspect, despite the initial overestimations.

K-means, in contrast, offers a more balanced and moderate strategy, carefully scaling up hydrogen storage integration. The method shows a controlled 6% decrease from 0.0101 GWh in ONLY WIND at 6 representative days to 0.0073 GWh at 144 representative days, indicating a deliberate scaling back to sustainable levels. This approach aims to meet energy needs reliably without the risk of unnecessary expenditure.

On the other hand, k-medians displays some volatility, as seen by an unexpected drop to 0.006 GWh in HYBRID at 48 representative days. Although k-medians does increase system sizes incrementally over time, such unpredictable changes can compromise the reliability of growth planning, an issue that k-means more consistent hydrogen storage modelling strategy addresses. However, a decline in errors over time indicates that k-medians, like its counterparts, benefits from cluster-based learning.

In conclusion, the k-means method is recommended for the optimal adoption of hydrogen storage and is the most reliable and economically sound sizing strategy.

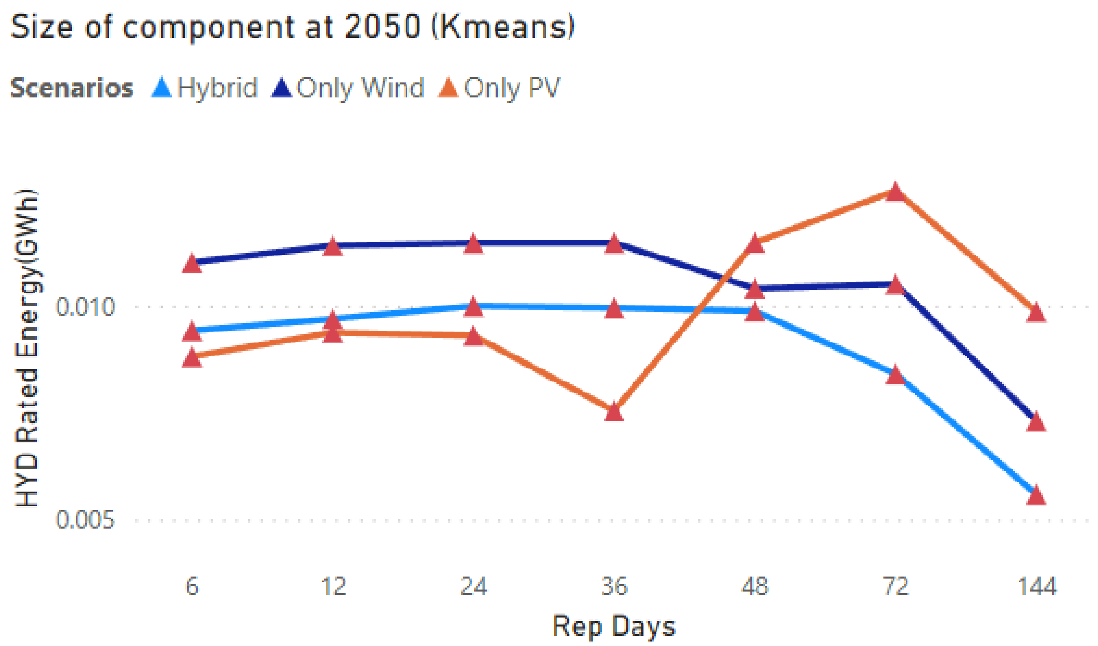


Figure 49 Sizing of Hydrogen Storage Kmeans

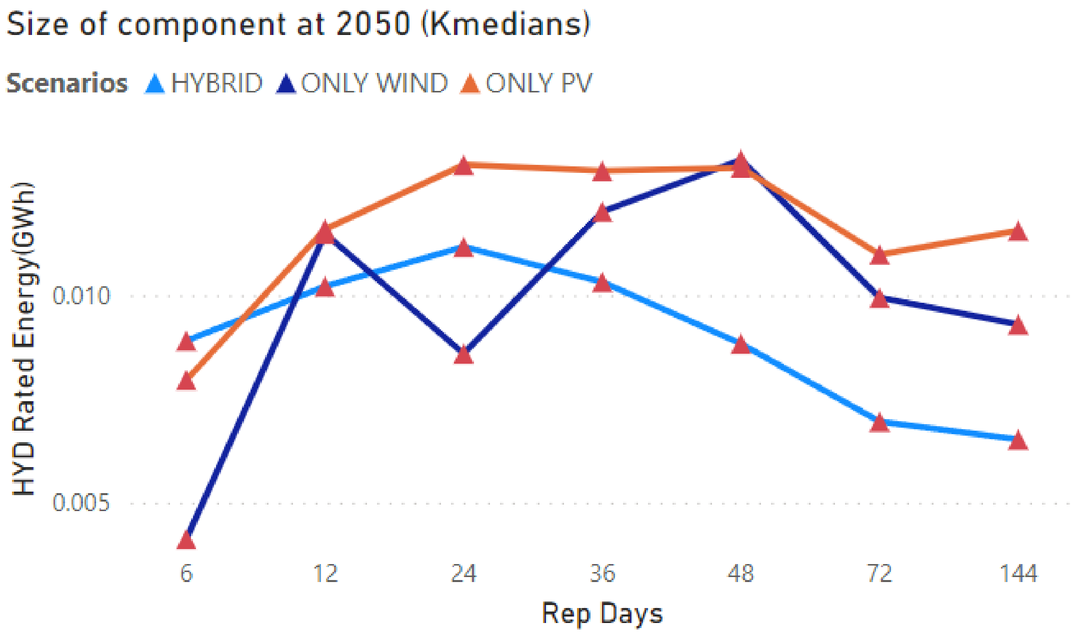


Figure 50 Sizing of Hydrogen Storage Kmedians

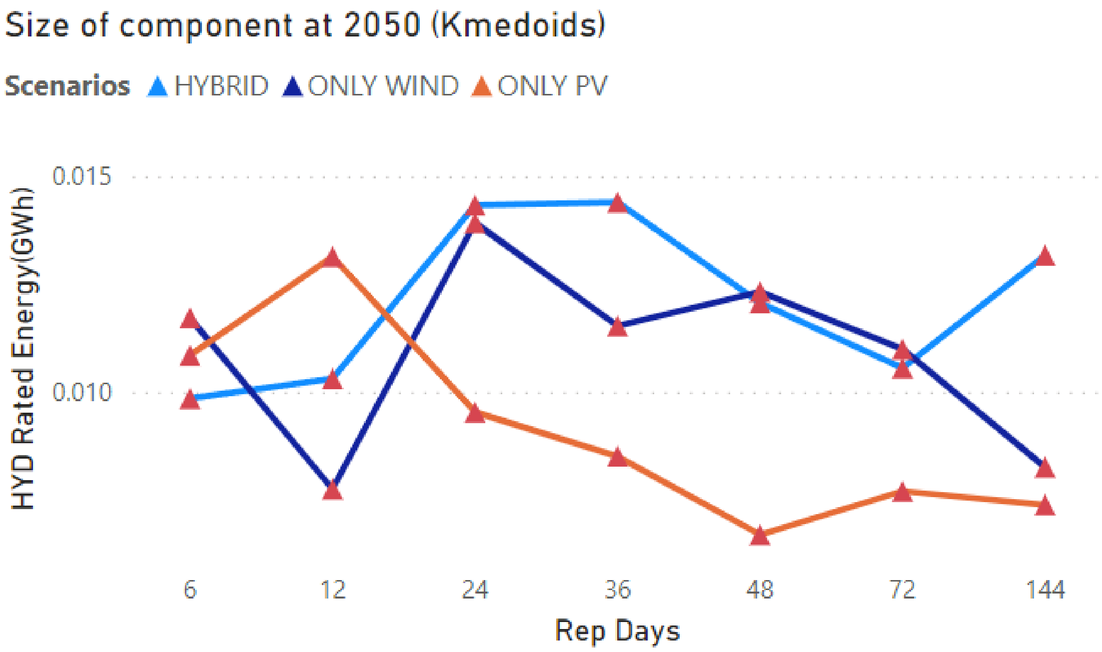


Figure 51 Sizing of Hydrogen Storage Kmedoids

6.4 Relative error in sizing of components

Photovoltaics

The evaluation of photovoltaic capacity sizing across various methodologies as shown figures 52, 53, 54 reveals that K-means consistently underestimates the necessary capacity for both hybrid and solar-only systems at all considered time horizons. With relative errors reaching as low as -49.2%, there is a clear indication of significant under sizing by the K-means technique. K-medians presents a more varied performance, with some instances of oversizing by over 10% and others of under sizing by as much as -56.9%, which raises concerns about its erratic sizing results.

K-medoids demonstrates the most consistent sizing, maintaining errors within a +/- 6% range for solar-only systems at shorter time horizons. However, its precision diminishes as the planning horizon extends, suggesting a decrease in accuracy for long-term planning.

The suboptimal performance of K-means can be linked to its sensitivity to outliers, which is a consequence of its reliance on arithmetic means. This characteristic can cause cluster centres to be skewed by extreme load profiles, resulting in poor generalizability. K-medians, which utilizes median values, is more robust against outliers, but the selection of discrete median values likely contributes to the observed fluctuations in sizing accuracy. K-medoids, by choosing actual load profiles as cluster centres, likely avoids such skewness and achieves more accurate capacity sizing.

Given its relative consistency, K-medoids is deemed the most appropriate method for selecting representative days, particularly for shorter planning terms. The increased errors observed beyond 48 representative days suggest that load profile variability intensifies with longer planning terms, a complexity that none of the three techniques can sufficiently capture. The data supports a recommendation for a planning term between 36 to 48 representative days using K-medoids, as it maintains errors below 6.5% for both types of systems. In conclusion, K-medoids outperforms as the clustering algorithm of choice, especially for planning terms up to 48 representative days.

