DELFT UNIVERSITY OF TECHNOLOGY

---

# Optimizing hydrogen supply transport for Dutch regional industry by use of multiple transport options

---

Master thesis submitted to Delft University of Technology
for
Complex System Engineering and Management
Faculty of Technology, Policy and Management

***Author:***
Sander Oostdijk (4727452)

***Graduation committee:***
Chair and first supervisor: Dr. ir. Petra Heijnen
Second supervisor: Prof. dr. Martijn Warnier

October 2, 2025

Associated code for this thesis input files can be found at
https://github.com/SanderOostdijk/Thesis_CoSEM

# Preface

This Master Thesis marks the end of a very important phase of my life: studying in Delft. Over the years I have learned a lot from my lectures, fellow students, and friends at student association VGSD. Not only in an academic way, but also through collaboration, organising fun events, and developing great friendships. I have enjoyed this period very much, and I am grateful to be able to conclude it by submitting this report.

After obtaining my bachelor's degree, I felt the urge for a new challenge, which led me to start the Complex Systems Engineering & Management Master at TBM. Combining social and technical skills is something I truly value, and this Master provided me with the opportunity to develop both. I am glad that I can finish this Master by conducting a study that challenged me to think in a systematic and academic way, while also allowing me to puzzle with improving the Optimal Network Layout Tool. Moreover, conducting research on how to accelerate the energy transition feels particularly valuable in this period of time.

I would like to thank Petra Heijnen, my first supervisor, for her support and guidance along the way. Thank you for all your advice, which helped me to keep going every week. Every meeting left me inspired about how to proceed, and you were always available to discuss improvements to the model, which I greatly appreciated. I also want to thank Martijn Warnier for your clear feedback from another perspective.

Additionally, I want to thank all friends and family who have supported me throughout this Master Thesis journey. Being able to let go of my thoughts and frustrations has helped me to move forward and bring this project to a good conclusion. Special thanks go to Annely, who has been there for me every step of the way. Thank you for your patience, time, and support, and for always being there for me!

I hope you enjoy reading this thesis!

# Executive summary

Combatting climate change is one of the greatest challenges of our time. The Dutch industry is the largest national emitter of greenhouse gases, responsible for 27% of all emissions (Centraal Bureau voor de Statistiek, 2025), mainly due to its heavy reliance on natural gas. Green hydrogen, produced by electrolysis with renewable electricity, is seen as a promising alternative for high-temperature processes and transport. To enable large-scale use, a national hydrogen backbone is currently being developed by Hynetwork, a subsidiary of Gasunie, which reuses part of the existing gas grid (Hynetwork, n.d.). Challenges in the development of this network are mostly located in keeping the network-costs as low as possible, while still ensuring safety and sustainability in the process. This network connects the five largest industrial clusters, but leaves many regional industries, together referred to as Cluster 6, without direct access. These companies are willing to reduce their footprint, but face difficulties in connecting to green energy infrastructure. Their diversity in size, sector and energy demand further complicates the design of suitable supply solutions, as one single transport option does not fit all. The core problem of this thesis is therefore to develop a hydrogen supply infrastructure tailored to the heterogeneous character of Cluster 6.

Analysis of existing literature shows that hydrogen transport and supply networks have been studied extensively, but mostly on a national scale and with pipelines as the dominant mode. While this provides insights into large-scale cost-optimal design, it overlooks the smaller, more fragmented demand of regional industry and the potential role of trucks and vessels. Knowledge is therefore missing on how regional companies, with their variety of characteristics, can be supplied in a cost-effective way. This study addresses that gap by developing a method that analyses smaller-scale areas in more detail, while including multiple transport modes. Socially, the study provides insights into how regional industries can be integrated into the energy transition, thereby reducing their dependency on the national backbone and accelerating decarbonisation.

To address this problem, a method is developed to model a cost-optimal hydrogen transport network layout using the Optimal Network Layout Tool (ONLT) (Heijnen, 2025a). First, the regional industry is identified, and literature is analysed to determine which characteristics most strongly influence network costs. These characteristics are then collected for all unique industry locations. To structure this diverse dataset, two clustering techniques are applied: K-means clustering groups industries with similar demand profiles, allowing tailored supply solutions per type of company, while DBSCAN clustering reduces complexity by combining nearby industries, making the final optimisation computationally feasible. After clustering and selecting relevant transport modes, several research areas are analysed with the ONLT. The tool is adapted to handle multiple transport options, enabling evaluation of pipelines, vessels and trucks in combination. In this way, the approach not only models cost-optimal networks but also tests how different transport options can complement each other.

Results from this study highlight the relevance of using a variety of transport options, which are fitted to the characteristics of the regional industry. By using a combination of pipelines, waterways (with use of vessels) and roads (with use of trucks), regional industry can be supplied of green hydrogen in a cost-optimal way, thereby accelerating the energy transition.
The first research steps showed that the amount of demand and distance to supply networks are the main variables which influence the cost of hydrogen supply infrastructure. For this study, a total of 291 companies is identified. Gathering all data for every industry location in Cluster 6 showed a wide variety in expected hydrogen demand, but the largest part (9̃0%) has a demand between 0-50 TJ of hydrogen energy per year. Besides that, some outliers exist in the range between 50-500 TJ per year. The largest sectors with regard to number of companies are the food and building materials sector. However, when size of a sector is determined by hydrogen demand, the chemical and building materials sectors become largest. In total, 17 sectors are identified within regional industry, confirming the variety again.

The use of pipeline, road and waterways provides flexibility of transport options, by having options not only over land, but also over water. Vessels on waterways have large capacities, therefore functioning as second large transporter besides pipelines. Last, road transport is used as flexibility option for small-scale transport, due to its high density network. Calculations of distance between industries and existing transport networks confirmed this as well, where 97% of all industry locations are located within 10 kilometres of the highways, and an average distance of 3.14 kilometre. The pipeline backbone and waterways are relatively far away from industry locations on average (11.60 km for pipeline backbone, 8.53 km for waterways), where 97% of all industry is located within 35 kilometres for the pipeline backbone and within 30 kilometre for waterways.

After all characteristics have been established, categorizing of all industry locations could be executed. Using K-means clustering, a set of six different clusters has been constructed, which described the variety in regional industry best. These help to analyse specific areas and types of industries, without having to analyse every single industry location. Three different research areas have been chosen, based on the locations of industry clusters. Apart from K-means clustering, a DBSCAN clustering has been executed which clustered all industry locations within 2000 meters, to reduce complexity and computational time for the ONLT.

Three different research areas are analysed after that, by using a modified ONLT, where mainly the cost function has been updated to calculate and optimise costs for the three different transport options. Overall, results showed that waterways and pipelines are the most expensive transport option, but can also take care of large quantities of hydrogen transport. These two options are used mostly to transport large amounts of hydrogen over larger distances, while road transport is used for small-scale, 'last-mile' deliveries. This is due to the smaller capacities of trucks, which therefore are less suitable for the transport of large amounts of hydrogen. Interesting to see is that research areas consisting of industry locations with high demand and close to all transport networks, use waterways and pipelines for a large share of transport. This is needed because of the high demand, but also increases average costs/km for these areas. Research areas with companies with low demand, and far away from both waterways and pipelines use road transport a lot, thereby lowering the average costs/km. Combined with the large amount of companies with low demand, these results show the opportunities for road transport of hydrogen for regional industries. Still, a hydrogen backbone is needed through pipelines or waterways, but the final distances can be delivered by truck, even reducing costs compared to using pipelines or waterways.

# Contents

**List of Figures** 6

**List of Tables** 8

**1 Introduction** 10
- 1.1 Description of Cluster 6 . . . 11
- 1.2 Stakeholder analysis . . . 12
- 1.3 Transport options for hydrogen . . . 13
- 1.4 Link to CoSEM program . . . 13
- 1.5 AI statement . . . 14

**2 Theoretical background and research questions** 15
- 2.1 Theoretical background . . . 15
- 2.2 Research questions . . . 17

**3 Methodology & data requirements** 19
- 3.1 Determining relevant characteristics for hydrogen transport optimization (SQ-1) . . . 19
- 3.2 Gathering regional industry data (SQ-2) . . . 20
- 3.3 Clustering regional industry (SQ-3) . . . 22
- 3.4 Selection of research areas (SQ-4) . . . 23
- 3.5 Network optimization for all transport options (SQ-5) . . . 24
  - 3.5.1 Research Flow Diagram . . . 27

**4 Identifying regional industry** 28
- 4.1 Literature review . . . 28
- 4.2 Relevant transport networks . . . 29
- 4.3 Conclusion . . . 29

**5 Data collection for regional industry** 30
- 5.1 Identification industrial companies . . . 30
  - 5.1.1 Demand data . . . 31
  - 5.1.2 Distance to transport networks . . . 32
- 5.2 Conclusion . . . 34

**6 Categorisation of regional industry** 35
- 6.1 Clustering on distance . . . 35
- 6.2 Clustering on characteristics . . . 37
  - 6.2.1 Execution of K-means clustering . . . 37
- 6.3 Conclusion . . . 39

**7 Selection of research areas** 40
- 7.1 Analysing grid-sections . . . 40
- 7.2 Description of selected research areas . . . 41
  - 7.2.1 Research area 1 - T15/16 . . . 42
  - 7.2.2 Research area 2 - T17/18 . . . 42
  - 7.2.3 Research area 3 - T19/20 . . . 43

**8 Model setup** **44**
8.1 Adjustments to ONLT . . . 44
8.2 Implementation of new transport options . . . 44
8.3 Cost function . . . 44
8.3.1 Pipeline costs . . . 44
8.3.2 Road/waterway costs . . . 45
8.4 Model assumptions . . . 47

**9 Results** **49**
9.1 Research area 1 - Low demanding industries with large distances to supply networks . . 49
9.2 Research area 2 - High demanding industries close to supply networks . . . 51
9.3 Research area 3 - Medium demanding industries close to pipelines, far from waterways 52

**10 Discussion** **54**
10.1 Interpretation of results . . . 54
10.2 Limitations of this study . . . 55
10.3 Verification and validation . . . 58

**11 Conclusions and recommendations** **59**
11.1 Answering research questions . . . 59
11.1.1 What are relevant characteristics of regional industries with regard to transport optimisation? . . . 59
11.1.2 How can the relevant characteristics be identified and prepared for clustering? . . 59
11.1.3 What characteristics-based clusters of regional industry can be identified? . . . . . 60
11.1.4 What are suitable research areas, where a specific regional industry cluster is strongly represented? . . . 60
11.1.5 How are different transport options used in a cost-efficient network model for different research areas? . . . 61
11.1.6 How can hydrogen supply transport for Dutch regional industry be cost-optimized? 61
11.2 Scientific & social relevance . . . 62
11.3 Recommendations for future work . . . 62

