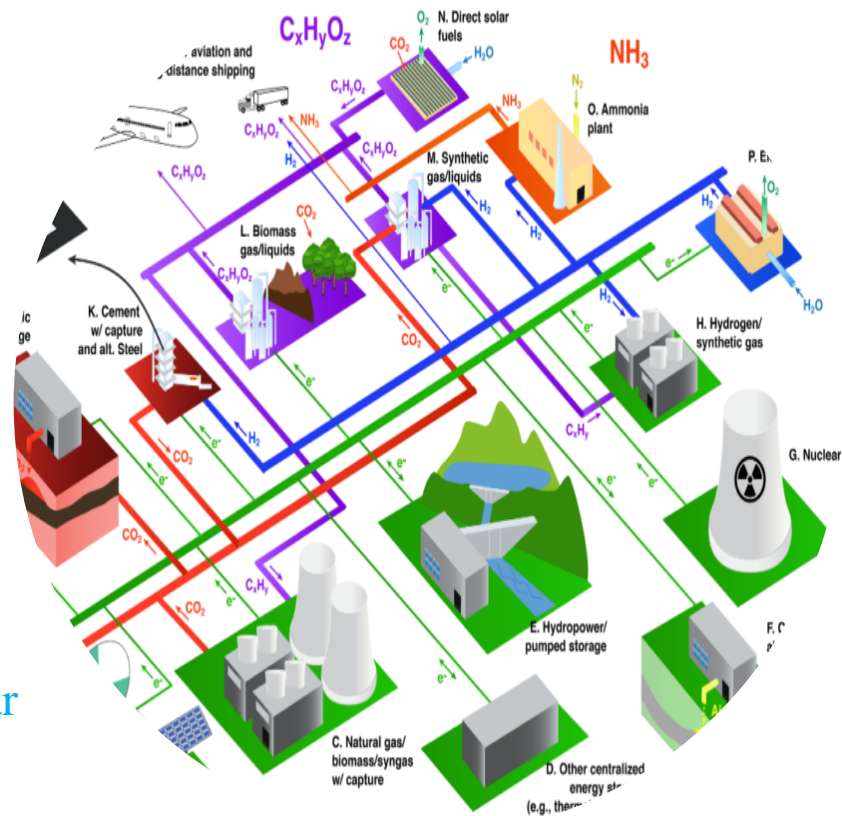


# Design and Application of MGA Analysis Package

A Python-based clustering package using machine learning  
algorithms to analyse near optimal energy systems



Author:

Gaurav Ulhas Khalegaonkar

©Gaurav Ulhas Khalegaonkar

Cover photo: (Davis, Lewis, Shaner, Aggarwal, & Arent, 2018)

All rights reserved.

# Design and Application of MGA Analysis Package

A Python-based clustering package using machine learning algorithms to analyse near optimal energy systems

Master thesis submitted to the Delft University of Technology in partial fulfilment of the requirements for the degree of

## **Master of Science in Sustainable Energy Technology**

Faculty of Electrical Engineering, Mathematics and Computer Science

By

Gaurav Ulhas Khalegaonkar

Student Number 5303125

To be defended in public on June 19, 2023

## **Thesis Committee**

First Supervisor: Dr Ir L.J. (Laurens) de Vries TU Delft

Second Supervisor: Dr A.F. (Aad) Correlje TU Delft

Daily Supervisor: Dr F. (Francesco) Lombardi TU Delft

An electronic version of this thesis is available at <http://repository.tudelft.nl/>



*The impediment to action advances action.  
What stands in the way becomes the way.*

Marcus Aurelius

Roman Emperor (161 AD to 180 AD)



## Preface

Dear Reader,

This report is written to present the research conducted as a part of the master's thesis. For the last eight months, I worked on my master's thesis, in which I developed a Python-based package to analyse the results of MGA-based energy system models. This report marks the end of my journey of becoming a master's in Sustainable Energy Technology.

The report presents the design and application of the MGA analysis package. While writing this report, I assume the readers have knowledge about energy transition and basic data analysis in Python. This report is divided into three parts: thesis definition, design and analysis. The thesis definition part lays the foundation of the thesis, while the design part explains the design choice made during the development of the package. The design of the package is explained by using a flow chart as well as by explaining the code. The application part elaborates on the case study carried out as a demonstration rather than an in-depth analysis.

Readers who are interested in the design of the MGA analysis method will find it in Chapter 4. Readers interested in interactive dashboard design will find it in Chapter 5. This report can be used as an in-depth user manual for the MGA analysis package. I hope this report will spark your interest in the field of machine learning applications for planning energy transition and integrated energy system design.

Gaurav Ulhas Khalegaonkar

May 17<sup>th</sup>, 2023





## Acknowledgement

First, praise to all mighty gods who are creators, preservers and destroyers of all. I am thankful for their grace and for bestowing mental and physical strength in me to walk this mortal plane.

Working on this thesis was the most challenging at the same time, most rewarding experience of my life. From the conceptualization to the current moment, it was a journey which had many ups and downs, many Eureka moments and many moments where I felt I was getting cornered. My two-year master's journey especially writing this thesis, taught me the value of patience, perseverance and strong will. I could not have done it without the help and support of all the wonderful people I met professionally and personally during the last two years.

I want to express my deepest gratitude to my daily supervisor Francesco Lombardi who was quintessential in my growth as an energy system modeller and analyst. Thank you for treating me as a fellow researcher and always challenging me to do better. We bounded over our affinity for energy systems and pretty graphs. I will miss our weekly meetings and discussions of the energy systems most. I also want to express my gratitude towards my supervisors, Professor Laurens De Vries and Professor Aad Correlje. Your comments and expert advice during the thesis review meetings shaped the thesis in the current format. I want to express special gratitude towards Professor Laurens De Vries, who played an instrumental role in my master's journey. I want to thank Anusha for providing her machine learning expertise and kick-starting my thesis. I also want to thank Shreya for providing knowledge about comparison frameworks.

I wish to thank my incredibly supportive family and friends. First time living away from my dear country was a challenging experience indeed. Thank you, Shloka, Shanti, Devansh, Janki, Ritik, Kishan and Megha, for making the Netherlands home away from home. Going on trips with you or sitting at our place discussing random things gave me the power to go through all. SET Karengé Milke (me, Shloka, Shanti and Devansh) is the best team I have ever worked with. Throughout the last two years, we have had moments of triumphant victories and also low points. Nevertheless, we persevered and came other the side. Megha, thank you for always being there for me as emotional support and guidance throughout my journey.

Saving the best for last, I want to thank my parents for their unconditional love and support and my lovely brother Aaditya who is always there for me. Thank you for believing in my dreams and giving wings to them. Thank you, dear Grandmother, for teaching me values and good qualities through your endless number of spiritual stories, which played an essential role in shaping me into the person I am today. I am dedicating this thesis to my grandmother, who sadly passed away during my thesis period.



## Executive summary

### Introduction

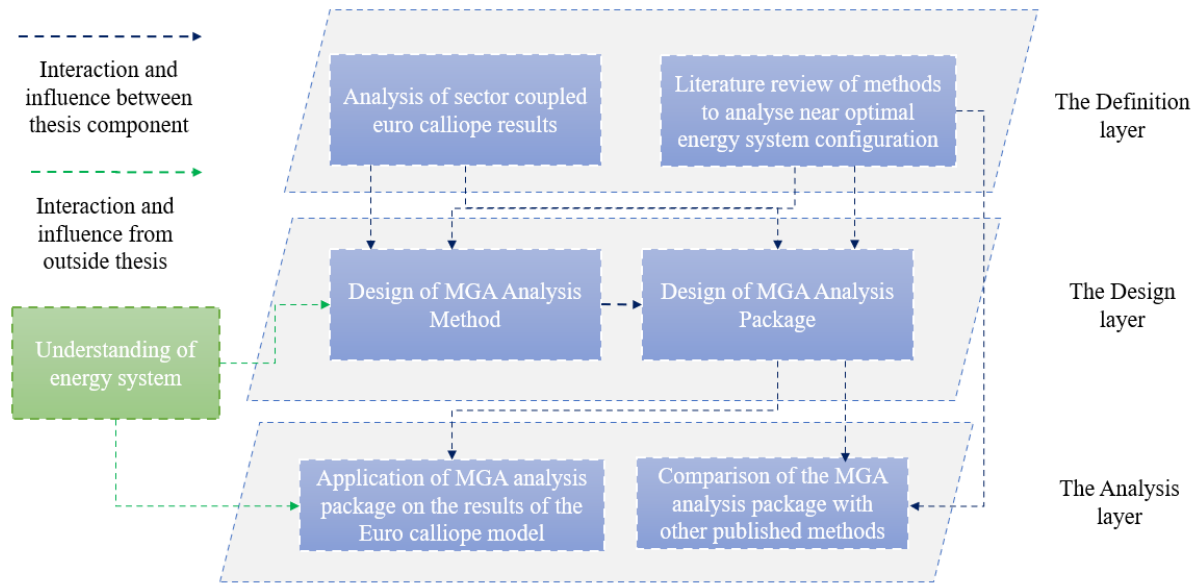
European energy systems must undergo significant transformation to achieve the goals set in the European green deal. The current energy system has to be converted into CO<sub>2</sub> neutral renewable energy-based system. Planning a new energy system is a crucial element in this transformation. The energy sector has hundreds of technologies and millions of actors working together to balance the system. Hence, catching the ripple effect of the change in one component of the entire system is challenging. Experts and policymakers are using computer-based energy modelling tools to tackle this problem.

Modelling to generate alternatives (MGA) is one such type of computer-based energy system model-solving method which produces near-optimal energy system configuration by changing total system cost. A key feature of the MGA algorithm is that it generates hundreds of equally possible energy system configurations. However, it becomes difficult to analyse the solutions of MGA-based energy system due to the sheer amount of data it produces. Which in turn limits the usage of the method. Hence this thesis aims to

*“Design a computer-based method to facilitate analysis of MGA-based energy system optimisation models results.”*

### Research framework

The research work of the thesis is divided into three layers: the foundation, design, and analysis layers. The definition layer produces the knowledge necessary for the design and analysis layer by conducting literature analysis. The design layer focuses on designing a new method to analyse the MGA-based energy system results (henceforth known as the MGA analysis method). The design layer also focuses on converting the MGA analysis method to an interactive Python-based package (henceforth known as the MGA analysis package), automating the MGA analysis method. The analysis layer focuses on demonstrating the application of the MGA analysis package. The analysis layer also compares the MGA analysis package with other published analysis methods for the results of MGA-based energy system models.



*The Research Framework*

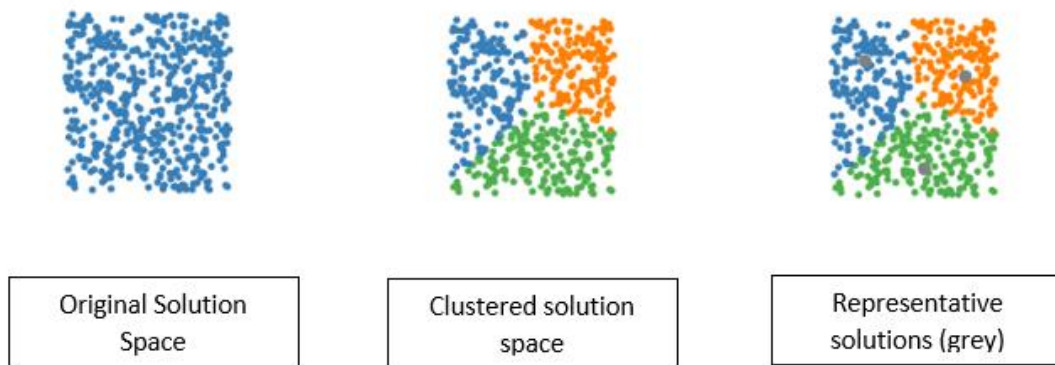
## Definition layer

Two separate literature analyses have been performed in the definition layer. The first one is the analysis of results sector couple euro calliope models solved using SPORES algorithms- a type of MGA algorithm to understand the complexity of the problem. Sector coupled euro calliope model has 98 locations, 35 energy technologies and 441 solutions. The installed capacity of each technology at each location is the model's primary result. There are 1.5 million rows of data in the primary results. Additionally, primary results can be extrapolated to find secondary results, such as electrification of the heating sector and consumer curtailment. Processing all data is a challenging task.

The second is a literature analysis to determine current methods of analysing MGA-based energy system model results. The second literature analysis concludes, "No standardised computer-based way is available to facilitate smooth analysis of MGA-based energy optimisation models". Knowledge generated in this layer is used in determining the features of the MGA analysis method and designing a comparison matrix.

## Design layer

The MGA analysis method is a three-step method to analyse the MGA-based energy system model results. In the first step, using Jenks's natural break algorithm, the MGA analysis method clusters the results into smaller groups with similar characteristics. The second step identifies representative solutions for each cluster formed in step one using a custom-made algorithm developed based on mean and distance in this thesis. The third step is determining the suitable number of clusters for the analysis using a custom-made algorithm created based on the variance in this thesis.



### *Key steps of MGA analysis method*

The MGA analysis method is automated by converting it into python based package to increase its usage. An analyst needs Python 3.9 or above version to use the MGA analysis package. Along with the MGA analysis package, an interactive dashboard is created in Jupiter Notebook to use the MGA analysis package with 4-5 user inputs.

### *Analysis layer*

The first task in the analysis layer was to demonstrate the application of the MGA analysis package. For this purpose, we used the Sector coupled Euro calliope model results (<https://edu.nl/3cba7>). For specificity, we analyse trade-offs between energy-generation technologies in CO<sub>2</sub>-neutral Europe. This analysis aimed to showcase the features of the method rather than develop new insights regarding the generation technologies of CO<sub>2</sub>-neutral Europe.

Key features of the package are decided by combing features of current methods of analysis of near-optimal energy system configuration and discussion with the MGA-based energy model expert. Seven key features, namely: automated method, groups solutions, gives representative solutions, shows boundaries of solution space, shows spatial diversity, interactive dashboard, shows interdependency of technologies, are identified. A comparison matrix is developed based on these features to compare the MGA analysis package with other published methods. The comparison matrix demonstrated the MGA analysis package's superiority over other published methods, where the MGA analysis package combines features of other analysis methods into one method.

## Conclusion

The case study in the analysis layer showcased that the MGA analysis package facilitates analysis of the MGA-based energy system model results by reducing the amount of data needed to analyse. Providing graphs in tandem with maps gives analysts a quick overview of the solution space. The dashboard provides maximum and minimum values of relevant parameters, which allows the analysts to understand the variety of options available for a given cluster. Which in turn leads to better planning of infrastructure. The MGA analysis package requires three input files in Excel format to run and can be used in Windows, Mac, and Linux operating systems by installing Python version 3.9 or higher, which increases the usability of the package. An interactive dashboard allows the analysts to analyse the results with 4-5 user input, making the MGA analysis package faster and more efficient.

Mainly there are three limitations of the package. The first one is the identification of input parameters, which carry huge weight in the analysis process. Identifying those parameters requires a deep understanding of the energy system models and real-world energy systems. The analyst has to be precise while defining the research question and its boundaries. The second limitation of the MGA analysis package is that the package cannot generate less than two clusters. Hence in some cases, the package is forced to generate two clusters even though there should be only one cluster. The third limitation of the MGA analysis package is that it gives an analyst overview of the results. While implementing those energy system configurations in the real world, an in-depth analysis of individual results has to be performed. Hence analysts should use the MGA analysis package as a preliminary analysis tool.

This thesis hypothesises that clustering the solutions of MGA-based energy system models and identifying a representative solution for each cluster is a possible method to analyse the results of MGA-based energy system models. Clustering or grouping solutions into sub-groups was the only way to reduce the amount of data to be analysed while keeping data. We used representative solutions and maximum/minimum values to analyse each cluster. Statistical parameters such as mean, max, min and variance combine to give a holistic idea about data distribution over a cluster. In the retrospective MGA analysis package, we combine all these features in one, which is one of the best ways to analyse the equally possible near-optimal energy system confirmation. Hence initial hypothesis of this thesis holds.

The MGA package is uploaded on the git-hub (<https://edu.nl/gmarp>) for application and further development.

# Table of Contents

Part 1: Thesis Definition .....	1
Chapter 1: Introduction .....	2
1. 1 Problem Definition .....	2
1.2 Research Gap.....	3
1.3 Research design.....	4
1.3.1 Research Approach.....	5
1.3.2 Research Method .....	5
1.4 Thesis structure .....	7
Chapter 2: Need for a MGA analysis package.....	9
2.1 Need for the energy system modelling.....	9
2.2 Linear optimization energy system models and their limitation .....	10
2.3 Incorporation of uncertainties in energy system models.....	10
2.4 Prominent ways of analysing near optimal energy systems.....	11
2.4.1 Expert picking solution based on experience .....	11
2.4.2 Comparing Extreme ends of results along different uncertainty parameter .....	12
2.4.3 Scenario discovery.....	12
2.5 Development of comparison matrix to compare MGA analysis package.....	13
Chapter 3: Sector coupled Euro-calliope model .....	15
3.1 Sector coupled Euro calliope model.....	15
3.2 Results of sector coupled Euro calliope model .....	16
Part 2: Design.....	20
Chapter 4: Design of MGA analysis method .....	21
4.1 Three steps of MGA analysis method .....	21
4.2 Clustering .....	22
4.2.1 How does Jenks natural break (JNB) algorithm works? .....	23

4.2.2 Application of JNB algorithm and storing clustered data .....	24
4.3 Finding representative solutions (FRS) algorithm .....	25
4.4 Finding suitable number of clusters (FSNC) algorithm .....	28
4.5 Regional equality index as technology specific parameter .....	30
<b>Chapter 5: Design of MGA analysis package and interactive dashboard.....</b>	<b>32</b>
5.1 Key features of MGA analysis package .....	32
5.2 Format specification of input files .....	33
5.3 Two modes of MGA analysis package .....	35
5.3.1 Manual mode of operation.....	36
5.3.2 Auto mode of operation.....	37
5.4 MGA analysis dashboard .....	38
5.4.1 Conveying clustered data via dashboard .....	39
5.4.2 Conveying technology specific parameters via dashboard.....	41
<b>Part 3: Analysis .....</b>	<b>44</b>
<b>Chapter 6: Application of MGA analysis package .....</b>	<b>45</b>
6.1 Data input .....	45
6.1.1 Primary technology data.....	45
6.1.2 Data of key technology .....	45
6.1.3 Technology specific parameter.....	46
6.2 Analysis of generation technologies .....	49
6.2.1 Wind energy generation analysis.....	49
6.2.2 Solar energy generation analysis .....	57
6.3 Biofuel energy generation analysis.....	62
6.3 Comparing MGA analysis package with other MGA analysis studies .....	65
<b>Chapter 7: Conclusion.....</b>	<b>68</b>
7.1 Answers to the research questions .....	68
7.2 Discussion of research steps of thesis .....	71



7.3 Critical reflection on the thesis.....	74
7.2.1 Personal Learning .....	74
7.2.2 Advice for further research.....	74
Bibliography .....	76
Appendix.....	79
Appendix A Maps of representative solutions.....	79

## List of Figures

Figure1 Steps of design process.....	6
Figure 2 Research flow diagram .....	8
Figure 3 Example of method expert picking solution based on experience .....	12
Figure 4 Example of comparing Extreme ends of results along different uncertainty parameter (Neumann & Brown , 2021) .....	12
Figure 5 Example of scenario discovery.....	13
Figure 6 Indicator for the analysis matrix.....	14
Figure 7 Blank comparison matrix for comparing MGA analysis package .....	14
Figure 8 Sector coupled Euro calliope model map (Pickering, Lombardi, & Pfenninger, 2022) .....	16
Figure 9 Primary results of Sector coupled euro calliope model.....	17
Figure 10 Secondary results of sector coupled euro calliope model .....	18
Figure 11 Grid transfer capacity results.....	18
Figure 12 Overview of MGA analysis package.....	21
Figure 13 Legend for logic flow chart .....	22
Figure 14 Jenks Natural Break (flow chart).....	23
Figure 15 Clustering algorithm (flow chart).....	24
Figure 16 Finding representative solution algorithm (flow chart).....	26
Figure 17 Finding suitable number of clusters algorithm. (flow chart).....	28
Figure 18 constituents MGA analysis package.....	32
Figure 19 Data cleaning in excel. ....	34
Figure 20 Installed capacity (location wise) input file format.....	34
Figure 21 Total installed capacity input file format.....	35
Figure 22 Technology specific parameters input file format.....	35
Figure 23 Two modes the MGA analysis package .....	36
Figure 24 Key features of MGA analysis dashboard.....	39
Figure 25 Example of clustered data (graph).....	39
Figure 26 Example of clustered data (text).....	40
Figure 27 Example of conveying technology specific parameter.....	41
Figure 28 Example of converging technology specific parameter (text).....	42
Figure 29 Comparing technology specific parameters boundaries along different cluster .....	42

Figure 30 Technology specific parameter matrix for wind energy.....	47
Figure 31 Technology specific parameter matrix for solar energy.....	48
Figure 32 Technology specific parameter matrix for biofuel energy .....	49
Figure 33 Clustered data of wind onshore technology .....	50
Figure 34 Parameter values of representative solution (wind onshore).....	50
Figure 35 Technology specific parameter (comparative analysis wind onshore).....	52
Figure 36 Regional equality index (Map) .....	52
Figure 37 National average import and electrolysis capacity (Map) .....	53
Figure 38 Grid transfer capacity (Map) .....	54
Figure 39 Clustered data of wind offshore technology.....	54
Figure 40 Parameter values of representative solution (wind offshore).....	55
Figure 41 Technology specific parameter (comparative analysis wind offshore).....	56
Figure 42 Clustered data of open field solar PV technology .....	57
Figure 43 Parameter values of representative solution (open field solar PV) .....	58
Figure 44 Technology specific parameter (comparative analysis open field solar PV) .....	59
Figure 45 Clustered data of roof mounted solar PV technology .....	60
Figure 46 Parameter values of representative solution (roof mounted solar PV).....	61
Figure 47 Technology specific parameter (comparative analysis roof mounted solar PV).....	61
Figure 48 Clustered data of biofuel supply technology .....	62
Figure 49 Parameter values of representative solution (biofuel supply technology).....	63
Figure 50 Technology specific parameter (comparative analysis biofuel supply technology).....	64
Figure 51 Filled comparison matrix for comparing MGA analysis package.....	66

## List of Tables

Table 1 Clustering algorithm (code explanation) .....	24
Table 2 Finding representative solution algorithm (code explanation) .....	26
Table 3 Finding suitable number of clusters (code).....	29
Table 4 Application of suitable number of cluster algorithm (code).....	29
Table 5 regional equality index (code explanation).....	31
Table 6 Data cleaning in python (code explanation) .....	33
Table 7 cluster manual mode (code) .....	36
Table 8 cluster auto mode (code).....	37
Table 9 Displaying clustered data (code explanation) .....	40

## Glossary of Acronyms

Abbreviation		Full form
1.	MGA	Model to generate alternatives
2.	IPCC	International panel on climate change
3.	IEA	International Energy Agency
4.	IRENA	International renewable energy Agency
5.	SPORES	Spatially explicit practically optimal results
6.	JNB	Jenks natural break
7.	FRSA	Finding representative solutions algorithm
8.	TSP	Technology specific parameters
9.	FNSCA	Finding suitable number of clusters algorithm
10.	GVF	Goodness of Variance fit
11.	REI	Reginal equality index



# Part 1: Thesis Definition

# Chapter 1: Introduction

This chapter introduces the research challenge of the thesis and the research approach to tackle that challenge. Section 1.1 discusses the general background behind the research, and section 1.2 elaborates on the main research question. Furthermore, section 1.3 conveys the research approach with reasoning. Finally, section 1.4 presents the structure of the thesis and reading guidelines.

## 1. 1 Problem Definition

Achieving the European Union's commitments under the European Green Deal, the Energy Union Strategy and the Paris Agreement requires a significant transformation of current energy systems into carbon-neutral and renewables-based systems (Krumm, Süsser, & Blechinger, 2021). One of the key components of every energy system is energy generation technologies. To achieve ambitious targets in the European green deal, the European energy system has to shift to CO<sub>2</sub> neutral energy generation technologies. However, the trade-offs of different CO<sub>2</sub> neutral energy generation configuration on energy system is still under research.

The energy sector has hundreds of technologies and millions of actors working together to balance the system (Pickering, Lombardi, & Pfenninger, 2022). Researchers are using computer-based models to understand the techno-economic impacts on the energy system due to changes in one or more energy system components.

Currently, four types of computer-based energy models are prominent in the field of energy transition research. These models are namely the energy system optimization model, energy system simulation model, integrated assessment model, and general computable equilibrium model. Each model is used to address a specific need of the energy transition. (Krumm, Süsser, & Blechinger, 2021). Each model has advantages and disadvantages, and one has to choose a specific model based on the specific needs of the question that one has to answer.

Linear optimization models are subsets of energy optimization models where entire energy system components are linked via a series of linear equations. Linear optimization models are popular in the energy system analysis sector and are used by an organization such as IPCC (international panel on climate change), IEA (International Energy Agency), and IRENA (International renewable energy Agency). In the case of linear optimization models, a solution is guaranteed as the model is built upon the linear optimization concept (Ommen, Markussen, & Elmegaard, 2014). A common linear optimization model's output is a single solution that is easy to convey and build a discussion around.



Generally, linear optimization models are centred around optimizing total system cost (Pfenninger, Hawkes, & Keirstead, 2014). The main constraints are capacity constraints for each resource and generation equals consumption and storage constraints. The output of linear optimization models is a single cost-optimized solution specifying each energy resource and carrier's capacity and location. However, stakeholders and recent literature reported that a single economically efficient solution has little to no use in the real world for several reasons (Gaschnig, et al., 2021). The first reason is that cost assumption in the model is nothing but numbers and expert judgment on how system cost will change in the future. Degrees of uncertainty associated with cost assumption increase as the time horizon increases. Accordingly, any feasible solution within the cost uncertainty range of the cost-optimal solution could be deemed just as economically viable (Lombardi, Pickering, & Pfenninger, 2023) (DeCarolus, 2011). The second reason is that optimal cost solutions may not always be desirable in practice, say because of sustainability or geopolitical concerns or due to social acceptability issues (Church & Baez, 2020) (Neumann & Brown, 2021). If the solution is not accepted by society for any reason, there is very little possibility of implementing it in the real world.

The 'Modelling to Generate Alternative' (MGA) method is one method to solve issues related to the single solution method and deal with unmodelled objectives (Pickering, Lombardi, & Pfenninger, 2022). The MGA method finds equally feasible solutions near the economic optimum. The SPORES algorithm is a recent advancement of the MGA method to find near-optimal solutions that look for spatially and technically distinctive feasible system designs while improving search efficiency and computational tractability.

## 1.2 Research Gap

Despite several advantages of MGA-based optimization methods over conventional cost-optimization, the modelling community still needs to increase the usage of the MGA method. Among the reasons that have so far prevented MGA from becoming a new standard is the overwhelming amount of information that it entails - hundreds to thousands of feasible system designs - compared to the conventional provision of a handful of arbitrary illustrative scenarios (Pickering, Lombardi, & Pfenninger, 2022) (DeCarolus, 2011). From the point of view of an analyst, it becomes difficult to provide in-depth analysis for future action due to the sheer amount of data. In other words, while analyzing the results of the MGA-based model, one of its strengths becomes its weakness.

For instance, a recent application of the SPORES algorithm (a type of MGA algorithm) generated 441 near-economically-optimal, equally possible energy system configurations for CO<sub>2</sub>-neutral Europe (hereafter referred to as “Euro-SPORES”) (Pickering, Lombardi, & Pfenninger, 2022). The model has 98 locations and 35 energy technologies. Each solution has a distinct energy infrastructure configuration. The primary solutions of the model are the installed capacity of each technology at each location, which is around 1.5 million rows of data. Furthermore, primary results can be extrapolated to

get the secondary results such as capital cost, operation and maintenance cost, electrification of the heating and transport sector, import and export of energy, demand curtailment and more. Along with that, some parameters, such as grid transfer capacity and import/export of energy, are different from others. These parameters have two locations rather than one, which adds to the analysis challenge due not uniformity. Analysing the results one by one manually is a challenging and overwhelming task.

In the literature ( (Pickering, Lombardi, & Pfenninger, 2022), (DeCarolis, 2011), (Neumann & Brown , 2021), (Bryant & Lempert, 2009), (Church & Baez , 2020)) three prominent ways of analyzing near-optimal energy system configurations exist. The first one is by selecting a few solutions based on the model expert's expertise. The second is by analyzing solutions at the boundaries (for example: based maximum and minimum installed capacity of certain technology), and the last is by using computer-based algorithms to analyze the solutions. Each way has its advantages and disadvantages. The conclusion of the literature analysis of prominent ways of analyzing near-optimal energy system configuration is that no standardized computer-based method is available to facilitate the analysis of MGA-based energy optimization models.