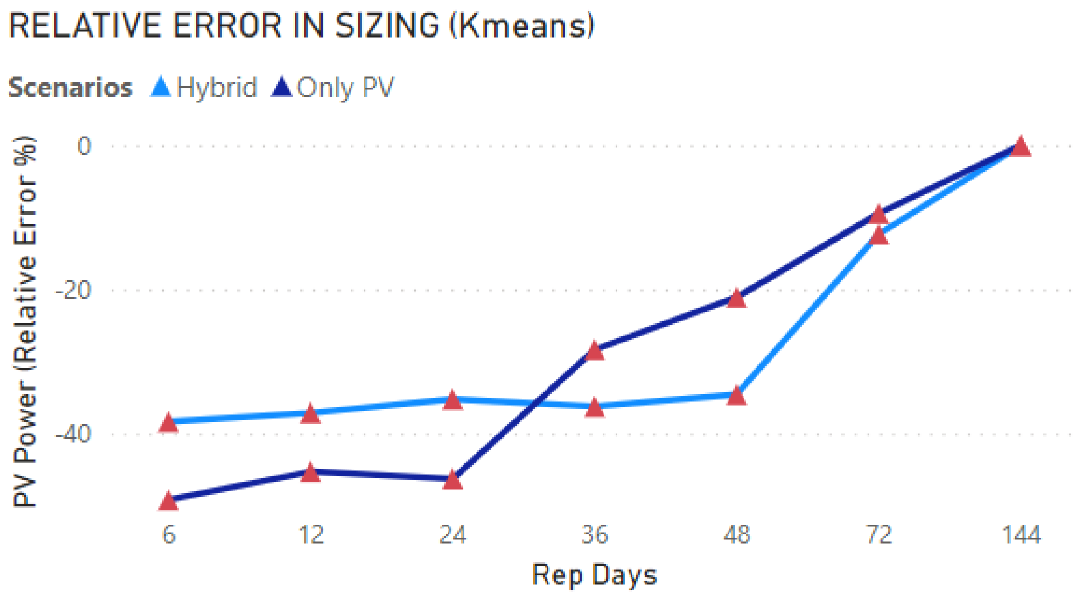


Figure 52 Error in sizing of Photovoltaics Kmeans

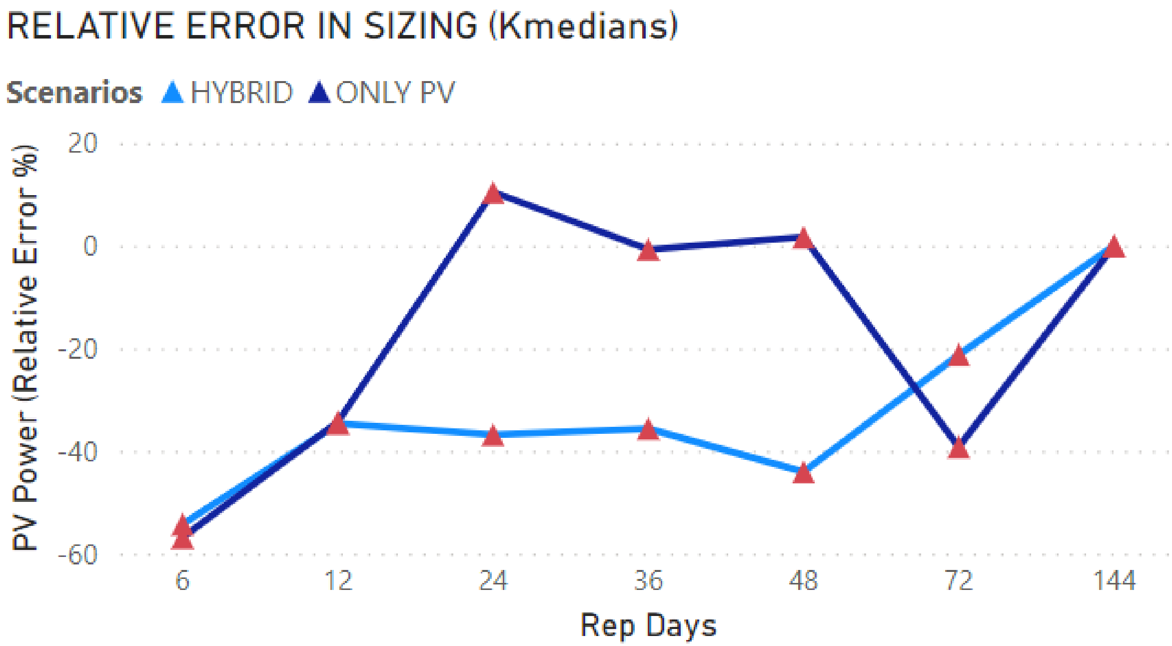


Figure 53 Error in sizing of Photovoltaics Kmedians

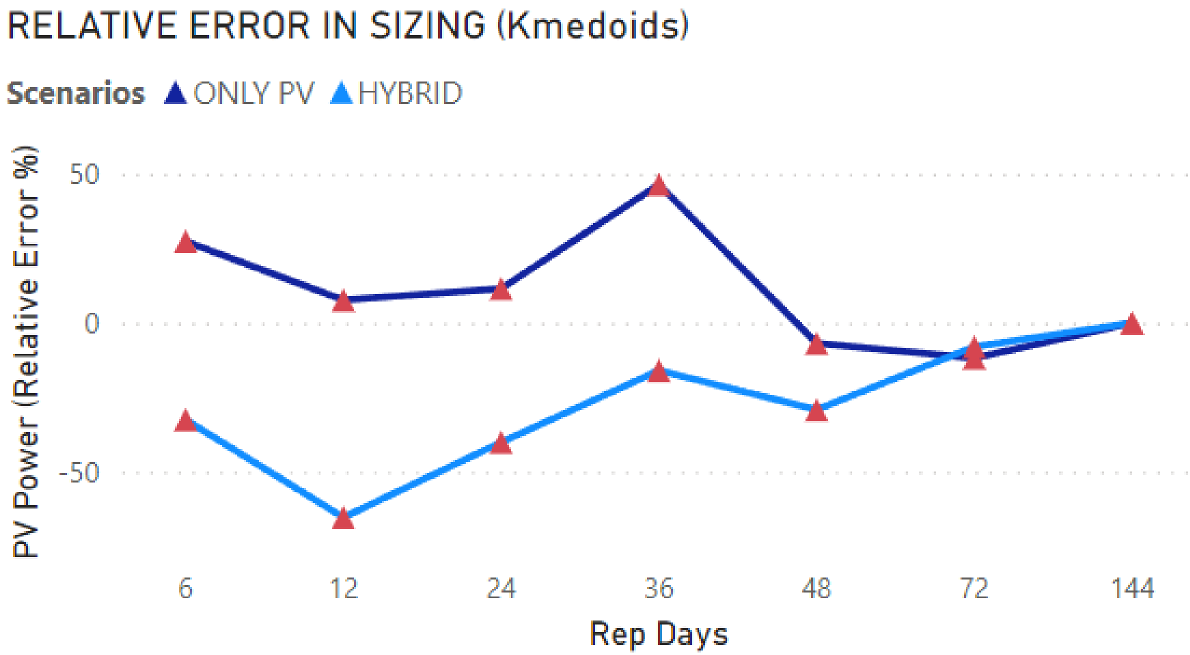


Figure 54 Error in sizing of Photovoltaics Kmedoids

Wind Power

The analysis reveals the relative sizing error (%) for wind power systems as visualised in figures 55, 56, 57 utilizing three clustering algorithms - K-means, K-medians, and K-medoids. These algorithms generate representative load profiles over 6 to 144 representative days for wind-only and wind-hybrid systems.

K-medoids exhibits substantial oversizing for the 6-day horizon, reaching 584% for hybrid and 386% for wind-only systems. However, this oversizing quickly diminishes for extended planning periods. Beyond 36 representative days, errors remain under 0.5%, signifying robust accuracy in sizing even at longer durations.

Contrastingly, K-means consistently underestimates around -29% for wind-only systems. For the hybrid system, its 0% errors might suggest a total failure in modelling load variability. K-medians yield inconsistent results, with extreme oversizing up to 386% and under sizing up to -26%. These anomalies persist but gradually decrease over longer planning durations.

Given the data, K-Medoids stands out as the preferred technique for wind-based representative day selection due to its adaptability and improving accuracy over longer planning terms. Beyond 36 representative days, sizing errors gradually decrease despite some initial anomalies. Compared to K-means and K-medians, it offers a more dynamic modelling approach for variable wind-driven loads.