**Bibliography** **64**

**A K-means clustering steps** **67**

**B Cost function** **73**

**C ONLT plots for every model step** **77**

# List of Figures

1.1 Cluster 6 industries . . . 11

2.1 Results of every subquestion . . . 18

3.1 Visualisation SQ-1 . . . 19
3.3 Visualisation SQ-2 . . . 20
3.4 Research Flow Diagram - SQ2 . . . 20
3.5 Visualisation SQ-3 . . . 22
3.6 Research Flow Diagram - SQ3 . . . 23
3.7 Visualisation SQ-4 . . . 23
3.8 Research Flow Diagram - SQ4 . . . 24
3.9 Visualisation SQ-5 . . . 24
3.10 Graph theory examples . . . 25
3.11 Examples of ONLT steps . . . 25
3.12 Research Flow Diagram - SQ5 . . . 26
3.13 Research Flow Diagram . . . 27

5.1 Companies per sector . . . 30
5.2 Energy demand for hydrogen . . . 31
5.3 Demand & number of companies for every sector . . . 32
5.4 Relevant networks in the Netherlands . . . 33
5.5 Distance distribution for transport networks . . . 34

6.1 Industry locations (coloured to matching sector and sized to demand) . . . 35
6.2 Example of industry locations close to each other . . . 36
6.3 Example DBSCAN clustering (Taken from QGIS, labels representing demand) . . . 37
6.4 Analysis plots for k-values . . . 38
6.5 Cluster overview . . . 39

7.1 Grid for selecting research areas . . . 40
7.2 Research area 1 & 2 . . . 42
7.3 Research area 3 . . . 43

9.1 Results Research Area 1 . . . 49
9.2 Share of length and cost for transport options - Research area 1 . . . 50
9.3 Results Research Area 2 . . . 51
9.4 Share of length and cost for transport options - Research area 2 . . . 52
9.5 Results Research Area 3 . . . 52
9.6 Share of length and cost for transport options - Research area 3 . . . 53

10.1 Example of number of trucks limitation . . . 57

A.1 Scatter-plot normalised data (complete dataset) . . . 67
A.2 Scatter-plot clusters (k=3) . . . 68
A.3 Scatter-plot final dataset . . . 68
A.4 Analysis plots for k-values . . . 69
A.5 Detailed silhouette scores for k (4-7) . . . 70
A.6 Standardized cluster values for final k-values . . . 71

C.1 Legend for maps . . . 77
C.2 Step 1 - RA-1 . . . 77

C.3 Step 2 - RA-1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
C.4 Step 3 - RA-1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78
C.5 Step 4 - edge improved tree - RA-1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78
C.6 Step 4 - final - RA-1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78
C.7 Step 1 - RA-2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
C.8 Step 2 - RA-2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 79
C.9 Step 3 - RA-2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
C.10 Step 4 - RA-2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
C.11 All ONLT steps Research area 3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 81

# List of Tables

2.1 Literature overview . . . 15

4.1 Relevant variables for network optimisation . . . 28

5.1 Statistics of distances for different networks . . . 34

6.1 Overview performance k-values on silhouette score . . . 38
6.2 Characteristics of final clusters . . . 39

7.1 Number of cluster nodes for each tile (n>0) . . . 41

8.1 Pipeline parameters . . . 45
8.2 Truck parameters . . . 46
8.3 Vessel parameters . . . 47

A.1 Overview performance k-values on silhouette score . . . 70
A.2 Characteristics of final clusters . . . 72

# Nomenclature

**Acronyms and Concepts**

| | |
|---|---|
| ABM | Agent Based Modelling |
| CAPEX | Capital Expenditures |
| DBSCAN | Density Based Spatial Clustering of Applications with Noise |
| Edge | Connection between two nodes |
| LHC | Liquid Hydrogen Carriers |
| MILP | Mixed-Integer Linear Programming |
| Node | A point in a network or graph that represents a location or object. Nodes are connected to each other through edges. |
| ONLT | Optimal Network Layout Tool |
| OPEX | Operational Expenditures |
| PDOK | Public Services On the Map (Publieke Dienstverlening Op de Kaart) |
| QGIS | Quantum Geographical Information System |

**Model Variables and Parameters**

| | |
|---|---|
| $C$ | Costs (€) |
| $C^{\text{inv}}$ | Investment per vehicle (€) |
| $C^{\text{pipe}}$ | Total pipeline costs (€) |
| $C^{\text{pipe}}_{\text{capex}}$ | Pipeline capital costs (€) |
| $C^{\text{pipe}}_{\text{opex}}$ | Pipeline operational costs (€ per year) |
| $C^{\text{rw}}$ | Total road/waterway transport costs (€) |
| $C^{\text{rw}}_{\text{capex}}$ | Capital costs for trucks/vessels (€ per year) |
| $C^{\text{rw}}_{\text{fix}}$ | Fixed operational costs (€ per year) |
| $C^{\text{rw}}_{\text{opex}}$ | Operational costs for trucks/vessels (€ per year) |
| $C^{\text{rw}}_{\text{var}}$ | Variable operational costs depending on distance (€ per year) |
| $c^{\text{var}}$ | Variable costs per meter of transport (€/m) |
| $D$ | Pipeline diameter (m) |
| $L$ | Length of the edge (m) |
| $L_v$ | Vehicle lifetime (years) |
| $N_{\text{load}}$ | Number of required loads |
| $N_{\text{rep}}$ | Number of vehicle replacements within $T$ |
| $N_v$ | Number of required vehicles |
| $Q_d$ | Hydrogen demand (kg) |
| $Q_v$ | Vehicle capacity (truck or vessel) (kg) |
| $T$ | Calculation period (years) |

# 1

# Introduction

Green hydrogen is one of the promising solutions to decarbonise, in order to combat climate change. Hydrogen can be utilized for different applications, such as energy-intensive industry or fuel for vehicles. It can be produced using renewable energy sources, such as wind or solar power. Next to production and consumption of hydrogen, supply infrastructure is a relevant component of the complete hydrogen system. A hydrogen network is needed to distribute the hydrogen from producers to consumers. Since 2023, Hynetwork, a subsidiary of Gasunie, has started with the realisation of the national hydrogen network. This network focuses on connecting the five largest industry clusters in the Netherlands (Hynetwork, n.d.). This is done by building new parts of pipeline network, but also by using existing natural gas pipelines, which will become available due to the phase out of natural gas.

However, not all Dutch industry is located in those five clusters. The so called 'Cluster 6' consists of the regional industry, which is spread out through the country. The industry in this sixth cluster is responsible for approximately 12.5% of the total emissions of the Dutch industry (Cluster 6, 2025) and developed large amount of sustainability goals. However, there are some challenges to achieve these goals. These smaller industry clusters aren't directly connected to the national network, which makes it harder to distribute green hydrogen to these clusters. Some local industry is relatively far from the national network, which could lead to high connections costs (Cluster 6, 2025). Next to that, natural gas pipelines aren't as available in regional areas as in the areas of the national network, because those are needed to supply residential areas (DEP, 2024). This is different to the five main clusters, where the energy-infrastructure is (almost) fully available for the industry and less competition is present. This might broaden the scope for investigating other transport options as well. It is not clear beforehand which solution fits all regional industry, due to the wide variety of industry-sectors within Cluster 6. It is however necessary to develop infrastructure for more sustainable industry in cluster six, because gaining access to green energy-infrastructure is an extra challenge for local industries.

This master thesis will develop a method to determine the cost-optimal hydrogen transport network for regional industries. Identifying the characteristics of these industries, determining matching transport options and designing a network layout will be the approach to design hydrogen supply infrastructure for Cluster 6.

## 1.1 Description of Cluster 6

This study focusses on network optimisation for regional industry in the Netherlands, located outside the five main clusters. The factors described in the previous section are taken mostly from studies outside the Netherlands, focused on 'general' network optimisation, not highlighting the specific challenges of the companies/demand nodes they include. This study, however, will execute an analysis for the specific companies located in Cluster 6. It is, therefore, relevant to investigate the specific characteristics of Cluster 6 companies as well.
Cluster 6 consists of nine different sectors. Figure 1.1 shows the distribution of the companies within those sectors (Geluk & Koopman, 2020). The black circles represent the five main clusters in the Netherlands. In contrast to the main clusters, Cluster 6 consists of many different types of companies with their own type of processes and demand for energy. A data-centre for example, has low use of natural gas in its process, but a high use of electricity. On the other hand, chemical companies use a significant amount of natural gas due to production processes which require high temperatures. In total, industry in Cluster 6 is responsible for approximately 30% of scope 1[1] fossil $CO_2$-emissions of the complete industry in the Netherlands (WaterEnergySolutions, 2022).

**Figure 1.1:** Cluster 6 industries

The National Cluster Energy Strategy - CES (Cluster 6, 2025) mentions the characteristics of these companies and their plans to reduce their emissions. A typical company in Cluster 6 has a high energy use, which can originate from natural gas (more than one million $m^3$ each year) or electricity (more than 10 GWh each year). In order to reduce the environmental emissions, Cluster 6 is trying to reduce their fossil energy use. This can be achieved mainly by reducing the use of natural gas and other fossil fuels, and capturing $CO_2$. The CES gathered the plans of the Cluster 6 companies, which combined could lead to a 50% reduction of $CO_2$-emissions in 2030 by Cluster 6 companies, compared to 2022.

**Challenges for sustainable improvements**

The energy transition of Cluster 6 companies is not without challenges. Missing infrastructure for green energy supply, competition and challenging areas are factors which a hydrogen network design should account for.
The lack of sufficient infrastructure, amongst other causes, delays sustainability improvements. Access to sustainable electricity, green hydrogen and $CO_2$ capture and storage are likely to not be available on a short term period. Approximately 73% of scheduled sustainability plans which need a new or extended connection to energy, cannot be executed due to the lack of energy-infrastructures (Cluster 6, 2025). Because of congestion on the energy-grid, in combination with a rising demand for energy, connection to the national hydrogen grid is unlikely to be possible before 2030. Due to the large distances between regional industry and the hydrogen network, the connection costs are expensive. Smaller industries are therefore dependent on regional distribution grids for hydrogen (Geluk & Koopman, 2020)
Another challenge, compared to the five main clusters, is the fact that the average Cluster 6 company needs to compete with other users in their environment for energy connections, such as housing construction and the mobility sector. This can lead to an increase of bottlenecks of energy supply on specific locations (WaterEnergySolutions, 2022). Large clusters, on the contrary, mostly have full access to the available energy-infrastructure.