Clustering the results and finding a representative solution will divide results into groups with similar characterises. Analysing representative solutions rather than analysing all the solutions will reduce the workload of analysts. Furthermore, converting the analysis method into a package with an interactive dashboard will increase the speed and customizability of data processing for the analyst. Then the main research question of this thesis becomes-

*“How to build a computer-based clustering package to analyse the results of MGA based energy system models?”*

This research aims to design a clustering method and Python package based on the designed method to analyze the results of the MGA-based optimization model. In doing so, increase the usage of the MGA model to analyze the energy systems. The package is designed for the MGA based energy system model users to analyze and post-process their results while reflecting on the model's design choices.

### 1.3 Research design

Research design is the ordering of the key elements of the research, linking the empirical data to the research objectives of the study and subsequently to the conclusion.

The research process is divided into three parts. The first part of the research is dedicated to identifying the need for clustering methods and current ways of analysing near-optimal energy system configurations. The second part of the research is to design the clustering method and Python package based on the clustering method. The third part of the research is the demonstration Python package on the results of Euro-SPORES and compares the clustering method with other published methods.

### 1.3.1 Research Approach

A quantitative approach is chosen to design the process of conveying the MGA-based optimization model results. Results from the Euro-SPORES study are chosen as a data set sample to build the process. Concepts such as data analysis, machine learning, and designing will be used to achieve the goal of this part. The application of the clustering method is demonstrated by analyzing trade-offs between energy generation technologies from the results of the Euro-SPORES model. Concepts such as critical analysis, system analysis and energy system design are used to accomplish the goal of this part. For comparison, literature analysis of similar methods is used to build a comparison matrix. The comparison matrix shows the advantages of the clustering method compared to other published methods. Concepts such as new product development and product comparison matrix are used to develop the comparison matrix.

### 1.3.2 Research Method

Main research question

*“How to build a computer-based package to analyse the results of MGA based energy system models?”*

*Clustering the solutions and identifying a representative solution for each cluster is a possible method to analyse the results of MGA-based energy system models* is the hypothesis of this thesis. On the one hand, clustering the solutions and identifying representative solutions for each cluster will reduce the amount of data to be analysed. On the other hand, an underlying group of similar, equally feasible solutions will allow for communicating the extent to which each representative solution may be tweaked further to adapt to unmodelled real-world necessities. The newly developed combine process of clustering and finding a representative solution is referred to as the MGA analysis method, and the developed Python package based on the MGA analysis method is referred to as the MGA analysis package in this thesis.

To answer the main research question, six sub-questions are derived. For each sub-question, we identify the function of the question, the type of knowledge, the type of research activity, the data type(s) needed, and the method(s) to answer it. The first research question discusses prominent methods of analysing near optimal energy system configuration. The second and third questions elaborate on the design of the MGA analysis method. The fourth question explains the construction of the Python package from the MGA analysis method. The fifth question details the design choices of the interactive dashboard, and, the sixth question guides the demonstration of the MGA analysis package on the results of the Euro-SPORES model.

*SQ1: What are prominent methods to analyse the results of near optimal energy systems configuration?*

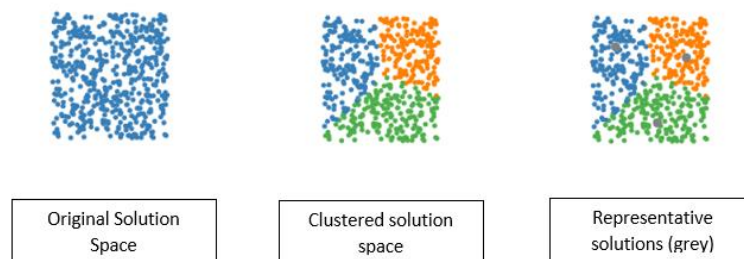
The function of this question is to learn about prominent ways of analysing results of near optimal energy system configuration. Furthermore, based on the learnings determine the key feature of MGA analysis package. For this purpose, literature review of the methods of analysing near-optimally energy system configuration has been performed.

*SQ2: Which algorithm to use to cluster the results MGA-based energy system model?*

Figure 1 shows the steps involved in the MGA analysis method. The first step of the MGA analysis method is the clustering process. The function of this research question is to determine best possible method to cluster the results of MGA based energy system model. The main research activity in this step is the application of various clustering methods on the results of Euro-SPORES. One- and two-dimensional clustering algorithms are used to cluster the results to determine the suitable clustering algorithm.

*SQ3: How to identify a representative solution for the clusters created by the clustering process?*

The MGA analysis method's second step is finding a representative solution for each cluster. The function of this question is to develop a method which is suitable for the clusters of energy system configurations. The main research activities are desk research of various statistical methods to find a representative solution and designing a method to identify a representative solution for the clusters created in step one.



*Figure1 Steps of design process*

*SQ4: How to build a computer-based package based on the designed MGA analysis method.*

The next step in the research is to convert the MGA analysis method into a customizable computer-based package. Parameters such as customizability, accessibility and ease of use dictated the choice in developing the computer-based package. Function of this question to guide the construction of MGA analysis package. The main research activity during this period was, implementing a machine learning algorithm and newly designed algorithms in Python, developing a Python package and hosting the package on git-hub.

*SQ5: Which parameters to display on the interactive dashboard to facilitate the analysis of MGA-based energy system model results?*

Clustered data must be conveyed to analysts in a format that makes it easy to analyse the results. The function of this question is to determine the features of the MGA analysis dashboard and explain the design choice made for the MGA analysis dashboard. The main research activity for this question is the development of an interactive dashboard on top of the MGA analysis package.

*SQ6: What are the impacts of change in energy generation installed capacities on the CO<sub>2</sub>-neutral European energy generation and balancing systems?*

The function of this question is to demonstrate the application of the MGA analysis package and its dashboard. The main research activity during this phase is the analysis of the results of Euro-SPORES using the MGA analysis package.

Finally, the MGA analysis package is compared with other published MGA energy system studies based on key features identified in sub-research question one.

## 1.4 Thesis structure

The thesis is grouped into three parts. The first part is the definition; part two is the design and part three is the application. Figure 2 shows the research flow diagram of the thesis.

The thesis is divided into seven chapters. Chapter 1 provides an overview of the research problem and introduces the aim and objectives of the research. It also briefly explains the research approach and methodology. Chapter 2 explains the need for MGA analysis package by analysing current ways to analysis of MGA based energy system model's results. The output of this chapter is a parameter matrix used to develop MGA analysis package. Chapter 3 elaborates the sector-coupled euro calliope model whose results are analysed in this thesis. Output of this chapter is preliminary analysis of results of Euro-SPORES. The first three chapters conclude the thesis definition part.

Chapter 4 explains the design of MGA analysis method and key choices made during development process. Chapter 5 elaborates on the design choices made to MGA analysis package and build an interactive dashboard.

Chapter 6 demonstrates the application of the MGA analysis package on the results of Euro-SPORES and compares the MGA analysis package with the published MGA result analysis method. In Chapter 7, the research aims and questions are reviewed, the research questions are answered and alternate paths for each step are discussed in brief. After that we reflect on the main hypothesis of the thesis to answer the question how successful we were in achieving the thesis target. The ending of chapter 7 concludes part three of the thesis and, with that, the thesis.

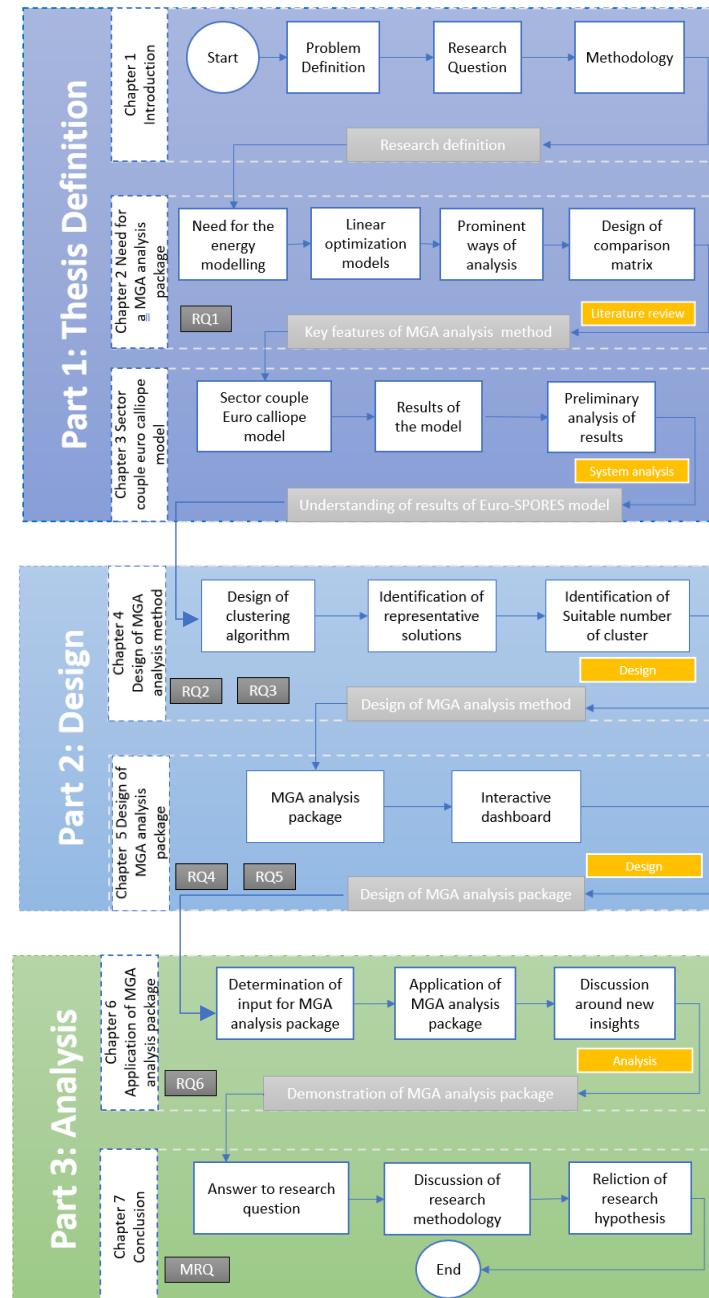


Figure 2 Research flow diagram

## Chapter 2: Need for a MGA analysis package

This chapter aims to explain the need for the MGA analysis package and the key expected features of the MGA analysis package. Section 2.1 elaborates on the broad categories of energy modelling. Then, section 2.2 details the linear optimization system model and discusses the advantages and disadvantages. Section 2.3 discusses the type of uncertainties in energy system modelling and methods to incorporate those uncertainties in the model. Then section 2.4 analyses the prominent ways of analyzing system model results, and section 2.5 elaborates development of a comparison matrix which is used to compare the MGA analysis package with other published MGA analysis methods.

### 2.1 Need for the energy system modelling.

The energy sector has hundreds of technologies and millions of actors working together to balance the system (Pickering, Lombardi, & Pfenninger, 2022). Researchers are using computer-based models to understand the techno-economic impacts on the energy system due to changes in one or more energy system components.

Currently, five types of computer-based energy models are prominent in the field of energy transition research. These models are namely the energy system optimization model, energy system simulation model, integrated assessment model, agent-based model, and general computable equilibrium model. Each model is used to address a specific need of the energy transition. (Krumm, Süsser, & Blechinger, 2021). Each model has advantages and disadvantages, and one has to choose a specific model based on the specific needs of the question that one has to answer.

The design and operation of energy systems with minimal environmental and economic impacts are highly complex issues; energy supply and demand must be spatially and temporally – balanced, with an ever-increasing set of generation units, storage technologies and transmission operations, and load management alternatives (Kotzur, et al., 2021). According to (Banos, et al., 2010), the analytical solving of these problems is no longer feasible and instead requires using a mathematical energy system optimization model (ESOMs) to identify the optimal design and operation. This need for mathematical computation models leads to forming computer-based energy system optimization models. These models' main purpose was not to generate numbers but to provide insights into current and future energy systems (Huntington, Weyant, & Sweeney, 1982).

## 2.2 Linear optimization energy system models and their limitation

Linear optimization energy system models are a type of energy system optimization model in which variables denote the capacities of technologies, and linear equations denote the relationship between technologies and the system's energy demand. Linear optimization models are popular in the energy system analysis sector and are used by an organization such as IPCC (international panel on climate change), IEA (International Energy Agency), and IRENA (International renewable energy Agency). In the case of the linear optimization model, a solution is guaranteed as the model is built upon the linear optimization concept (Ommen, Markussen, & Elmegaard, 2014).

Generally, linear optimization models are centred around optimizing total system cost (Pfenninger, Hawkes, & Keirstead, 2014). A linear optimization model has three main components—the first objective function- around which a system is optimized. As stated above, the energy optimization model objective function is the total system cost equation. The second decision variable is used to denote the capacity of each technology or related parameter. Finally, linear equality and inequality constraints are used to denote the relationship between the components of the system. The main constraints are capacity constraints for each resource and generation equals consumption and storage constraints. The output of the common linear optimization equation is a single solution containing the installed capacity of each technology in the system when the cost is minimized. A single solution facilitates the easy analysis of the system and planning. Due to an increase in computational power, energy system models have evolved to model systems with high temporal and spatial resolution. These advancements allowed energy system modellers to model the energy system in greater detail and policymakers to make more informed decisions.

However, stakeholders and recent literature reported that a single economically efficient solution has little to no use in the real world for several reasons (Gaschnig, et al., 2021). The first reason is that cost assumption in the model is nothing but numbers and expert judgment on how system cost will change in the future. Degrees of uncertainty associated with cost assumption increase as the time horizon increases. Accordingly, any feasible solution within the cost uncertainty range of the cost-optimal solution could be deemed just as economically viable ( Lombardi, Pickering, & Pfenninger, 2023) (DeCarolis, 2011) The second reason is that optimal cost solutions may not always be desirable in practice because of sustainability or geopolitical concerns or social acceptability issues (Church & Baez , 2020) (Neumann & Brown , 2021). If society does not accept the solution for any reason, there is very little possibility of implementing it in the real world.

## 2.3 Incorporation of uncertainties in energy system models

Energy system modellers started to incorporate uncertainties into their models to generate multiple feasible solutions. The literature addresses two types of uncertainties (DeCarolis, 2011) (Neumann & Brown , 2021). The first one is parametric uncertainties, which deal with variations in the input



parameter, such as the cost of coal availability of wind and solar energy. Examples of energy optimization models based on parametric uncertainties are scenario analysis, global sensitivity analysis, Monte Carlo analysis and stochastic programming (DeCarolis, 2011).

The second one is structural uncertainties, which deal with the imperfect and incomplete nature of the equations describing the system. Modelling to generate alternatives (MGA) is a type of energy optimization technique which addresses structural uncertainties. MGA is a two-step algorithm to find near-optimal solutions by modifying the objective function of the problem. The first step is to compute the optimal solution for the system. The optimal solution works as a base case for the second step. In the Second step, the objective function is changed by multiplying with a constant known as slack. Then algorithm finds all the energy system configurations with the updated objective function. MGA-based energy system models generate hundreds of equally feasible energy system configurations. The next challenge is to analyze these hundreds of equally feasible, near-optimal energy system configurations and generate insights about the energy system.

The SPORES is a type of MGA algorithm developed by (Lombardi, Pickering, & Pfenninger, 2023). The SPORES algorithm is a recent advancement of the MGA method to find near-optimal solutions that look for both spatially and technically distinctive feasible system designs while improving search efficiency and computational tractability. The workflow of the SPORES algorithm is detailed in the same publication. Results used as a test case in this thesis are the Euro calliope model's results solved using the SPORES algorithm.

## 2.4 Prominent ways of analysing near optimal energy systems

From the literature (Pickering, Lombardi, & Pfenninger, 2022), (Neumann & Brown , 2021), (Bryant & Lempert, 2009) three prominent ways of analyzing hundreds of near-optimal energy system configurations have been identified. This section elaborates on all three ways and discusses the advantage and disadvantages of each.

### 2.4.1 Expert picking solution based on experience

In the first method, experts chose representative solutions based on their experience to build a discussion around. This method allows easy communication as the analyst does not have to go through the entire solution space. Limited information delivery is the strength of this method. However, this method is not standardized, meaning representative solutions will change from expert to expert. Figure 3 demonstrates the expert picking solution based on experience pictorially.

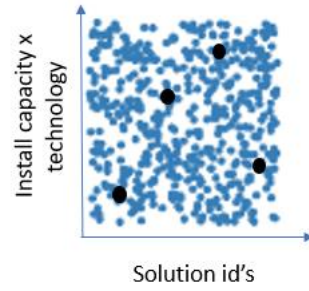


Figure 3 Example of method expert picking solution based on experience

#### 2.4.2 Comparing Extreme ends of results along different uncertainty parameter

In the second method, the boundaries of solutions are compared along different structural uncertainties. Figure 4 shows an approach in which authors compare the maximum and minimum offshore wind installed capacity along different structural uncertainty values. This method is a standardized method which gives over all boundaries of solution. However, this method fails to convey the spatial diversity of solution space and distribution of solution in the solution space.

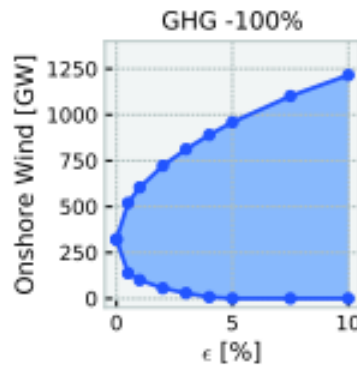


Figure 4 Example of comparing Extreme ends of results along different uncertainty parameter (Neumann & Brown , 2021)

#### 2.4.3 Scenario discovery

Scenario discovery uses data mining algorithms to find critical input parameters for a group of solutions (Bryant & Lempert, 2009). Standardizing computer-based method data is the primary advantage of this method as this method uses data mining algorithms, increasing the scope of analysis. The key characteristic of MGA-based energy system models results is that all solutions' input parameters are the same. Hence, determining critical parameters is not possible in the case of MGA-based energy system model results, which makes scenario discovery unsuitable for analysis results of MGA-based energy system models. Figure 5 demonstrates the scenario discovery method pictorially.

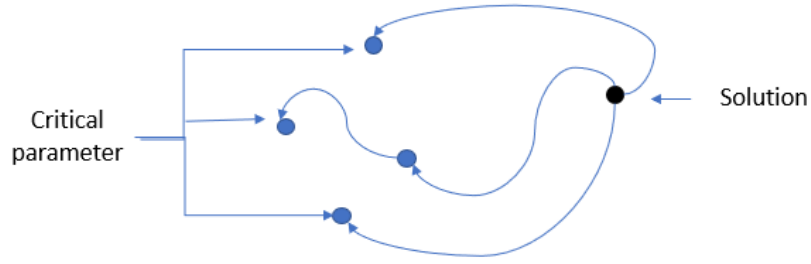


Figure 5 Example of scenario discovery

The literature analysis concludes that no standardized computer-based method is available to facilitate smooth analysis of MGA-based energy optimization models. Hence, this thesis aims to design a computer-based method to facilitate smooth analysis of MGA-based energy system optimization models.

The conclusion of the literature analysis of prominent ways of analyzing near-optimal energy system configuration is that no standardized computer-based method is available to facilitate the analysis of MGA-based energy optimization models. Hence, this thesis aims to build a computer-based package to analyze the results of MGA based energy system model.

Key features of the package are decided by combing features of current methods of analysis of near-optimal energy system configuration and discussion with the MGA-based energy model expert. Seven key features, namely: automated method, groups solutions, gives representative solutions, shows boundaries of solution space, shows spatial diversity, interactive dashboard, shows interdependency of technologies, are identified. Automation of the method makes the method more accessible. At the same time, grouping the results and providing representative solutions ensured minimal data delivery. Showing boundaries of solution space gives an idea of boundaries of solution space. Providing spatial diversity information is crucial as most MGA models have high spatial variation. Showing the impact of one technology on another facilitates system-wide analysis. Finally, an interactive dashboard makes it easy for users to use the method.

These seven features were a guideline for developing the MGA analysis method and package. Furthermore, the same features compare the MGA analysis package against other published methods.

## 2.5 Development of comparison matrix to compare MGA analysis package

The MGA analysis package method has to compare with current prominent published methods to determine the edge of the MGA analysis package over other methods. For this purpose, a comparison matrix has been developed using identified key features in section 2.4. The comparison matrix's column has a different research publication method, and each row has the expected key feature of MGA analysis methods. If a method has a particular feature, the associated cell will be assigned as green. If a method

does not have a particular characteristic, the cell associated with them will be assigned as red, and if any feature does not apply to a particular method, the cell associated with them will be assigned as grey.

Five case studies used for comparison were chosen from 11 case studies of application MGA analysis of future energy systems. These studies are (Pickering, Lombardi, & Pfenninger, 2022), (Neumann & Brown , 2021), (DeCarolis, Babae, Kanungo, & Li, 2015), (Sasse & Trutnevyte, 2020), (Price & Keppo, 2017). The application of the MGA analysis package is demonstrated by analysing the results of the sector couple euro calliope model, which has 98 nodes and 441 solutions. Hence Chosen case studies for comparison have to be multi-nodal. These selection criteria allowed us to compare the analysis methods more fairly. Figure 6 shows the indicator colours used in the comparison matrix, and Figure 7 shows the blank comparison matrix, which will be used during the application part.



Figure 6 Indicator for the analysis matrix

Research papers Features	Pickering et al	Neumann et al	<u>DeCarolis</u> et al	Price et al	<u>Sasse et</u> al	MGA analysis package
Automated method						
Groups the solution space						
Gives representative solution						
Shows boundaries of solution space						
Shows spital diversity						
Interactive dashboard						
Shows interdependency of technologies						

Figure 7 Blank comparison matrix for comparing MGA analysis package

# Chapter 3: Sector coupled Euro-calliope model

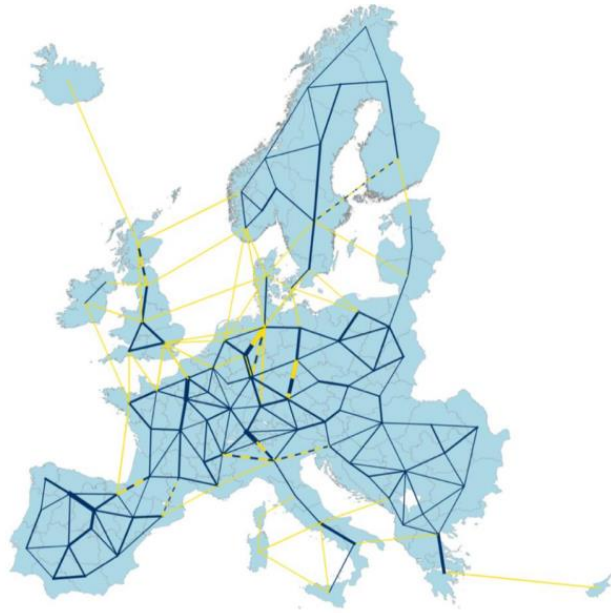
This chapter introduces the model and results used as a test case for the MGA analysis package developed in this thesis. Chapter 3 will convey the complexity of handling the MGA-based energy optimisation model results. In this chapter, Section 3.1 elaborates on the design of the sector-coupled euro calliope model. Furthermore, section 3.2 shows the model's primary and secondary results and discusses the results' complexity.

## 3.1 Sector coupled Euro calliope model

This section details the sector-coupled Euro calliope model whose results are used as a test case for the package developed in this thesis. The results used are the solution of the sector-coupled euro calliope model solved using SPORES algorithms (type of MGA method) with a slack of 10 (Pickering, Lombardi, & Pfenninger, 2022).

Sector coupled Euro calliope model is built in the open-source energy modelling software called Calliope (Lombardi & Pfenninger, 2023). The model represents 35 European countries: the EU-27 (minus Malta), Norway, Iceland, Switzerland, Bosnia and Herzegovina, Montenegro, North Macedonia, Serbia, Albania, and the United Kingdom (Pickering, Lombardi, & Pfenninger, 2022). There are 98 nodes in the model. Nodes function as a place where energy can be generated, stored and consumed. Energy can be exchanged between the nodes; however, energy cannot be imported or exported from the boundaries of the model. Figure 8 shows the map of the Euro calliope model with all the nodes and country borders.

There are 38 energy technologies in the model which can either generate, store or consume energy. As mentioned above, energy is produced, stored and consumed at nodes. Energy can be transported to other nodes via energy carriers. There are 13 carries in the model, namely electricity, hydrogen, CO<sub>2</sub>, liquid and gaseous hydrocarbon (kerosene, methanol, diesel and methane), solids (residual biofuel and municipal waste), low-temperature heat (combined space heat and hot water, and cooking heat), and vehicle distance (heavy-and light-duty road vehicles) (site the joules paper). These carriers are also used energy technology to generate, store, or consume energy.



*Figure 8 Sector coupled Euro calliope model map (Pickering, Lombardi, & Pfenninger, 2022)*

The demand data for the Euro-Calliope are sourced from Eurostat, JRC-IDEES and Open Power System Data databases (Pickering, Lombardi, & Pfenninger, 2022). The supply data, such as technology costs, are sourced from the Danish Energy Agency technology catalogue (Pickering, Lombardi, & Pfenninger, 2022). The demand and supply data used for generating results is from 2018. The model is solved on a 2-hour resolution using the SPORES algorithm with a slack of 10 percent. The model generated 441 near economically viable energy systems for Europe. These results are made available to the public via an interactive dashboard to encourage further research (Pfenninger & Lombardi, 2022)

Three key findings of this research are: first, there is a diversity of options available to reach carbon neutrality by 2050; second, there are some must-have technologies without which reaching carbon neutrality in Europe is not possible, third there exists a great deal of flexibility in terms of infrastructure deployment.

### 3.2 Results of sector coupled Euro calliope model

The results of the Euro-SPORES model are taken from the Zendo repository (Pickering, 2022). The primary results of the Euro SPORES are installed capacities of technologies. Figure 9 shows those results, which are stored in the Excel sheet. The spore column represents the solution id, the techs column shows the technology, the locs column represents the installation location of the given technology, and the nameplate\_capacity column shows the installed capacity of the given technology at a given location.

As we know, the sector-coupled euro calliope model has 441 solutions, 35 technologies and 98 different locations. The number of rows in the data sheet roughly equates to  $441 \times 35 \times 98$ , equating to around 1.5 million rows of data.

	A	B	C	D	E	F	G	H
1	spore	techs	locs	carriers	unit	nameplate_capacity		
2	0	battery	ALB_1	electricity	tw	9.30E-06		
3	0	battery	AUT_1	electricity	tw	6.82E-06		
4	0	battery	AUT_2	electricity	tw	7.10E-06		
5	0	battery	AUT_3	electricity	tw	6.68E-06		
6	0	battery	BEL_1	electricity	tw	6.58E-06		
7	0	battery	BGR_1	electricity	tw	9.31E-06		
8	0	battery	BIH_1	electricity	tw	8.26E-06		
9	0	battery	CHE_1	electricity	tw	6.56E-06		
10	0	battery	CHE_2	electricity	tw	6.55E-06		
11	0	battery	CYP_1	electricity	tw	1.03E-05		
12	0	battery	CZE_1	electricity	tw	6.48E-06		
13	0	battery	CZE_2	electricity	tw	6.94E-06		
14	0	battery	DEU_1	electricity	tw	7.14E-06		
15	0	battery	DEU_2	electricity	tw	7.23E-06		
16	0	battery	DEU_3	electricity	tw	8.34E-06		
17	0	battery	DEU_4	electricity	tw	7.45E-06		
18	0	battery	DEU_5	electricity	tw	7.70E-06		
19	0	battery	DEU_6	electricity	tw	7.62E-06		
20	0	battery	DEU_7	electricity	tw	7.99E-06		
21	0	battery	DNK_1	electricity	tw	8.08E-06		
22	0	battery	DNK_2	electricity	tw	6.24E-06		
23	0	battery	ESP_1	electricity	tw	5.59E-06		
24	0	battery	ESP_10	electricity	tw	5.91E-06		
25	0	battery	ESP_11	electricity	tw	5.93E-06		

Figure 9 Primary results of Sector coupled euro calliope model

Primary results can be extrapolated to find secondary results. Secondary results such as electrification of the heating and transport sector, import-export of energy, grid transfer capacity curtailment of energy production, and percentage of electric vehicles can be extracted from the model. Figure 10 shows the secondary results of the model. The SPROE column shows the solution id, the metric column shows secondary results from the model, and the paper metric column gives the numeric value of the secondary result. Furthermore, Figure 11 shows the grid transfer capacity results. These are unique results as grid transfer capacity has to have two nodes, importing and exporting.