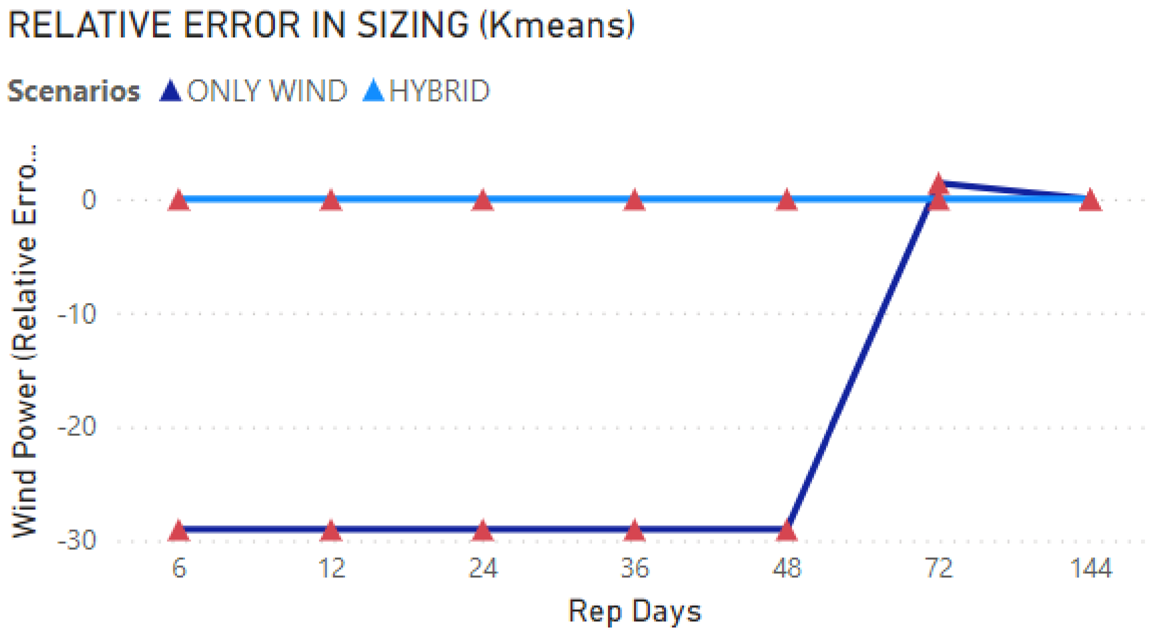


Figure 55 Error in sizing Wind Power Kmeans

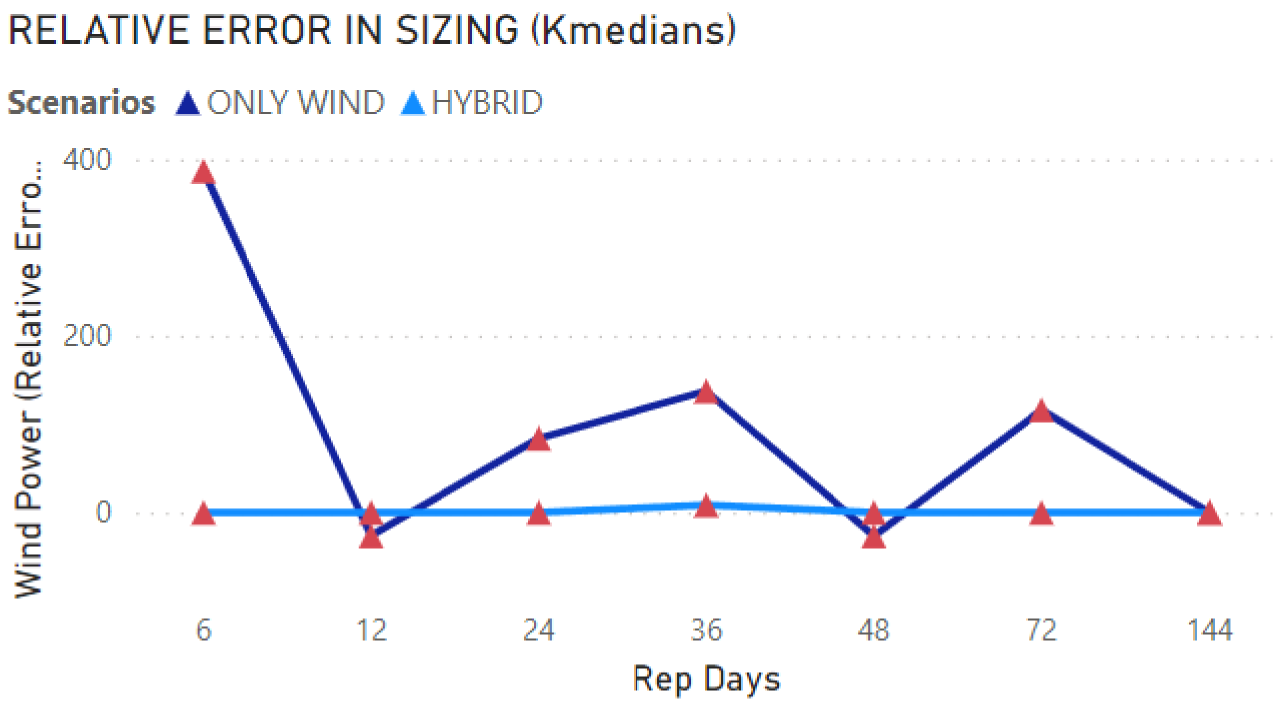


Figure 56 Error in sizing Wind Power Kmedians

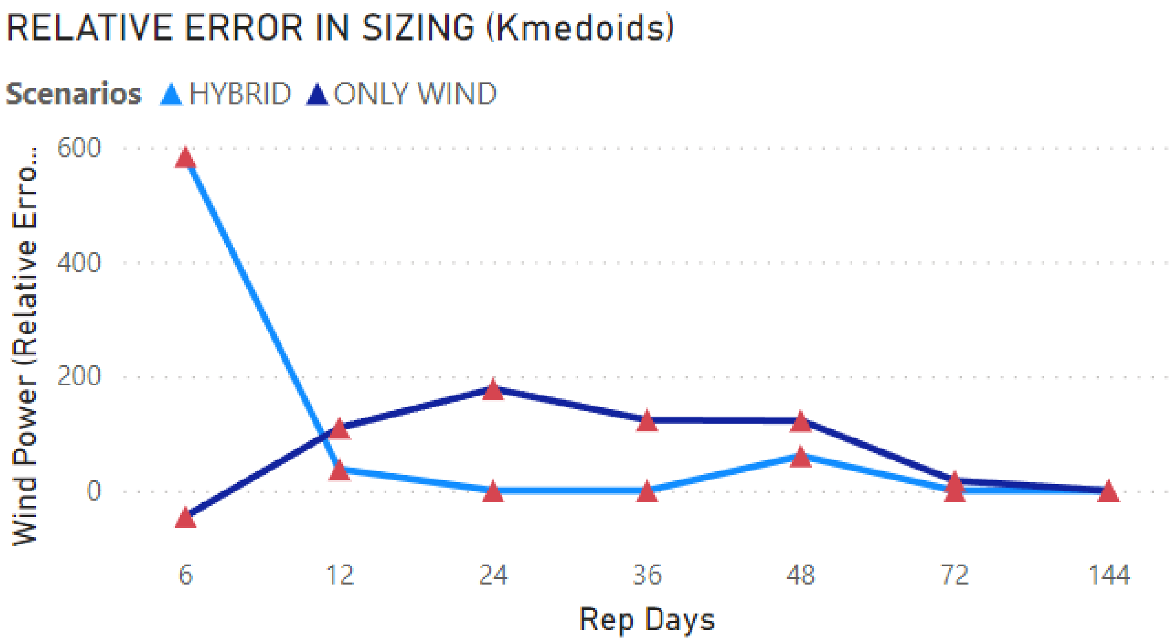


Figure 57 Error in sizing Wind Power Kmedoids

Lithium-ion Battery Technology

The data illustrates the relative sizing error (%) for lithium-ion battery systems as shown in figures 58, 59, 60 utilizing K-means, K-medians, and K-medoids clustering algorithms over 6 to 144 representative days.

For most representative days, K-medoids exhibits the most consistent performance, with errors falling within a +/- 60% range for standalone battery systems. When applied to hybrid renewable energy configurations, the accuracy of K-medoids significantly improves beyond 36 representative days, with sizing errors less than 16%. Contrarily, K-means and K-medians persistently struggle with over or under estimation issues. Extreme outliers, characterized by oversizing up to 300% and under sizing up to 100%, underscore these algorithms' limitations in accurately representing day selections.

Given its consistent performance across a variety of system configurations, K-medoids is recommended as the preferred technique. Its ability to stabilize battery sizing accuracy from a 36 representative day planning duration underscores the effectiveness of using past demand patterns for profiling. Compared to other methods, a maximum error of less than 50% demonstrates K-medoid's capability in capturing load changes.