---

[1] Scope 1 emissions exist of greenhouse gases than an organisation emits from sources it owns or controls directly. Other scopes are 2 and 3, and are indirect emissions(McKinsey & Company, 2024).

## 1.2 Stakeholder analysis

The subject of this study affects a lot of different organisations, companies, governmental institutions, and ultimately the whole society. In order to identify the most relevant stakeholders and their interest for this study, a stakeholder analysis is done. This will also support in identifying relevant metrics on which the results can be measured. Cluster 6 companies are already described in the previous section and will therefore not be analysed again for this analysis.

**Network Operators and Infrastructure Providers**

The network operator Gasunie and her subsidiary Hynetwork develops the hydrogen pipeline backbone in the Netherlands. Their goal is to develop and operate the hydrogen pipelines, in order to supply green energy to the Netherlands. This is done by order of the Dutch government, which will be discussed later. Gasunie (2024) mentions that they are focussed on sustainability, safety in construction and use of pipelines and reducing investment costs by reusing natural gas pipelines. Hynetwork has large power in the development of this network, by designing the exact network-layout and building it. Their interest in the development of hydrogen supply transport for regional industry is expected to be large as well, because this may influence the use of their national hydrogen backbone.

**Government**

The Dutch government is assigning Hynetwork to develop the Dutch hydrogen backbone. This is done to decarbonise the Dutch industry, because it is responsible for 25% of all $CO_2$ emissions. Their goal is to finish this network in 2033, as latest updates show (NU.nl, 2025). These updates also show that the costs are doubled compared to the first estimations. This large increase of costs show the need to reduce investment and operational costs wherever possible. The government's budget was initially set at 1.5 billion euros to invest in the hydrogen network, but now rose to 3.8 billion euros. According to Gasunie, the increase in investment costs have to be paid by companies which are connected to the hydrogen network (NU.nl, 2025). The government has large interest in the succession of the hydrogen backbone, so that the Netherlands will decarbonise as much as possible, so that climate agreements can be achieved. Their power is mostly placed in policies and fundings, which can support hydrogen backbone developments.

**Transport and Logistics Companies**

Besides Hynetwork, other transport companies might be needed as well to transport hydrogen from supplier to consumer. Research shows that hydrogen can also be transported by other means of transport, such as trucks or vessels (Busch et al., 2023). These companies will probably be interested in supply needs, layout of supply routes and planning in order to deliver the right amount of hydrogen to the right places. Investments might be needed in trucks or vessels, or they might be needed to modify current vehicles to be able to transport hydrogen. This stakeholder does have large interests, as described, but their power or influence is smaller, because the extent to which they are needed is determined by other stakeholders, which design or inventorise supply needs for these transport companies.

**Energy Producers and Suppliers**

Hydrogen can only be supplied if there is sufficient production capacity available. This production mainly comes from electrolysers, which convert renewable electricity into hydrogen, but also from potential hydrogen imports via ports or other international connections. Energy producers and suppliers therefore play a crucial role in providing the input for the entire hydrogen supply chain. Their interests are twofold: on the one hand, they need a reliable and growing market for green hydrogen, in order to make their investments in production capacity profitable; on the other hand, they benefit from clear and stable policies and infrastructure planning, so that hydrogen can actually reach the consumers. Their influence is significant, since the availability and price of hydrogen directly depend on production and supply. Without sufficient production, even the best-designed infrastructure cannot function effectively.

**Conclusion of stakeholder analysis**

In summary, all stakeholders described above share a common interest in the successful development of a cost-efficient and reliable hydrogen supply infrastructure. For the government, this mainly relates to achieving climate targets and reducing emissions. For network operators, the focus is on designing and operating a safe and financially viable backbone. Transport and logistics companies see opportunities in developing new supply routes, while energy producers depend on a stable market to justify their investments in production capacity. Although their roles and influence variate, their interests come together in one central objective: enabling the large-scale use of green hydrogen to accelerate the decarbonisation of Dutch industry.

## 1.3 Transport options for hydrogen

Hydrogen can be transported in various ways, but is mostly done in liquid or gaseous form (DEP, 2024). Liquid hydrogen is achieved by cooling hydrogen to -253°C, under atmospherical pressure. This process requires relatively large amounts of energy, which increases the costs. However, due to the high purity of liquid hydrogen, it can be applied in different processes. Liquid hydrogen can be transported by trucks, for example. Various studies show the benefits of using Liquid Hydrogen Carriers (LHC) (Busch et al., 2023; Pellegrini et al., 2024). Trucks, vessels or trains with liquid hydrogen serve a large part of demand, next to pipelines, as modelled by Busch et al. (2023).

Gaseous hydrogen, on the other hand, is mostly transported through pipelines. In the Netherlands, Gasunie (2024) already has some experience in transporting gaseous hydrogen through pipelines. The goal is to use existing natural gas pipelines as much as possible, in order to reduce costs. The expectation is that natural gas will be phased out over time, which leads to a large existing network of pipelines which can be used for the transport of hydrogen. Some modification of the pipelines is needed, but this is much cheaper than building new pipelines. However, it is not totally clear when and which parts of existing pipelines are available, due to uncertainty in natural gas deliveries from other countries, such as Russia, or delays in the development of sustainable energy-infrastructure. Because less existing pipelines are available for hydrogen in the Netherlands, the costs of a national hydrogen network rose (Hynetwork, 2025), which reveals the uncertainty of using existing pipelines. Because of the uncertainty of natural gas pipelines and time-consuming work to prepare the right dataset of natural gas pipelines, this option will not be used as a possibility to use for the transport of hydrogen to regional industry. The transport options are further worked out in chapter 5.1.2.

## 1.4 Link to CoSEM program

This thesis subject is closely connected to the MSc Complex Systems Engineering and Management (CoSEM) program, which focuses on the analysis and design of complex socio-technical systems that combine multiple stakeholders, technologies and interests. Hydrogen supply infrastructure for regional industry can be seen as such a system, with both technical components (pipelines, vessels, trucks) and institutional aspects (policies, company strategies). The design of a cost-optimal hydrogen network requires not only technical and economic modelling, but also awareness of its societal impact. Decarbonising regional industry is essential for reaching national climate goals and reducing greenhouse gas emissions, while at the same time ensuring the long-term sustainability and competitiveness of Dutch industry. By applying clustering methods, network modelling and multi-modal transport analysis, this thesis directly reflects the interdisciplinary approach of CoSEM and contributes to decision-making for one of today's most urgent societal challenges: the energy transition.

## 1.5 AI statement

Artificial Intelligence tools were used to support parts of this research. ChatGPT (OpenAI) was used to assist in coding, debugging, and reviewing text sections. LanguageTool was applied to check grammar and style. ResearchRabbit was used to explore and identify relevant scientific literature. All results, interpretations, and conclusions remain the responsibility of the author. Every part of text and code is checked and reasoning of this study has been completely done by the author.

# 2

# Theoretical background and research questions

This chapter will analyse the current academic literature in order to identify a knowledge gap. When the knowledge gap is clear, research questions are drawn, which will be based on the problem statement from the previous chapter and the identified knowledge gap from literature.

## 2.1 Theoretical background

Research on green hydrogen in general is significantly increasing over the past decade, being recognised as one of the promising solutions to a sustainable future. In order to know what research has brought so far, which knowledge this delivers and how this report can be of added value to the large amount of research, a literature review is performed. For this report, research sources Scopus and ResearchRabbit have been used to identify relevant papers, next to citations of relevant papers, which deal with the same sort of topic.

| ID | Source | Regional/local industry? | Network analysis? | Method |
|---|---|---|---|---|
| 1 | Namazifard et al. (2024) | - | Yes, robustness, clustering | MILP |
| 2 | Lafeber (2024) | - | Yes, robustness | RBPM |
| 3 | Mendler et al. (2025) | Yes | - | Spatial framework |
| 4 | Vergara et al. (2024). | - | - | Literature review |
| 5 | Franco and Giovannini (2023) | - | - | General analysis |
| 6 | Neuwirth et al. (2022) | - | - | Data analysis |
| 7 | Parolin et al. (2024) | Yes | Yes, multiple transport options | MILP |
| 8 | Parolin et al. (2022) | Yes | Yes | MILP, GIS |
| 9 | Busch et al. (2023) | - | Yes, shortest path theory | Energy Transformation Pathway Optimization Suite |
| 10 | Hammond et al. (2025) | Yes | Yes, including obstacles | Graph theory |
| 11 | Liu et al. (2024) | Yes | - | ABM |
| 12 | Kountouris et al. (2024) | - | - | Sector-coupled energy system model |
| 15 | Pellegrini et al. (2024) | - | - | Literature review |
| 16 | Oberle and Klopstein (2024) | Yes | - | Pandapipes |
| 17 | Lipiäinen et al. (2023) | - | - | Cost calculations |
| 18 | Kim et al. (2023) | - | - | LCA |
| 19 | Kotek et al. (2023) | - | - | Model based scenario analysis |
| 20 | Cheng and Cheng (2023) | - | - | Ranking method |
| 21 | M. Yang et al. (2023) | - | - | Comparison |
| 22 | Bique and Zondervan (2018) | - | - | MILP, AIMMS |
| 23 | Timalsina et al. (2025) | - | - | MILP, ROA |
| 24 | Beagle et al. (2023) | - | - | Monte Carlo |

**Table 2.1:** Literature overview

Various search terms within Scopus has been used, resulting in approximately 75-80 papers. Next to Scopus, Researchrabbit is used to find similar research compared to selected papers from the Scopus results. Papers from Scopus were selected based on year of publishing, the use of multiple transport options and optimization methods. Finally, a selection of 24 papers is reviewed. Table 2.1 provides an overview of these papers, including characteristics of the papers.

The final selection shows a variance of papers, with limited focus on regional or local industry and a combination of network optimization. Mostly, methods such as MILP are used to determine cost optimal solutions, but those methods do not include a network layout for specific areas.

**Discussion of literature**

The amount of research on hydrogen infrastructure has grown strongly in recent years. Many studies focus on analysing transport options or estimating the costs of different supply scenarios. These studies can roughly be divided into three categories: cost-optimisation analyses for different transport modes, evaluations of potential hydrogen infrastructures, and research on the applications of hydrogen.