	A	B	C	D	E	F
1	spore	metric	unit	paper_metrics		
2	0	curtailment	percentage	0.608657		
3	0	electricity_production_gini	fraction	0.651432		
4	0	average_national_import	twh	30.10306		
5	0	fuel_autarky_gini	fraction	0.810345		
6	0	storage_discharge_capacity	tw	1.255737		
7	0	transport_electrification	percentage	100		
8	0	heat_electrification	percentage	98.46999		
9	0	ev_flexibility	fraction	0.86525		
10	0	biofuel_utilisation	percentage	85.79324		
11	1	curtailment	percentage	2.590369		
12	1	electricity_production_gini	fraction	0.677328		
13	1	average_national_import	twh	30.12293		
14	1	fuel_autarky_gini	fraction	0.936996		
15	1	storage_discharge_capacity	tw	1.277774		
16	1	transport_electrification	percentage	60.70534		
17	1	heat_electrification	percentage	92.06902		
18	1	ev_flexibility	fraction	0.766944		
19	1	biofuel_utilisation	percentage	99.94445		
20	2	curtailment	percentage	3.099465		
21	2	electricity_production_gini	fraction	0.59504		
22	2	average_national_import	twh	22.31212		
23	2	fuel_autarky_gini	fraction	0.763133		
24	2	storage_discharge_capacity	tw	7.699765		
25	2	transport_electrification	percentage	99.79716		

Figure 10 Secondary results of sector coupled euro calliope model

1	spore	techs	exporting_unit	carriers	importing_grid_transfer_capacity		
2	0	ac_transmission	ALB_1	tw	electricity	GRC_1	0.0008
3	0	ac_transmission	ALB_1	tw	electricity	MKD_1	0.0007
4	0	ac_transmission	ALB_1	tw	electricity	MNE_1	0.001684
5	0	ac_transmission	ALB_1	tw	electricity	SRB_1	0.002002
6	0	ac_transmission	AUT_1	tw	electricity	AUT_2	0.0063
7	0	ac_transmission	AUT_1	tw	electricity	CHE_1	0.0009
8	0	ac_transmission	AUT_1	tw	electricity	CHE_2	0.0015
9	0	ac_transmission	AUT_1	tw	electricity	DEU_6	0.0028
10	0	ac_transmission	AUT_1	tw	electricity	DEU_7	0.0025
11	0	ac_transmission	AUT_1	tw	electricity	ITA_1	0.002378
12	0	ac_transmission	AUT_2	tw	electricity	AUT_1	0.0063
13	0	ac_transmission	AUT_2	tw	electricity	AUT_3	0.024425
14	0	ac_transmission	AUT_2	tw	electricity	DEU_7	0.015969
15	0	ac_transmission	AUT_2	tw	electricity	SVN_1	0.00336

Figure 11 Grid transfer capacity results

The preliminary analysis of the results of Euro-SPORES shows the sheer amount of data generated by the Euro-SPORES in the form of results. As mentioned in section 1.2, from the point of view of an analyst, it becomes difficult to provide in-depth analysis for future action due to the sheer amount of data. Results have to be grouped together to minimize the amount of data the analyst has to analyze,



which dictates the choice of use of the clustering process as a first step of the MGA analysis package. The decision to convert the MGA analysis method into a computer-based customizable package was based on the need for faster and automatic data processing. An interactive dashboard was added for the MGA analysis package to perform data processing with less user inputs.

# Part 2: Design

# Chapter 4: Design of MGA analysis method

This chapter aims to explain the design of the MGA analysis method, along with the procedure of automating the MGA analysis method by coding it in a Python environment. Section 4.1 elaborates on the structure of the MGA analysis method with key steps involved. Section 4.2 explain the clustering method, and section 4.3 details finding representative solutions. Finally, section 4.4 explains the design choice made during building, finding a representative solution algorithm.

## 4.1 Three steps of MGA analysis method

There are three main steps of the MGA analysis method. Figure 12 shows the overview of the MGA analysis package. Clustering the solution space is the first step in the MGA analysis package. Jenks's natural algorithm is used in the first step. Jenks's natural algorithm (Ahmad, 2019) is a machine learning clustering algorithm, also known as a one-dimensional K-means clustering algorithm. Finding representative solutions is the second step of the MGA analysis package. A custom algorithm based on mean and distance is developed for this step. The last step of the MGA analysis package is to identify a suitable number of clusters. For this step, a new algorithm based on variance is designed to identify suitable clusters.

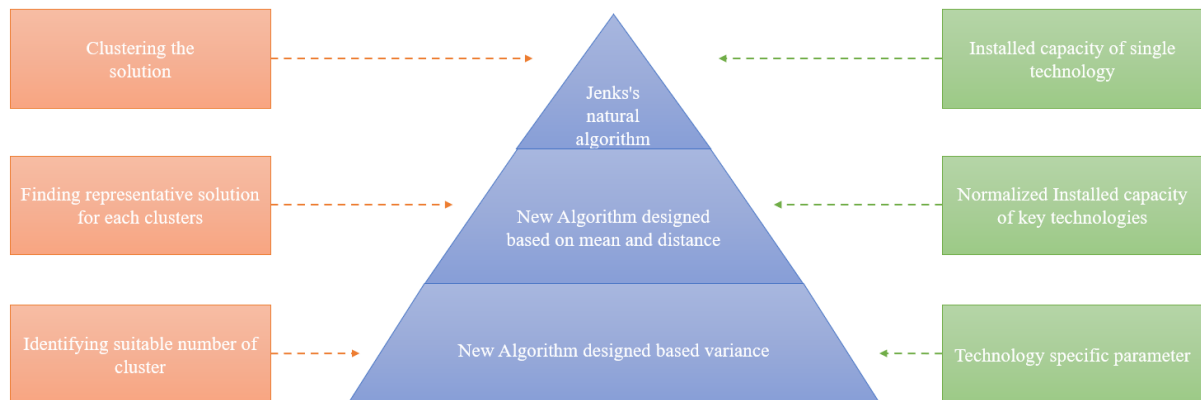


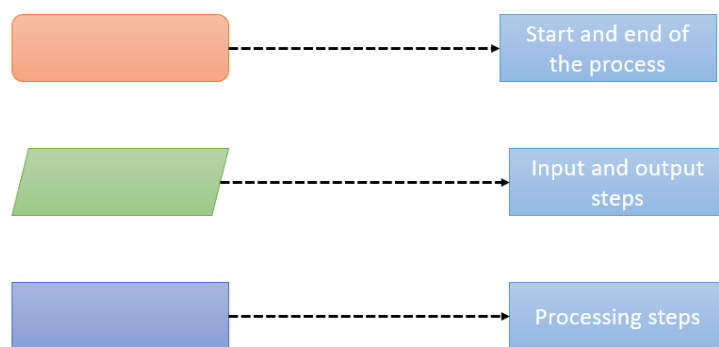
Figure 12 Overview of MGA analysis package

The first step uses data one technology for clustering. At the start of the process, an analyst has to choose one technology to analyse whose data is used in the clustering step. For the Euro SPORES model, installed capacity results are the primary results hence installed capacity is used in the clustering step in the application step. However, the MGA analysis method sees the installed capacities as just a stream

of data. Hence another parameter can be used based on the models whose primary results are not installed capacities.

The second step uses data from key technologies in the system. Key technology depends on the type of research question one wants to address using the model. Key technologies in the system have to be determined before the start of analysis by a model expert and technology expert. The data of the same parameter (in the case of thesis installed capacity) has to be used for all the key technologies. The third step uses the data of secondary results of the model and primary data of technologies which impacts the primary technology most. The data of relevant secondary results and data technology that impacts primary technology are called technology-specific parameters. Technology-specific parameters also have to be determined before the start of analysis by model experts and technology experts. Each subsequent step has more data than the previous step, making the analysis more comprehensive.

In the explanation of the package, first logic flow chart of the given process is presented, and then the Python code for that process is explained. Some standard symbols are used in the flow chart. Figure 13 shows the legend for the logic flow chart.



*Figure 13 Legend for logic flow chart*

## 4.2 Clustering

The choice of clustering algorithm was heavily influenced by the algorithm choices of finding a representative solution. We decided to use Measures of central tendencies to find the representative solution as those parameters give an aggregate idea of a data set. The three most common measures of central tendency are the mean, median and mode. Mode measures the repetition of data, and median shows the data's midpoint. Mean shows the average value of data. Combining mean with variance (a measure of dispersion which shows the spread of data) and boundary value of data will give analysts a better idea of the cluster. Hence the decision to use the mean as the main component of finding a representative solution is made.

Therefore, clustering has to be done so that variance within the cluster would be minimal, but the variance between clusters would be maximum. Jenks's natural break (JNB) (Ahmad, 2019) and K-means clustering (Sharma, 2023) work on the principle of minimizing variance inside the cluster and

maximizing variance between the clusters. The key difference is that K-means clustering is a two-dimensional clustering algorithm, whereas JNB is a one-dimensional clustering algorithm. In the case of the K-means algorithm, users have to provide centre points around which cluster forms; however, in the case of JNB, users do not have to provide centre points. If the user has to provide a centre point, finding a representative solution step becomes irrelevant, as the user will dictate the centre point. Hence the decision to use JNB for the first step is taken.

#### 4.2.1 How does Jenks natural break (JNB) algorithm works?

Jenks natural break (JNB) algorithm is an iterative process in which data is divided into random groups using different sets of breaks to find which set of breaks which has the smallest in-class variance (Ahmad, 2019). Breaks are the boundaries between two clusters. In case of this thesis the number of clusters and installed capacity of primary technology are input for the JNB algorithm. The process starts with arranging the data in ascending order and then adding breaks at arbitrary places. Then the following two steps are repeated.

**Step 1:** Calculate the sum of squared deviation from the cluster means.

**Step 2:** Shift data points from one cluster to another cluster and repeat.

In the end combination of breaks with minimum in-class variance and maximum inter-class variance is selected as the most suitable break combination. Figure 14 shows the flow chart of Jenks natural break algorithm.

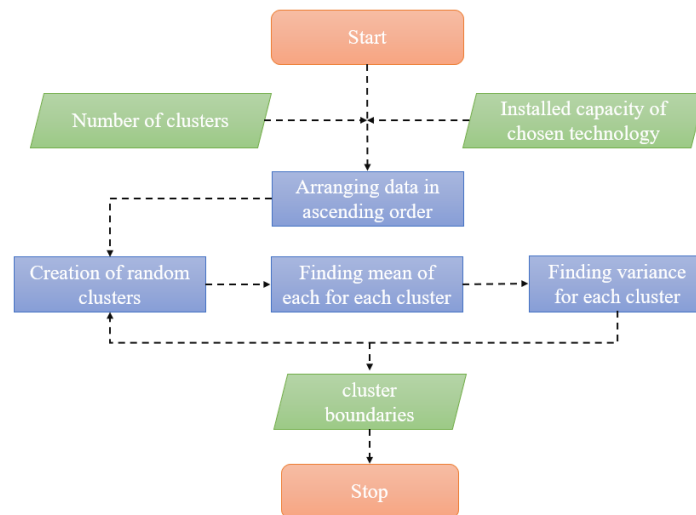


Figure 14 Jenks Natural Break (flow chart)

JNB algorithm is then coded into a function name `jenks_breaks_function` to call and use multiple times for specific technology.

### 4.2.2 Application of JNB algorithm and storing clustered data

The standard Outputs of the JNB algorithm are the cluster boundaries. The MGA analysis method needs to store the values of installed capacity for further usage in subsequent steps. Hence function name clustered\_data\_function is created, which can apply the JNB algorithm to a specific technology and store the installed capacity values and SPOERS id in separate data sets. The first part of the function applies the JNB algorithm to data. The second part stores the cluster data in a multi-level dictionary called clustered\_data. Figure 15 shows the flow chart of the application of the clustering data function, and Table 2 explains the code of the clustering data function.

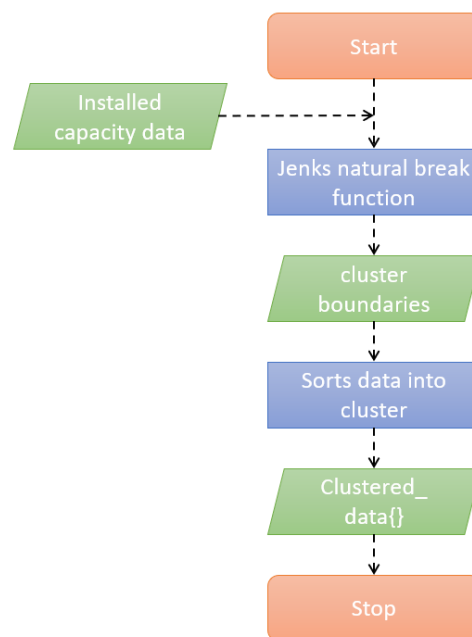


Figure 15 Clustering algorithm (flow chart)

Table 1 Clustering algorithm (code explanation)

Function	Code
Define the function	<pre>def clustered_data_function (ins_cap,gini_ins,paper_metrics, display_technology,cluster_number):</pre>
Creating a dataframe to store cluster boundaries.	<pre>cluster_boundry=pd.DataFrame()</pre>
Creating a list of all the key technologies in the system by extracting first row of ins_cap dataframe	<pre>technology=ins_cap.columns.values.tolist() spore=technology[0] technology.pop(0)</pre>
Application of Jenks natural break algorithm on the data set and storing breaks values in the cluster_boundry dataset.	<pre>for i in technology:     tech=ins_cap[i].tolist()     breaks = get_jenks_breaks (tech, cluster_number)</pre>

	<code>cluster_boundry[i]=breaks</code>
Initializing new data frame to store clustered data	<code>clustered_data={}</code>
Creating two empty list. One to store installed capacity another to store corresponding SPORES id	<code>value_1=[] index_1=[]</code>
Storing installed capacity and SPORES id in two lists if installed capacity falls between cluster boundaries	<code>if cluster_boundry[j][k]&lt;= ins_cap[j].values[i]&lt;cluster_boundry[j][k+1]:  value_1.append(ins_cap[j].values[i]) index_1.append(ins_cap[j].index[i])</code>
Storing two list created in last step in the clustered_data dictionary.  Last two steps run in a loop to store all the data.	<code>clustered_data[j][k+1]={}  clustered_data[j][k+1]['installed_capacity'] =value_1 clustered_data[j][k+1]['SPORES']=index_1</code>

The outcome of this step is a multi-level dictionary named `clustered_data` which contains a dictionary for each cluster. Each cluster-specific dictionary has SPORES ids and installed technology capacities for the specific cluster.

### 4.3 Finding representative solutions (FRS) algorithm

The idea was that the mean of data of chosen technology would represent a solution for the cluster. However, the energy system is a complex system, with hundreds of technologies working together to balance the system. We decided to include data from other technologies in this system to find a representative solution. This way, the representative solution will be for the system, not for a single technology.

Hence a custom algorithm named “Find Representative Solutions” (FRS) has been designed to find a representative solution for the clustered data using mean as a base for identification. Input for the FRS algorithm is a data frame with a normalized capacity of all key technologies. The package creates the normalized capacity data frame using the installed capacity data of key technology. Figure 16 shows the flow chart of the FRS algorithm. Furthermore, table 3 explains the code of FRS in detail.

FRS algorithm is a two-step algorithm. The first step of FRS is to find a mean cluster solution. In this step, the algorithm collects the normalized capacity of all SPORES which are part of the cluster. Then find the mean of normalized capacity for key technologies and store the values as a list.

For particular cluster

$$Mean_{sol\ cluster1} = [Mtech_1, Mtech_2, Mtech_3, Mtech_4 \dots Mtech_n]$$

Where Meansol cluster1 is a matrix containing the mean solution of key technologies for cluster 1. Mtech<sub>i</sub> is the mean normalized installed capacity of technology ‘i’ for the cluster.

In the second step, FRS finds a deviation of each solution from the mean solution. The deviation is calculated by adding the absolute difference between the mean solution and SPORE solution of each key technology. The SPORES with the least deviation is considered the representative solution.

$$Deviation = [|Mtech_1 - X_mtech_1| + |Mtech_2 - X_mtech_2| + \dots |Mtech_n - X_mtech_n|]$$

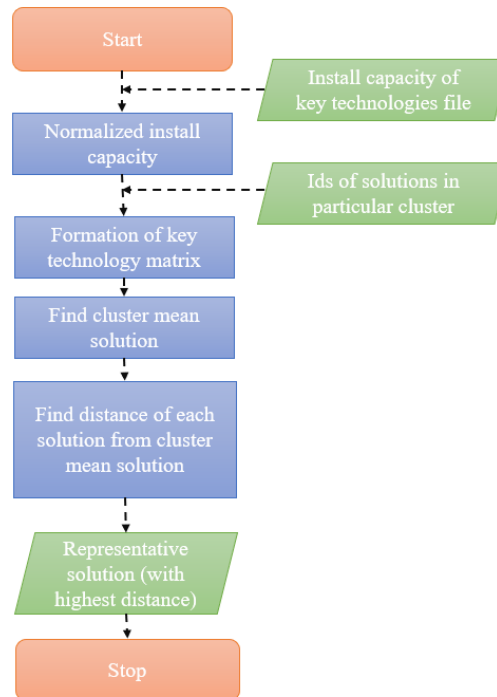


Figure 16 Finding representative solution algorithm (flow chart)

Where Mtech<sub>i</sub> is the mean normalized installed capacity of technology ‘i’ for the cluster, and X<sub>m</sub>tech<sub>i</sub> is the normalized installed capacity of technology ‘i’ with SPORES id ‘m’.

Table 2 Finding representative solution algorithm (code explanation)

Function	Code
Calculate normalized installed capacity for each technology	<pre>for j in technology:     x_scaled =min_max_scaler.     fit_transform(ins_cap[[spore,j]])</pre>
Storing normalized installed capacity in the data frame named norm_cap	<pre>norm_cap.loc[:,j] = x_scaled[:,1]</pre>



Creating two empty list. One to store normalized capacity another to store corresponding SPORES id	<pre>norm_1=[] index_1=[]</pre>
Storing normalized capacity and SPORES id in two lists if installed capacity falls between cluster boundaries	<pre>if cluster_boundry[j][k]&lt;= ins_cap[j].values[i]&lt;cluster_boundry[j][k+1]:  norm_1.append(norm_cap[j].values[i]) index_1.append(ins_cap[j].index[i])</pre>
Storing two list created in last step in the clustered_data dictionary.	<pre>clustered_data[j][k+1]={} clustered_data[j][k+1][ 'normalized_capacity']=value_1 clustered_data[j][k+1]['SPORES']=index_1</pre>
Creating list for each key technology to store normalized installed capacity for particular cluster.	<pre>for j in technology:     for k in range (0, len(cluster_boundry)-1):         mul_list_dict = collections.defaultdict(list)</pre>
Store normalized install capacity in the list	<pre>for i in clustered_data[j][k+1]['SPORES']: for l in technology: mul_list_dict[l].append (norm_cap[l].values[i])</pre>
Finds mean solution for each technology	<pre>for l in technology: g=mul_list_dict[l] mean_sol.append(mean(g))</pre>
Finds deviation of each SPORES solution of mean solution	<pre>for j in technology: for k in range (0, len(cluster_boundry)-1): abs_val_list=[] for i in clustered_data[j][k+1]['SPORES']: abs_val=0 count=0 for l in technology: temp_list=[] for n in range(len(clustered_data[j] [k+1]['mean_sol'])):  temp_list.append(abs(clustered_data[j][k+1] ['mean_sol'][n]-norm_cap[l].values[i]))  abs_val=abs_val+temp_list[count] count=count+1 abs_val_list.append(abs_val)</pre>
Finds minimum deviation and the SPORES with minimum deviation. Store the SPORES id with minimum deviation as a representative solution.	<pre>x=pd.Series(abs_val_list).idxmin()  clustered_data[j][k+1]['rep_sol']= clustered_data[j][k+1]['SPORES'][x]</pre>

The outcome of this step is the updated multi-level dictionary from the last step. The second step adds the id of the representative solution to each cluster-specific dictionary.

#### 4.4 Finding suitable number of clusters (FSNC) algorithm

JNB algorithms can determine the suitable number of clusters using the installed capacity of primary technology by finding the Goodness of Variance fit (GVF). GVF is defined as the ratio between the difference of variance of data from the mean of data and the variance of clustered data from the cluster mean to the variance of data from the mean of data. The drawback of this method is limited information usage in determining the suitable number of clusters. For this purpose, a new algorithm has been designed to determine the suitable cluster number. The new algorithm uses the data of technology-specific parameters to find the suitable number of clusters. Integrating data of technology-specific parameters (TSP) in the finding suitable number cluster make cluster determination algorithms specific for each technology.

The algorithm initiates with the number of clusters equal to 2 and iterates then check if any two parameters along different clusters are within the range of 1.5% of each other. If more than 1/3 of the parameters are found to be in the range of 1.5% of each other, then the suitable number of the cluster is one less than used for the current iteration. The algorithm uses technology-specific parameters to determine the suitable number of clusters. A function named `cluster_check_function` is defined to automate the process. Figure 17 shows the flow chart of finding a suitable number of clusters, and then Table 6 explains the code of the cluster check function.

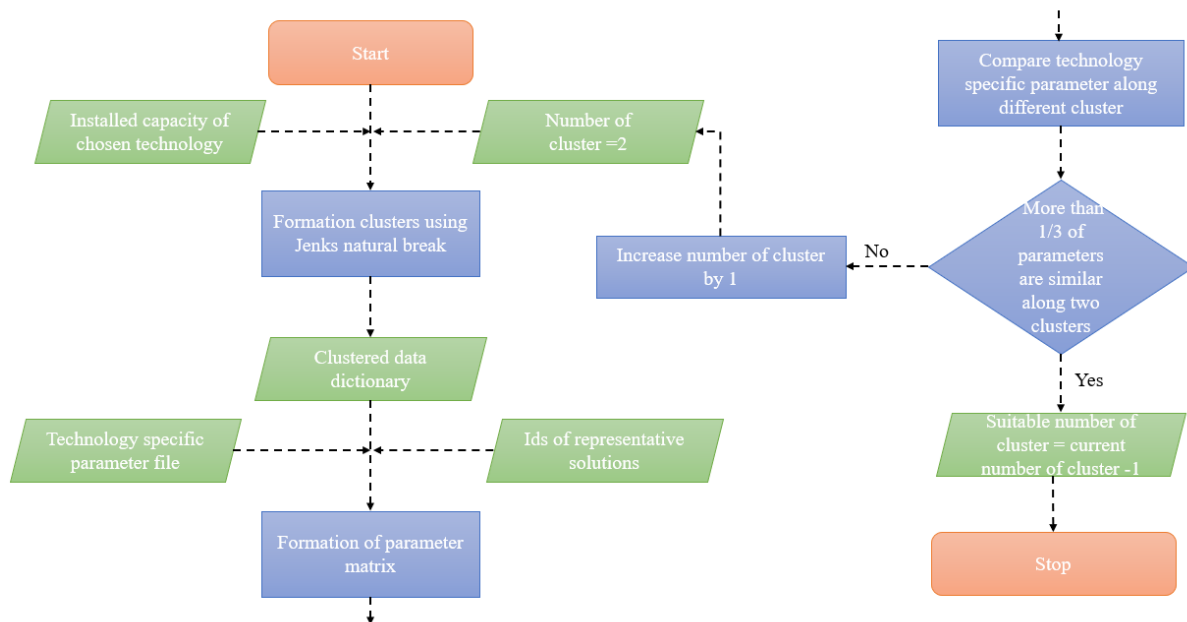


Figure 17 Finding suitable number of clusters algorithm. (flow chart)

Table 3 Finding suitable number of clusters (code)

Function	Code
Defining cluster_check_function	<pre>def cluster_check_function(ins_cap,gini_ins, paper_metrics,display_technology,cluster_number):</pre>
Cluster_data_function use to cluster data	<pre>[clusters,cluster_boundry]=clustered_data_ function(ins_cap,gini_ins,paper_metrics, display_technology,cluster_number)</pre>
Defining technology specific parameter	<pre>parameter=paper_metrics.columns.values.tolist() parameter.pop(0) u=['Regional_eq_index'] parameters=u+parameter cluster_list=list(range(1, cluster_number+1))</pre>
Storing representative solution parameters value in the list	<pre>for j in display_technology: mul_list_dict = collections.defaultdict(list) for k in range (0, len(cluster_boundry)-1): m=clusters[j][k+1]['SPORES'].index (clusters[j][k+1]['rep_sol'])  for l in parameters: mul_list_dict[l].append(clusters[j][k+1][l][m])</pre>
Finding minimum difference between technology specific parameter along cluster.	<pre>for n in parameters: for u in range(0,len(cluster_list)): for v in range(0,len(cluster_list)): if u!=v: col.append(abs(mul_list_dict[n][u]- mul_list_dict[n][v]))  div_para.append((min(col)))</pre>
Check if parameters are within 1.5 percent of each other.	<pre>for s in range(len(div_para)): if div_para[s]&lt;0.015: check_mat.append(1) else: check_mat.append(0)</pre>
Returns if number of cluster-1 if more than 1/3 parameter are with 1.5 percent of each other along different cluster	<pre>if mean(check_mat)&gt;0.3333: return(cluster_number-1) break</pre>

Cluster\_check\_function is called in the main code. Using for loop, the function is iterated to find the suitable number of clusters. Table 7 explains the application finding the suitable number of cluster algorithms.

Table 4 Application of suitable number of cluster algorithm (code)

Function	Code

Iterate cluster_check_function to find suitable number of cluster.	<pre> for i in range(2,10): num_of_cluster=cluster_check_function (ins_cap,gini_ins,paper_metrics, display_technology,i)if isinstance(num_of_cluster,int)==True: print("ideal number of cluster:", num_of_cluster) break </pre>
--	---

FSNC algorithm starts with two as the base number of clusters. Then checks for a suitable number of clusters sequentially. Hence the minimum number of clusters which the algorithm can form is two. This limitation is due to the limitation of the JNB algorithm, which accepts two or more two as input for a number of clusters. As a result, the MGA analysis method has to form two clusters even though a suitable number of clusters determined by the FSNC algorithm is one.

#### 4.5 Regional equality index as technology specific parameter

As stated previously, there are 98 nodes in the model, and most of the solutions vary based on the spatial arrangement of technology. The inclusion of parameter which shows spatial diversity is crucial for the analysis of the results of MGA-based energy system models. Along with Representative solutions providing some parameters describing how concentrated energy infrastructure is will give analyst a better understanding of the option available regarding infrastructure deployment. In the article on the regional impact of electricity system transition, the authors used the Gini index to showcase the spatial variation (Sasse & Trutnevyte, 2020). With the same idea, we designed a method to find the regional equality index for each SPORES result using installed capacity data.

We defined the term regional equality index based on the well know the statistical term- Gini index. The Gini index represents how concerted data is distributed over space. In this thesis, the regional equality index is one minus the Gini index. The following equation shows the formula of reginal equality index expressed using gini index.

$$\text{Regional equality index} = 1 - \text{Gini index}$$

$$\text{Gini index} = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n^2 \bar{x}}$$

The value of the regional equality index varies from 0 to 1. A regional equality index equal to one denotes that technology is installed uniformly across all the regions, whereas a regional equality index equal to zero denotes that technology is concentrated in some parts. In Euro-SPORES results, some technology does not exist for all the solutions. In that case, we assume regional equality index to be 1 for that particular technology in that particular solution. The following table shows the python code to find the regional equality index for critical technologies. Table 4 explains the code for finding the regional equality index.

Table 5 regional equality index (code explanation)

Function	Code
Importing relevant libraries	<pre>import pandas as pd import warnings  warnings.filterwarnings('ignore')</pre>
Importing installed capacity data in the data frame	<pre>data_1='installed_capacity_loc.csv' cost=pd.read_csv(data_1)</pre>
Storing installed capacity of particular technology of particular SPORES in data frame named x	<pre>for j in techs: df=cost[(cost[cost.columns[1]] ==j)] for k in range(len(set(cost[cost.columns[0]]))):     df1=df[df[cost.columns[0]]==k]  x=df1[cost.columns[5]].values.tolist()</pre>
Implementation of Gini index formula	<pre>for i in range(len(x)):     for l in range(len(x)):         temp=abs(x[l]-x[i])         numm=numm+temp      lenx=len(set(cost['locs']))     meanx=sum(x)/lenx     den=2*meanx*lenx*lenx</pre>
Finding regional equality index from Gini index	<pre>if den==0:     gini=0 else:     gini=numm/den  reg_eq=1-gini</pre>
String regional equality index in the data frame and storing data frame in the form of CSV file	<pre>gini_cost.loc[k,j] = reg_eq gini_cost.to_csv('gini.csv')</pre>

This code saves the regional equality index in the CSV file name gini in the local directory. Regional equality index is one of technology specific parameter for the case study in this thesis and used to identify suitable number of clusters.

The end of the third step concludes the MGA analysis process. The next step in the thesis is to combine all three steps in the Python package and develop an interactive dashboard to use the MGA analysis process with fewer user inputs.

# Chapter 5: Design of MGA analysis package and interactive dashboard

This chapter aims to discuss the design of the MGA analysis package from the MGA analysis process. Section 5.1 discuss key features of the MGA analysis package. Section 5.2 explains the format of input files for the package, and section 5.3 details two modes of the MGA analysis package. Section 5.4 elaborates on the design choices of the interactive dashboard for the MGA analysis package.