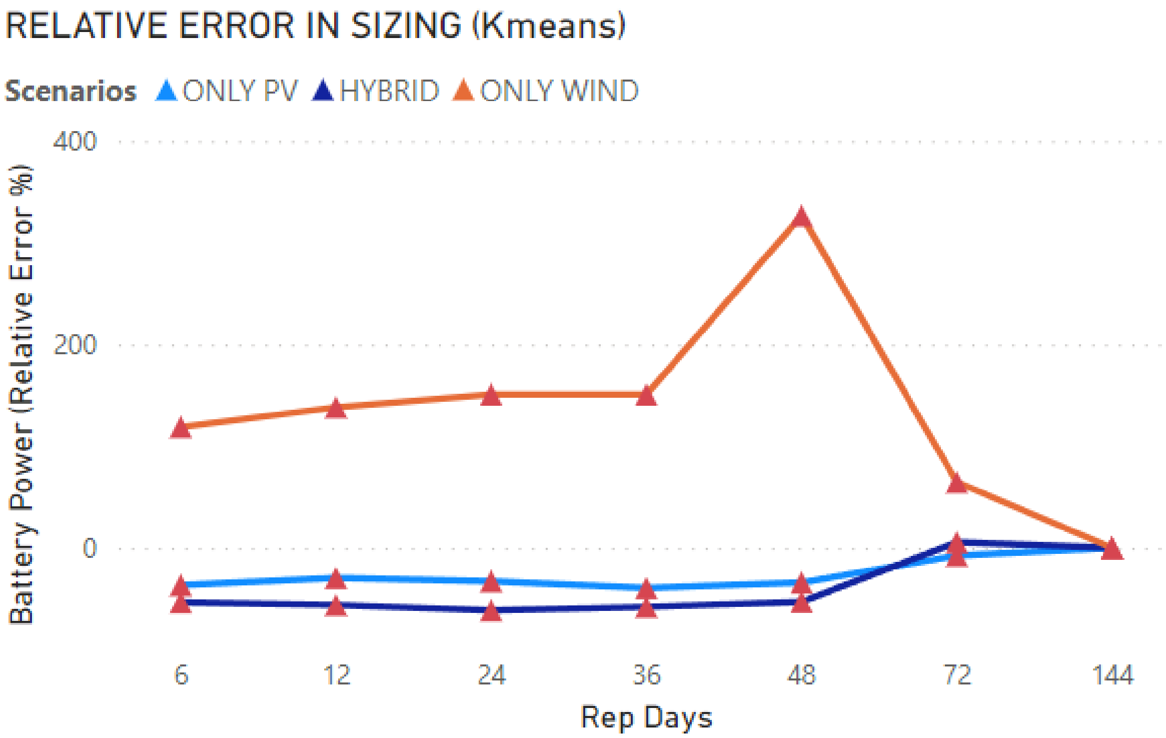


Figure 58 Error in sizing Battery Technology Kmeans

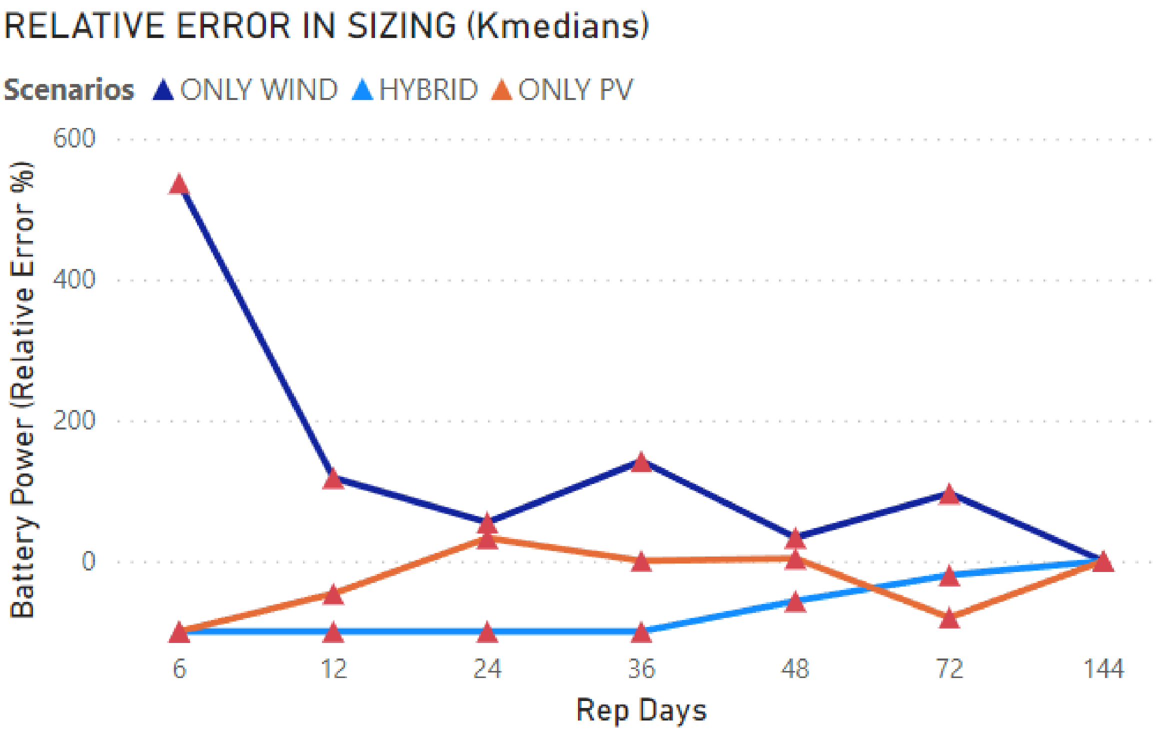


Figure 59 Error in sizing Battery Technology Kmedians

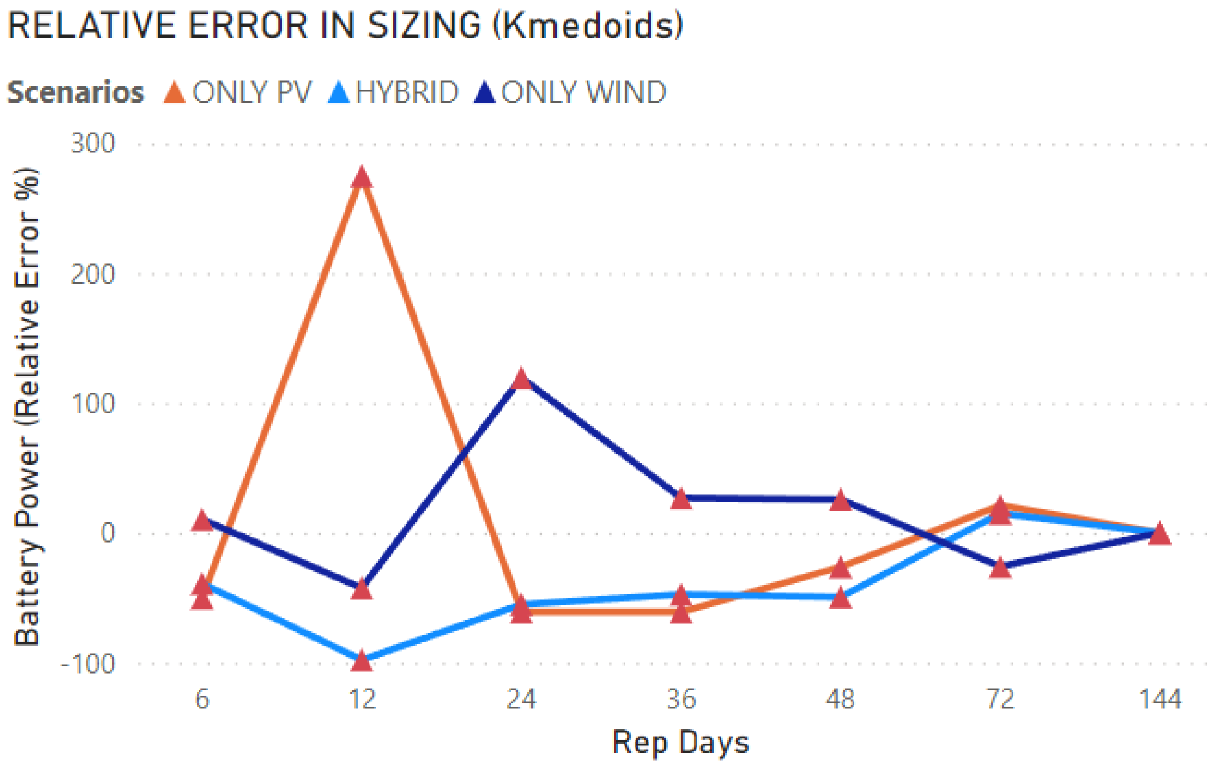


Figure 60 Error in sizing Battery Technology Kmedoids

Lithium-ion Battery Storage

The data illustrates the relative sizing error (%) for lithium-ion battery storage systems as shown in figures 61, 62, 63 utilizing K-means, K-medians, and K-medoids clustering algorithms.

Among the three clustering algorithms, K-medoids exhibits the most consistent performance across various storage systems and planning durations. For instance, in the case of a battery-only system, sizing errors remain within the range of +/- 50%, even at shorter time horizons. The accuracy of K-medoids further enhances for the hybrid configuration, with less than 25% misestimation observed beyond 24 representative days. Contrarily, K-means and K-medians continuously display over or under sizing across storage system types, with extreme instances exceeding 600%.

Given its dependable sizing estimates across a variety of storage scenarios, K-medoids emerges as the preferred technique. Its swift performance stabilization from 24 representative day terms suggests effective load profiling using historical demand patterns. When compared to the other methods, the controlled errors underscore K-medoids adaptability in modelling renewable-storage dynamics.