Most of these studies cover relatively large geographical areas, such as Germany (Bique & Zondervan, 2018; Busch et al., 2023), (parts of) Europe (Kotek et al., 2023; Kountouris et al., 2024; Neuwirth et al., 2022), Belgium (Namazifard et al., 2024), or Canada (Cheng & Cheng, 2023). In other cases, the analysis is not tied to a specific area at all, but rather discusses general transport options and their costs. For example, Bique and Zondervan (2018) compared a base and a green scenario for Germany, showing that a green hydrogen supply is significantly more expensive than the base case. They also concluded that liquid hydrogen distribution is preferable, particularly when transported by rail. Similarly, Busch et al. (2023) showed that for liquid hydrogen, rail is the primary mode of transport, followed by vessels and trucks. Their results also indicate that higher demand for liquid hydrogen decreases the relative need for pipelines, although the total length of pipeline infrastructure still increases in high-demand scenarios.

A common methodological approach in this literature is Mixed Integer Linear Programming (MILP), which focuses on mathematical cost minimisation. Timalsina et al. (2025), for example, used MILP to determine the optimal transport mix for hydrogen. Their results suggest that phased strategies, such as first building new pipelines and later upgrading to repurposed gas pipelines or trucks, can provide cost-optimal solutions. These kinds of studies highlight cost dynamics well, but usually lack detail on local conditions.

Research that takes network optimisation into account is less common and often applied only on a smaller scale (Busch et al., 2023; Hammond et al., 2025). Moreover, demand nodes are usually defined as large industry clusters or cities, and distances are often based on Euclidean measures, without considering obstacles or site-specific constraints. Hammond et al. (2025) is an exception: they proposed a spatially explicit optimisation model that integrates GIS-based locations with production and consumption values, while also considering geographical obstacles. Similarly, Parolin et al. (2022) showed that including geographical constraints and combining multiple transport modes can lead to more realistic and often more cost-efficient infrastructure solutions.

A few papers can be matched (partly) to the subject that this study will investigate. These papers are discussed in detail. Hammond et al. (2025) uses graph theory to investigate the routing of pipelines, because routing specifically is underdeveloped in hydrogen infrastructure models. This paper focusses on using Obstacles Genetic Algorithm, by which it is prevented that uniform costs surfaces are assumed, and this will make bespoke routes. This is of added value for hydrogen supply infrastructure for Cluster 6, because bespoke routes are needed for the regional industry, which is located at different types of locations and areas in the whole country. However, the study of Hammond focusses only on the use of pipelines, whereas it was concluded in section 1.3 that various transport options are possible for the supply of hydrogen. This decreases the flexibility in supply, which is needed for the wide range of different companies within Cluster 6.

A study that does include multiple transport options, on regional scale is Parolin et al. (2024). It aims to assess the infrastructural needs to supply the regional demand from transport industrial sectors. Transport options used are compressed hydrogen trucks, liquid hydrogen trucks and pipelines. The results showed that the use of multiple transport options is relevant, even when large hydrogen flows are needed, for which pipelines are most suitable. This study can be of high value for the hydrogen development for regional industry such as Cluster 6. Downside of this report, however, is that the consumers were quite broad. Different sectors were analysed, such as industry, aviation and different consuming transport options. In order to develop a bespoke network layout, specifically for every company in regional industry, Parolin et al. (2024) lacks detailed analysis of consuming industry companies.

Following the analysis of these papers, there remains a clear gap in research. Existing studies mostly focus on broad national or continental scales, use simplified distance assumptions, or apply cost-optimisation models without explicitly accounting for local industrial sites and geographical characteristics. These models mostly use methods such as MILP, whereas graph theory would be more suitable if the goal is to account for regional conditions. Research that does combine multiple transport options with regional industry locations and their specific conditions is limited.
This thesis contributes to closing that gap by developing a method that optimises hydrogen transport for regional industries, while explicitly considering multiple transport modes and characteristics of demanding industries. In this way, it adds to the existing literature by moving from broad and generic analyses towards a more detailed, regionally applicable approach to hydrogen network design.

## 2.2 Research questions

First, the research questions are established, based on the literature review described in the previous section. This review mentioned the research gap found, where current research lacks providing sufficient insight into transport networks for hydrogen supply, based on local conditions, such as availability of supply networks. Specific and realistic plans are hard to determine, because of lacking frameworks or models to support decisions. The following research question and subquestions will support in finding answers for this research gap.

**How can hydrogen supply transport for Dutch regional industry be cost-optimized?**

This research question is answered by the following subquestions:

1. What are relevant characteristics of regional industries with regard to hydrogen transport optimization?
2. How can the relevant characteristics be identified and prepared for clustering?
3. What characteristics-based clusters of regional industry can be identified?
4. What are suitable research areas, where a specific regional industry cluster is strongly represented?
5. How are different transport options used in a cost-efficient network model for different research areas?

Figure 2.1 provides a schematic overview of the results of every subquestion.

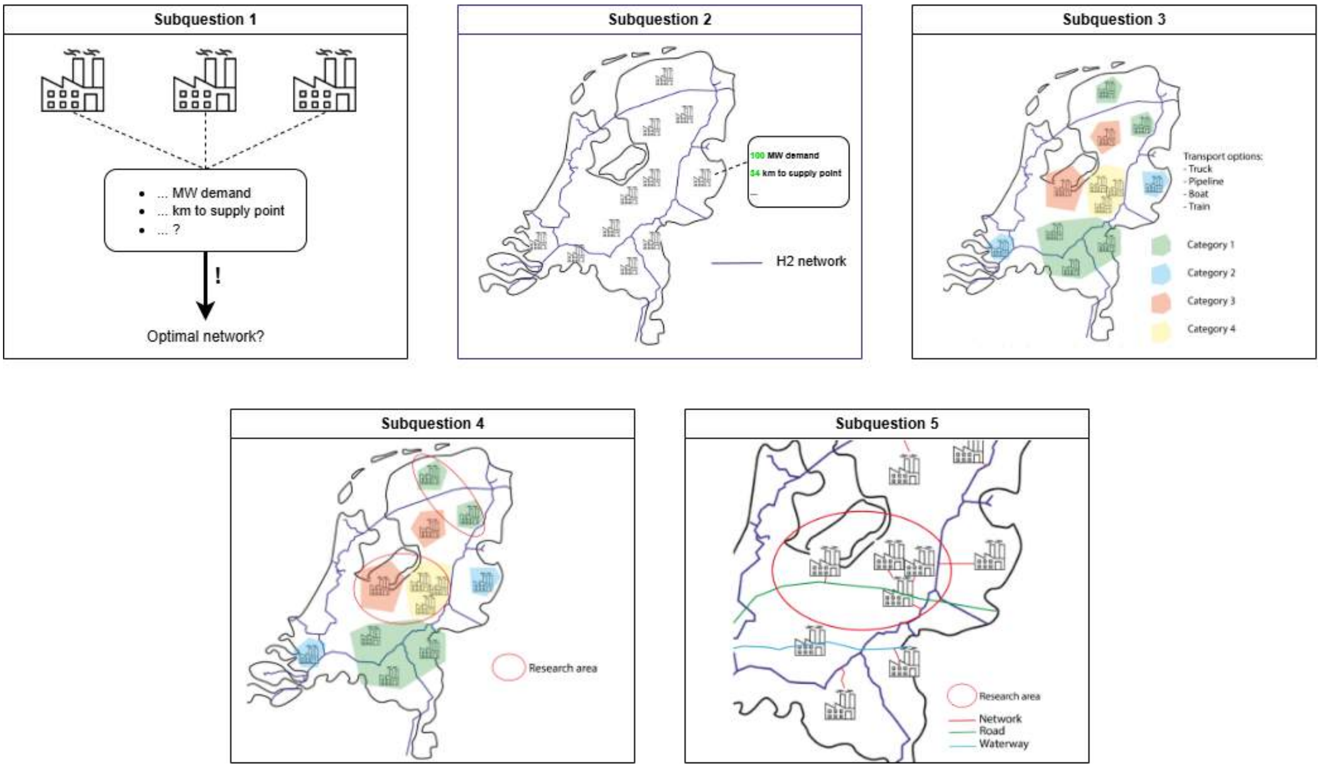


**Figure 2.1:** Results of every subquestion

These subquestions will answer the main research question. They can be divided into two parts. Subquestion 1-3 are used to identify and visualize the characteristics of the regional industry, so that it's clear what characteristics are most relevant for network optimisation. As shown in section 1.1, the industry in Cluster 6 is quite diverse. It is therefore important to know what type of industries are present, so that hydrogen supply transport can be suited to these specific companies. As concluded in section 2.1, current research on hydrogen supply networks is mostly done on large scales, missing a level of detail. By including these first three questions, this study will contribute to the current research by including analysis on characteristics for all companies in the network. Figure 2.1 visualises the outcomes of these subquestions in the box for subquestion 3 (top-right). A map with identified clusters will be drawn, so that it is visually interpretable what Dutch regional industry looks like.

The second part of this research will model an optimal network layout, based on the information required in subquestion 1-3, using the Optimal Network Layout Tool (Heijnen, 2025a). First, subquestion 4 determines the relevant research areas, based on the identified clusters. Research areas that mainly contain one specific industry cluster are especially interesting to analyse, because optimal transport options for a specific cluster can be identified. Subquestion 5 will then use the ONLT to calculate the cost-optimal network layout for regional industries, for every research area. Based on the outcomes of the ONLT, conclusions can be drawn which hydrogen supply transport is cost-optimal for the various identified regional industries in the Netherlands. By answering these questions, it will be known how regional industry can be connected to hydrogen supply, so that all industries, not only large consumers, can use sustainable fuel.

# 3

# Methodology & data requirements

This section will describe how the defined research questions will be answered. First, the methods and data required for every subquestion are described. Second, the research flow diagram shows the structure of this thesis. The complete research flow diagram is shown in section 3.5.1.

## 3.1 Determining relevant characteristics for hydrogen transport optimization (SQ-1)

**Method**
The first subquestion will focus on determining the relevant characteristics, which influence transport cost-optimization. Within the description of the theoretical background (chapter 2, different papers which analyse how transport optimization is influenced were found. Using existing literature and desk research, a set of characteristics is determined. The stakeholder analysis (section 1.2 highlighted the different interests present for this study. Sustainability and costs are the main drivers for these stakeholders. Sustainability can be achieved by developing and building the hydrogen transport network, but costs are impacted by the exact network layout and use of different transport options. The latter will be investigated by this study, as is shown in the research questions (section 2.2) The influence on cost of transport will therefore be the main variable on which the influence of the characteristics will be measured.
This subquestion will also select transport options suited for this study, based on their characteristics and diversity, so that different types of transport are available.