## 5.1 Key features of MGA analysis package

The MGA analysis package is coded in Python, as Python is one of the best platforms for developing data analysis and machine learning packages. MGA analysis package uses Python 3.9 as an interrupter with basic libraries such as Pandas, Numpy, Matplotlib, Collection, Statistics, IPython, and sklearn. Installation of these libraries in the Python environment is necessary to use the MGA analysis package.

Figure 18 shows the constituents of the MGA analysis package. The MGA analysis package requires three input files. The first file contains the total installed capacity of key technology in the system. The second file contains the installed capacity of key technology with their location data. The third file contains the data on technology-specific parameters. Three steps of MGA analysis process are coded in the separate functions, which are also input of MGA analysis package. There are two modes of operation MGA analysis package, auto and manual. Each mode is initiated by its *.ipynb* file in the package.

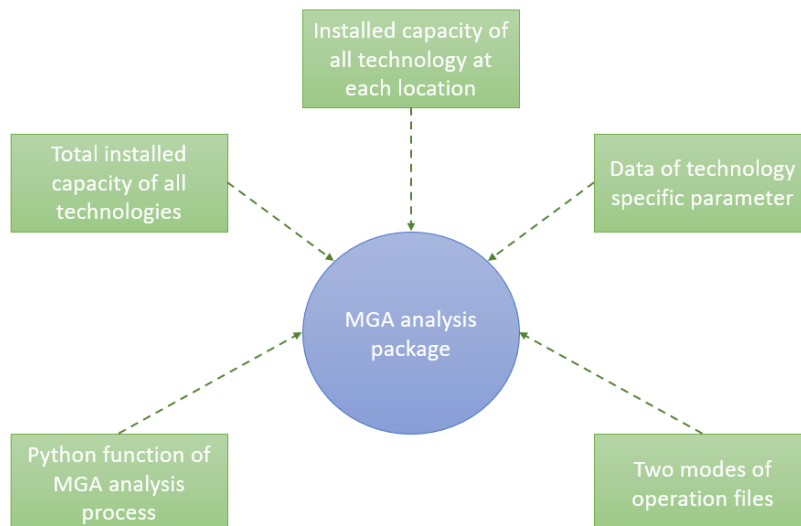


Figure 18 constituents MGA analysis package

## 5.2 Format specification of input files

The primary output file of Euro-SPORES model contains 1.5 million rows of data. That data must be converted into a specific format to input it the MGA analysis package. First, correct data has to be extracted in a uniform format. In this thesis, data processing is done in two steps, first, in Python, where data of key technology is extracted and then in Excel, where data is converted into a uniform format.

### Data processing results in Python

The results of the Euro-SPORES model contain 38 different energy technologies. As mentioned in the previously, data of key technologies are extracted for the MGA analysis package. Table 1 shows the python code of for selecting key technologies.

Table 6 Data cleaning in python (code explanation)

Function	Code
Initializing libraries require for the job	<pre>import numpy as np import pandas as pd import warnings warnings.filterwarnings('ignore')</pre>
Loading CSV file in the environment and converting it into dataframe name df	<pre>data='nameplate_capacity.csv' df=pd.read_csv(data)</pre>
Extracting installed capacity of selected technologies and storing them in dataframe name df_1	<pre>df_1=df.loc[df['techs'].isin(['wind_onshore', 'wind_offshore','roof_mounted_pv', 'open_field_pv','biofuel_supply', 'waste_supply', 'electrolysis','battery', 'heat_storage_big', 'heat_storage_small','hydrogen_storage'])]</pre>
Storing results in a csv file	<pre>df_1.to_csv('installed_capacity_loc.csv')</pre>
Adding installed capacity SPORE and technology wise to get net installed capacity	<pre>df_2=df_1.groupby(['spore', 'techs']).sum()</pre>
Storing results in a csv file	<pre>df_2.to_csv('net_installed_capacity.csv')</pre>

### Processing results in Excel.

Data from the python data cleaning process has specific technology needed for analysis. However, each technology is not present in each SOPRE. To make data uniform, one has to add the missing SOPRE number for all the technology with zero installed capacity value. Second step is done manually in Excel. Figure 19 shows data cleaning in the excel.

19	wind_offshore	0.030640884		19	wind_offshore	0.030640884
20	wind_offshore	1.021837925		20	wind_offshore	1.021837925
22	wind_offshore	1.336157018		21	wind_offshore	0
23	wind_offshore	0.968699044		22	wind_offshore	1.336157018
24	wind_offshore	0.470595016		23	wind_offshore	0.968699044
25	wind_offshore	1.148362366		24	wind_offshore	0.470595016

Figure 19 Data cleaning in excel.

After processing data in Excel, data is converted into a respective file format. The input files for the MGA analysis package have been in a specific format to operate the package. All three files must be in a .CSV format and arranged in a specific way. The first file is the installed capacity of key technology with location data. Figure 20 shows the format of the installed capacity file. The spore column contains the SPORES id of the solution, the tech column contains the name of the technology, the locs column contains the location of the technology, and the nameplate\_capacity contains the installed capacity of a particular technology at a particular location. When a user inputs his data, the sequence of the title row must remain the same. Users should sequentially add key technologies data. The user should name the file *installed\_capacity\_loc.csv*.

The second file contains the total installed capacity of key technologies. Figure 21 shows the format of the total installed capacity file. The first row contains the name of all the key technologies in the system. The number and name of key technologies in both files should be matched. The first column contains the SPORES id of the solution. The user should name the file *net\_installed\_capacity.csv*. Figure 22 shows the format of the technology-specific parameter file. The first column contains the SPORES id, and the subsequent column contains the technology parameter. The user should name the file based on primary technology, for example, *wind\_onshore.csv*.

1	spore	techs	locs	carriers	unit	nameplate_capacity
2	0	wind_onshore	ALB_1	electricity	tw	1.86E-05
3	0	wind_onshore	AUT_1	electricity	tw	1.13E-05
4	0	wind_onshore	AUT_2	electricity	tw	9.86E-06
5	0	wind_onshore	AUT_3	electricity	tw	8.13E-06
6	0	wind_onshore	BEL_1	electricity	tw	0.07825
7	0	wind_onshore	BGR_1	electricity	tw	0.105463
8	0	wind_onshore	BIH_1	electricity	tw	2.90E-05
9	0	wind_onshore	CHE_1	electricity	tw	5.01E-05
10	0	wind_onshore	CHE_2	electricity	tw	4.81E-06
11	0	wind_onshore	CYP_1	electricity	tw	8.72E-06
12	0	wind_onshore	CZE_1	electricity	tw	5.86E-06
13	0	wind_onshore	CZE_2	electricity	tw	0.171196
14	0	wind_onshore	DEU_1	electricity	tw	1.34E-05
15	0	wind_onshore	DEU_2	electricity	tw	0.091638
16	0	wind_onshore	DEU_3	electricity	tw	0.024011

Figure 20 Installed capacity (location wise) input file format



1	SPORES	wind_onshore	wind_offshore	roof_mounted_pv	open_field_pv
2	0	3379.83	0.05	324.36	1641.73
3	1	3167.7	154.05	0.03	1682.03
4	2	3095.8	30.73	54.91	2665.79
5	3	3079.34	30.54	54.41	2673.66
6	4	2914.41	421.13	0.01	2840.46
7	5	1788.86	1283.36	46.43	1685.24
8	6	1713.3	901.3	147.51	1528.85
9	7	1777.84	1314.52	65.97	1457.05
10	8	2067.22	959.42	16.4	1358.32
11	9	1745.39	1344.19	27.29	1510.21
12	10	1923.29	1048.39	67.75	1147.57
13	11	2264.69	1151.88	0.07	1102.36
14	12	3095.11	29.13	56.64	2648.77
15	13	3166.17	9.34	12.89	2259.57
16	14	2048.83	1077.7	53.81	2455.63

Figure 21 Total installed capacity input file format

1	SPORE	Total wind capacity	Electrolysis capacity	Transport electrification	Heat electrification
2	0	0.785019178	0.191	1	0.984699912
3	1	0.764297834	0.354	0.607053379	0.920690231
4	2	0.694708625	0.403	0.997971649	0.710497876
5	3	0.688773473	0.399	0.987369992	0.709598848
6	4	0.769213494	0.76	1	0.054679562
7	5	0.67534898	0.265	1	0.940473812
8	6	0.512223204	0.276	1	0.571792135
9	7	0.682528196	0.288	1	0.960226537
10	8	0.659101279	0.044	1	0.997897051
11	9	0.681537222	0.283	1	0.947934784
12	10	0.639509931	0.251	0.999833902	0.906341164
13	11	0.7980979	0.306	1	0.98197204
14	12	0.693892319	0.404	0.985726396	0.714105142
15	13	0.712168309	0.382	0.943459534	0.927677436
16	14	0.694708625	0.3	1	0.978775928
17	15	0.674040751	0.29	1	0.933231228
18	16	0.630462835	0.04	1	0.997460363

Figure 22 Technology specific parameters input file format

### 5.3 Two modes of MGA analysis package

The MGA analysis package offers two modes of application. Figure 23 shows two modes of the MGA analysis package. The first mode is the auto mode, in which the package decides a suitable number of clusters for the given technology. The second mode is the manual mode, in which the user decides the number of clusters. Auto mode is for the analysis of results of MGA-based energy systems. In contrast, a manual model is created to develop the MGA analysis package further and improve algorithms in it. By selecting respective. ipynb file user can choose the mode of operation of the MGA analysis package.

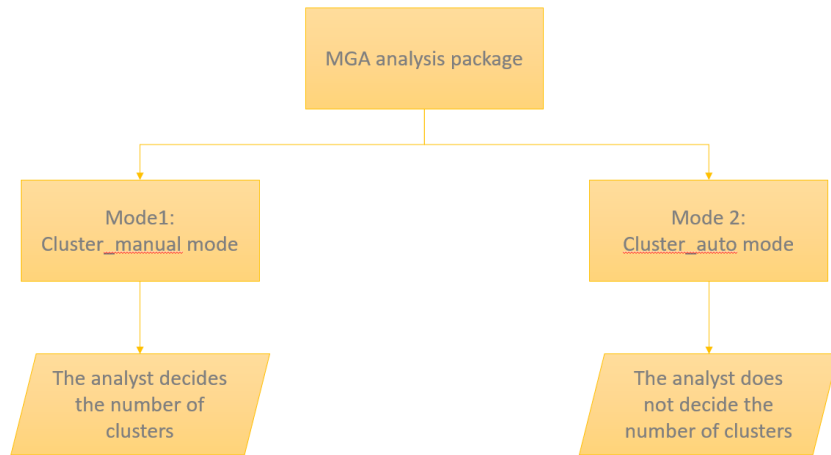


Figure 23 Two modes the MGA analysis package

### 5.3.1 Manual mode of operation

In the manual mode user input the technology to analyse and a number of clusters to form. Input files for the manual mode are the net installed capacity of each key technology, technology-specific parameters for each technology and a file which contains the regional equality index for each key technology. Table 8 presents the code of the manual mode of operation.

Table 7 cluster manual mode (code)

Function	Code
Importing libraries and function	<pre>import pandas as pd  from clustered_data_function import clustered_data_function from cluster_plot_function import cluster_plot_function from parameter_plot_function_clusterwise import parameter_plot_function_clusterwise from parameter_plot_function import parameter_plot_function</pre>
Importing installed capacity and regional equality index file and assigning them to data frame.	<pre>data_1='net_installed_capacity.csv' ins_cap=pd.read_csv(data_1) data_3='gini.csv' gini_ins=pd.read_csv(data_3)</pre>
Takes user input how many clusters to create.	<pre>cluster_number=int(input("Number of cluster you want: "))</pre>
Display key technologies in the system	<pre>technology=ins_cap.columns.values.tolist() technology.pop(0) print('Key technologies in the system are: ', technology)</pre>

Takes user input which technology to analyze	<code>display_tech=input("Which one technology clusters do you want to see: ")</code>
Imports technology specific parameter file	<code>data_2='display_parameters/{}.csv'.format(display_technology[0]) paper_metrics=pd.read_csv(data_2)</code>
Implementation of clustered_data_function	<code>[clusters,cluster_boundry]=clustered_data_function(ins_cap,gini_ins,paper_metrics,display_technology,cluster_number)</code>
Implementation of cluster_plot_function and parameter_plot_function	<code>cluster_Figureure=cluster_plot_function(display_technology,cluster_boundry,clusters) parameter_Figureure=parameter_plot_function(paper_metrics,display_technology,cluster_boundry,clusters,cluster_number)</code>
Taking user input which cluster to analyze in depth and parameter_plot_function_clusterwise	<code>cluster_id=int(input("which cluster do you want to analyze: "))  parameter_Figureure=parameter_plot_function_clusterwise(paper_metrics,display_technology,cluster_boundry,clusters,cluster_id)</code>

### 5.3.2 Auto mode of operation

In the auto mode user only input the technology to analyse. Input files for the auto mode are the net installed capacity of each key technology, technology-specific parameters for each technology and a file which contains the regional equality index for each key technology. Table 9 presents the auto model of operation.

Table 8 cluster auto mode (code)

Function	Code
Importing libraries and function	<code>import pandas as pd  from clustered_data_function import clustered_data_function from cluster_plot_function import cluster_plot_function from parameter_plot_function_clusterwise import parameter_plot_function_clusterwise from parameter_plot_function import parameter_plot_function from cluster_check_function import cluster_check_function</code>
Importing installed capacity and regional equality index file and assigning them to data frame.	<code>data_1='net_installed_capacity.csv' ins_cap=pd.read_csv(data_1) data_3='gini.csv' gini_ins=pd.read_csv(data_3)</code>

Display key technologies in the system	<pre>technology=ins_cap.columns.values.tolist() technology.pop(0) print('Key technologies in the system are: ', technology)</pre>
Takes user input which technology to analyze	<pre>display_tech=input("Which one technology clusters do you want to see: ")</pre>
Imports technology specific parameter file	<pre>data_2='display_parameters/{}.csv'.format (display_technology[0]) paper_metrics=pd.read_csv(data_2)</pre>
Determining number of suitable cluster for particular technology	<pre>for i in range(2,10):     num_of_cluster=cluster_check_function     (ins_cap,gini_ins,paper_metrics,display_technol     ogy,i)     if isinstance(num_of_cluster,int)==True:         print("ideal number of cluster:",         num_of_cluster)</pre>
Implementation of clustered_data_function	<pre>[clusters,cluster_boundry]=clustered_data_ function(ins_cap,gini_ins,paper_metrics, display_technology,cluster_number)</pre>
Implementation of cluster_plot_function and parameter_plot_function	<pre>cluster_Figureure=cluster_plot_function (display_technology,cluster_boundry,clusters) parameter_Figureure=parameter_plot_function (paper_metrics,display_technology,cluster_bound ry, clusters,cluster_number)</pre>
Taking user input which cluster to analyze in depth and parameter_plot_function_cluster wise	<pre>cluster_id=int(input("which cluster do you want to analyze: "))  parameter_Figureure=parameter_plot_function_ clusterwise(paper_metrics,display_technology, cluster_boundry,clusters,cluster_id)</pre>

## 5.4 MGA analysis dashboard

The main purpose of the MGA analysis dashboard is to give the users an interactive platform from which the user can access the MGA analysis method and MGA analysis package with 4-6 user commands. Figure 24 shows the key feature of the MGA analysis package. MGA analysis dashboard shows user-clustered data in graphical as well as text format. Along with that dashboard provide the facility to do a comparative analysis between the clusters by providing graphs of technology-specific parameters of representative solution and also at the cluster's boundaries.

Jupyter Notebook has built-in functions to take input from the user in various formats. Along with that, Jupyter Notebook allows users to run code block-wise and see the input of each block separately. Hence the decision to construct an interactive dashboard as a *.ipynb* file instead of a *.py* file is made.

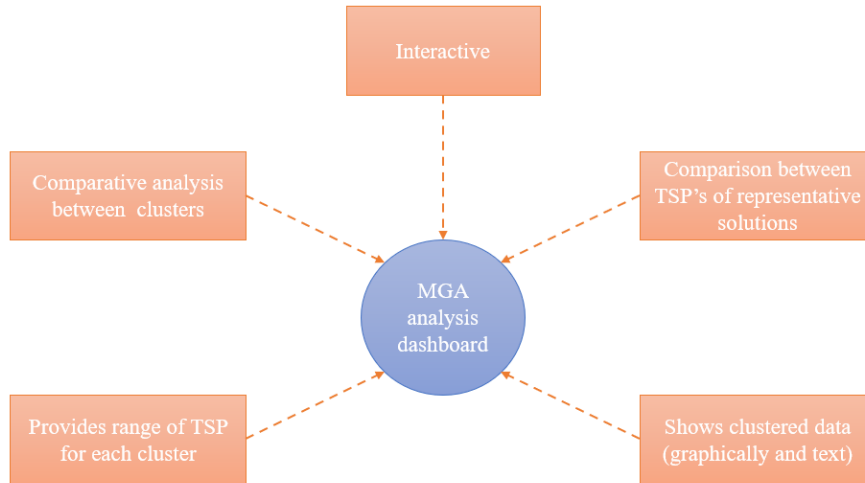


Figure 24 Key features of MGA analysis dashboard

#### 5.4.1 Conveying clustered data via dashboard

The idea behind building a interactive dashboard conveying cluster solutions was to present the results in a way that provides a maximum understanding of the system with minimal information. The use of graphical tools such as scatter plots and statical tools such as minimum, maximum and mean value is used to achieve the goal of interactive dashboard. Following is a sample of how to convey clustered results. Figure 25 and 26 shows the clustered data in pictorially and text format.

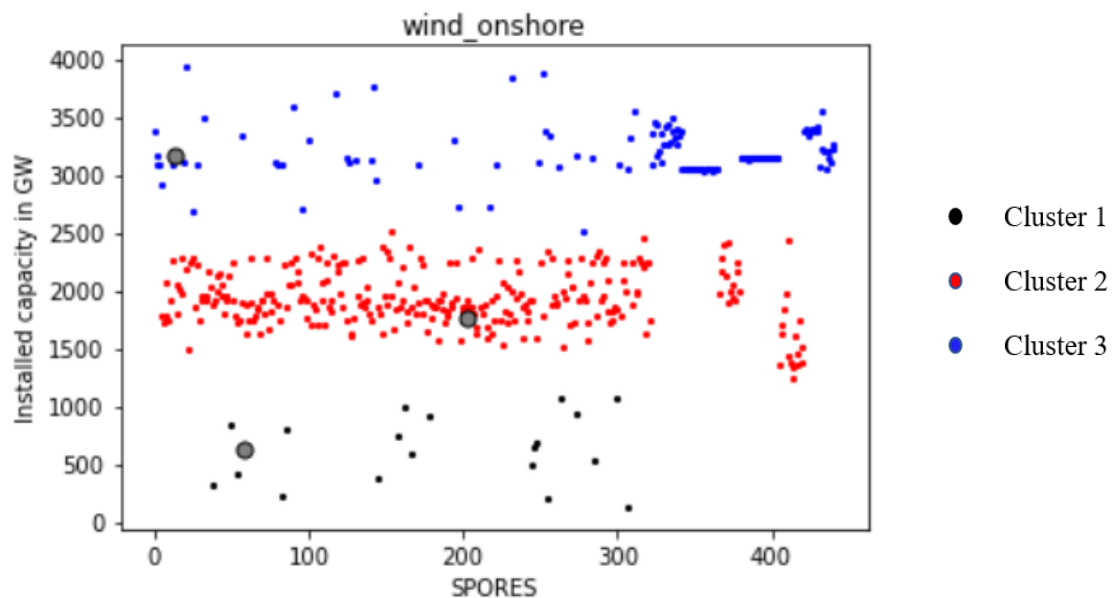


Figure 25 Example of clustered data (graph)

```

Technology: wind_onshore
cluster ID: 1
Number of SPORES: 20
Id of representative SPORE: 58
Installed capacity of representative SPORE: 625.2GW
Minimum install capacity in the cluster: 128.64GW
Maxmium install capacity in the cluster: 1071.62GW

cluster ID: 2
Number of SPORES: 287
Id of representative SPORE: 203
Installed capacity of representative SPORE: 1752.05GW
Minimum install capacity in the cluster: 1238.0GW
Maxmium install capacity in the cluster: 2516.01GW

cluster ID: 3
Number of SPORES: 133
Id of representative SPORE: 13
Installed capacity of representative SPORE: 3166.17GW
Minimum install capacity in the cluster: 2519.11GW
Maxmium install capacity in the cluster: 3928.05GW

```

Figure 26 Example of clustered data (text)

To display scatter plot and allied information function name `cluster_plot_function` is built. Input to function are key technology, cluster boundary and clustered data dictionary. Table 5 explains the logic and code of `cluster_data_function`.

Table 9 Displaying clustered data (code explanation)

Function	Code
Define the function	<pre>def cluster_plot_function(display_technology, cluster_boundry,clustered_data):</pre>
Initiate a list with color name	<pre>col= ['k', 'r', 'b', 'c', 'm', 'y']</pre>
Plot cluster data with different color	<pre>for j in display_technology:     Figure = plt.Figureure()     print('Technology: '+str(j))     for k in range (0, len(cluster_boundry)-1):         m=clustered_data[j][k+1]['SPORES'].         index(clustered_data[j][k+1]['rep_sol'])         for l in range(len(clustered_data[j][k+1]         ['installed_capacity'])):             plt.scatter(clustered_data[j][k+1]             ['SPORES'][l],clustered_data[j][k+1]             ['installed_capacity'][l],c=col[k],s=10, linewidth=0)</pre>
Plot representative solution with grey dot	<pre>plt.scatter(clustered_data[j][k+1]['SPORES'][m] ,clustered_data[j][k+1]['installed_capacity'][m] ,c='tab:gray',s=50, edgecolor='black', linewidth=1)</pre>

Displaying information to the stakeholders	cluster to the	<pre> print('cluster ID: '+ str(k+1)) print('Number of SPORES: ' +str(len(clustered_data[j][k+1]['SPORES']))) print('Id of representative SPORE: ' '+str(clustered_data[j][k+1]['rep_sol'])) print('Installed capacity of representative SPORE: ' +str(clustered_data[j][k+1]['installed_capacity'] [m])+ 'GW') print('Minimum install capacity in the cluster: ' + str(min(clustered_data[j][k+1]['installed_capacity'])) +'GW') print('Maxmium install capacity in the cluster: '+ str(max(clustered_data[j][k+1]['installed_capacity']) )+'GW' ) </pre>
Return Figure as a output of function		<pre> plt.show() return (Figure) print() </pre>

#### 5.4.2 Conveying technology specific parameters via dashboard

As discussed in section 3.4, technology-specific parameters are closely related to a particular technology in the model. As mentioned in section 4.5 regional equality index is one of the technology-specific parameters present in all technology. The technology-specific parameter has to convey so that they will be easy to compare along the clusters and convey their variation inside the cluster. This method has been designed to plot two graphs, one which gives variation inside the cluster and another which gives variation between clusters. To show variation inside the cluster, maximum, minimum, and representative solution value is plotted as a scatter plot. On the other hand, a line graph of the representative solutions has been used to show variation between clusters. Figure 27 compares the technology-specific parameters of the representative solution of each cluster. Figure 28 shows the value of technology-specific parameters of the representative solution and the maximum and minimum values for the cluster.

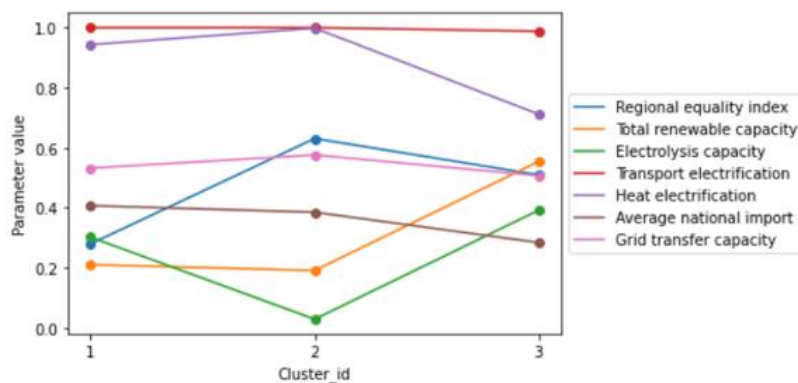


Figure 27 Example of conveying technology specific parameter



Parameter: Regional equality index  
Value for representative SPORE: 0.15  
Minimum value for the cluster: 0.06  
Maximum value for the cluster: 0.91

Parameter: Total wind capacity  
Value for representative SPORE: 0.29  
Minimum value for the cluster: -0.0  
Maximum value for the cluster: 0.4

Parameter: Electrolysis capacity  
Value for representative SPORE: 0.22  
Minimum value for the cluster: 0.17  
Maximum value for the cluster: 1.0

Parameter: Transport electrification  
Value for representative SPORE: 1.0  
Minimum value for the cluster: 1.0  
Maximum value for the cluster: 1.0

Parameter: Transport electrification  
Value for representative SPORE: 1.0  
Minimum value for the cluster: 1.0  
Maximum value for the cluster: 1.0

Parameter: Heat electrification  
Value for representative SPORE: 0.79  
Minimum value for the cluster: 0.63  
Maximum value for the cluster: 0.85

Parameter: Average national import  
Value for representative SPORE: 0.33  
Minimum value for the cluster: 0.27  
Maximum value for the cluster: 0.42

Parameter: Grid transfer capacity  
Value for representative SPORE: 0.36  
Minimum value for the cluster: 0.36  
Maximum value for the cluster: 0.47

Figure 28 Example of converging technology specific parameter (text)

Figure 29 shows the comparison of technology specific parameters between the clusters by providing values for representative solution as well as minimum and maximum value for the cluster.

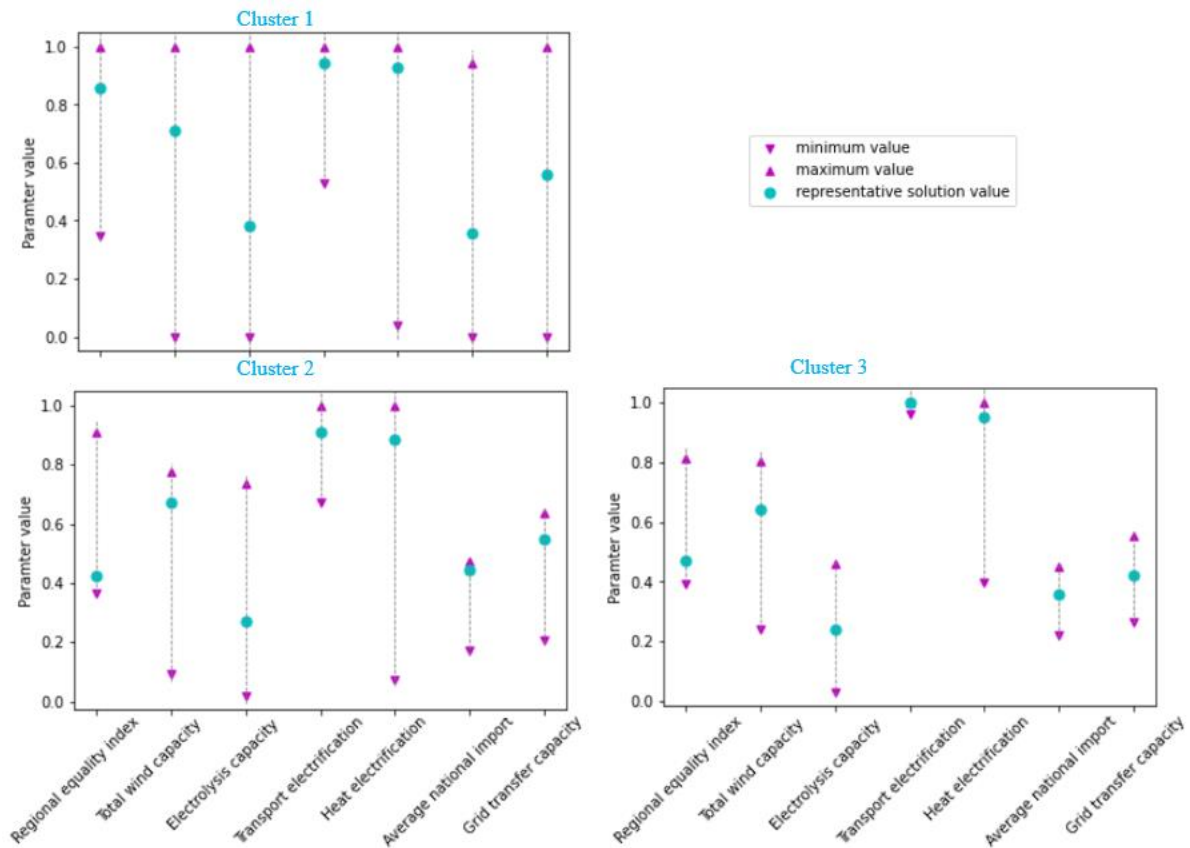


Figure 29 Comparing technology specific parameters boundaries along different cluster



This chapter showcased one way to convert MGA analysis method into MGA analysis package in the Python environment. Python is used as the base environment as it contains many libraries for data analysis like sci-kit Learn, Pandas, Matplotlib and NumPy. The MGA analysis package takes input in the form of three Excel files and has two modes of operation. Auto mode is used to analysis of the results where the analyst does not input the number of clusters. In manual mode, the analyst has to enter the number of clusters. Manual mode is used for the development of the MGA analysis package. The MGA analysis package automates the MGA analysis method and increases the accessibility of the method.