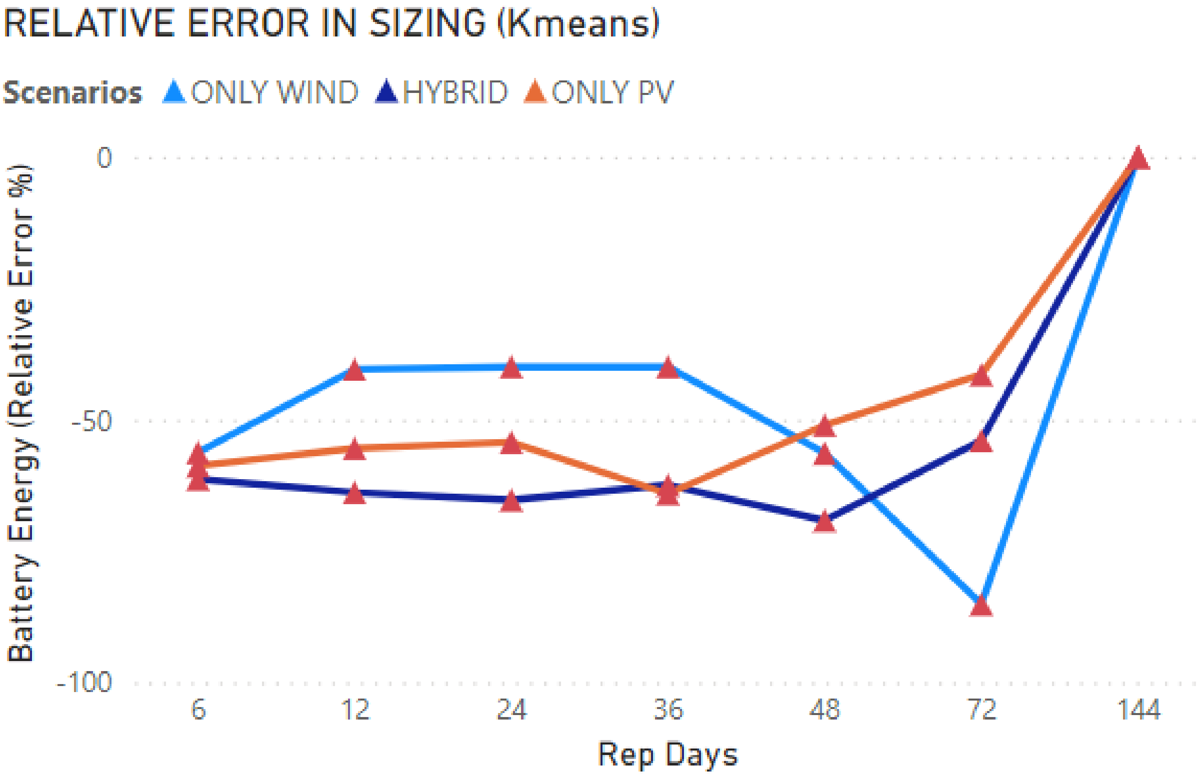


Figure 61 Error in sizing Battery Storage Kmeans

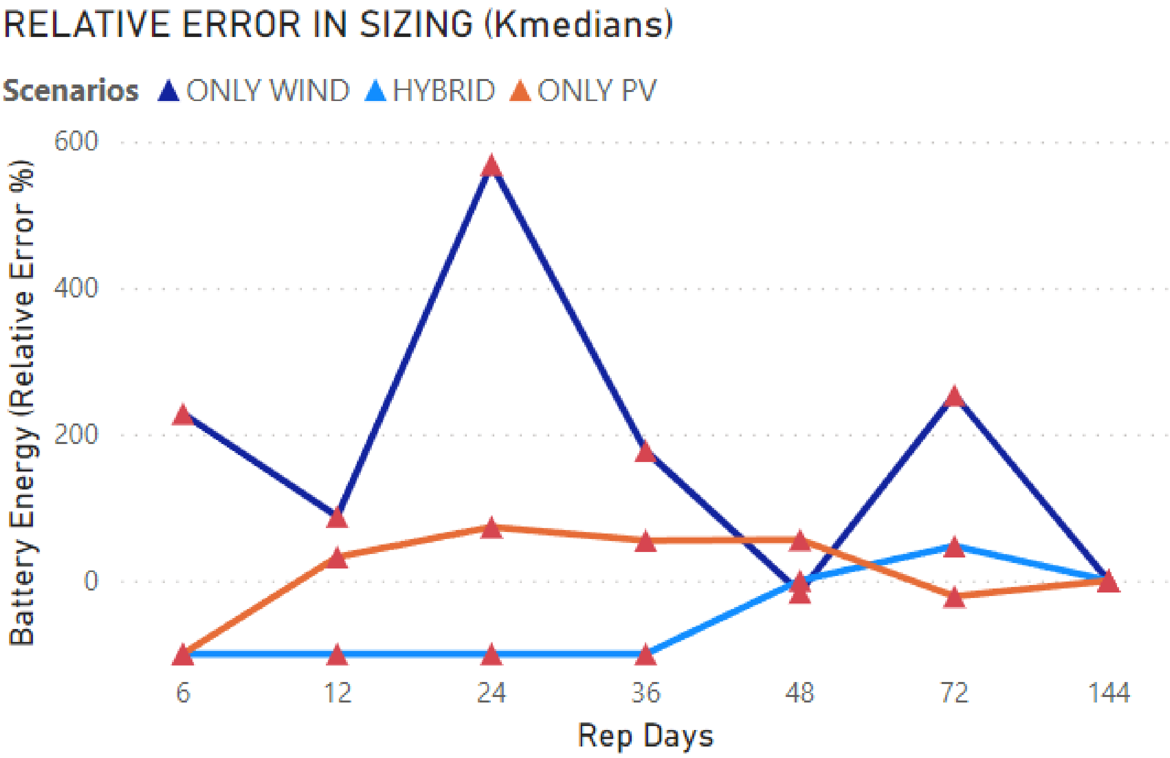


Figure 62 Figure 61 Error in sizing Battery Storage Kmedians

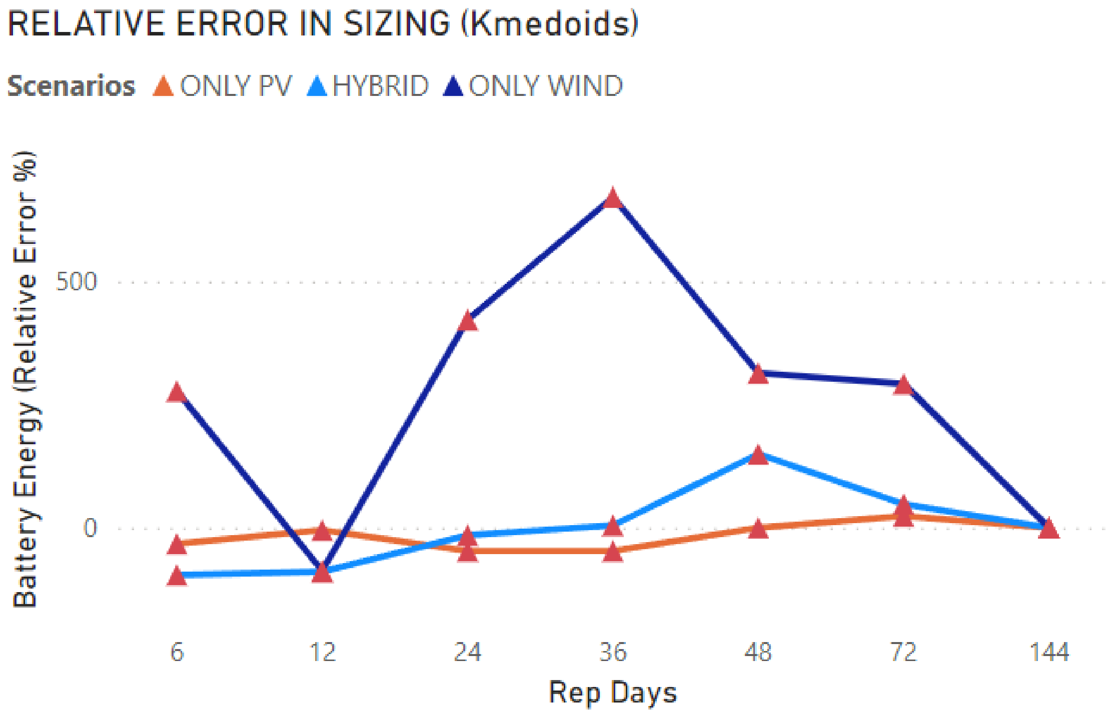


Figure 63 Figure 61 Error in sizing Battery Storage Kmedoids

Hydrogen Fuel Cell

The data illustrates the relative sizing error (%) for hydrogen fuel cell as shown in figures 64, 65, 66 utilizing K-means, K-medians, and K-medoids clustering.

K-medoids exhibits the most stable performance, with sizing errors typically falling within +/- 30%. Beyond 36 representative days, there is a consistent improvement in accuracy, often around +/- 5%, indicating the technique's effectiveness for long-range modelling. Contrarily, K-means and K-medians display recurring over or under estimation issues, occasionally producing extreme outliers exceeding 70% even at extended planning ranges.

K-Medoids is better due to its use of actual load profiles rather than statistical averages. This allows for a more accurate adaptation to the variability and long-term shifts in renewable generation and fuel cell loads. In contrast, the fixed means and medians used by K-Means and K-Medians struggle to capture these seasonal dynamics.