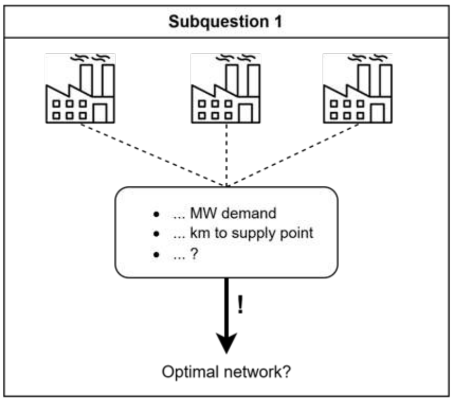


**Figure 3.1:** Visualisation SQ-1

**Data requirements**
For this subquestion, literature sources from the theoretical background will be used mostly. Cluster 6 reports can describe specific characteristics for this case, and academic literature can describe characteristics which are proven to be relevant in earlier studies. Besides this, complementary desk research will be done. Together, sufficient data can be found to establish relevant characteristics with regard to transport optimization.

## 3.2 Gathering regional industry data (SQ-2)

**Method**

The results of subquestion 1 provided the relevant characteristics, for which the exact data can be gathered in subquestion 2. Each industry location will therefore be analysed based on these characteristics, resulting in a dataset including data for every location. The exact data used is described in the following section.
To determine distances between supply networks (e.g. the hydrogen backbone) and demand nodes, a combination of QGIS and the Geopandas package in Python will be used. QGIS allows different maps, layers and links to be visualised and drawn, while also enabling distance calculations. The national hydrogen backbone and other relevant supply networks will be drawn in QGIS, and the industry nodes will be marked. This will result in a map with streets, roads and buildings as the base layer, combined with the hydrogen backbone and industry nodes.

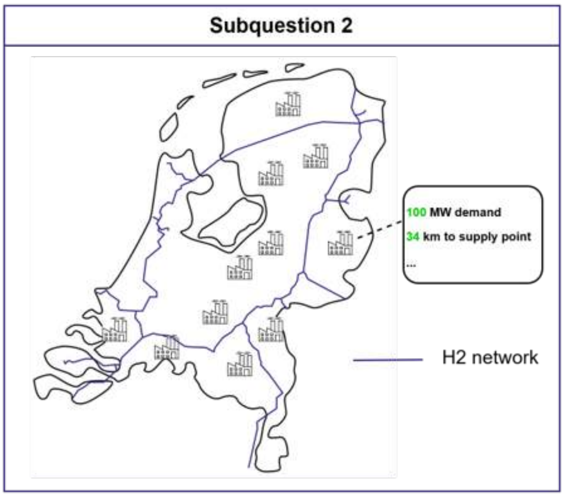


**Figure 3.3:** Visualisation SQ-2

In addition to demand nodes, which represent industry locations, supply nodes also need to be identified. This is done in order to know from what location supply will originate, which is important to know for the final model. In this model, specific starting locations for supply need to be allocated, so that every transport options has a starting supply point. Possible supply options will be located through literature and desk research. This will involve some simplification, because the future number and locations of hydrogen production sites are still uncertain. Currently, only one hydrogen factory is operating in the Netherlands (TenneT, 2024).

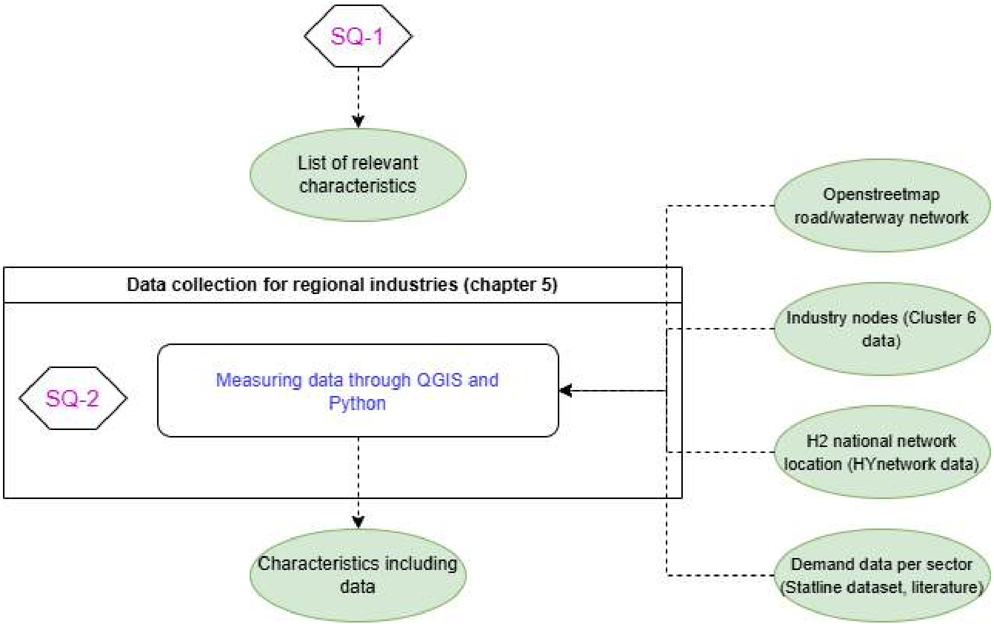


**Figure 3.4:** Research Flow Diagram - SQ2

**Data requirements**

This step is crucial, because it determines the characteristics that form the dataset for all following subquestions. Figure 3.4 displays the inputs and outputs of this subquestion.
First, the locations of industry nodes and the transport networks are required. The sections of the transport networks will be gathered through the use of PDOK datasets, which consists of all major networks in the Netherlands. Although exact coordinates of the Dutch hydrogen backbone are not yet available, the planned route sections are included in the roll-out document of Hynetwork (2024). These sections can be drawn in QGIS to create a usable data layer. For other supply networks such as roads, waterways or train tracks, datasets are available online.

In order to gather coordinates of industry locations, Cluster 6 reports will be used, which provide maps of connected industries, labelled per subsector (Cluster 6, 2025). In addition, Geluk and Koopman (2020) lists all participating companies in a table. Exact geographical coordinates are not directly available, but combining these sources with Google Maps allows the identification and marking of company locations in QGIS, which can then be used as demand nodes.

Second, hydrogen demand data is required for each sector. This part is more complex, because future demand is uncertain and not yet available at company level. On a higher level, however, estimates exist. For example, Jovicic and West (2023) analysed future production scenarios and resulting hydrogen demand for Northwest Europe between 2030–2050. These scenarios are general and cover only a few sectors, which do not fully align with the subsectors of Cluster 6. Other studies, such as Lamboo et al. (2024), provide an overview of current hydrogen use in Dutch industry, but mainly for small to medium-sized users, which makes them less relevant for Cluster 6.
Various regional reports are also available. For example, WaterEnergySolutions (2022) describes expected energy use, hydrogen use, and emission reduction for most Dutch provinces in 2025 and 2030. Although this is not directly aligned with the specific sectors in Cluster 6, it can still provide useful input for defining demand per sector. The most consistent dataset, however, originates from CBS Statline (2025), which provides sectoral energy use data until 2024. This dataset does not include future projections, but it can be used as a baseline for current energy demand.

Because sector-level demand is more readily available than company-level data, some generalisation is needed. Total sectoral energy use will be divided across the number of companies in Cluster 6 within each sector. Additionally, an assumption must be made regarding the share of energy demand expected to be fulfilled by hydrogen in the future. This step allows a consistent dataset of hydrogen demand to be constructed for all demand nodes.
For production locations, it is assumed that sufficient hydrogen supply is available at any time. Supply starting locations will be determined in subquestion 3.

## 3.3 Clustering regional industry (SQ-3)

**Method**

First, the clustering of the gathered data from subquestion 2 will be performed. This will be done in Python. In order to compare the different metrics, the data will be scaled using the StandardScaler package (SciKitLearn, n.d.). This method makes sure that every data-point has the same weight. This is done by standardizing the data, so that the new average is 0 en standard deviation is 1. The following formula is used to calculate new standardized values:

$$z = (x - \mu)/\sigma \tag{3.1}$$

where:

- $x$ = original value
- $\mu$ = average of original dataset
- $\sigma$ = standard deviation of original dataset

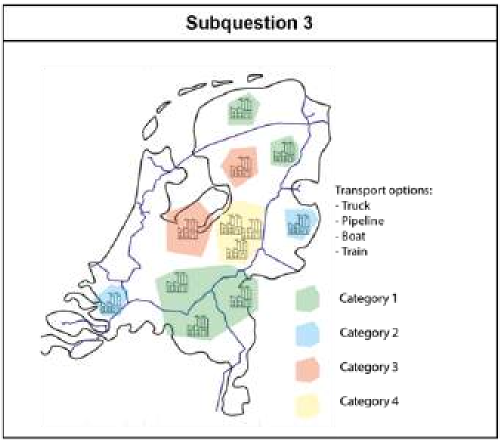


**Figure 3.5:** Visualisation SQ-3

The clustering is done using the K-means clustering method (MacQueen, 1967). K-means clustering is an unsupervised learning algorithm, which can be used to group unlabelled data into groups or clusters (Kavlakoglu & Winland, 2024). The algorithm is a so-called exclusive clustering method, which means that every data point can only exist in one cluster. In order to cluster every data point, K-means clustering uses an iterative process, where the dataset is partitioned into similar groups, based on the distance between the centroids of the clusters. Finally, the objective is to minimize the sum of distances between data points and their assigned clusters. K-means is used for this study, because it is suitable to identify clusters with similar characteristics, which is the goal of this subquestion. Besides that, K-means uses numeric variables and can easily be scaled to large numbers of data points, which both are present in this study.
A cluster is formed with all data points that are nearest to a centroid. Next, the mean of all points within a cluster is calculated and the cluster centre or centroid is relocated. This process is repeated until the maximum number of iterations has been reached, or the centroid positions have reached convergence (Kavlakoglu & Winland, 2024). The amount of clusters that is generated is not determined beforehand, but will be optimized, dependent on the quality of the categories. The quality will be evaluated using the elbow method, silhouette score and ability to describe a cluster. After this, established clusters are visualized on the earlier modified map in QGIS, so that it is visually clear which categories are present in what areas, and in what numbers.