An interactive dashboard is created to facilitate the use of MGA analysis package where analyst can use MGA analysis package with 4-5 inputs. Dashboard shows clustered data in the graph as well as in text format. Along with that dashboard provides boundaries of each cluster. Furthermore, analyst can analyse each cluster in depth by comparing technology specific parameters for representative population and maximum and minimum value of technology specific parameter for that cluster.

## Part 3: Analysis

# Chapter 6: Application of MGA analysis package

This chapter aims to demonstrate the application of the MGA analysis package on the result of Euro-SPORES. Section 6.1 elaborates on the input selection for the analysis, and section 6.2 analyses the trade-off between different energy generation technology of CO<sub>2</sub>-neutral Europe. In the end, section 6.3 compares the MGA analysis package with other published MGA analysis methods to determine the edge of the MGA analysis package over other methods.

As mentioned in chapter one, the guiding research question for the application step is “What *trade-offs can be analysed between energy generation technology of 2050’s CO<sub>2</sub>-neutral Europe using the MGA analysis package?*”. Results of Euro-SPORES are used to determine the trade-offs between energy generation technologies of 2050’s CO<sub>2</sub>-neutral Europe (Pickering, 2022). The description of the sector-coupled euro calliope model is presented in chapter 3. MGA analysis package is used via an interactive dashboard to analyse the results of Euro-SPORES.

## 6.1 Data input

MGA analysis method needs three data inputs. The first input is the data of primary technology for the clustering algorithm; the second input is the data of key technologies in the system—finally, the data of technology-specific parameters for each primary technology. All the end is selected by interacting with the model expert of Euro-SPORES model.

### 6.1.1 Primary technology data

There are six generation technologies in the sector coupled euro calliope model- namely wind power, solar power, hydropower, nuclear power, energy from biofuel, and energy from waste. This thesis demonstrates the application of the MGA analysis package for wind power, solar power and energy from biofuel. In the model of nuclear power, hydropower operates and energy from waste specific determined capacity for almost all the solutions. Hence the decision not to consider those technologies for analysis is taken.

### 6.1.2 Data of key technology

Key technologies in the model vary most along different SPORES results and have the largest share in the system. As mentioned in section one, energy generation technologies in the system are wind power, solar power, hydropower, nuclear power, biofuel energy, and waste energy.

The nuclear power installation cap is based on countries' current policies on nuclear power. Most countries have committed a certain amount of nuclear power to their national agenda. That data is used

as basic in the model. As a result, the installed capacity of nuclear power is almost the same for all SPORES. As per the model, hydropower installations in most countries are at their peak already. Installing further hydropower will cost more due to terrain constraints. As a result, hydropower installation is almost the same in all SPORES results. Hence these two technologies are not considered as input to the package. Hence, key generation technologies in the system are wind power, solar power, energy from biofuel, and energy from waste.

Key balancing technologies in the model are electrolysis, hydrogen storage, battery storage, and heat storage. These technologies are used to negate the impact of the intermittent nature of renewable energy sources.

### 6.1.3 Technology specific parameter

The third input is technology-specific parameters. Technology-specific parameters are the parameters in the model which affect chosen technology the most. These parameters are identified by disusing with the model expert of Euro-SPRES model. There are five technology-specific parameters for all three primary technologies. Regional equality is one of the social parameters that will impact the deployment of renewable energy infrastructure (Sasse & Trutnevyte, 2019). Average notational import/export will act as a proxy to determine the energy independence of countries. Wind and solar are renewable energy sources which produce energy in the form of electricity. Conversely, biofuel energy produces energy in the form of heat. Hence the impact of changes in the installed capacity of these technologies on the electrification of the heating and transport sector is crucial to determine. Intermittent nature and rapid expansion of solar and wind cause more frequent system imbalances and congestion (Stawska, Romero, Weerdt, & Verzijlbergh, 2020). Hence, the impact of integrating different types of renewable on the grid capacity is a crucial factor. All the technology specific parameters are determined by disusing with model expert of Euro-SPORES.

Parameters such as grid transfer capacity and import/export of energy between regions are normalized to facilitate comparison. At the same time, transport and heat electrification were expressed in percentages and converted on a scale of 0-1 by dividing each quality by 100. The range of the regional equality index is from 0-1. Hence it is used as it is.

### Wind energy specific parameters matrix

Figure 30 shows the parameter matrix for wind energy. Technology-specific parameters for wind energy are electrolysis capacity, total wind capacity, regional equality index, grid transfer capacity, heat electrification, transport electrification, and average national import. There are two type of wind technology in the model: wind off-shore and wind-onshore. Total wind capacity monitors changes in one wind technology capacity due to changes in other wind technology capacities. Wind energy especially offshore-wind energy is used to generate green hydrogen to overcome intermittent nature of wind (Singlitico, Østergaard, & Chatzivasileiadis, 2021). The parameter of installed electrolysis

capacity is used to determine the correlations between wind technology and electrolysis technology. The regional equality index is used to monitor the concentration of wind energy technology infrastructure deployment along different clusters. The heat and transport electrification parameters are used to monitor the impact of wind energy technology on the electrification of the heat and transport sector. Electricity grid expansion is one of the crucial parameters for the success of renewable energy. Hence grid transfer capacity is used to determine the correlation between wind energy and grid transfer capacity. National average import parameters monitor nodes' energy independence changes along different wind energy capacities.

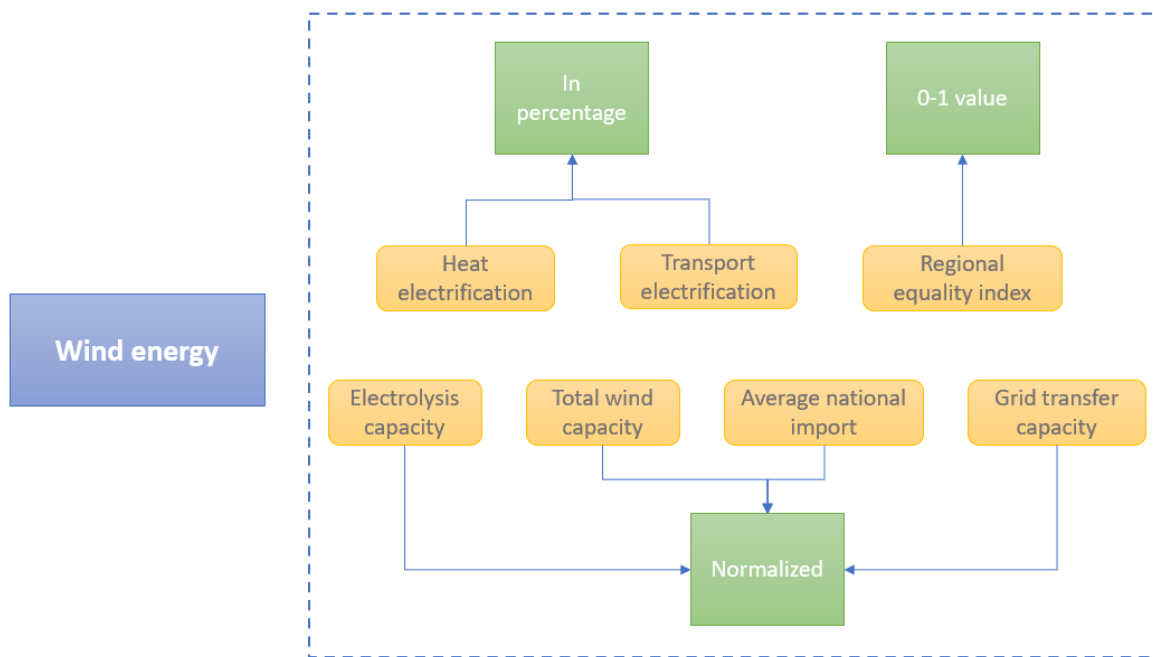


Figure 30 Technology specific parameter matrix for wind energy

### Solar energy specific parameter matrix

Figure 31 shows the parameter matrix for solar energy. Technology-specific parameters for solar energy are battery capacity, total solar capacity, regional equality index, grid transfer capacity, heat electrification, transport electrification, and average national import. There are two types of solar technology in the model: open field solar and roof mounted solar. Total solar capacity monitors changes in one solar technology capacity due to changes in other solar technology capacities. Excess solar energy is stored in the batteries. Parameter of installed battery capacity is used to determine the correlations between solar technology and battery technology. The regional equality index is used to monitor the concentration of solar energy technology infrastructure deployment along different clusters. The heat and transport electrification parameters are used to monitor the impact of solar energy technology on the electrification of the heat and transport sector. Electricity grid expansion is one of

the crucial parameters for the success of renewable energy. Hence grid transfer capacity is used to determine the correlation between solar energy and grid transfer capacity. National average import parameters monitor nodes' energy independence changes along different solar energy capacities.

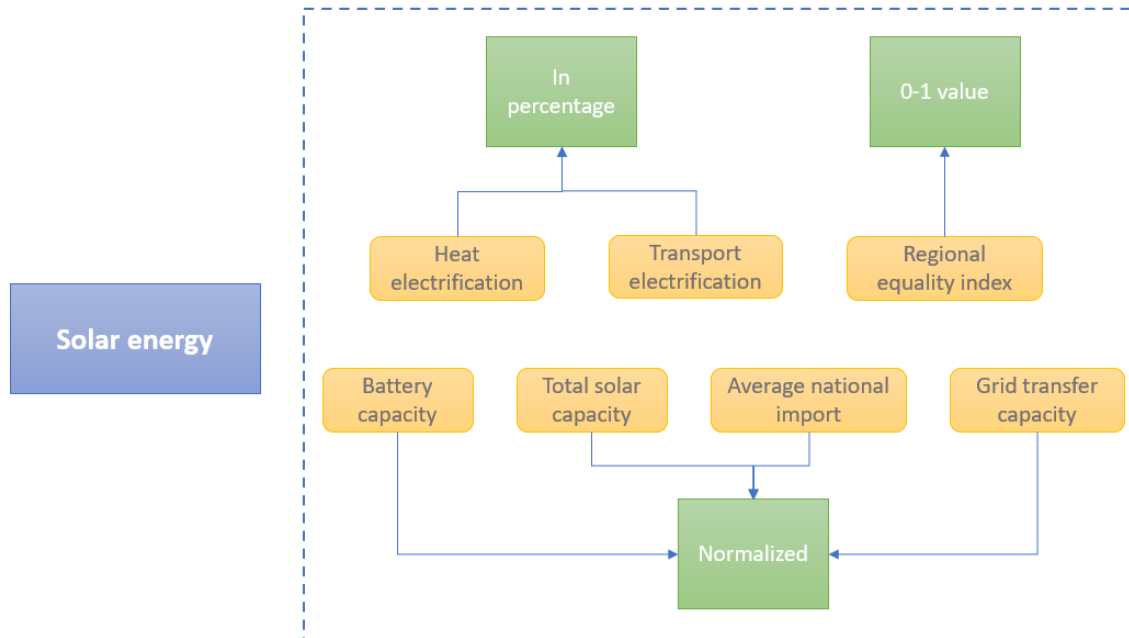


Figure 31 Technology specific parameter matrix for solar energy

### Biofuel energy specific parameters

Figure 32 shows the technology-specific parameters for biofuel. Technology-specific parameters for biofuel energy are the regional equality index, electrolysis capacity, total renewable capacity, grid transfer capacity, heat electrification, transport electrification, and average national import. The regional equality index is used to monitor the concentration of biofuel energy technology infrastructure deployment along different clusters. Parameter of installed capacity of solar and wind is used to monitor the effect of wind and solar capacity changes on the biofuel supply. In the system, synthetic fuel can be produced either via electrolysis or using biofuel via process gasification. Hence electrolysis capacity is monitoring the impact of an increase in biofuel supply on the synthetic fuel production. The heat and transport electrification parameters are used to monitor the impact of biofuel energy technology on the electrification of the heat and transport sector. Electricity grid expansion is one of the crucial parameters for the success of renewable energy. As biofuel is used as a substitute for electrical energy, grid transfer

capacity determines the correlation between biofuel supply and grid transfer capacity. National average import parameters monitor nodes' energy independence changes along different biofuel usage.

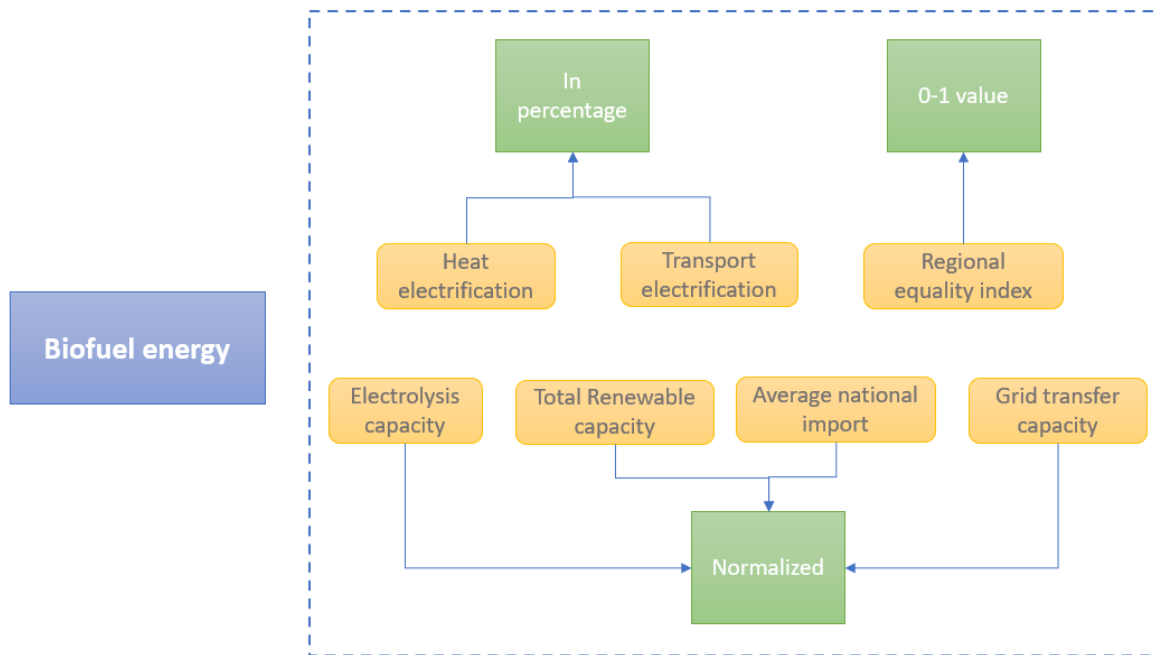


Figure 32 Technology specific parameter matrix for biofuel energy

## 6.2 Analysis of generation technologies

This section analyzes the results of sector coupled euro calliope model using MGA analysis package to find trade-offs between energy generation technologies of 2050's CO2 neutral Europe.

### 6.2.1 Wind energy generation analysis

As mentioned in the section 6.1.3 wind energy has two types of technology in the model. First is technology type being wind onshore and second technology type is wind offshore.

#### Wind onshore technology

In the case of onshore wind technology, three is the suitable cluster number. Figure 33 shows the clustered wind onshore technology data with representative solutions. Minimum onshore wind capacity in the system is 625.2 GW and maximum wind onshore capacity in the system is 3908.05 GW.

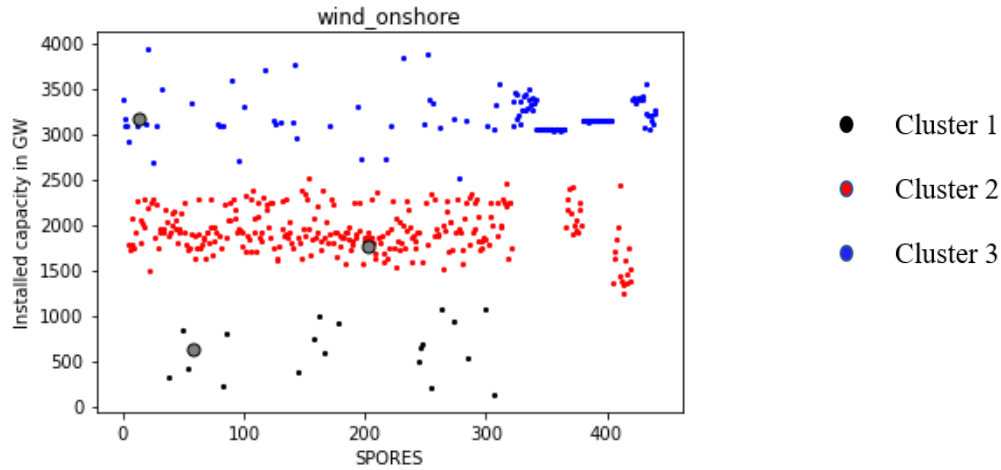


Figure 33 Clustered data of wind onshore technology

The first cluster contains 20 SPORES solutions, and the representative solution for the first cluster is a solution with SPORES id 58. The maximum and minimum onshore wind capacities are 1071.62 GW and 128.64 GW. The installed capacity of the representative solution is 625.20 GW. The second cluster contains 287 SPORES solutions. The id of the representative solution is 287, with an installed capacity of wind onshore equal to 1752 GW. The maximum and minimum installed capacity of wind onshore for the second cluster are 2516.01 GW and 1238 GW, respectively. The third cluster contains 133 SPORES solutions. The id of the representative solution is 13, with the installed capacity of wind onshore equal to 3166.17 GW. This cluster's maximum and minimum installed capacity of wind onshore are 3928.05 GW and 2519.11 GW, respectively.

### Comparative analysis between representative solutions

Figure 34 shows the representative solutions value for all the clusters and all technology specific parameters.

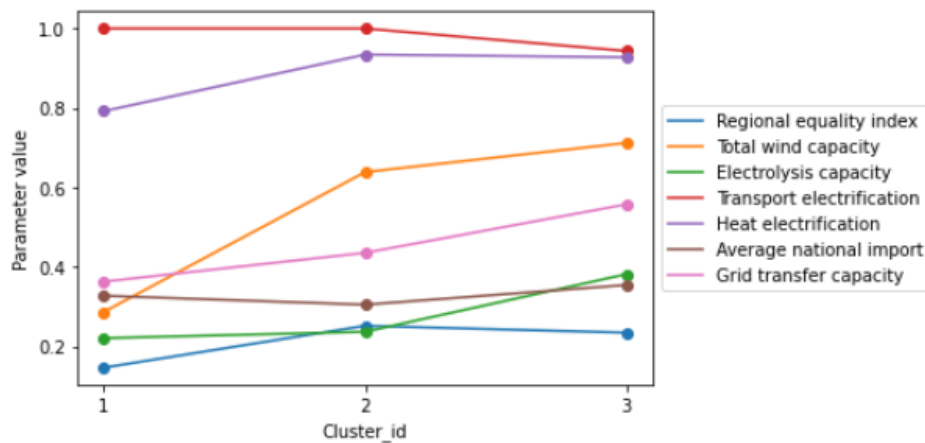


Figure 34 Parameter values of representative solution (wind onshore)



The value regional equality index of the representative solution for all clusters lies between 0.15 to 0.25 indicating that wind technology development is concentrated in few places. The regional equality index increases from cluster one to cluster two and then reduces slightly in cluster three. This trend shows that with an increase in wind production, first system starts to diversify the deployment of wind farms; however, after a certain limit again starts to concentrate on the deployment of wind farms.

The average national import increases first and then decreases, correlates with the regional equality index, indicating a focus on disturbed energy infrastructure with a medium amount of wind energy and concentrated infrastructure with high wind energy in the system. Grid transfer capacity increases slightly from cluster one to cluster two; however, there is a significant increase from cluster two to cluster three as a high percentage of electricity is present in the energy mix.

The electrolysis capacity stays the same for cluster 1 and cluster 2 but significantly increases in the case of cluster 3, indicating that with high wind capacity, electrolysis is being used to balance technology. The value of transport electrification decreases as wind energy production increases. This indicates that transport infrastructure becomes more robust regarding fuel choice with high wind energy installation. In the case of high wind capacity, excess electricity is converted into hydrogen, which is converted into synthetic fuel used in the transport sector. The value of the heating electrification parameter increases with an increase in wind capacity in the system. As wind capacity increases, the percentage of electricity in the energy mix increases. Due to excess of electricity, previously inefficient electrical heating in some sectors becomes a viable option in case of representative solution.

## Cluster wise analysis

Figure 35 shows the graphs for technology-specific parameters with maximum, minimum and representative solution value.

### A. Regional equality index, Grid transfer capacity and average national import

Figure 36 shows energy generation locations on the European map. The upper limit of the regional equality index decreases slightly from cluster one to cluster two. However, the decrease from cluster two to cluster three is more significant, showing that wind onshore infrastructure gets concentrated at a high install capacity of wind. Due to the more distributive nature of wind installation in clusters one and two, the average national import values are significantly low compared to cluster three, making nodes more energy independent. Consequently, due to low wind install capacity and less need for the exchange of electricity between the nodes, grid transfer capacity is relatively less in cluster one and cluster two compared to cluster three. On the other hand, with high wind capacity in the system, system design becomes more robust regarding grid expansion and national import. Figure 37 shows a map of national average import and electrolysis capacity for Europe. Whereas Figure 38 shows the required grid transfer capacity for Europe.

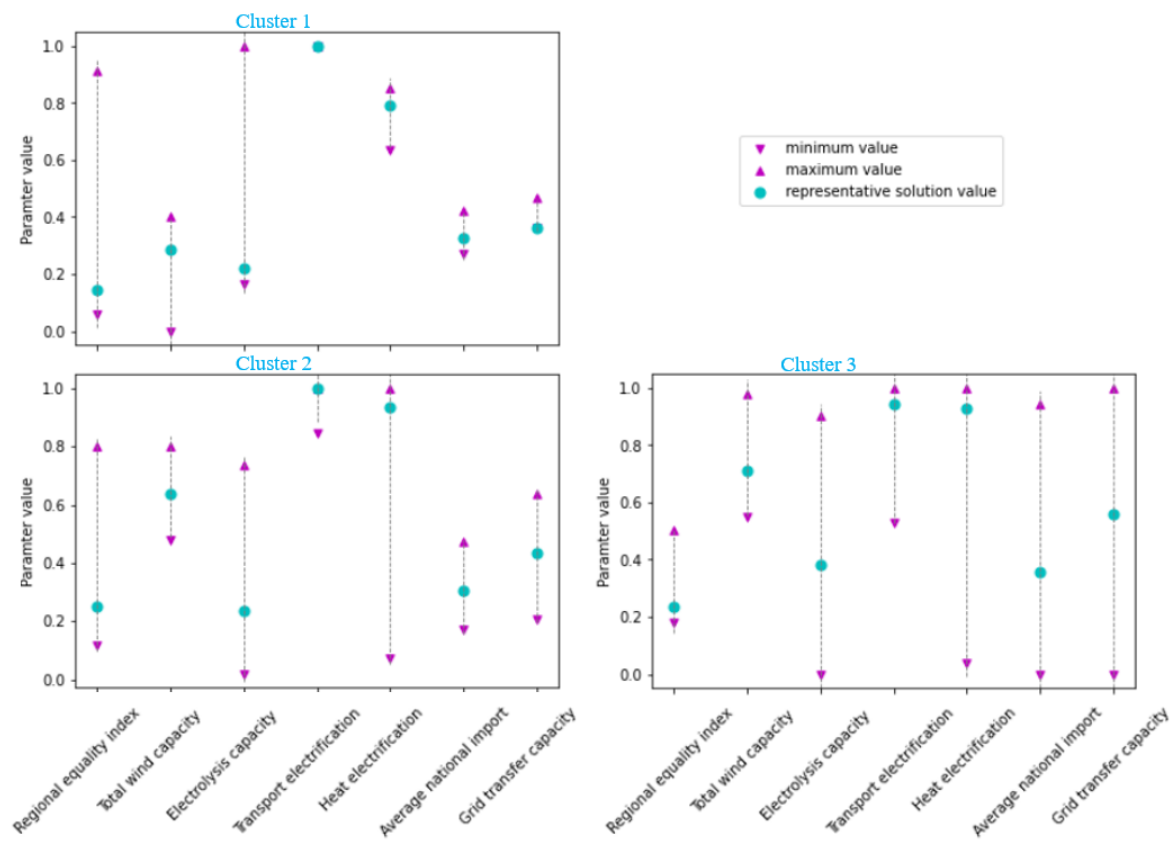


Figure 35 Technology specific parameter (comparative analysis wind onshore)

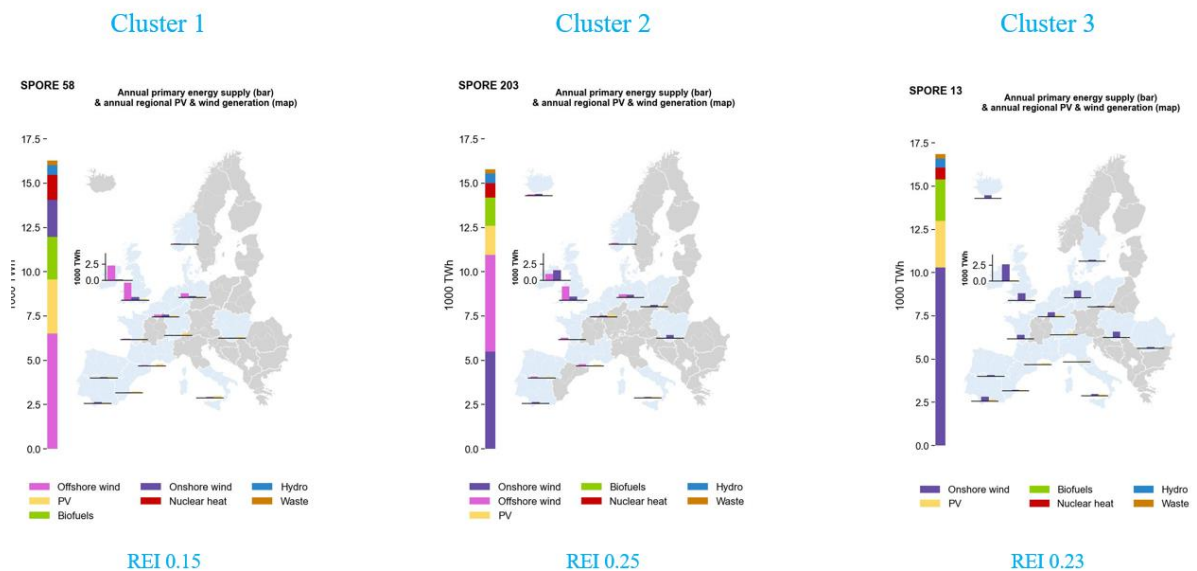


Figure 36 Regional equality index (Map)

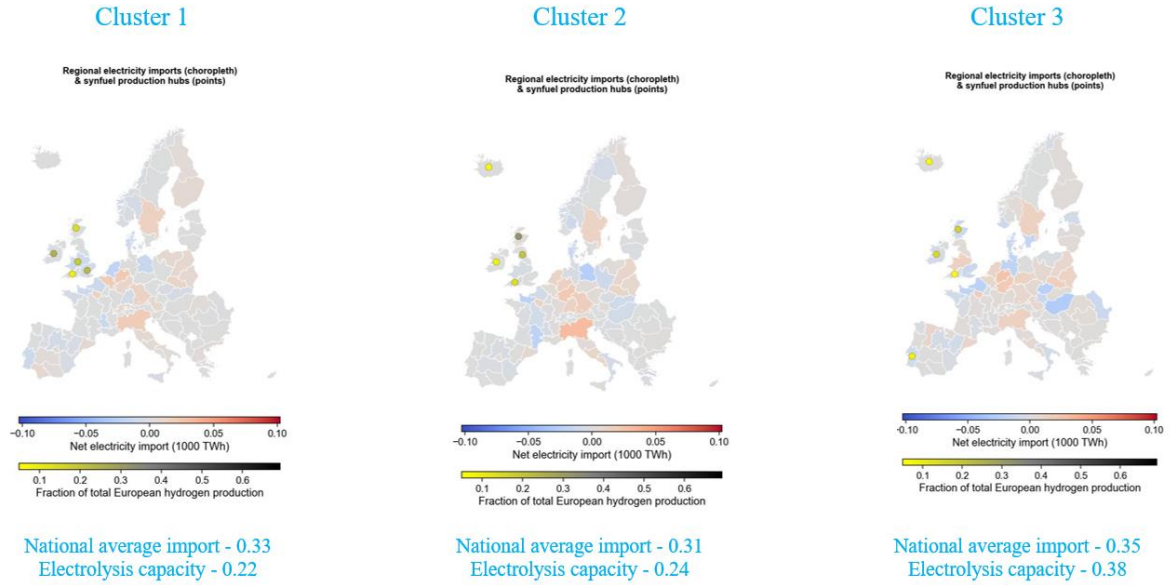


Figure 37 National average import and electrolysis capacity (Map)

## B. Electrification of transport and heating sector

In the case of electrification of the transport sector in cluster one, all three values, minimum, maximum, and representative solution, are 1. In cluster one, electricity has to be directly utilized as wind capacity in cluster one is relatively less compared to sectors two and three. This prompts the electrification of easily electrifiable sectors, one of which is the transport sector. The lower limit of electrification of the transport sector reduces from cluster one to cluster three, providing more options for the transportation sector and the possibility of synthetic fuel utilization.