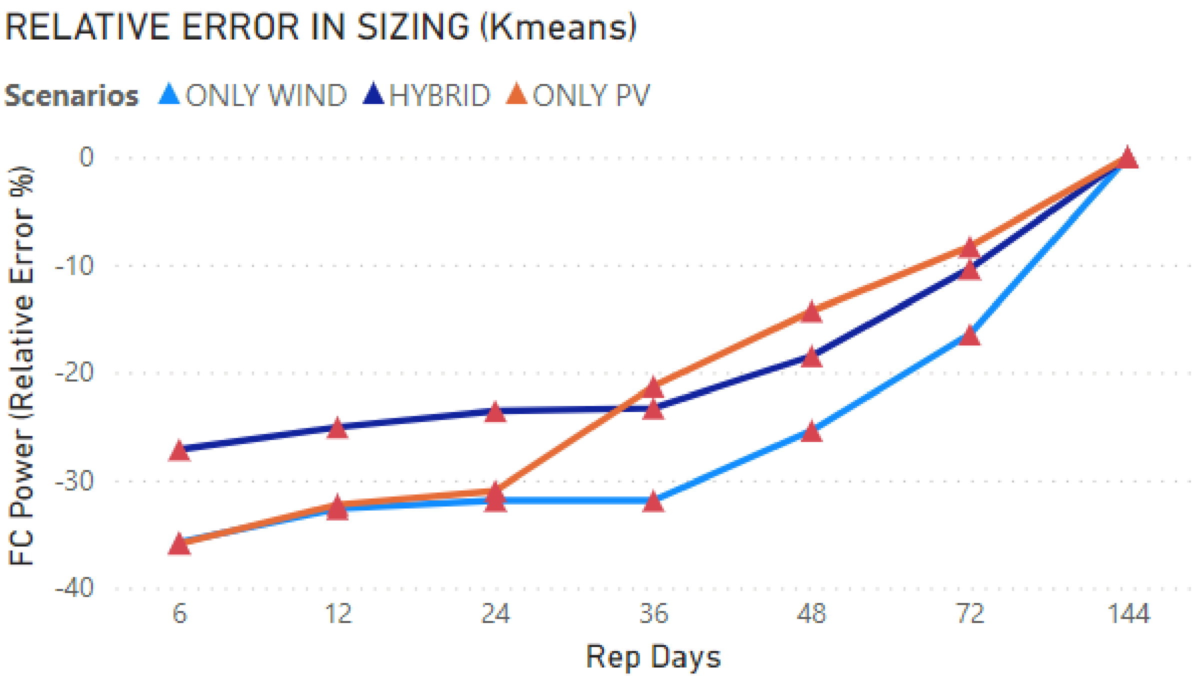


Figure 64 Error in sizing Hydrogen Fuel Cell Kmeans

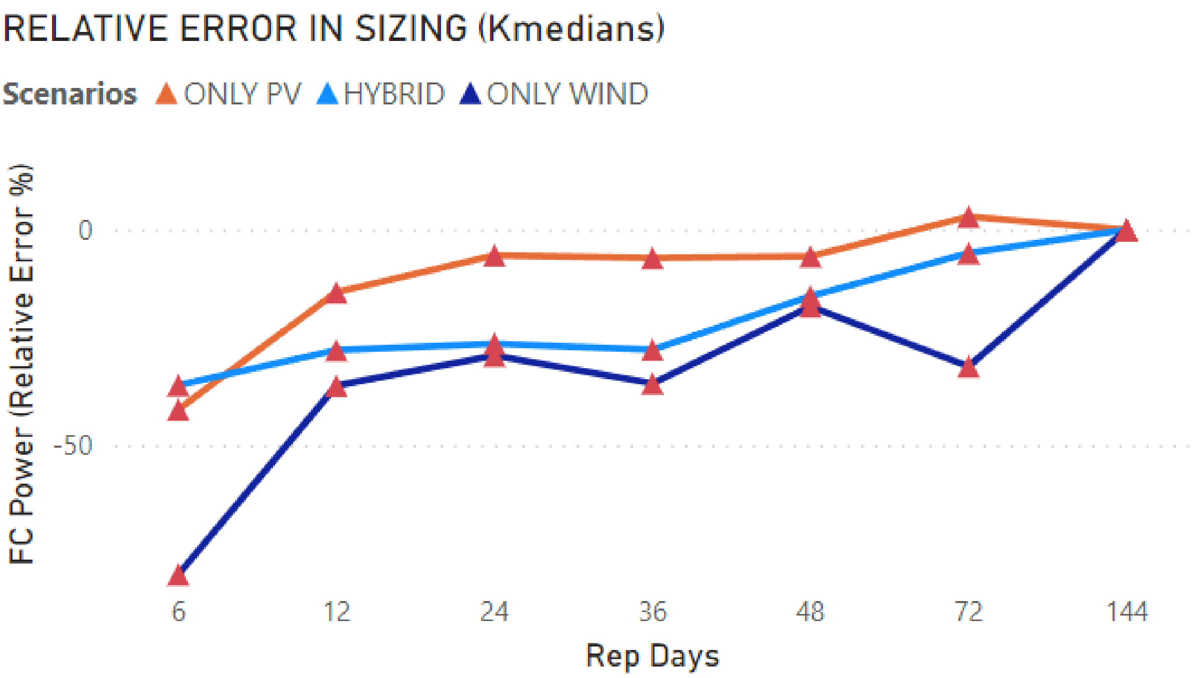


Figure 65 Error in sizing Hydrogen Fuel Cell Kmedians

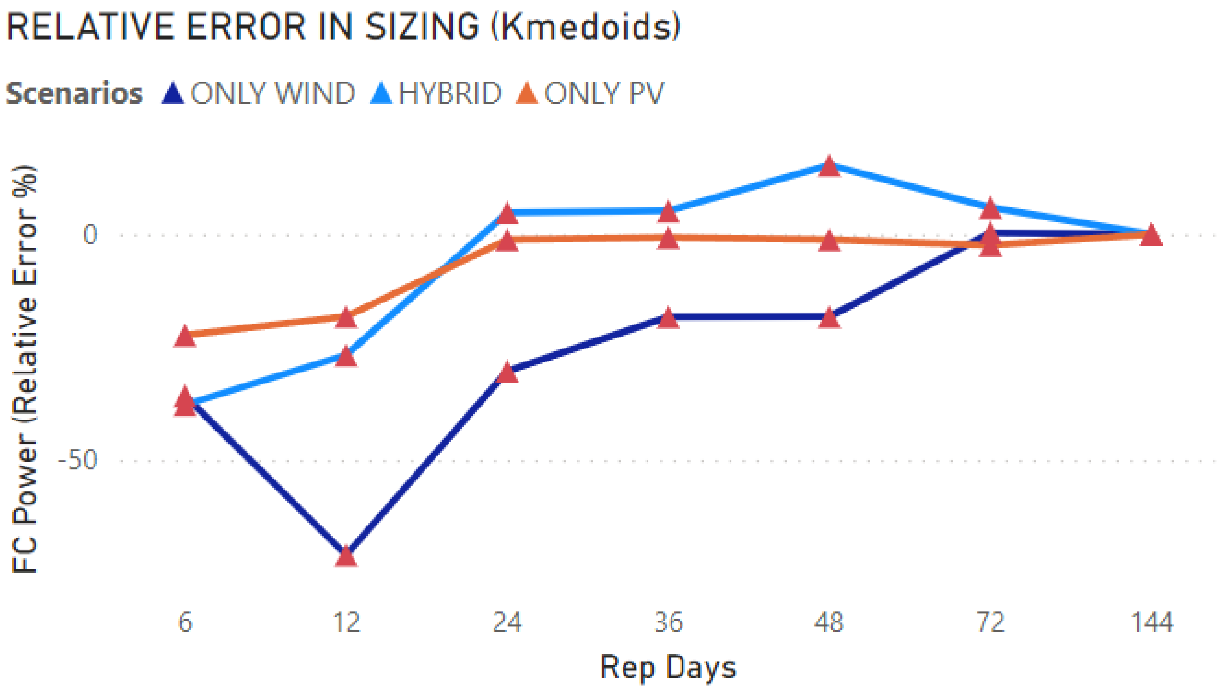


Figure 66 Error in sizing Hydrogen Fuel Cell Kmedoids

Hydrogen Electrolyser

The data illustrates the relative sizing error (%) for hydrogen fuel cell as shown in figures 67, 68, 69 utilizing K-means, K-medians, and K-medoids clustering.

Among the various system types and planning durations, K-medoids exhibits the most stable performance, with sizing errors predominantly falling within the +/- 10% range. Beyond 36 representative days, the accuracy continues to improve, settling at +/- 5%, indicating its proficiency in long-term modelling. Conversely, K-means and K-medians display recurring over or under sizing trends across configurations, with outliers reaching up to 80% even at extended planning ranges.

Given its dependable estimates across a variety of renewable-hydrogen systems, K-medoids emerges as the preferred method. The gradual reduction of errors from 24 representative day terms suggests effective load profiling using historical data. When compared to the other techniques, the controlled accuracy underscores K-Medoids' aptitude for capturing the complexities inherent in electrolyser modelling.