Secondly, the industry locations will be clustered on physical distance, using the DBSCAN clustering method Ester et al. (1996). DBSCAN stands for Density-Based Spatial Clustering of Applications with Noise. It clusters data-points which are in a certain 'epsilon', which is the size of the circle around a single point. This is also called the neighbourhood. Shortly said, this means that if two data-points are within a certain distance from each other, the method will cluster them together and creates a new data-point which represent the clustered data-points. (Thailappan, 2024). Besides that, the least number of data points which the method wants to see within the range of a data point can be chosen. Therefore, outliers can be accepted by this method. This will be done to reduce computation time of the final model used in subquestion 5 by reducing complexity through reducing the number of locations.

**Data requirements**

No specific data for this step is needed, except earlier mentioned literature and the data collected in subquestion 2.

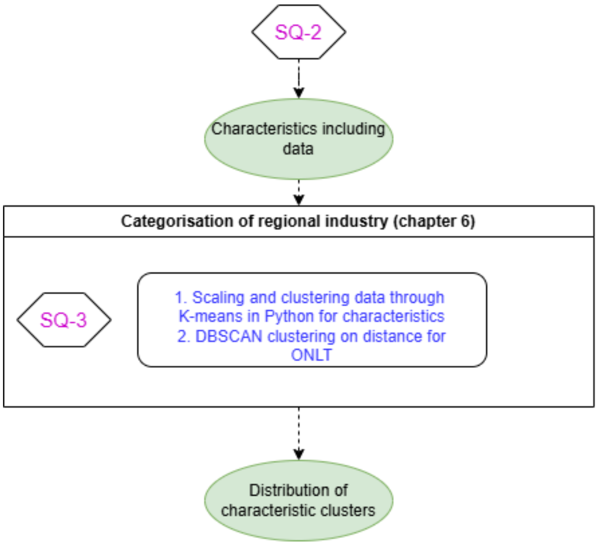


**Figure 3.6:** Research Flow Diagram - SQ3

## 3.4 Selection of research areas (SQ-4)

**Method**

Ideally, the Netherlands would be modelled completely at once, to make sure the full context of the country can be accounted for. However, in order to apply the optimisation model efficiently, it will not be possible to run the model for the entire country at once. This would result in very high computation times and would make it harder to analyse the influence of specific industry types. Therefore, a selection of research areas will be made, so that representative parts of the Netherlands can be analysed in more detail.

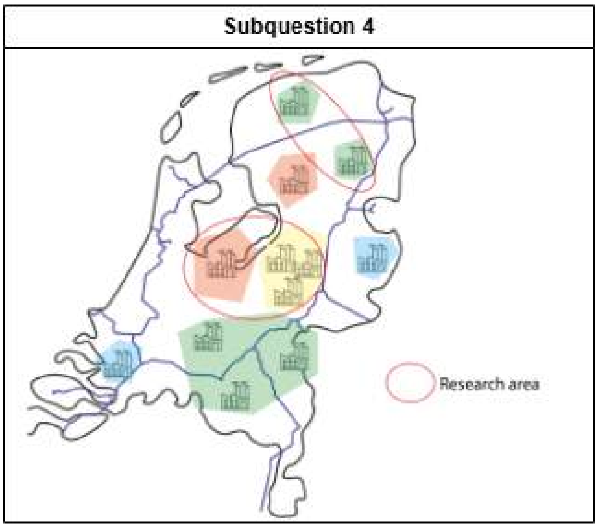


**Figure 3.7:** Visualisation SQ-4

The selection will be done by creating a grid over the Netherlands, dividing the country into separate tiles. This will be done using QGIS. For each tile, the number and assigned K-means cluster of industry locations will be analysed. By using the tile statistics from QGIS and Excel, it will become clear in which tiles certain K-means clusters are most present. Areas with a large concentration of one or two clusters will be considered most relevant, since they allow conclusions to be drawn for specific K-means clusters, i.e. the different type of companies identified in the previous step. To be clear: this step only uses the cluster results from the K-means clustering (clustering on distance- and demand characteristics). The results from the other clustering method (DB-SCAN, clustering on location) is used in the final model. This means that in the final model, some data-points are merged into one new data-point, through DB-SCAN clustering. It is expected that this won't impact the final model negatively, because the DB-SCAN method still accounts for all demand and distance. Because only data-points within 2000 meters are clustered in DB-SCAN, the K-means cluster-label for these data-points is expected to be similar, because of similar distance-characteristics.

From this analysis, a limited number of areas will be chosen as case studies for the optimisation model. Each of these areas is expected to represent a different characteristic that could influence the final optimal hydrogen supply network.

**Data requirements**

Data for this step is already available from previous subquestions. The allocation of cluster labels to every industry location is needed, which come from subquestion 4. Further data isn't needed.

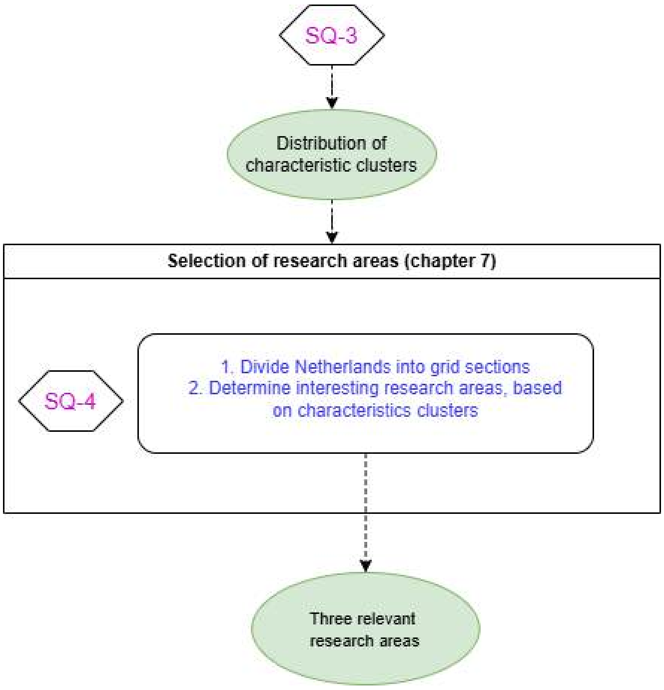


**Figure 3.8:** Research Flow Diagram - SQ4

## 3.5 Network optimization for all transport options (SQ-5)

**Method**

Now that the input data is clear, the Optimal Network Layout Tool (ONLT) (Heijnen, 2025a) in python will be used to determine the cost-optimal network layout. In order to run the tool for different research areas, it needs to be prepared and tested to be ready. A number of modifications to the model are needed to fit this study. A cost function in which the capital costs and operational costs are included for different transport options is needed. The exact form of the cost function will be determined, based on available data for costs on transport options. Besides that, the behaviour of the transport options is needed to be implemented.

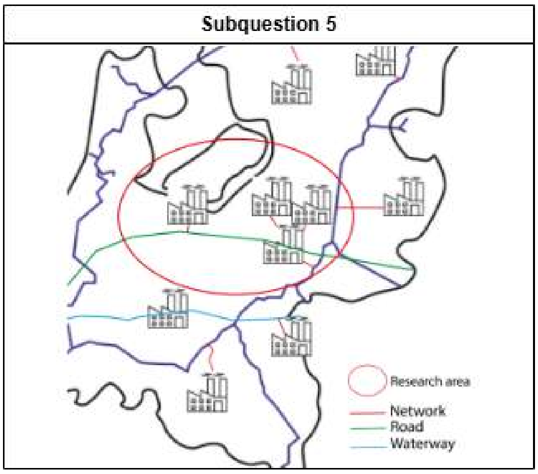
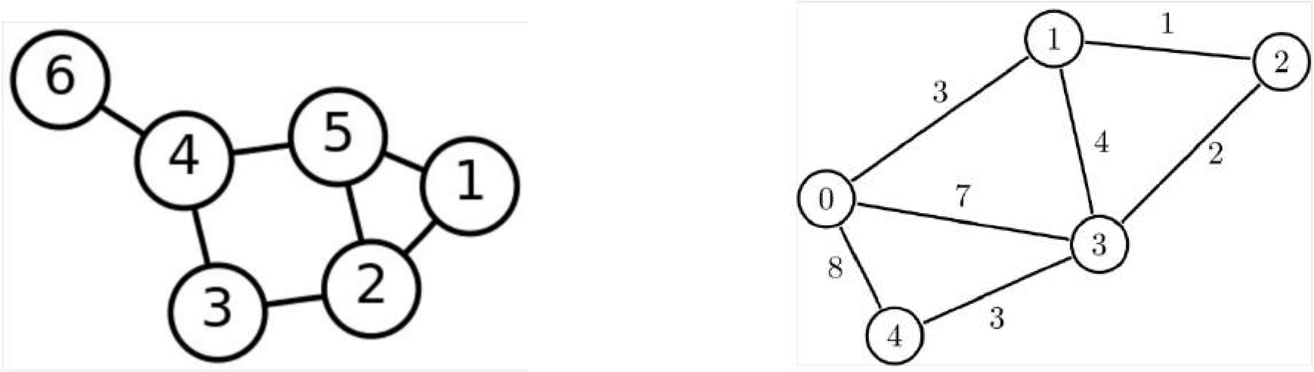


**Figure 3.9:** Visualisation SQ-5

**Graph theory**

The ONLT is build on the use of Graph theory. This section will shortly describe the main concepts of this theory.

A graph is a mathematical structure, mostly representing networks such as social networks, physical routing networks or communication systems, for example. Graph theory is mostly used to understand structures of graphs, and analyse problems such as pathfinding, connectivity and network optimization. Graph theory consists of two fundamental elements, called nodes and edges. Nodes represent an individual point within a graph, where edges represent the connection or relationship between two nodes (Pykes, 2024). Figure 3.10 visualises two examples of graphs, which are explained further below.