Consecutively electrification of the heating sector follows a similar path from clusters one to three. In cluster one feasible percentage of electrification in the heating sector is from 63% to 85%. As wind capacity increases from cluster one to cluster three, the number of alternatives regarding the heating sector's electrification increase, giving more possibility to design an energy system.

## C. Electrolysis and total wind install capacity.

Total wind installed capacity increased from cluster one to cluster three, showing that increasing onshore wind capacity increases overall wind capacity in the system. Though the value of the representative solution increases from cluster to cluster three, the upper limit decreases from cluster one to cluster three and increases again in cluster three. This trend indicates that the low to medium wind capacity sector, which has high electrification possibility, will get electrified; however, after a certain limit, electrolysis is used as a balancing technology to store the excess wind energy.

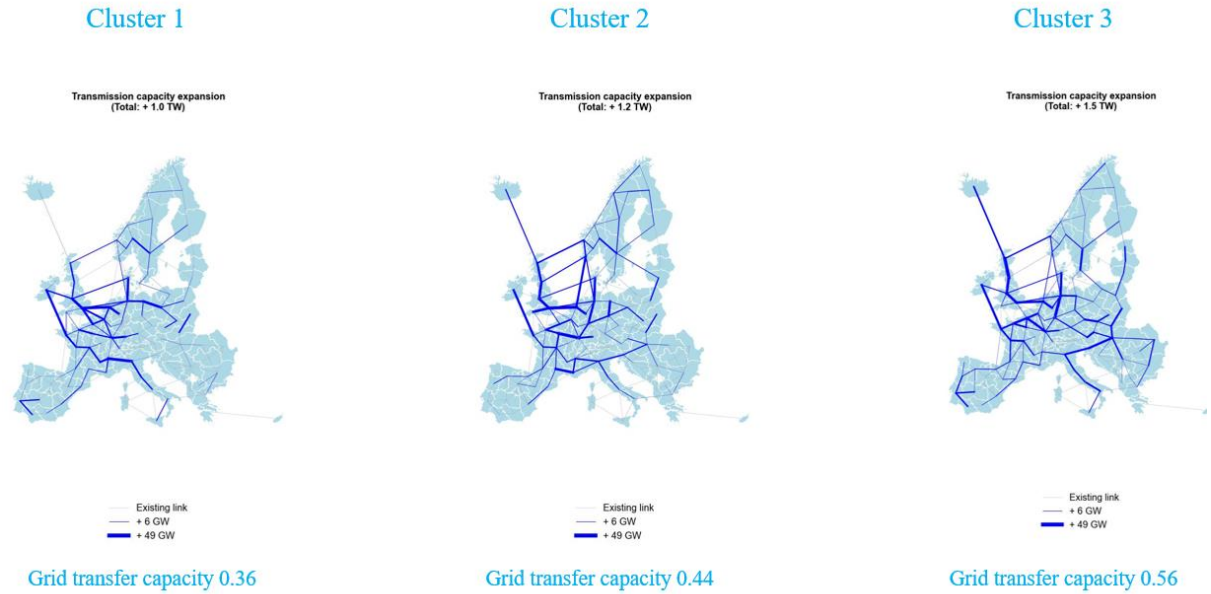


Figure 38 Grid transfer capacity (Map)

Onshore wind technology analysis shows that combining maps of technology distribution and grid expansion from original study (Pickering, Lombardi, & Pfenninger, 2022) coupled with representative solutions and cluster data helps analysts to understand the full picture of model results. However, only maps of representative solutions are not enough, as the available options for any cluster vary significantly. The cluster-wise analysis and analysing boundaries of solution space are the crucial step in understanding trade-offs in the different energy systems.

### Wind offshore technology

In case of wind offshore technology three is the suitable numbers of clusters. Figure 39 shows the clustered offshore wind technology data with representative solutions. The minimum offshore wind capacity in the system is 0.00 GW and maximum offshore wind capacity in the system is 1906.34 GW.

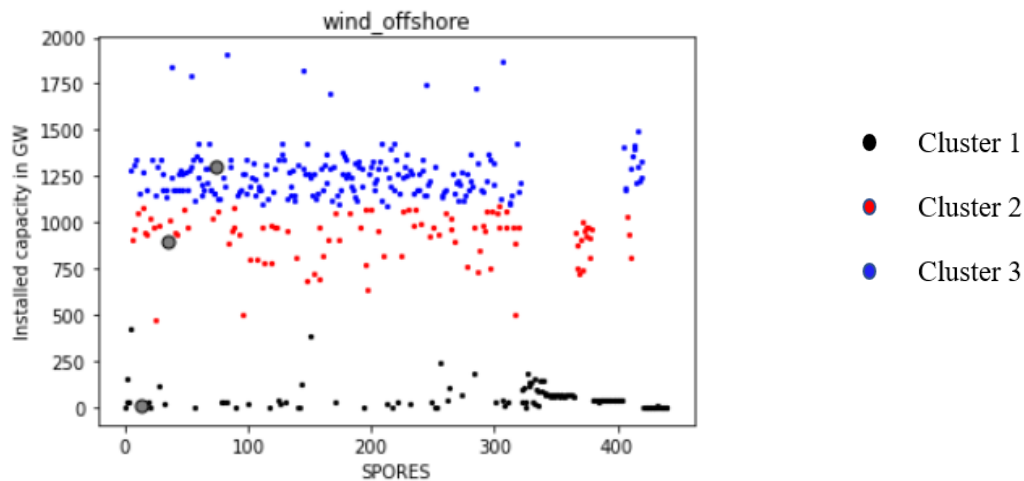


Figure 39 Clustered data of wind offshore technology

The first cluster contains 132 SPORES solutions, and the representative solution for the first cluster is a solution with SPORES id 13. The maximum and minimum offshore wind capacities are 421.12 GW and 0.00 GW. The installed capacity of the representative solution is 9.34 GW. The second cluster contains 98 SPORES solutions. The id of the representative solution is 35, with an installed capacity of wind offshore equal to 899.14 GW. The maximum and minimum installed capacity of wind offshore for the second cluster are 1085.93 GW and 470.60 GW, respectively. The third cluster contains 210 SPORES solutions. The id of the representative solution is 74, with the installed capacity of wind offshore equal to 1302.57 GW. This cluster's maximum and minimum installed capacity of offshore wind are 1906.34 GW and 1090.36 GW, respectively.

### Comparative analysis between representative solutions

Figure 40 shows the representative solutions value for all the clusters and all technology specific parameters.

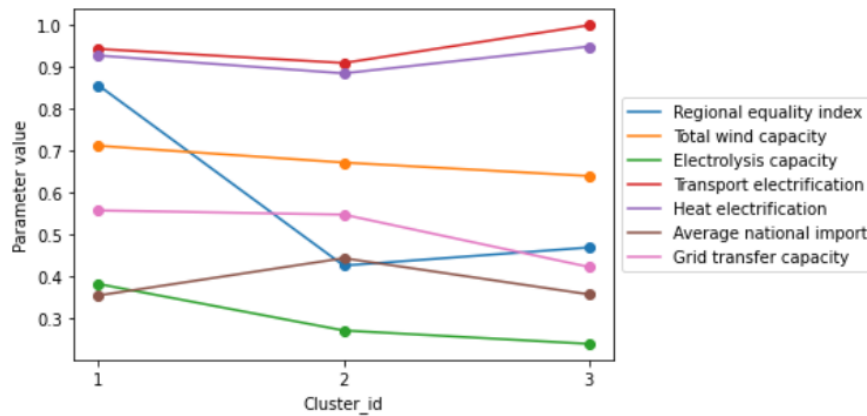


Figure 40 Parameter values of representative solution (wind offshore)

The value of the regional equality index drop significantly drops from cluster one to cluster two, indicating that technology deployment is more uniform between the nodes at low offshore installed capacity. Furthermore, the value of the regional equality index increases from cluster two to cluster three but not significantly.

However, total wind installed capacity stays nearly the same in all three clusters, indicating that an increase in wind offshore capacity reduces offshore wind capacity for the representative solution. Furthermore, the installed capacity of electrolysis reduces from cluster one to cluster two and stays the same from cluster two to cluster three. The given trend indicates that increasing offshore wind capacity pushes the system to produce less synthetic fuel. The graph shows the graph, which increases the electrification of the heating and transport sector.

## Cluster wise analysis

Figure 41 shows the graphs for technology-specific parameters with maximum, minimum and representative solution value.

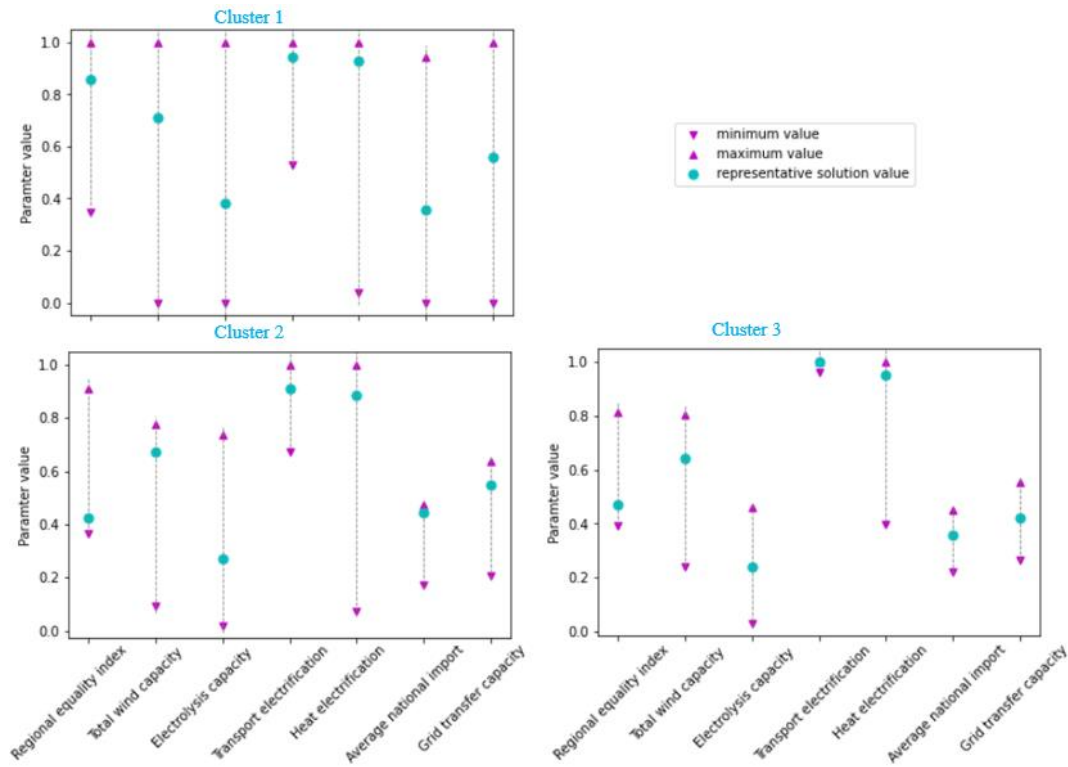


Figure 41 Technology specific parameter (comparative analysis wind offshore)

### A. Electrolysis and total wind install capacity.

Total wind installed capacity stayed nearly the same in all three clusters; however, the upper cap is reduced from cluster one to cluster two and stays the same from cluster two to cluster three. The given trend indicates that the increase in offshore wind capacity system prefers less total wind capacity. The claim made in section 6.1 that electrolysis is used as a balancing mechanism is further reinforced by the decrease in the upper limit of electrolysis capacity significantly from cluster one to cluster three.

### B. Regional equality index, Grid transfer capacity and average national import

The upper limit of the regional equality index sharply decreased from cluster one to cluster two. However, the decrease from cluster two to cluster three is insignificant, showing that there are no options for uniform distribution of offshore wind infrastructure at high offshore wind capacity. Furthermore, the need for grid transfer capacity and average national is reduced from clusters one to three. The given trend indicates that offshore wind installation does not impact grid expansion capacity, mainly due to offshore wind being the dominant energy source in the system.

### C. Electrification of transport and heating sector

In the case of offshore wind technology, electrification of the transport and heating sector shows a reverse trend that of offshore wind technology. With increasing wind, offshore capacity electrification of transport increases to the point in cluster three percentage of electrification becomes 100. At the same time, limits for electrification of the heating sector reduce from 100 percent to 60 percent. Further evidence is that offshore wind is the dominant energy between wind technologies.

#### 6.2.2 Solar energy generation analysis

As mentioned in section 3.2.2, solar energy has two types of wind technology in the model. First is open-field solar technology, and the second is roof top solar technology.

##### Solar open field PV technology

In case of solar open field technology four is the suitable numbers of clusters. Figure 42 shows the clustered solar open field technology data with representative solutions. Minimum open field PV capacity in the system is 0.00 GW and maximum open field PV capacity in the system is 4784.76 GW.

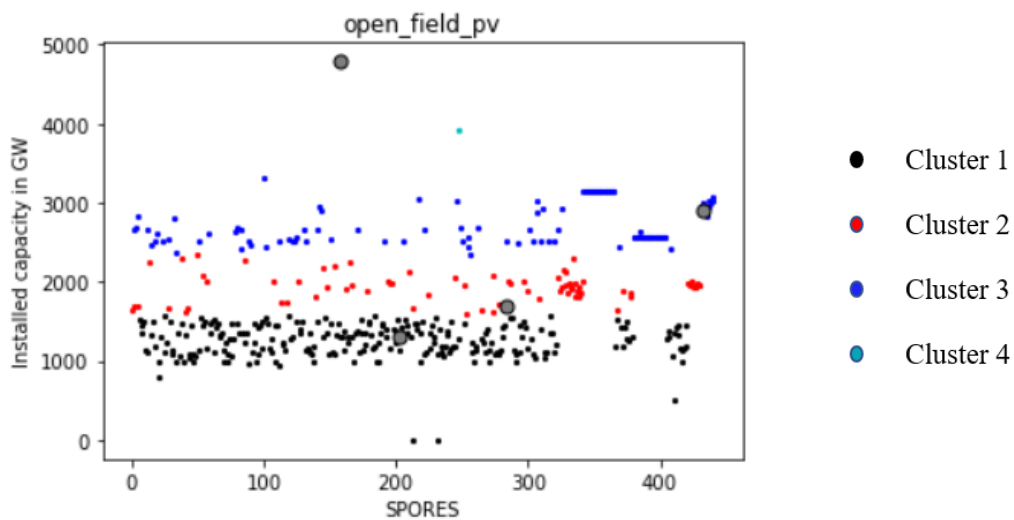


Figure 42 Clustered data of open field solar PV technology

The first cluster contains 246 SPORES solutions, and the representative solution for the first cluster is a solution with SPORES id 203. The maximum and minimum open field solar PV capacities are 1581.39 GW and 0.00 GW. The installed capacity of the representative solution is 1299.33 GW. The second cluster contains 75 SPORES solutions. The id of the representative solution is 283, with an installed capacity of open field solar PV equal to 1692.01 GW. The maximum and minimum installed capacity of open field solar PV for the second cluster are 2340.39 GW and 1590.66 GW, respectively. The third cluster contains 117 SPORES solutions. The id of the representative solution is 433, with the installed capacity of open field solar PV equal to 2901.17 GW. This cluster's maximum and minimum installed



capacity of open field solar PV are 3316.76 GW and 2345.34 GW, respectively. The fourth cluster has only two SPORES solutions with installed capacities of open field solar PV 4784.76GW and 3908.80 GW. For the purpose analysis cluster four solution will be consider as outliers and studied separately. The ids of two outlier SPORES are 158 and 258.

### Comparative analysis between representative solutions

Figure 43 shows the value of the representative solution for all the clusters and all technology-specific parameters. Total solar and battery storage capacity plots follow the same trend, indicating that in the model, battery storage is the primary storage technology for storing solar energy. Furthermore, as open-field PV capacity increases, solar capacity also increases, indicating no substitutional effect in this case.

The regional equality index increases with an increase in open field solar PV capacity indicating that open field solar PV installation gets distributed with increased capacity. For two outlier solutions, the regional equality index lies between the regional equality index of the representative solution for cluster 2 and cluster 3. Average national import is least in the case of cluster 3, indicating that solar installation nodes become more energy independent with high open field installation capacity. Along with the regional equality index graph and average national import, graphs mirror each other, indicating the strong interdependence of the system on open-field solar PV.

Heat electrification parameters decrease with an increase in solar capacity. At the same time, the transport electrification parameter decreases from cluster one to cluster two and then increases from cluster two to cluster three. Maximum and minimum values for each cluster have to be analyzed to understand the correlation between electrification and open field solar PV installation.

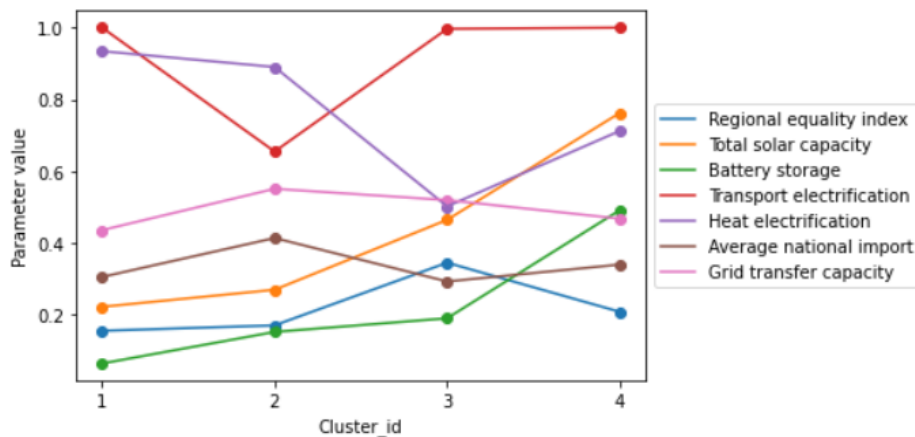


Figure 43 Parameter values of representative solution (open field solar PV)

### Cluster wise analysis

Figure 44 shows the graphs for technology-specific parameters with maximum, minimum and representative solution value.



### A. Battery storage capacity and total solar install capacity.

In the case of total solar capacity, upper and lower limits increase from clusters one to three. The lower limit of battery storage remains the same for all three clusters; however, the upper limits increase and decrease from cluster to cluster three. The reduction in an upper limit of storage capacity for cluster three indicates that with high solar capacity system prefers to not to store energy in the battery. In the case of two outlier solutions, total solar capacity and storage capacity increases even more.

### B. Electrification of transport and heating sector

As the capacity of open-field PV increases option for electrification of the transport sector reduces. Battery storage and electric transport has same underlying input assumption. At high amount of open field PV, system tends to electrify as much as transport sector available. Unlike wind technology where excess of energy is converted into hydrogen which can be easily converted into synthetic fuel. In the case of electrification of the heating sector, more option becomes available as open-field PV capacity increases in the system, with more electricity in the system, which can be further converted into synthetic fuels via electrolysis.

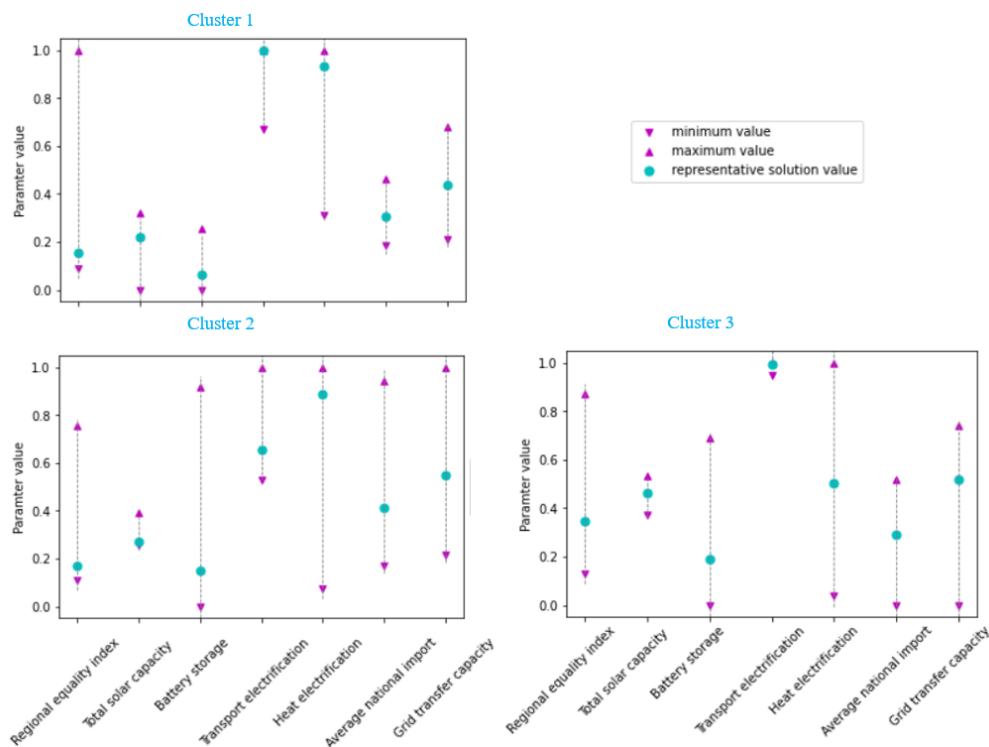


Figure 44 Technology specific parameter (comparative analysis open field solar PV)

### C. Regional equality index, Grid transfer capacity and average national import

As with wind offshore and onshore, increase in installed capacity leads to the concertation of development. In case of open filed PV, the range of regional equality index reduces sharply from

cluster one to cluster two stays the same for cluster three. As the capacity of open field PV increases, the option available for average national import increases. However, there is less need for energy import and grid transfer capacity for high open field solar PV installation than for medium open-field PV capacity.

### Solar roof-mounted PV technology

In case of solar roof mounted technology two is the suitable numbers of clusters. Figure 45 shows the clustered solar roof-top PV technology data with representative solutions. Minimum roof-top solar in the system is 0.00 GW and maximum roof-top solar capacity in the system is 700.85 GW.

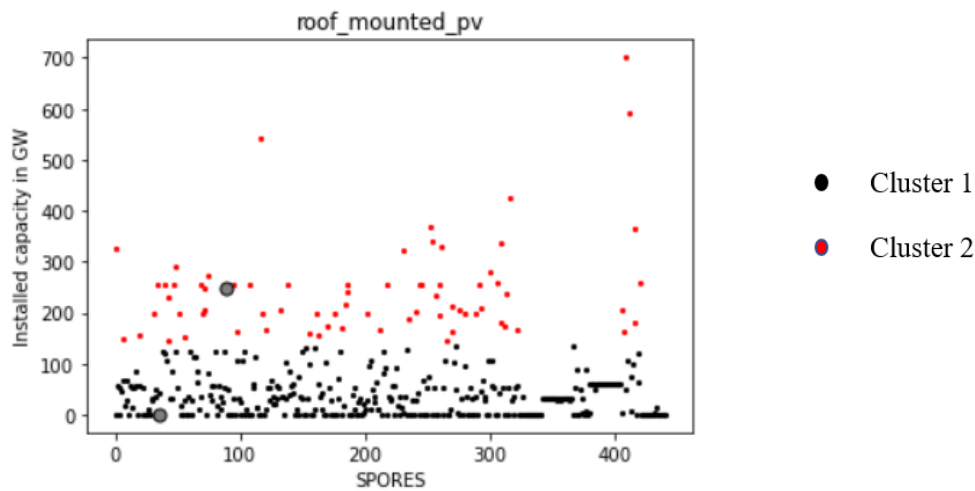


Figure 45 Clustered data of roof mounted solar PV technology

The first cluster contains 368 SPORES solutions, and the representative solution for the first cluster is a solution with SPORES id 35. The maximum and minimum roof-top solar PV capacities are 136.01 GW and 0.00 GW. The installed capacity of the representative solution is 0.14 GW. The second cluster contains 72 SPORES solutions. The id of the representative solution is 89, with an installed capacity of roof-top solar PV equal to 249.59 GW. The maximum and minimum installed capacity of roof-top solar PV for the second cluster are 700.85 GW and 143.30 GW, respectively.

### Comparative analysis between representative solutions

Figure 46 shows the representative solutions value for all the clusters and all technology specific parameters.

The regional equality index reduces sharply from cluster one to cluster two. The regional equality index of the representative solution for cluster one is 0.8 and for cluster 2 is 0.09 showing a high degree of concentration with high rooftop PV capacity. Heat electrification and transport electrification increases slightly from cluster one to cluster two. While every other parameter, namely

total solar capacity, battery storage, average national import and grid transfer capacity, stayed almost the same.

The given trend shows that the system heavily avoids installing rooftop PV due to its limitation over open-field PV; as a result system is not getting changed to a degree with changes in rooftop PV.

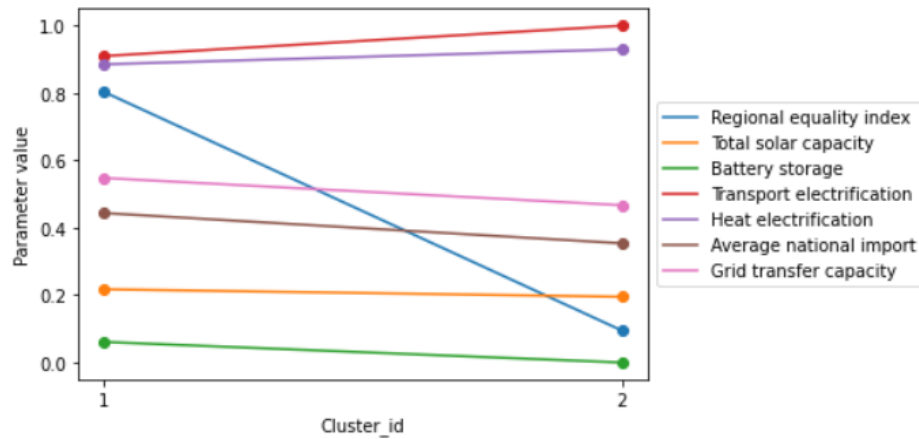


Figure 46 Parameter values of representative solution (roof mounted solar PV)

## Cluster wise analysis

Figure 47 shows the graphs for technology-specific parameters with maximum, minimum and representative solution value.

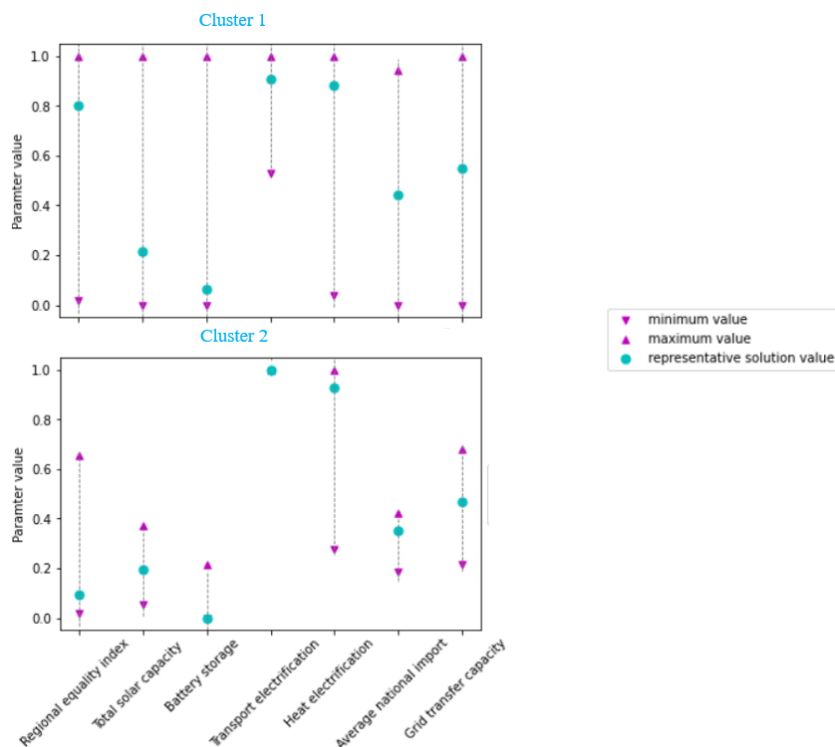


Figure 47 Technology specific parameter (comparative analysis roof mounted solar PV)

All parameters have the entire range of options available in case of low rooftop PV installation. In the case of transport electrification, the graph shows that the transport sector has to be 50 percent electrified. However, it is the entire solution space available for all the solutions.