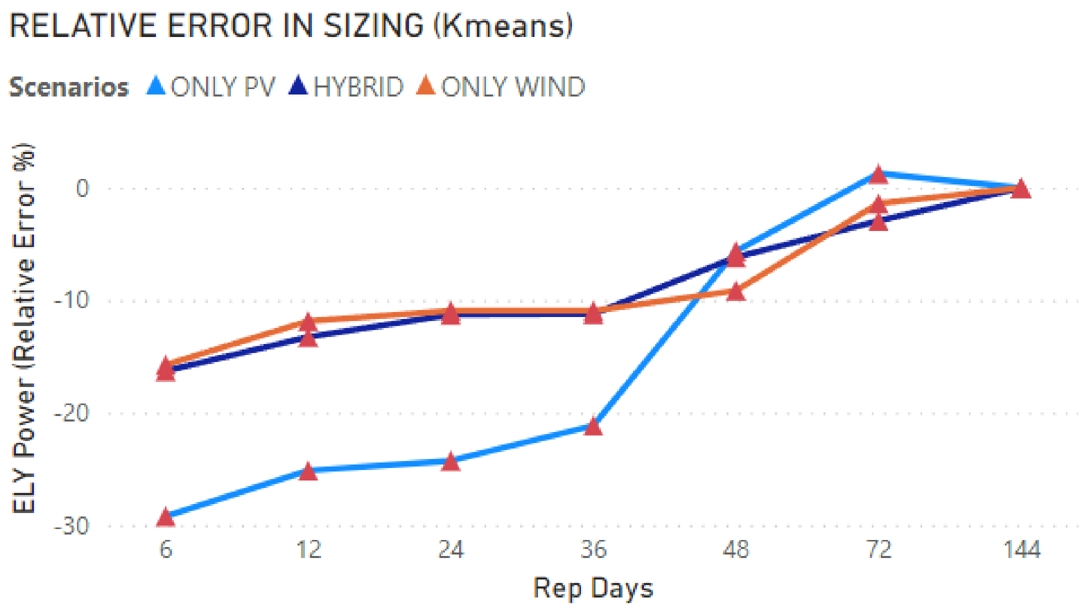


Figure 67 Error in sizing Hydrogen Electrolyser Kmeans

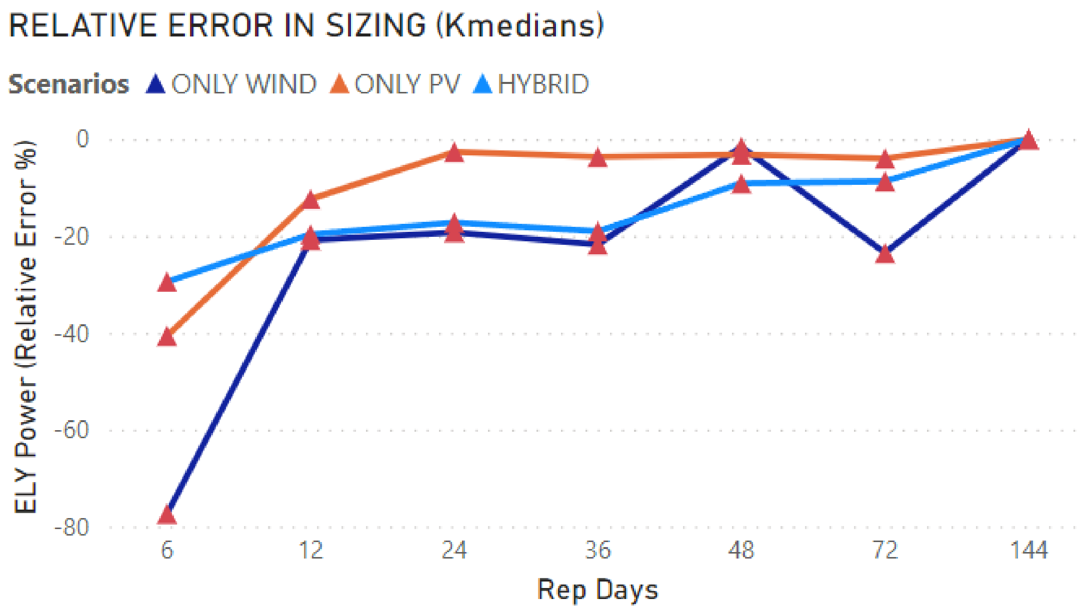


Figure 68 Error in sizing Hydrogen Electrolyser Kmedians

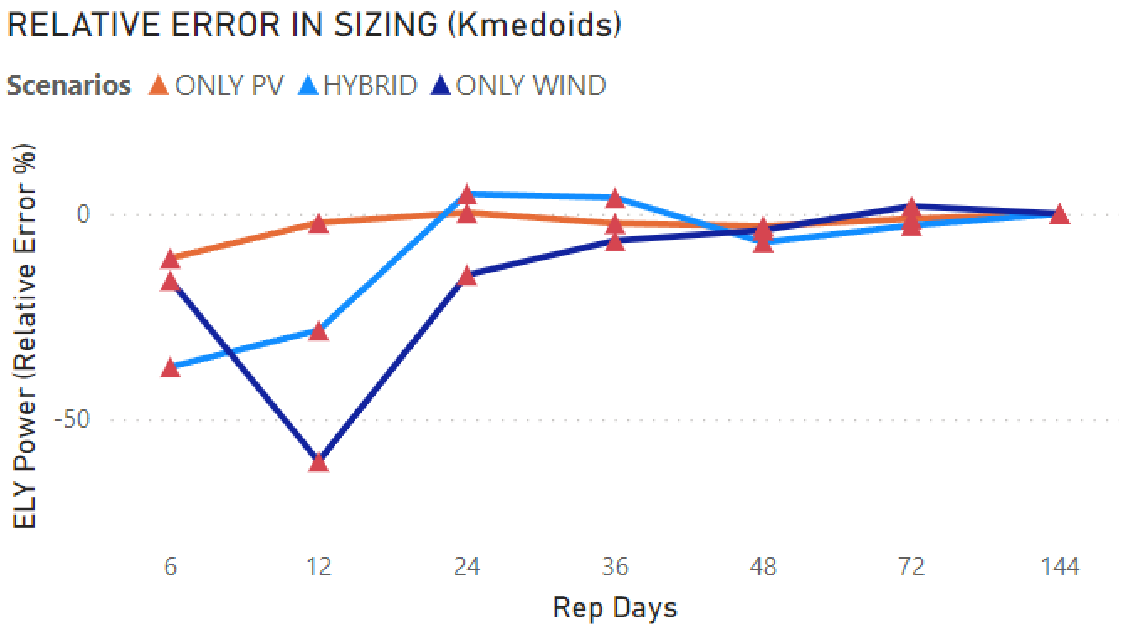


Figure 69 Error in sizing Hydrogen Electrolyser Kmedoids

Hydrogen Storage

The data illustrates the relative sizing error (%) for hydrogen storage systems employing K-means, K-medians, and K-medoids clustering.

In terms of system types and planning durations, K-medoids exhibits the most stable performance, with sizing errors predominantly falling within +/-50%. Beyond 36 representative days, the accuracy continues to improve to +/-10%, signifying its efficacy for long-term modelling. Conversely, K-means and K-medians reveal recurrent over or under sizing trends across storage scenarios, with outliers exceeding 70% even at extended planning ranges.

Given its dependable estimates across a variety of renewable-hydrogen storage systems, K-medoids stands out as the optimal technique. The reduction in errors from 24-day terms suggests adequate load profiling based on historical data patterns. When compared to other methods, the confined accuracy underscores K-Medoids' aptitude for handling the complexities inherent in storage modelling.

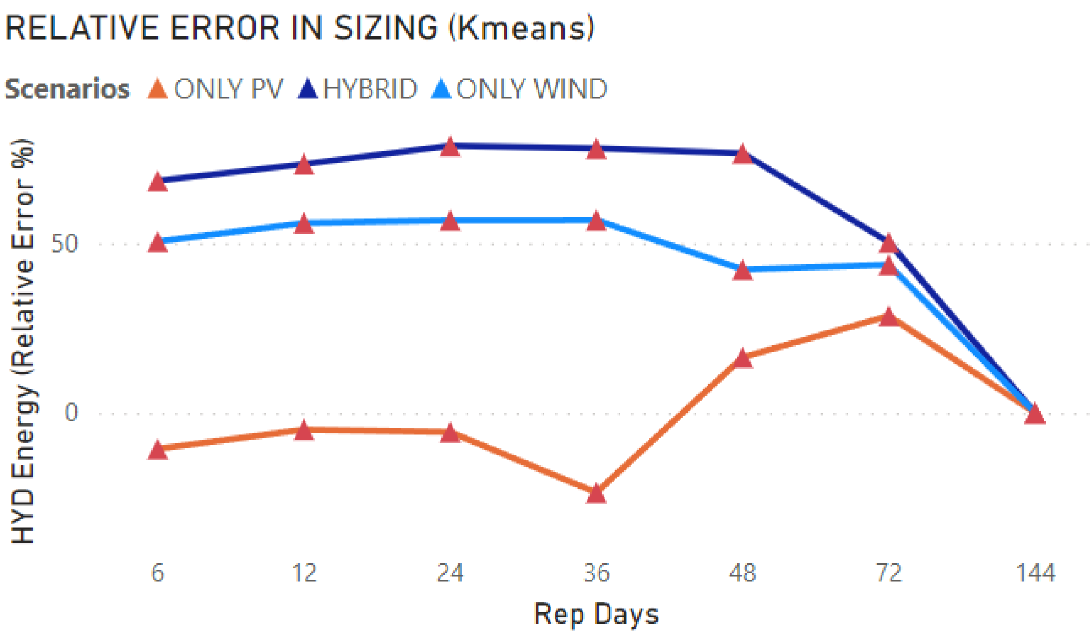


Figure 70 Error in sizing Hydrogen Storage Kmeans

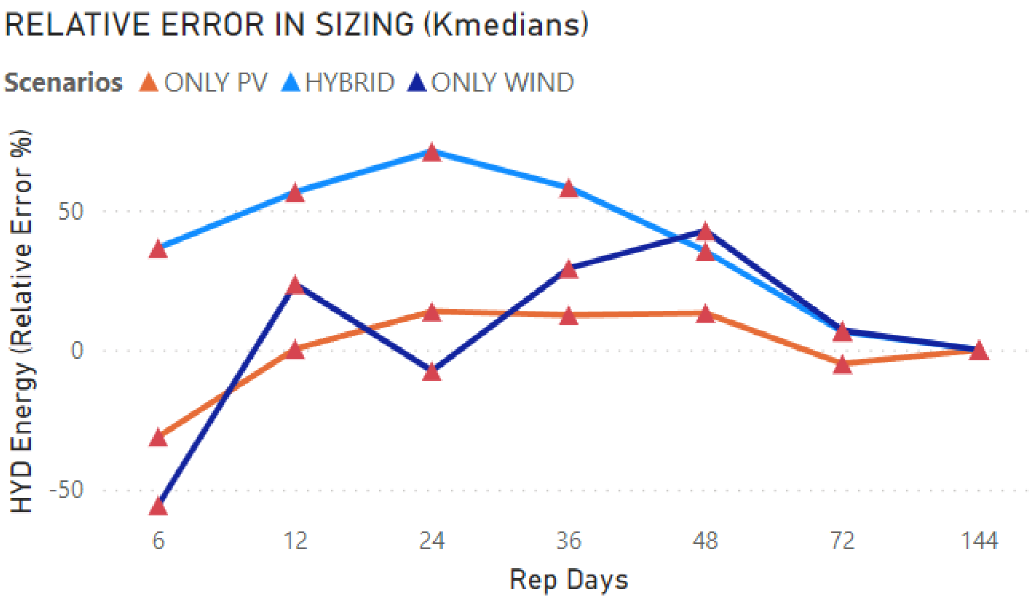


Figure 71 Error in sizing Hydrogen Storage Kmedians

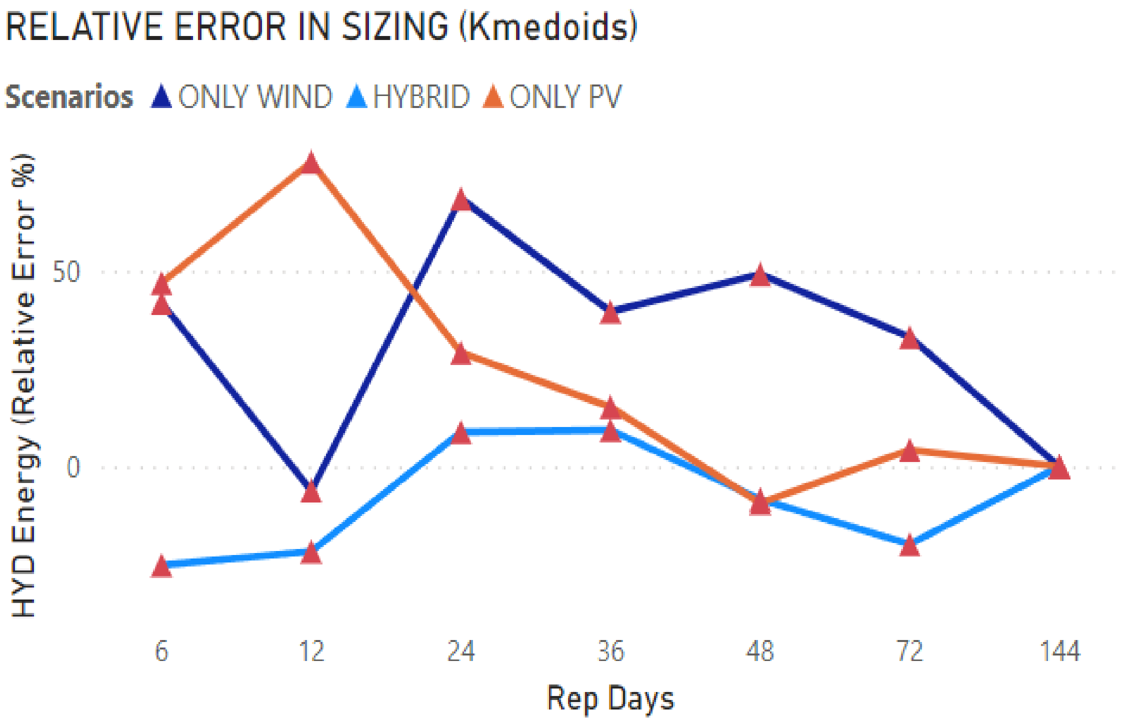


Figure 72 Error in sizing Hydrogen Storage Kmedoids

7. CONCLUSION

To sum up, this investigation significantly advances our understanding of energy system modelling, particularly in relation to island energy systems such as Favignana. The comprehensive assessment of various clustering algorithms in terms of computational efficiency, financial implications, and their ability to accurately depict energy systems under a range of conditions offers substantial insights. The findings from the investigation are listed below:

- The K-means algorithm consistently proves to be a more efficient option in terms of computational duration and cost-efficiency.
- K-means algorithm is the best suited for modelling total costs. Its suitable for scenarios where a steady prediction of costs is crucial.
- The K-medoids algorithm is more efficient when the data is robust and has many variations.
- Also, when it comes to modelling storages considering representative days the K-medoids algorithm suits the best as it gives less error when compared to the actual number of days.