**(a)** Example of simple graph

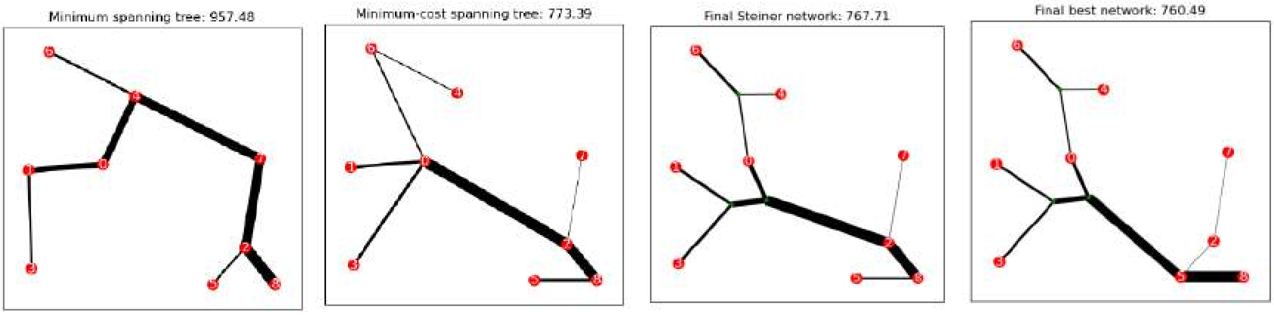
**(b)** Example of weighted graph

**Figure 3.10:** Graph theory examples

Figure 3.10a visualises a simple graph, including six different nodes with connections. This is a simple graph, because the graph doesn't have multiple edges between any pair of nodes or a loop, which connect a single node to itself. An example can be a basic social network, where the friends are represented by the nodes, and friendships by the edges.
Figure 3.10b shows a weighted graph. The basic principles are the same as a simple graph, but in this case, the edges are weighted. The weights can represent multiple things, for example, the length, capacity or costs of and edge. A weighted graph is mostly used in network optimization problems and will be used in this study as well.
Besides simple and weighted graphs, directed and undirected graphs are two key concepts of graph theory. In a directed graph, the edges are assigned with a direction, like one-way streets. An undirected graph exists of edges with no direction, like two-way streets.

### Optimal Network Layout Tool (ONLT)

The optimal network layout tool is a model based on graph theory, which finds a minimal cost network that connects multiple sources (producers) with multiple sinks (consumers), taking into account that demand and supply patterns can vary over different time steps (Heijnen, 2025a). Producers and consumers (representing nodes) are defined beforehand, as can existing connections, specific routing networks which need to be followed, storage locations or obstacles the network needs to consider. The connections have different properties, such as length and capacity.
The model exists of five steps to achieve a cost-optimal network. The first step is a preparation step and imports and reads the input file including all data needed, such as coordinates of all locations and networks, demand, etcetera. After this step, four steps are executed to determine the cost optimal network. Figure 3.11 provides examples of a network formed through these four steps (Heijnen, 2025b).



**(a)** Step 1

**(b)** Step 2

**(c)** Step 3

**(d)** Step 4

**Figure 3.11:** Examples of ONLT steps

First, a *minimum-spanning-tree* is made, which calculates the minimum-length network that connects all nodes. Besides that, capacity of edges is assigned in such a way that there is just enough capacity to supply all demand nodes, if there is sufficient supply.

The second step copies the network from step one, and takes costs into account to build the network-layout. By rewiring the edges in a heuristic process, the total network costs might decrease. The network resulting from this step is called a *minimum-cost-spanning-tree*. After this, the third step evaluates the network from step two for better solutions by looking for possible split points (green points in figure 3.11c) on edges. This can reduce the length of edges, which could decreases costs. These split points are called Steiner nodes, and therefore the network resulting from this step is called the *minimum-cost-Steiner-tree*. In the final step, the previous two steps are repeated as long as better results (lower costs) are found.

*Data requirements*
The only extra data that is required for this subquestion is cost data for the all transport options. These can be gathered through the same literature as used in the theoretical background. For all other data, it's taken from the results of previous subquestions.

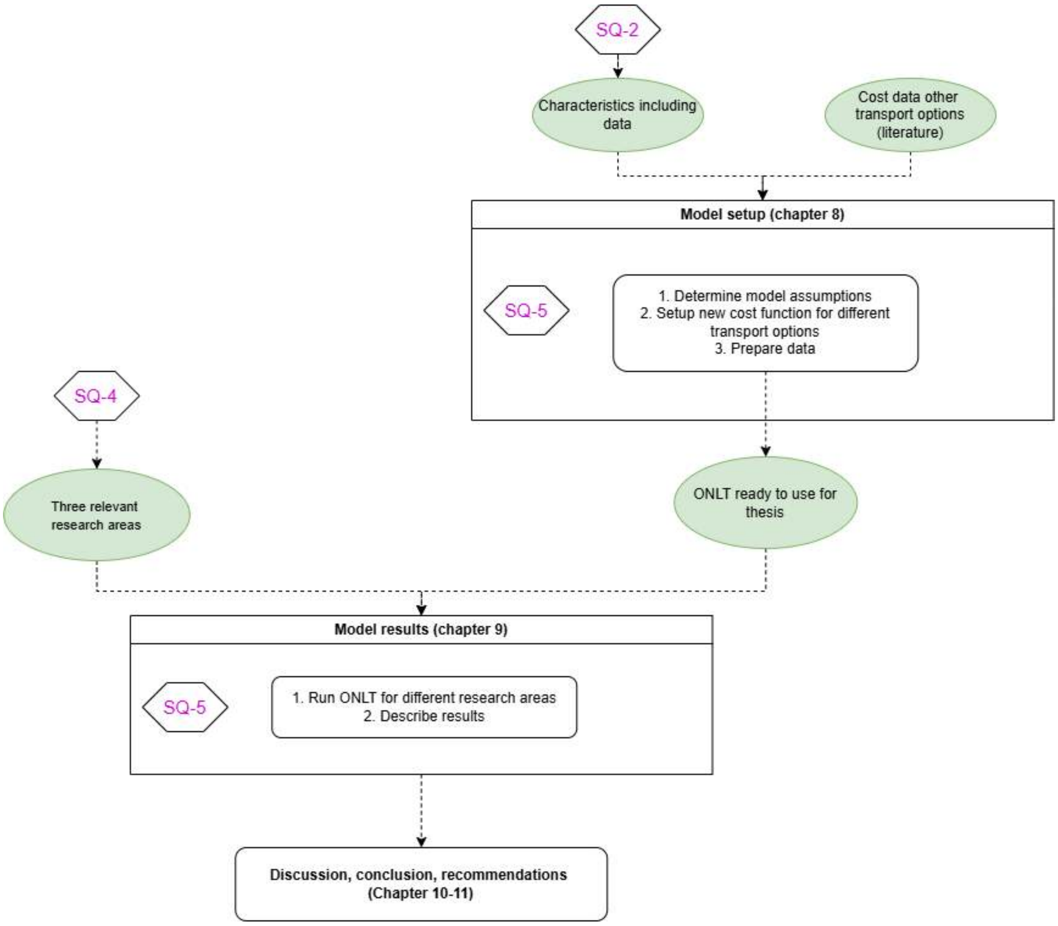


**Figure 3.12:** Research Flow Diagram - SQ5

### 3.5.1 Research Flow Diagram

The figure below visualises the entire Research Flow Diagram, for all subquestions. All separate parts are described in previous sections.

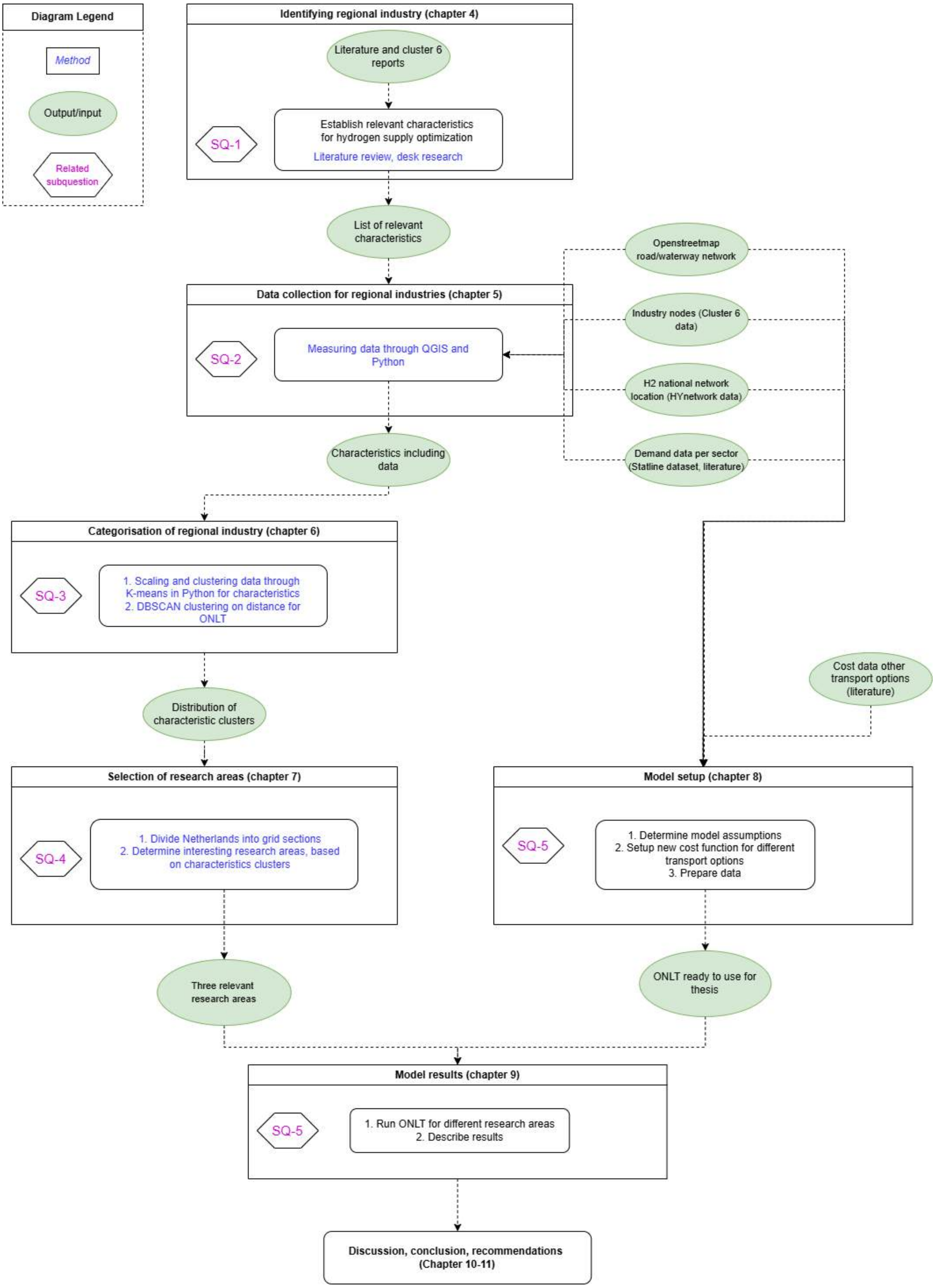


**Figure 3.13:** Research Flow Diagram

# 4

# Identifying regional industry

Optimizing a (hydrogen) network cannot be done without knowing what variables influence the optimal layout of the network. The diversification within Cluster 6 is one of the challenges highlighted for the energy transition (WaterEnergySolutions, 2022). The companies investigated are quite diverse, regarding their locations, demand of hydrogen and industry-sector. This subquestion analyses which variables influence the cost-optimization of a hydrogen network. This is done by analysing multiple literature sources, including reports from the Cluster 6 organisation itself, previously discussed literature and a specific literature search for factors influencing hydrogen networks. The literature is analysed by looking specifically for methods used in other optimization reports, in which it should be clear which variables are relevant for the cost-optimal solution. First, literature on cost-optimization for hydrogen networks is discussed. Second, the transport options for this study will be determined.