Cluster two has significantly fewer options available in regard to technology-specific parameters. The main reason for this is the significant difference between the number of SOPRE solutions in each cluster. As mentioned in section 6.2.2 number of SOPRE solutions in cluster one is 368, and in cluster two is 72. In this case, the MGA analysis package starts to show its limitation. MGA analysis package has to divide the data into at least two clusters. With some technologies, only one cluster may exist due to the similarity of available options. However, this will be missed by MGA analysis package as iteration to determine suitable number of clusters starts with two.

### 6.3 Biofuel energy generation analysis

In case of biofuel energy technology three is the suitable numbers of clusters. Figure 48 shows the clustered biofuel technology data with representative solutions. Minimum biofuel energy capacity in the system is 0.00 GW and maximum biofuel energy technology capacity in the system is 4784.76 GW.

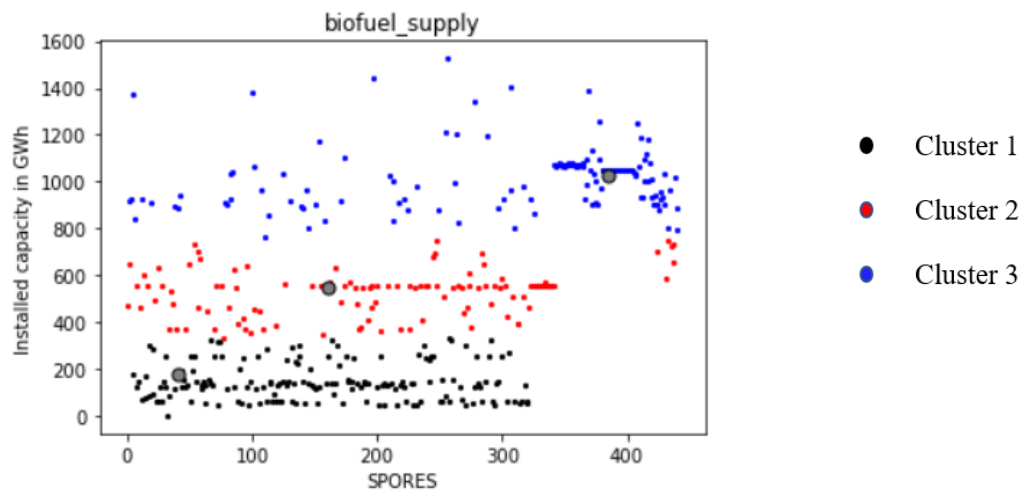


Figure 48 Clustered data of biofuel supply technology

The first cluster contains 172 SPORES solutions, and the representative solution for the first cluster is a solution with SPORES id 40. The maximum and minimum biofuel supplies are 331.94 GWhr and 0.00 GWhr. The biofuel supply of the representative solution is 176.37 GWhr. The second cluster contains 121 SPORES solutions. The id of the representative solution is 160, with biofuel supply equals to 551.18 GWhr. The maximum and minimum biofuel supplies for the second cluster are 751.16 GWhr and 332.20 GWhr, respectively. The third cluster contains 147 SPORES solutions. The id of the representative solution is 384, with the biofuel supply equals to 1022.30 GWhr. This cluster's maximum and minimum biofuel capacities are 1525.73 GWhr and 761.10 GWhr, respectively.

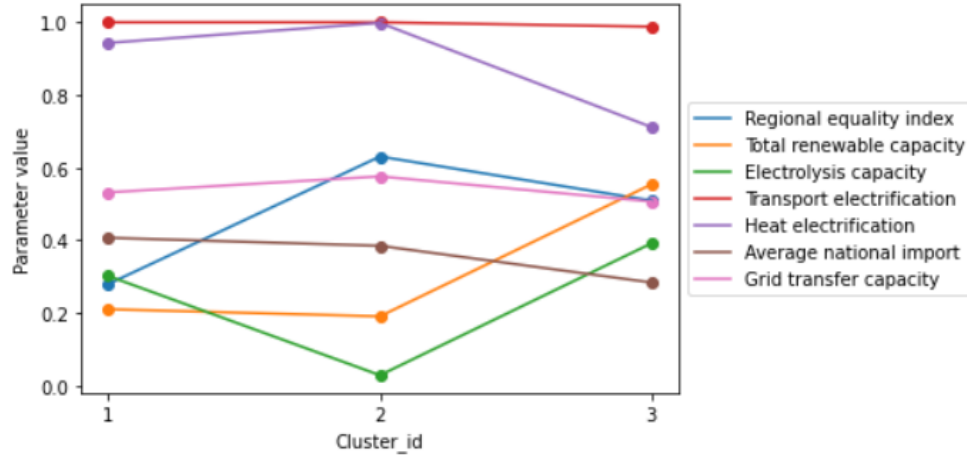


Figure49 Parameter values of representative solution (biofuel supply technology)

Figure 49 shows the value of the representative solution for all the clusters and all technology-specific parameters.

The regional equality index increases significantly from cluster one to cluster two, and decreases slightly in cluster three, indicating that biofuel production diversifies over the region at higher biofuel usage. An increase in biofuel supply does not significantly impact the grid transfer capacity; however, average national import decreases from cluster one to cluster three. An increase in the regional quality index is one reason for reducing average national imports at higher biofuel supply.

The heat electrification parameter sharply decreases from cluster two to cluster three, indicating that with high biofuel capacity need for the electrification of the heating sector decreases. An increase in biofuel is used mostly in the heating sector rather than the transport sector. As a result, degree of electrification of the transport sector stays the same.

Total renewable and electrolysis capacity increases with an increase in biofuel supply from medium to high. The given trend is counterintuitive, as biofuel is an alternative energy source to other renewable energy sources. Analysis of boundaries has to be performed to understand this trend.

### Cluster wise analysis

Figure 50 shows the graphs for technology-specific parameters with maximum, minimum and representative solution value.

#### A. Regional equality index, Grid transfer capacity and average national import

The available range of the regional equality index decreases from cluster one to cluster two, indicating that the distribution of biofuel supply is concentrated in some nodes. The available range of the Grid transfer capacity and average national import increases from cluster to cluster two and

three. Even though the available range of grid transfer capacity and average national import is the same for clusters two and three, the upper and lower limits are higher in the case of cluster two. The given trend indicates that medium biofuel supply nodes depend on each other for energy.

### B. Electrification of transport and heating sector

Available option for electrification of transport and heating sector increases from cluster one to cluster three. The trend shows that at higher biofuel supply system becomes more robust in regards to the electrification of transport and heating sector.

### C. Total renewable capacity and electrolysis capacity

Available option installation renewable energy installation capacities for clusters one and two are around 70 % of the available option. Whereas in the case of the third cluster, entire possibilities of options are available for the installation of renewable energy. Electrolysis capacity follows the trend of total renewable capacity, indicating that the system becomes more robust with a high biofuel supply.

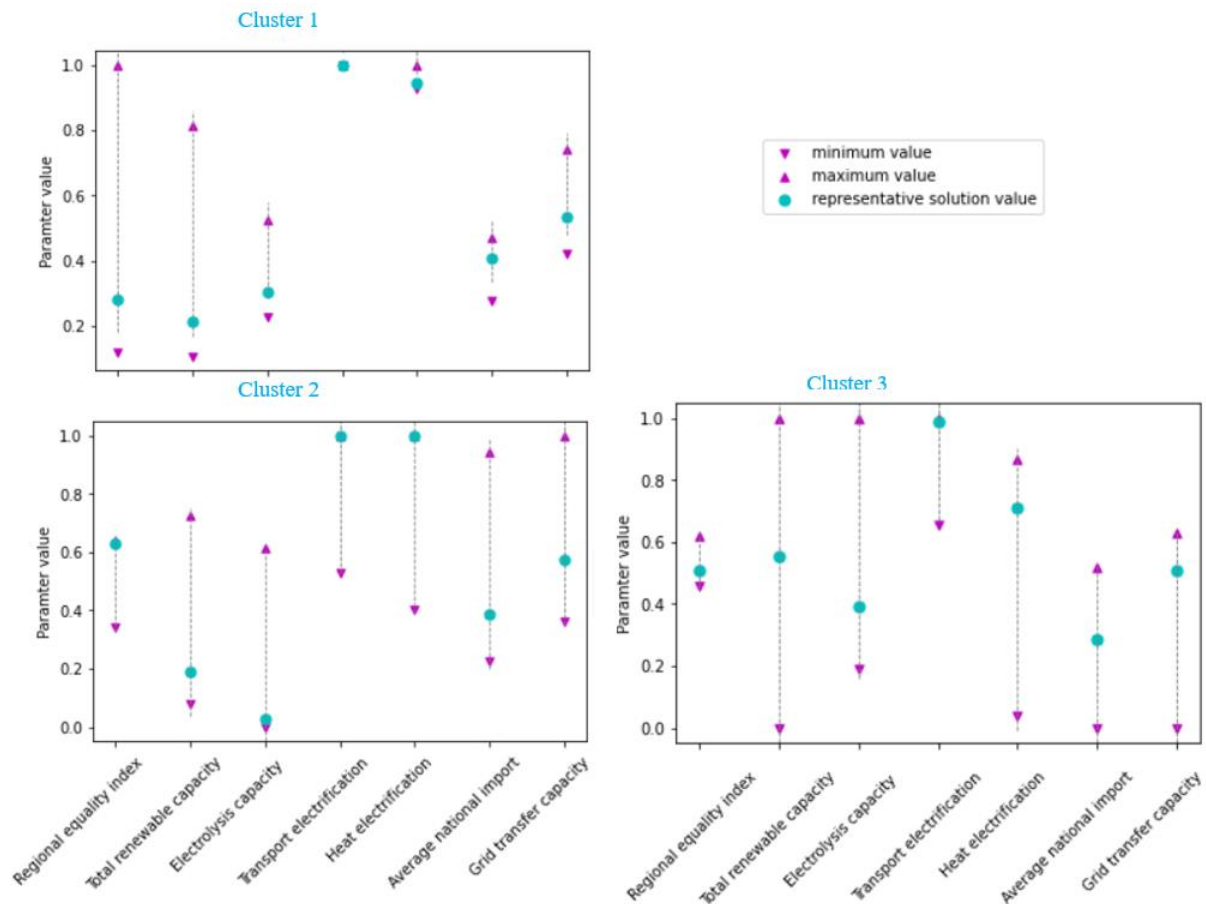


Figure 50 Technology specific parameter (comparative analysis biofuel supply technology)

This section aimed to demonstrate the application of the MGA analysis package. In this section, we discussed and analysed the trade-offs between generation technologies of CO2 neutral Europe using data and graphs generated by the MGA analysis package. This case study shows that the MGA analysis

package facilitates analysis of the MGA based energy system model results by reducing the amount of data needed to analyse. Providing graphs in tandem with maps gives analysts a quick overview of the solution space. However, maps are not part of the MGA analysis package. Maps for open-field solar, roof-mounted solar, offshore wind, and biofuel technology are presented in the Appendix A. The dashboard provides maximum and minimum values of technology-specific parameters, which allows the analysts to understand the variety of options available for a given cluster. Which in turn leads to better planning of infrastructure.

The MGA analysis package requires three input files in Excel format to run and can be used in Windows, Mac, and Linux operating systems by installing Python version 3.9 or higher, which makes MGA analysis package versatile. An interactive dashboard allows the analysts to analyse the results with 4-5 user input, making the MGA analysis package faster and more efficient.

Mainly there are three limitations of the package. The first one is the identification of key technologies and technology-specific parameters, which carry huge weightage in the analysis process. Identifying those parameters requires a deep understanding of the energy system models and real-world energy systems. The analyst has to be precise while defining the research question and its boundaries. The second limitation of the MGA analysis package is that the MGA analysis package cannot generate less than two clusters. As shown in the case of roof-mounted PV, there should be only one cluster in case of roof mounted PV. However, there are two clusters due to the limitations of the package. The third limitation of the MGA analysis package is that it gives an analyst overview of the results. While implementing those energy system configurations in the real world, in-depth analysis has to be performed. Hence analysts should use the MGA analysis package as a preliminary analysis tool.

### 6.3 Comparing MGA analysis package with other MGA analysis studies

Section 2.5 explains the design of the matrix, which is designed to compare the MGA analysis package with other prominent research studies published in the literature. Figure 51 shows the filled matrix after the application.

The given matrix compares MGA analysis package based on eight characteristics- standardized method, automated method, groups the solution space, gives the representative solution, shows boundaries the of solution space, shows spatial diversity, interactive dashboard, shows the interdependence of technologies. The significance of each characteristic is explained in section 2.5.

The results of Pickering et al. are used as a test case in this thesis. In that research, study authors analyze their results using maps of solutions based on scenarios- namely low storage, high PV, lowest storage, lowest biofuel lowest curtailment. The given procedure is standardized, gives representative solutions, and shows the interdependence of technologies by comparing installed capacity. Maps give an idea about spatial diversity. However, the method is not automated and does not group the solution space.

The paper's authors created an interactive dashboard to showcase their results, but the dashboard is not used to analyze the solution

Research papers Features	Pickering et al	Neumann et al	DeCarol is et al	Price et al	Sasse et al	MGA analysis package
Automated method						
Groups the solution space						
Gives representative solution						
Shows boundaries of solution space						
Shows spatial diversity						
Interactive dashboard						
Shows interdependency of technologies						

Figure 51 Filled comparison matrix for comparing MGA analysis package

The analysis method presented by Neuman et al. shows second the prominent way of analyzing multiple near-optimal energy system Configuration, which is discussed in section 2.4.2. In this study, authors compare boundaries of solution space along different cost parameters. This method is a standardized method which is easy to replicate and compare technological variation in the system. Along with that shows boundaries of solution space and provides maps to give an idea about spatial diversity. However, the process is not automated and does not provide an interactive dashboard to analyze the results. Along with that process also does not group the solution space and gives a representative solution.

DeCarol is et al. study published on the topic of alternatives option for the future of the US electricity system has only 28 solutions and a single node. The authors analyze their results based on different CO2 scenarios. The analysis procedure is standardized and easy to replicate. Along with that study also show the interdependency of technologies and gives an idea about extremes. However, the method is not automated, does not group the solution, and does not show the representative solution. The characteristics of spatial diversity are not applicable to this study as a single node-based study.

Price et al. published a global-level MGA energy system study that has 16 nodes in it. The prominent analysis procedure in this study was comparing fractional variability relative to optimal. The given procedure is standardized and gives an idea about the boundaries of solution space by means of



comparing fractional variability relative to optimal. Along that study shows the interdependency of technologies. However, the given procedure does not group the solution space, does not give a representative solution and does not show spatial diversity. Along with that method is automated and does not have an interactive dashboard to analyze the results.

Sasse et al. published a study on Switzerland's future renewable electricity generation portfolio. In the study, the authors focus on regional equality of distribution of renewable energy infrastructure. The formula used to calculate the regional equality described in section 5.2 is taken from the study. For the analysis, the authors created and compared the scenarios based on regional equality. The analysis procedure was standardized and easy to replicate, showing a representative solution. Along with that, the analysis also showed the spatial diversity and boundaries of solution space by scenario analysis. Maps are used to convey the spatial diversity of different technologies and the interrelation of different technologies. However, the analysis procedure is not automated and does not have an interactive dashboard to analyze the results.

The MGA analysis package developed in this thesis has all the comparison matrix characteristics, making the method more suitable for analysis. The strongest points of the MGA analysis package are the automated method, interactive dashboard and grouping of the solution, which is unavailable in any other method. All the analysis methods are standardized and show the impact of technologies on each other. The design of the MGA analysis package method is built to implement the strongest points of each analysis method in the literature study, which makes the MGA analysis package the best method to analyze the results of MGA-based optimization energy models.

## Chapter 7: Conclusion

In the preceding chapters, we discussed the design of the MGA analysis package and its application to the results of the sector Euro calliope model. This chapter answers the research questions in section 7.1. Section 7.2 discusses the choices made during each step of the research and reflect upon them. Section 7.3 reflects on the thesis in general by discussing personal learning and recommendation for future work.

### 7.1 Answers to the research questions

This thesis aimed to design a computer-based clustering method to facilitate analysis of MGA-based energy system optimization models. Six sub-research questions were designed to guide the process of achieving the aim of the thesis. First research question is literature-based question used to generate knowledge necessary for the thesis. Second and third research questions guided the design choices for MGA analysis method, while fourth and fifth questions guided the key design choices of MGA analysis package and interactive dashboard respectively. Aim of the sixth question was to guide the demonstration of MGA analysis package.

*SQ1: What are prominent methods to analyse the results of near optimal energy systems configuration?*

In the literature, there are three prominent ways of analysing near optimal energy system configuration. The first one is by selecting a few solutions based on the model expert's expertise. The second is by analyzing solutions at the boundaries (for example: based maximum and minimum installed capacity of certain technology), and the last is by using computer-based algorithms to analyze the solutions. Each way has its advantages and disadvantages. The conclusion of the literature analysis of prominent ways of analyzing near-optimal energy system configuration is that no standardized computer-based method is available to facilitate the analysis of MGA-based energy optimization models.

Key features of the package are decided by combing features of current methods of analysis of near-optimal energy system configuration and discussion with the MGA-based energy model expert. Seven key features, namely: automated method, groups solutions, gives representative solutions, shows boundaries of solution space, shows spatial diversity, interactive dashboard, shows interdependency of technologies. Key features acts as guidelines for design process.

*SQ2: Which algorithm to use to cluster the results MGA-based energy system model?*

The choice of clustering algorithm was heavily influenced by the algorithm choices of finding a representative solution. We decided to use mean as a bases for identification of representative solution. Combining mean with variance of data will give analysts a better idea of the cluster data. Hence, we

decide to use variance-based clustering algorithms cluster the solutions. Therefore, clustering has to be done so that variance within the cluster would be minimal, but the variance between clusters would be maximum. Jenks's natural break (JNB) (Ahmad, 2019) and K-means clustering (Sharma, 2023) work on the principle of minimizing variance inside the cluster and maximizing variance between the clusters. The key difference is that K-means clustering is a two-dimensional clustering algorithm, whereas JNB is a one-dimensional clustering algorithm. In the case of the K-means algorithm, users have to provide centre points around which cluster forms; however, in the case of JNB, users do not have to provide centre points. If the user has to provide a centre point, finding a representative solution step becomes irrelevant, as the user will dictate the centre point. Hence the decision to use JNB for the first step is taken.

*SQ3: How to identify a representative solution for the clusters created by the clustering process?*

We decided to use Measures of central tendencies to find the representative solution as those parameters give an aggregate idea of a data set. The three most common measures of central tendency are the mean, median and mode. Mode measures the repetition of data, and median shows the data's midpoint. Mean shows the average value of data. Combining mean with variance (a measure of dispersion which shows the spread of data) and boundary value of data will give analysts a better idea of the cluster. Hence the decision to use the mean as the main component of finding a representative solution is made. However, the energy system is a complex system, with hundreds of technologies working together to balance the system. We decided to include data of multiple technologies in this system to find a representative solution. This way, the representative solution will be for the system, not for a single technology.

*SQ4: How to build a computer-based package based on the designed MGA analysis method?*

The MGA analysis method is coded using Python programming language to make the MGA analysis package. Python is used as the base environment as it contains many libraries for data analysis like sci-kit Learn, Pandas, Matplotlib and NumPy. The MGA analysis package takes input in the form of three Excel files and has two modes of operation. Auto mode is used to analysis of the results where the analyst does not input the number of clusters. In manual mode, the analyst has to enter the number of clusters. Manual mode is used for the development of the MGA analysis package. The MGA analysis package automates the MGA analysis method and increases the accessibility of the method.

*SQ5: Which parameters to display on the interactive dashboard to facilitate the analysis of MGA-based energy system model results?*

An interactive dashboard is created to facilitate the use of MGA analysis package where analyst can use MGA analysis package with 4-5 inputs. Dashboard shows clustered data in the graph as well as in text format. Along with that dashboard provides boundaries of each cluster. Furthermore, analyst can

analyse each cluster in depth by comparing technology specific parameters for representative population and maximum and minimum value of technology specific parameter for that cluster.

*SQ6: What are the impacts of change in energy generation installed capacities on the CO<sub>2</sub>-neutral European energy generation and balancing systems?*

The sixth research question was designed to guide the MGA analysis package's application on the sector-coupled euro calliope model results. The first step was the determination of input parameters. The primary technologies are installed capacity of wind onshore, wind offshore, roof-mounted solar PV, open-field solar PV, and biofuel energy. Along with them, energy from waste, electrolysis capacity, battery capacity, hydrogen storage, and big and small heat storage are considered key system technologies for the research question. The technology-specific parameter matrix was designed for each primary technology.

At higher installed capacity for wind onshore, open field solar PV, and energy from biofuel, the available range for the secondary parameter such as regional equality index, electrification of heating and transport sector increases. However, in the case of higher installed capacity of roof-mounted solar PV and offshore wind, the available range for the secondary parameter such as regional equality index, electrification of heating and transport sector decreases. In the system, wind offshore and roof-mounted PV are more expensive and restrictive than their counterparts. This is one of the reasons for the observed phenomenon. However, the aim of this question was to demonstrate the application of MGA analysis package by providing systematic analysis which is demonstrated in chapter 6.

*MRQ “How to build a computer-based package to analyse the results of MGA based energy system models?”*

The output of this thesis is a Python package developed to analyze the results of MGA based energy system optimization model. As mentioned above, the MGA analysis package is a three-step process that includes more data in every succession step. MGA analysis package groups the solution space, finds representative solutions, and identifies a suitable number of clusters. The package has python based interactive dashboard, which can be used in the Jupiter notebook environment. The dashboard provides visuals of clustered data, representative solutions and boundaries of each cluster. Along with that dashboard also provide visuals of technology-specific parameter for the representative solution as well as the boundaries of the cluster.

All the information combined in visual and text format allows an analyst to grasp the MGA-based optimization model results faster. The MGA analysis package is compared against the method in the prominent MGA analysis study in literature. The MGA analysis method comes ahead of every other method. The package is hosted on the Git-hub website are free to download. By default, the results of the sector couple euro calliope model are uploaded in the model. However, users can enter their data in

the package to analyze their MGA-based energy optimization results. All these features allow the analysis of MGA based energy optimization model using the MGA analysis package.

## 7.2 Discussion of research steps of thesis

This section discusses choices made during each step of the thesis. We took a step back and analyzed all the work we have done during the development period of this thesis. This section reflects on the research approach of the thesis, design choices made during the design of the MGA analysis method, and features and limitations of the MGA analysis method as well as the MGA analysis package. Finally, we reflect on the personal and academic growth of the author during the thesis period.

### Discussion on the design of the method

After completing the literature analysis, we decided to group the solution space and identify a representative solution for each group are two crucial steps for the MGA analysis method. These two steps will allow user to analyze the system with least of data. For the first step, we decided to use machine learning clustering algorithms to form the groups of solution space. Two options were available for clustering the solution space- one-dimensional clustering and two-dimensional clustering. Three different clustering algorithms were tested: Jenks natural break, K-means clustering, and affinity progression. Jenks's natural break is a one-dimensional clustering algorithm, whereas K-means clustering, and affinity progression are two-dimensional clustering algorithms. Due to time constraints and the simplicity of the method one-dimensional clustering algorithm was selected as the de facto algorithm for the clustering process. As the chosen algorithm is one dimensional, data of single has to be input for clustering process.

The second step was to find the representative solution for each cluster. The first idea was that the mean of data of chosen technology would represent the cluster's solutions. However, the energy system is a complex system which includes hundreds of technologies working together to balance the system. We decided to include data from other technologies in this system to find a representative solution. This way, the representative solution will be for the system, not for a single technology.

However, if data from multiple technologies are not normalized, then the impact of the technology with the highest installed capacity will be the largest on determining a representative solution. Hence to find a representative solution, normalized data of technology has been chosen.

The next step was the determination of a suitable number cluster. Jenks's natural break algorithm has a second algorithm attached to it, which finds a suitable cluster number. However, that algorithm uses data from chosen technology to determine the suitable number of clusters. A special algorithm was needed for the complex system, such as an energy system. In this we developed algorithms to find the suitable number of clusters. Input for this algorithm was parameters which impact the chosen technology most. Large variations in those parameters will change the dynamics of the energy system.

The number of suitable clusters is determined based on the variance of those technology-specific parameters. This step completed the design process of the method.

As mentioned in the introduction of the thesis, the MGA-based energy system configuration varies in the spatial development of energy infrastructure. The statistical term Gini index was used to convey regional equality in deploying energy infrastructure—regional equality index for added to technology-specific parameter matrix.

One of the limitations of the MGA analysis method is identifying key technologies and technology-specific parameters, which carry huge weight in the analysis process. Identifying those parameters requires a deep understanding of the energy system models and real-world energy systems. The analyst has to be precise while defining the research question and its boundaries.

### Discussion on the design of the package

After completing the design of the MGA analysis method, the next step was to automate the process. Automation of processes leads to an increase in the usage of the method. For this purpose, we decided to use Python as a coding language. The main reason for using Python is the accessibility of Python language and free to use nature. All the design steps are coded in separate Python functions for easy access and a clean look.

Finally, an interactive dashboard was created by all the calling Python functions created in the last step in the Jupiter notebook. The dashboard displays clusters, representative solutions and technology-specific parameters, and boundaries of solution space graphically and in text. The user has to choose the technology to analyze. Other than that, the entire process is automated. In the trial version, when the dashboard was not the interactive package was used to display all the results for all the technology. The interactive dashboard gave the user command over the analysis process.

The package has two modes of operation, auto and manual. In auto mode, the user has to choose only technology; however, in manual mode, the user can also determine the number of clusters. Manual mode is created to explore the solution space at its own command and learn something new. The default mode of the MGA analysis package is auto.

By default, the MGA analysis package has the sector-coupled euro calliope model results, but users can insert their results in an Excel sheet to analyze them. Creating this feature took the most time during the building process MGA analysis package. Though the initial aim was to design a way to analyze the sector coupled euro calliope model results, the project's ambitions grew after identifying the need for an MGA analysis package after the literature review. The package is hosted on the git-hub website with detailed usage instructions. The MGA analysis package cannot generate less than two clusters. As shown in the case of roof-mounted PV, there should be only one cluster in case of roof mounted PV. However, there are two clusters due to the limitations of the package.

MGA analysis package should be used as a tool to get an initial understanding of the results of MGA based energy system optimization model. MGA analysis package gives modelers insight into their know model assumption and a chance to reflect on the model's design. Analysts should use this method as a lantern in the dark to guide further analysis.

The MGA package is uploaded on the git-hub (<https://edu.nl/gmarp>) for application and further development.

### Discussion on the case study

To address the case study question, we decided to use the sector-coupled euro calliope model's results for two main reasons. The first reason – sector coupled euro calliope model is Europe's most extensive sector couple energy system model. The second is the availability of model experts. One of the thesis committee members is a model expert of sector coupled euro calliope model. We need a model expert with all the knowledge about the model to run the MGA analysis package. The role of a model expert is to guide the data selection process for the analysis based on stakeholders' requests.

The same thesis committee member acts as a stakeholder and model expert for the case study. With the guidance of that thesis committee member, we identified the case study's primary and key technologies. We also designed a technology-specific parameter matrix for the primary technology.

The purpose of the case study was to show the application of the MGA analysis package, how to use the package, and what the package can do for you. Hence how to interpret the results of the MGA analysis package become secondary. The analysis provided in the case study is not extensive due to time limitations and constraints of the aim of the thesis. Further research is possible in that regard which is discussed in the advice for further research section.

### Discussion on matrix of comparison for MGA analysis method

The comparison matrix was designed to showcase the capabilities of the MGA analysis method. Eight key characteristics which the package and method need to fulfil were identified with the help of a model expert. Most of the characteristics were part of one or more previous case studies. However, combining all those characteristics was a challenge. A comparison matrix is able to deliver the message that we were able to stand to challenge and overcome it.

Five case studies which were used for comparison were chosen from a total of 11 case studies of application MGA analysis of future energy systems. The showcase of the application of the MGA analysis package was on the results of the sector couple euro calliope model, which has 98 nodes and 441 solutions. Hence Chosen case studies are multi-nodal and have more near-optimal solutions than non-chosen ones. This allowed us to compare the analysis methods more fairly.

### 7.3 Critical reflection on the thesis

This section provides a critical reflection on the thesis by analysing the extent to which the initial hypothesis of the thesis still holds. Then, we reflect on the personal and academic growth of the author during the thesis period. This section concludes with recommendations for further research.