- The K-medians is more suited for the datasets that has a significant non-normal distribution, especially in the PV data.
- The number of representative days modelled is directly proportional to the computational time and the quality of results.
- The examination of the ideal number of representative days reveals that employing 48 representative days can effectively encapsulate the requisite variability in renewable energy systems without incurring excessive computational burdens.

In essence, the research emphasizes the significance of choosing suitable clustering algorithms and the quantity of representative days in energy system modelling, especially for systems rich in renewable energy sources. These methodological decisions are instrumental in striking a balance between computational efficiency, cost, and accuracy, thereby facilitating the creation of robust and reliable energy models for policy and planning objectives.

REFERENCES

1. Energy data Europe: https://energy.ec.europa.eu/news/eu-energy-figures-main-eu-energy-data-your-pocket-2023-10-09_en
2. Energy Statistics: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Energy_statistics_-_an_overview
3. State energy union report: https://energy.ec.europa.eu/topics/energy-strategy/energy-union/eighth-report-state-energy-union_en
4. 2030 Climate energy framework: https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2030-climate-energy-framework_en
5. European green deal: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en
6. EU plan for green transition https://www.consilium.europa.eu/en/policies/green-deal/fit-for-55-the-eu-plan-for-a-green-transition/
7. RED II: https://joint-research-centre.ec.europa.eu/welcome-jec-website/reference-regulatory-framework/renewable-energy-recast-2030-red-ii_en
8. Energy Efficiency Directive: https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficiency-targets-directive-and-rules/energy-efficiency-directive_en
9. Global Enegy and Climate Model: https://www.iea.org/reports/global-energy-and-climate-model
10. World Energy Outlook:https://www.iea.org/reports/world-energy-outlook-2023/context-and-scenario-design
11. McCollum, D.L. Machine learning for energy projections. Nat Energy 6, 121–122 (2021). https://doi.org/10.1038/s41560-021-00779-9
12. M.G. Prina, D. Groppi, B. Nastasi, D.A. Garcia, Bottom-up energy system models applied to sustainable islands, Renew. Sust. Energ. Rev. 152 (2021), 111625, https://doi.org/10.1016/j.rser.2021.111625.
13. R. Loulou, G. Goldstein, A. Kanudia, A. Lettila, U. Remme, Documentation for the TIMES Model: Part I, 2016. https://iea-etsap.org/docs/Documentation_for_the_TIMES_Model-Part-I_July-2016.pdf, 2016.
14. M. Howells, H. Rogner, N. Strachan, C. Heaps, H. Huntington, S. Kypreos, A. Hughes, S. Silveira, J. DeCarolis, M. Bazillian, A. Roehrl, OSeMOSYS: the open-source energy modeling system. An introduction to its ethos, structure and development, Energy Policy 39 (2011) 5850–5870, https://doi.org/10.1016/j.
15. R. Novo, F.D. Minuto, G. Bracco, G. Mattiazzo, R. Borchiellini, A. Lanzini, Supporting decarbonization strategies of local energy systems by de-risking investments in renewables: a case study on Pantelleria Island, Energies 15 (2022), https://doi.org/10.3390/en15031103.

16. L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, Time series aggregation for energy system design: modeling seasonal storage, Appl. Energy 213 (2018) 123–135, https://doi.org/10.1016/j.apenergy.2018.01.023.
17. R. Novo, P. Marocco, G. Giorgi, A. Lanzini, M. Santarelli, G. Mattiazzo, Planning the decarbonisation of energy systems: the importance of applying time series clustering to long-term models, Energy Convers. Manag. X 15 (2022), 100274, https://doi.org/10.1016/j.ecmx.2022.100274
18. P. Gabrielli, M. Gazzani, E. Martelli, M. Mazzotti, Optimal design of multi-energy systems with seasonal storage, Appl. Energy 219 (2018) 408–424, https://doi.org/ 10.1016/j.apenergy.2017.07.142.
19. Thomas Schütz, Markus Hans Schraven, Marcus Fuchs, Peter Remmen, Dirk Müller, Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis, Renewable Energy, Volume 129, Part A, 2018, Pages 570-582, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2018.06.028.
20. PVGIS, https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis_en
21. Global Wind Atlas, https://globalwindatlas.info/en
22. Piano Energetico Ambientale della Regione Siciliana (PEARS2030) https://www.regione.sicilia.it/aggiornamento-piano-energetico-ambientale-regione-siciliana-pears-2030
23. Ministro Dell'Ambiente. Decreto 27 Dicembre 1991. url: https://www.mite.gov.it/sites/default/files/archivio/normativa/Egadi_dm-27-12- 1991.pd
24. Direttiva del Consiglio relativa alla conservazione degli habitat naturali e seminaturali e della flora e della fauna selvatiche. 1992. https://www.msn.unipi.it/wp-content/uploads/2013/03/DIR-CEE- 43-1992.pdf
25. Clean Energy Transition Agenda (CETA) Favignana, https://clean-energy-islands.ec.europa.eu/countries/italy/favignana
26. SEA Favignana, https://seafavignana.info/produzione/
27. Introduction to OSeMOSYS. url: https://osemosys.readthedocs.io/ en/latest/manual/Introduction.html (accessed: 01.09.2016)
28. CPLEX url: https://www.ibm.com/docs/en/icos/20.1.0? topic=cplex-what-does-do
29. A.H. Land and A.G. Doig. «An Automatic Method of Solving Discrete Programming Problems». In: (1960)
30. Operational life of Diesel Power Plant, https://openjicareport.jica.go.jp/pdf/10776995_02.pdf
31. Efficiency of Diesel Power Plant, Kanogˇlu, Mehmet & Işık, Süleyman & Abuşogˇlu, Ayşegül. (2005). Performance characteristics

of a Diesel engine power plant. Energy Conversion and Management. 46. 1692-1702. 10.1016/j.enconman.2004.10.005.

32. https://www.wartsila.com/docs/default-source/power-plants-documents/downloads/white-papers/general/wartsila-bwp-combustion-engine-power-plants.pdf
33. Renewables ninja website, https://www.renewables.ninja/
34. Nallapaneni Manoj Kumar, Srikar Dasari, Jagathpally Bhagwan Reddy,Availability factor of a PV power plant: evaluation based on generation and inverter running periods,Energy Procedia,Volume 147,2018,Pages 71-77,ISSN 1876 6102, doi.org/10.1016/j.egypro.2018.07.035.
35. 29th CIRP Life Cycle Engineering Conference Economic Lifetimes of Solar Panels Manbir Sodhia*, Lennart Banaszeka , Chris Mageeb , Mercedes Rivero-Hudec
36. Sebastian Pfaffel, Performance and Reliability of Wind Turbines: A Review, Energies 2017, doi.org/10.3390/en10111904, www.mdpi.com/1996-1073/10/11/1904
37. https://www.sciencedirect.com/science/article/pii/S2213138821002964
38. Li-Battery: Hector Beltran, Pablo Ayuso, and Emilio Perez. Lifetime Expectancy of Li-Ion Batteries used for Residential Solar Storage. 2019
39. Lithium-ion Technology url: https://sps.honeywell.com/content/dam/his-sandbox/products/gas-and-flame-detection/documents/Technical-Note-166_Lithium-Ion-Batteries_01-02.pdf
40. Lithium-ion storage, url: https://www.powertechsystems.eu/home/tech-corner/lithium-ion-battery-advantages/#:~:text=Lead%20acid%20batteries%20are%20less,of%20most%20lead%20acid%20batteries
41. HydrogenTechnology,url:https://www.nrel.gov/docs/fy21osti/78296.pdf
42. HydrogenCosts,url:https://iea.blob.core.windows.net/assets/a02a0c80-77b2-462e-a9d5-1099e0e572ce/IEA-The-Future-of-Hydrogen-Assumptions-Annex.pdf
43. Technology Brief: Analysis of Current-Day Commercial Electrolyzers,url:https://www.nrel.gov/docs/fy04osti/36705.pdf
44. H. Bohm, ¨ A. Zauner, D.C. Rosenfeld, R. Tichler, Projecting cost development for future large-scale power-to-gas implementations by scaling effects, Appl. Energy 264 (2020), https://doi.org/10.1016/j.apenergy.2020.114780.
45. Lucas Sens, Ulf Neuling, and Martin Kaltschmitt. «Capital expenditure and levelized cost of electricity of photovoltaic plants

and wind turbines - Development by 2050». In: Renewable Energy 185 (2022), pp. 525–537
46. https://www.nrel.gov/docs/fy21osti/78296.pdf
47. Hector Beltran, Pablo Ayuso, and Emilio Perez. Lifetime Expectancy of Li-Ion Batteries used for Residential Solar Storage. 2019
48. Combustion of Fuels - Carbon Dioxide Emission. url: https://www.engineeringtoolbox.com/co2-emission-fuels-d_1085.html
49. Power BI, https://www.microsoft.com/en-us/power-platform/products/power-bi