This thesis hypothesises that clustering the solutions and identifying a representative solution for each cluster is a possible method to analyse the results of MGA-based energy system models. Clustering or grouping solutions into sub-groups was the only way to reduce the amount of data to be analysed while keeping the impact of data. We used representative solutions and maximum/minimum values to analyse each cluster. Statistical parameters such as mean, max, min and variance combine to give a holistic idea of the group. In the retrospective MGA analysis package, we combine all these parameters in one, which is the best way to analyse the equally possible near-optimal energy system confirmation. Hence initial hypothesis of this thesis holds.

#### 7.2.1 Personal Learning

My affinity for designing integrated energy systems and a little bit of experience in energy modelling led me to choose my current thesis topic. But little did I know what I was getting into. With no machine learning experience and only basic knowledge of coding, the learning curve during the initial period of the thesis was steep. My first challenge was designing the thesis process and setting up the boundaries for the thesis. Understanding when to stop during research is a crucial thing I learn. This thesis is a multi-layer study, each requiring knowledge and expertise of its own. Apart from technical skills such as coding and data analysis, I gained expertise in web thinking, structured thinking and critical thinking as an energy system is a very complex system to design and balance. The MGA analysis package and this report are a product of eight months of full-time work. Conversion of research work into a report was one of my biggest challenges. The goal while writing the report was to write in such a way that non-expert persons could also understand and follow this research.

This thesis marks the culmination of my journey in the master's program in sustainable energy technologies (SET). Patience, perseverance and strong will are the qualities I acquired during this thesis and master's program. The SET study program is about designing sustainable energy systems for our future and includes aspects from designing new technologies to integrating those technologies into society. My master thesis gave me an opportunity to use all my course knowledge and create something which will be useful energy modelling community in the future. In doing so, I developed my expertise in designing integrated energy systems and energy modelling.

#### 7.2.2 Advice for further research

From this research limitations and insights, we suggest that the following research undertakings would be beneficial.



Mainly there are three limitations of the package. The first one is the identification of input parameters, which carry huge weight in the analysis process. The second limitation of the MGA analysis package is that the package cannot generate less than two clusters. Hence in some cases, the package is forced to generate two clusters even though there should be only one cluster. The third limitation of the MGA analysis package is that it gives an analyst overview of the results.

- The input parameters for the case study are decided based on the discussion with model experts and knowledge. A separate auxiliary process can be developed to identify/verify selected input parameters.
- Currently, the MGA analysis package is tested on the single MGA energy system analysis study results. A comparative study of the output of the application of the MGA analysis package on different MGA analysis studies will provide examples of developed auxiliary processes.
- Identifying a suitable number of algorithms has to modify to differentiate between one and two clusters. The code for Identifying a suitable number of algorithms has all data to differentiate between one and two clusters; however, it cannot process data and cannot differentiate between one and two clusters.
- Due to the constraint of the thesis, the interpretation of the results of the MGA analysis package took a secondary seat. For further research, a study can be developed to analyze the results of sector-coupled euro calliope results using multiple applications research questions.
- MGA analysis package requires the user to have Jupiter notebook and Python 3.0 version installed on their system. For the future development of the package, the package can be hosted on a web page. The web page can be designed so that users can add their results in Excel format and get the analysis on the same web page.

## Bibliography

- Ahmad, R. (2019, April 4). *Jenks Natural Breaks — The Best Range Finder algorithm*. Retrieved from Medium: <https://medium.com/analytics-vidhya/jenks-natural-breaks-best-range-finder-algorithm-8d1907192051>
- Banos, R., Manzano-Agugliaro, F., Montoya, F. G., Gil, C., Alcayde, A., & Gómez, J. (2010). Optimization methods applied to renewable and sustainable energy: A review. *Renewable and Sustainable Energy Reviews*, 1753-1766.
- Bryant, B. P., & Lempert, R. J. (2009). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting & Social Change*, 34-49.
- Church, R., & Baez, C. (2020). Generating optimal and near-optimal solutions to facility location problems. *Environment and Planning B: Urban Analytics and City Science* 47(6), 1014-1030.
- DeCarolis. (2011). Using modeling to generate alternatives (MGA) to expand our thinking on energy futures. *Energy Economics* 33(2), 145-152.
- DeCarolis, J. F., Babaei, S., Kanungo, S., & Li, B. (2015). Modelling to generate alternatives with an energy system. *Environmental Modelling & Software*, 300-310.
- Delft, T. (2022, 06 13). *TU Delft researchers develop model for green, energy independent Europe*. Retrieved from tudelft.nl: <https://www.tudelft.nl/en/2022/tu-delft/tu-delft-researchers-develop-model-for-green-energy-independent-europe>
- Gaschnig, H., Süsner, D., Ceglaz, A., Stavrakas, V., Vassilis, J., & Flamos, A. (2021). Better suited or just more complex? On the fit between user needs and modeller-driven improvements of energy system models. *Energy* 239, 32.
- Huntington, H., Weyant, J., & Sweeney, J. (1982). Modeling for Insights, not Numbers: the the Experiences of the Energy Modeling Forum. *Omega*, 10:449-62.
- Kotzur, L., Nolting, L., Hoffmann, M., Groß, T., Smolenko, A., Priesmann, J., . . . Robinius, M. (2021). A modeler's guide to handle complexity in energy systems optimization. *Advances in Applied Energy* 4, 1-19.

- Krumm, A., Süsser, D., & Blechinger, P. (2021). Modelling social aspects of the energy transition: What is the current representation of social factors in energy models? *Energy* 239, 17.
- Lombardi, F., & Pfenninger, S. (2023, 02 14). *Calliope: a multi-scale energy systems modelling framework (V0.6..10)*. Retrieved from [calliope.readthedocs.io](https://calliope.readthedocs.io/en/stable/): <https://calliope.readthedocs.io/en/stable/>
- Lombardi, F., Pickering, B., & Pfenninger, S. (2023). What is redundant and what is not? Computational trade-offs in modelling to generate alternatives for energy infrastructure deployment. *Applied Energy* 339, 121002.
- Neumann, F., & Brown, T. (2021). The near-optimal feasible space of a renewable power system model. *Electric Power Systems Research* 190.
- Ommen, T., Markussen, W. B., & Elmegaard, B. (2014). Comparison of linear, mixed integer and non-linear programming Comparison of linear, mixed integer and non-linear programming. *Energy Elsevier*, 109-118.
- Pfenninger, S., & Lombardi, F. (2022, June 13). *Carbon-neutral European energy system*. Retrieved from Explore Calliope : <https://explore.calliope/?spore-id::>
- Pfenninger, S., Hawkes, A., & Keirstead, J. (2014). Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews* , 74-86.
- Pickering, B. (2022, May 13). *Diversity of options to eliminate fossil fuels and reach carbon-neutrality across the entire European energy system*. Retrieved from Zenodo: <https://zenodo.org/record/6546817#.Y3zuRDjMKUk>
- Pickering, B., Lombardi, F., & Pfenninger, S. (2022). Diversity of options to eliminate fossil fuels and reach carbon neutrality across the entire European energy system. *Joule-cell press open access*, 1253-1276.
- Price, J., & Keppo, I. (2017). Modelling to generate alternatives: A technique to explore uncertainty in energy-environment-economy models. *Applied Energy*, 356-369.
- Sasse, J.-P., & Trutnevyte, E. (2019). Distributional trade-offs between regionally equitable and cost-efficient allocation of renewable electricity generation. *Applied Energy* 254, 113724.

- Sasse, J.-P., & Trutnevyte, E. (2020). Regional impacts of electricity system transition in Central Europe until 2035. *Nature communication*, 1-15.
- Sharma, N. (2023, April 19). *K-Means Clustering Explained*. Retrieved from MLOps Blog: <https://neptune.ai/blog/k-means-clustering>
- Singlitico, A., Østergaard, J., & Chatzivasileiadis, S. (2021). Onshore, offshore or in-turbine electrolysis? Techno-economic overview of alternative integration designs for green hydrogen production into Offshore Wind Power Hubs. *Renewable and Sustainable Energy Transition* 1, 100005.
- Stawska, A., Romero, N., Weerdt, M., & Verzijlbergh, R. (2020). Demand response: For congestion management or for grid balancing? *Energy Policy* 148, 111920.

# Appendix

## Appendix A Maps of representative solutions

Appendix A shows maps related to the regional equality index, National average import and electrolysis capacity and grid transfer capacity for open-filed solar, roof-mounted solar, offshore wind, and biofuel technology.

### 1. Wind offshore technology

#### a. Regional equivalent index

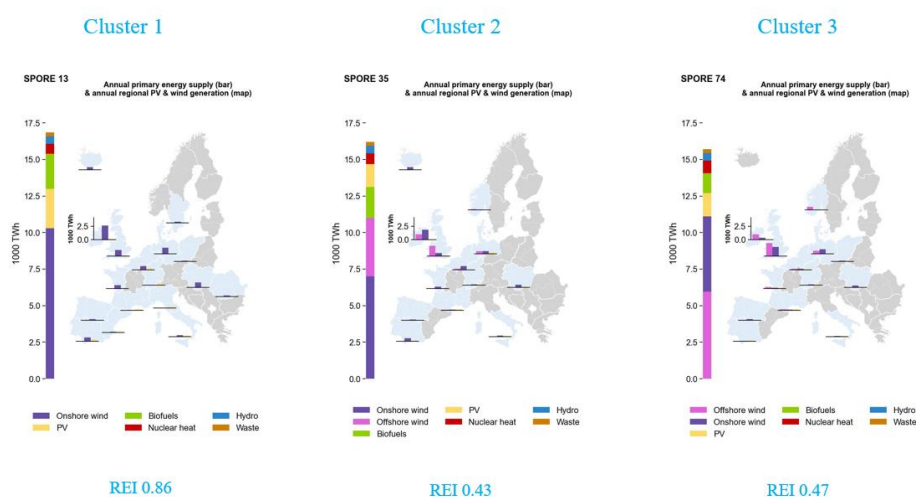


Figure 52 Regional equality index (Map)

#### b. National average import and electrolysis capacity (Map)

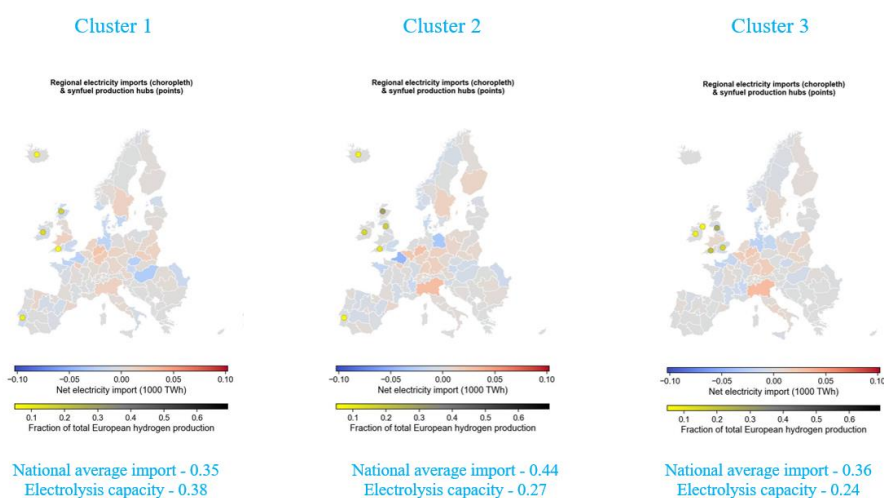


Figure 53 National average import and electrolysis capacity (Map)

### c. Grid transfer capacity (Map)



Figure 54 Grid transfer capacity (Map)

## 2. Open field PV technology

### a. Regional equivalent index

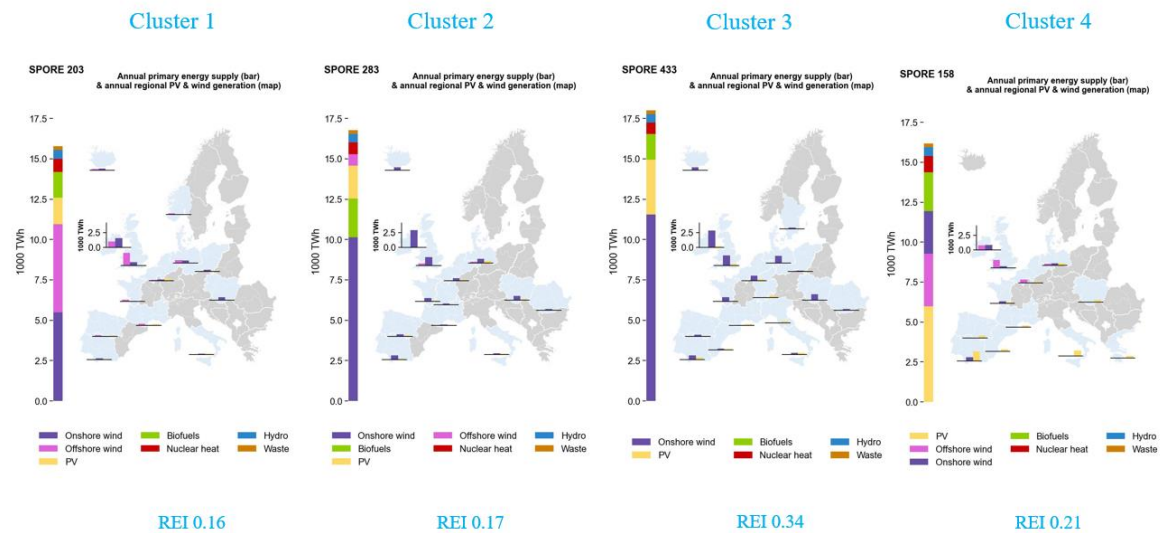


Figure 55 Regional equality index (Map)

b. National average import and electrolysis capacity (Map)

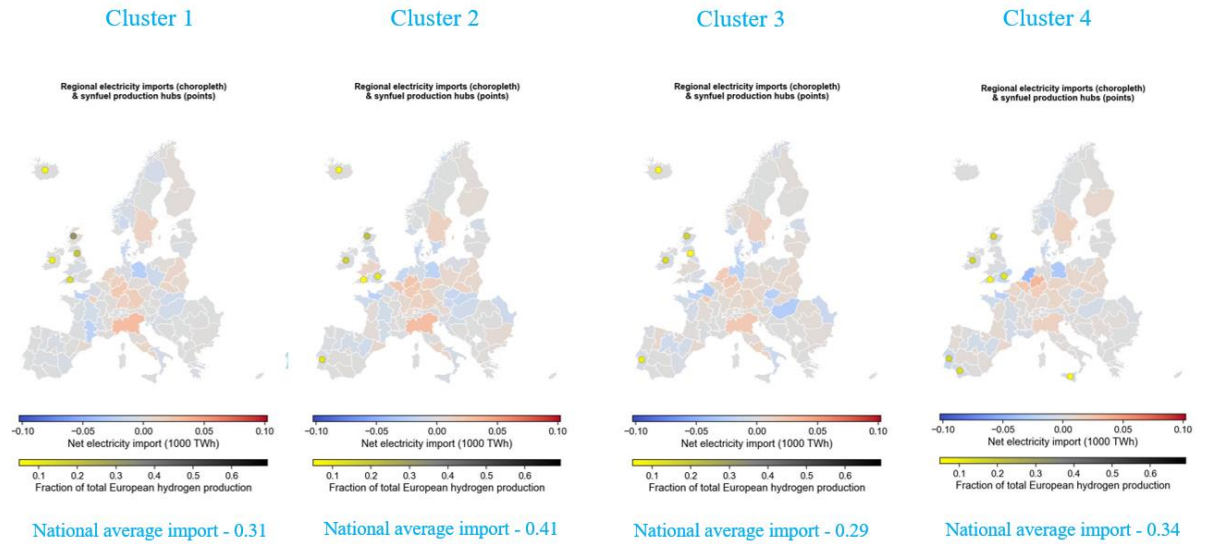


Figure 56 National average import and electrolysis capacity (Map)

c. Grid transfer capacity (Map)

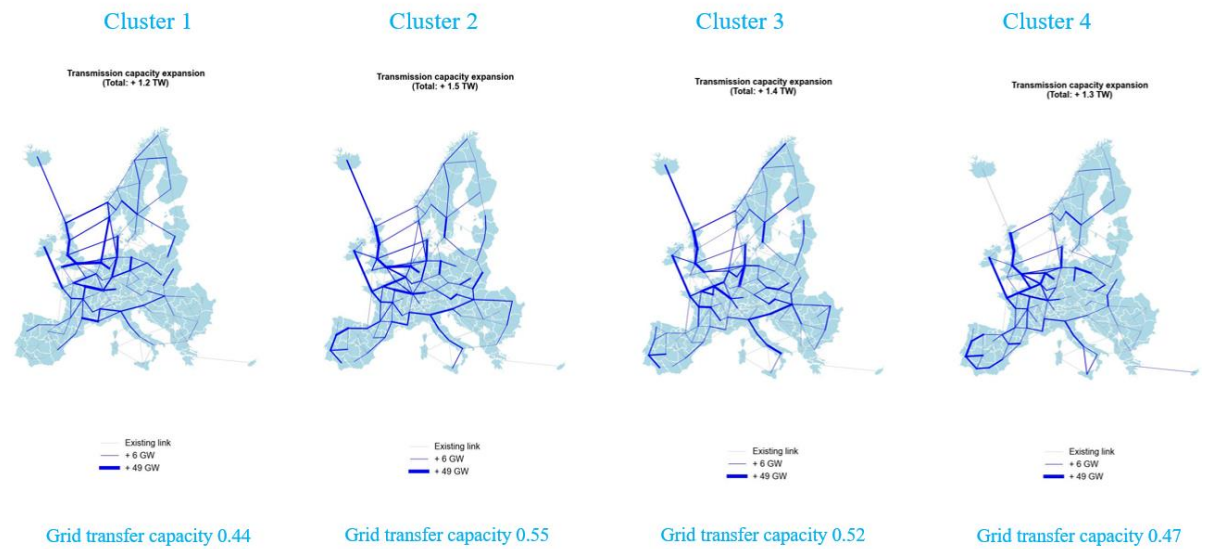


Figure 57 Grid transfer capacity (Map)

### 3. Roof mounted PV technology

#### a. Regional equivalent index

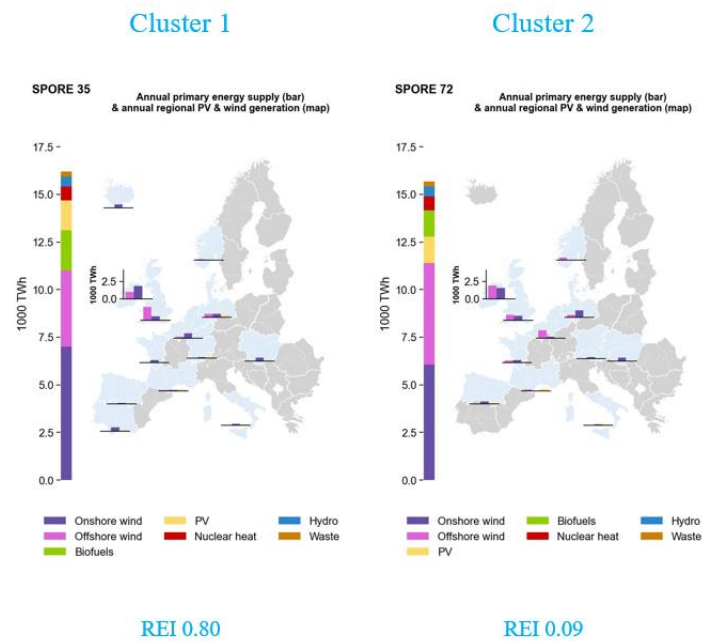


Figure 58 Regional equality index (Map)

#### b. National average import and electrolysis capacity (Map)

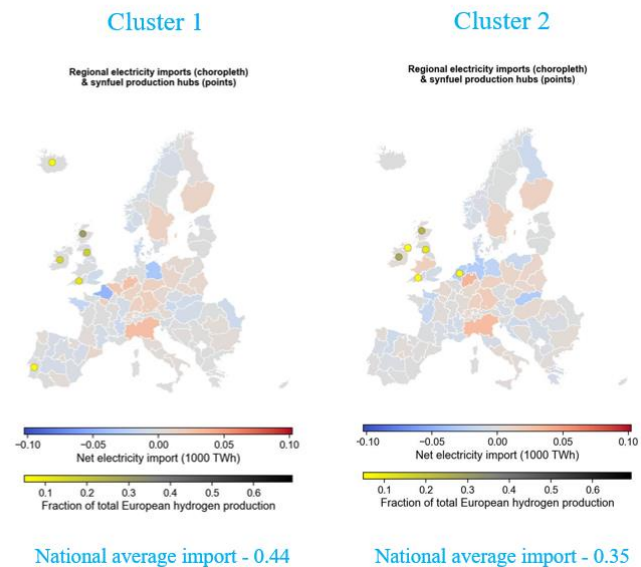


Figure 59 National average import and electrolysis capacity (Map)



### c. Grid transfer capacity (Map)

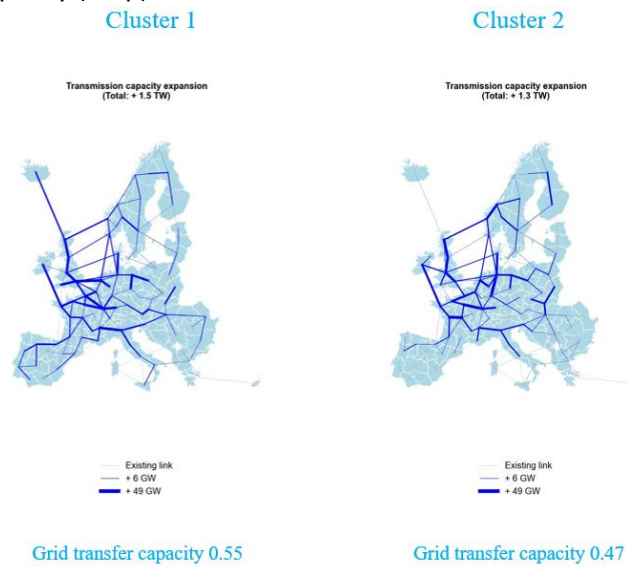


Figure 60 Grid transfer capacity (Map)

## 4. Roof mounted PV technology

### a. Regional equivalent index

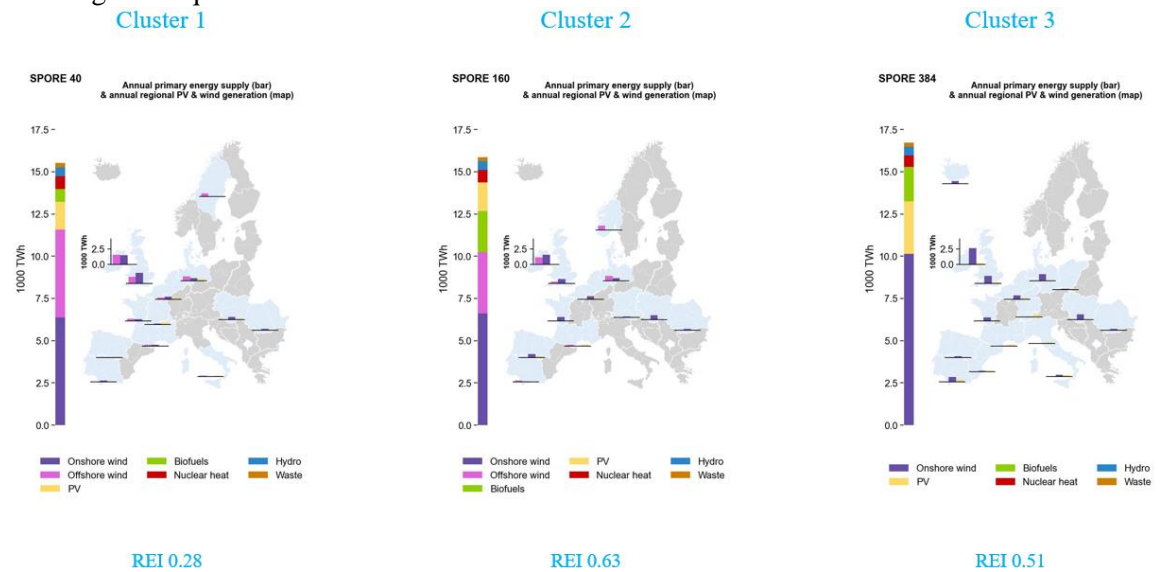


Figure 61 Reginal equality index (Map)

b. National average import and electrolysis capacity (Map)

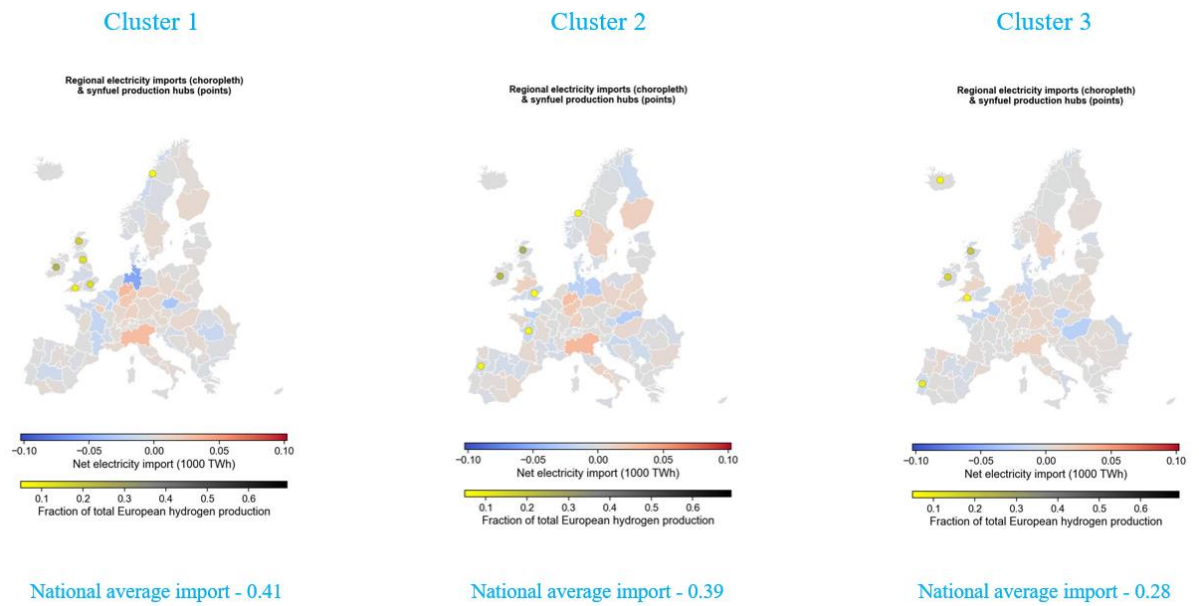


Figure 62 National average import and electrolysis capacity (Map)

c. Grid transfer capacity (Map)

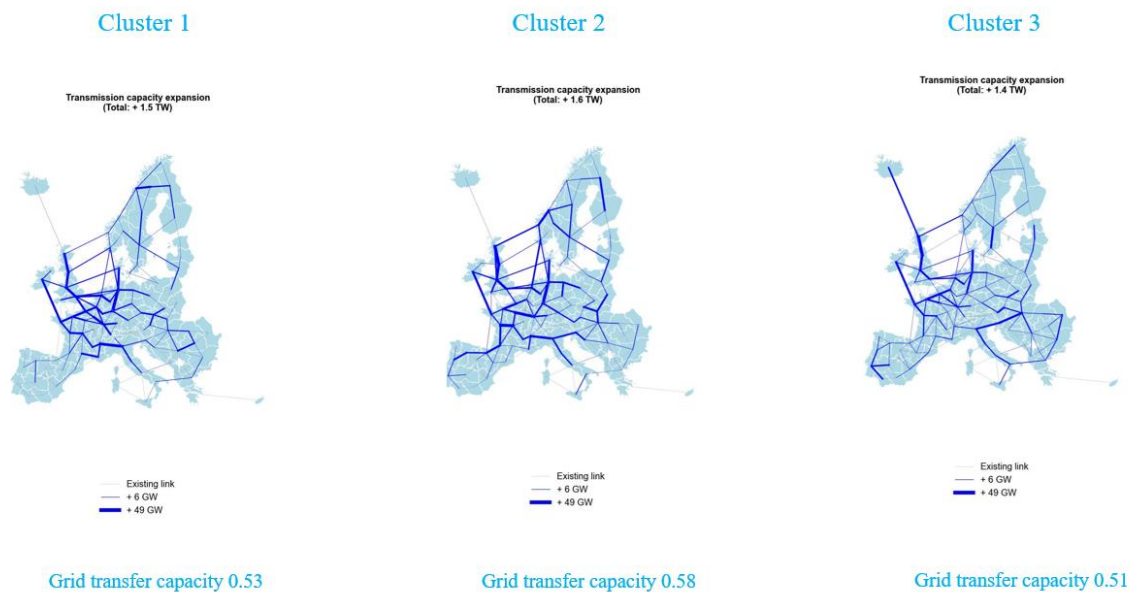


Figure 63 Grid transfer capacity (Map)