## 4.1 Literature review

Recent literature provides a large amount of network optimization analysis. Table 4.1 highlights the research papers used for this report, which are analysed because of their application in similar investigations or their connection to this case study (Cluster 6 industries). For every study, the factors which influenced the network cost are marked. The results show that demand and distance of transportation are the factors that are used most in network analysis. Demand is used because the amount of hydrogen determines the capacity needed of a transportation, for example the size of pipelines or the amount of trucks. The distance of transportation accounts for the other major part of costs when designing a transportation network. Both factors influence the capital/investment costs (CAPEX) and operational costs (OPEX).

| | **Author** | **Demand** | **Distance to supply** | **Characteristics industrial process** | **Obstacles** | **Transport option** |
|---|---|---|---|---|---|---|
| *Cluster 6 reports* | Geluk and Koopman (2020) | x | x | x | | |
| | WaterEnergySolutions (2022) | | | x | | |
| | Cluster 6 (2025) | x | x | | | |
| *Academic literature* | Namazifard et al. (2024) | x | x | x | | |
| | Lafeber (2024) | x | x | x | | |
| | Parolin et al. (2024) | x | | | x | |
| | Parolin et al. (2022) | x | x | | | x |
| | Busch et al. (2023) | x | x | x | | x |
| | Hammond et al. (2025) | x | x | | x | |
| | Lipiäinen et al. (2023) | x | x | | | |

**Table 4.1:** Relevant variables for network optimisation

Besides demand and distance, various reports mention the impact of the consuming locations and their process characteristics. These characteristics include the specific form of hydrogen demand (liquid, gaseous, possibility for electrolysis). Busch et al. (2023) for example, concludes that liquid hydrogen is only used for transportation if there is demand at industry locations for liquid hydrogen. Otherwise, hydrogen is transported in gaseous form. This influences the cost of transportation, because the form of hydrogen determines the amount of hydrogen that is needed for the same amount of energy, due to energy losses at conversion. Another relevant characteristic of a process is the temperature needed. Hydrogen can be used to fuel industrial processes which require high temperatures, but if the required temperature is lower (<°C200), electrification of the processes can a more suitable option (Geluk & Koopman, 2020). It is therefore important to investigate which industrial places require high temperature processes, and use natural gas as fuel. This distinction will be made in section 5.1.1

Hydrogen networks are build in different types of environments. When designing a realistic network, there can be obstacles in place, such as urban, Natura2000 or military areas. Next to that, topologies of water and road networks can influence the hydrogen network, as concluded by Hammond et al. (2025). Hammond mentions that a limited number of studies exist which account for complex geographical constraints. However, when geographical constraints are included by a obstacles genetic algorithm in the model, the infrastructure's predicted capital expenditure is reduced by 20 to 40%, compared to other models that assume uniform cost surfaces. This shows the relevance of modelling for real-life scenarios, where actual locations and areas are taken into consideration.

## 4.2 Relevant transport networks

Secondly, relevant transport networks are chosen, so that the networks which suits this study best are taken into account for further subquestions. The literature review from section 4.1 showed that, besides demand, distance to transportation influences the costs of transportation. The introduction in chapter 1.3 described various transport options suitable for hydrogen transport. The most common ways over land are pipelines (gas), trucks or trains (liquid) and over water vessels (liquid) can be used. This model will use three different transport options. First, pipelines are included, because there are already plans to build a hydrogen backbone from pipelines in the Netherlands. This is also done for Europe, which shows the relevance of this option (European Hydrogen Backbone, n.d.). Both the hydrogen backbone (which exists of new pipelines and reused natural gas pipelines), as the possibility for building new pipelines will be included in this study. As mentioned in the introduction, existing natural gas pipelines for new connections to regional industry will be out of scope. Besides pipelines, another transport option for land and one option for water is used. The study of Busch et al. (2023) showed that liquid trucks are used more than trains. Next to that, trucks are more flexible with regard to last mile deliveries, due to the fact that roads are leading to every industry location, where railroads are not. Last, vessels transporting liquid hydrogen will also be considered, to analyse the impact of transport over water. They can also serve as an alternative transport option with large capacities, when the hydrogen pipeline backbone isn't close to industries, but waterways are.

## 4.3 Conclusion

Cluster 6 reports show similar factors compared to the literature review, which are relevant for the optimisation of hydrogen supply networks, but are at the same time challenges to keep costs for infrastructure low. The larger the distances are for new pipelines, for example, the higher the costs. The same goes for hydrogen demand. The larger the demand, the higher the costs. Demand and distance of transport are clear variables which determine the final costs of a network. Besides that, the type of industrial processes determine if and how much hydrogen is suitable as sustainable substitute. The industrial processes can be identified by identifying the industry sector for each company. Finally, network layouts, and therefore costs, are influenced by the actual environment of the network. These parameters will be taken into account when developing the final network model.
Three different network options has been chosen to be used in this study, because they suit the needs of earlier established challenges (section 1.1) for Cluster 6. The hydrogen backbone, waterways and road network will be used, providing different type of transport options for different situations.

# 5

# Data collection for regional industry

In the previous chapter, it is established that hydrogen demand and distance to supply are the main variables which determine the costs for the hydrogen network. Besides that, Cluster 6 reports showed the variety in different companies, regarding their sector and locations. This chapter will describe how the data for those variables is collected, and how this data describes the variety of industries within Cluster 6.

## 5.1 Identification industrial companies

This section will describe which industrial companies are selected and what their relevant characteristics are. Cluster 6 exists of a large number of companies, spread throughout the whole country, as was visible in figure 1.1. Cluster 6 (2025) is a national cluster strategy for Cluster 6, which provides a list of participating companies. This list only included the name of every company, so multiple steps needed to be taken in order to complete the dataset.

First, every company name was searched online in order to determine the production locations of the specific company, which could be one or multiple. This resulted in a total of 374 production locations. Using Google Maps, the coordinates of every production location of every company could be determined. The second step included assigning every company location to one of the sectors which are present within Cluster 6. Most sectors include a list of related companies, from which every company in this dataset could be labelled with a sector. If a specific company wasn't mentioned by any sector, the company was allocated to a sector on own insights by matching company characteristics with sector characteristics. Figure 5.1 shows the amount of companies for every sector.

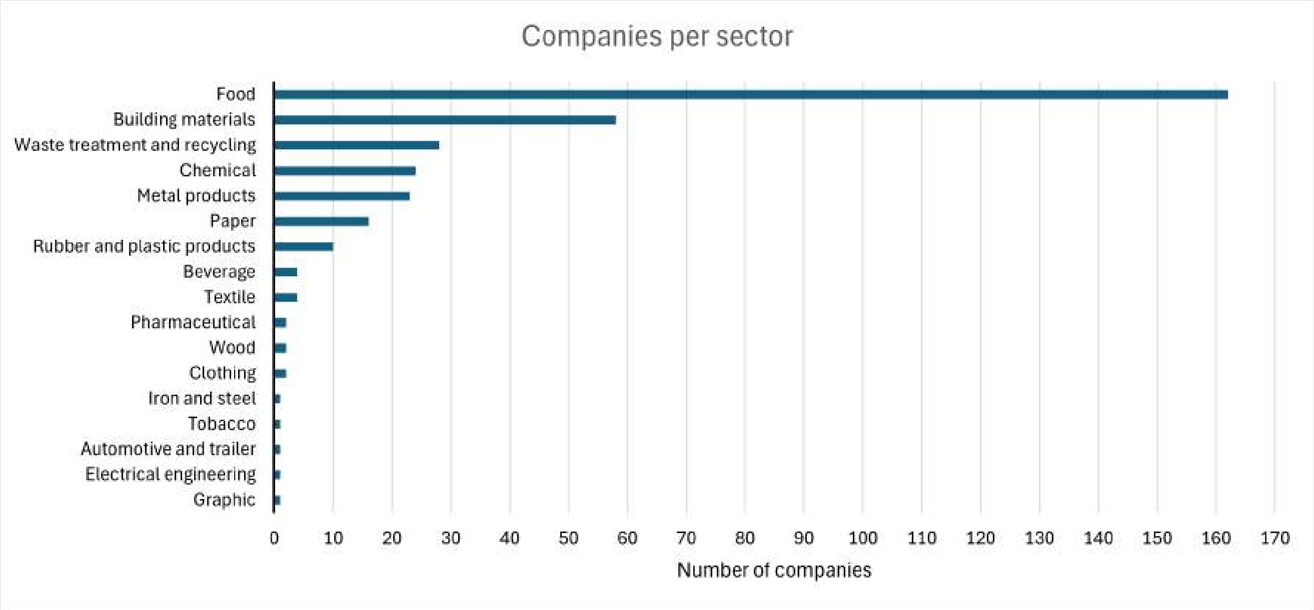


**Figure 5.1:** Companies per sector

It is clear that the food sector is the largest sector within Cluster 6, followed by building materials. Not every company could be assigned to a specific cluster. Some companies didn't exist any more, other companies didn't have a production location and some part of the companies didn't fit in any of the used sectors. Therefore, 34 locations were excluded from the selection, resulting in 340 locations.

### 5.1.1 Demand data

Now that the type of companies is known, demand data can be assigned to every company. CBS Statline (2025) provides the yearly energy demand for every sector. For this study, the 2024 demand for natural gas is used as a reference for how much energy is needed for every sector, for one year. It is likely that hydrogen will replace natural gas as a fuel for high-temperature processes (Geluk & Koopman, 2020). However, not all natural gas demand will be replaced by hydrogen, but other sustainable fuels can be used as well. Parolin et al. (2024) assumes that 50% of current natural gas consumption will be replaced by hydrogen, and the other 50% by energy sources such as electricity. This assumption is also used in this study. Therefore, the total energy demand fuelled by natural gas for every sector, is divided by 2 in order to achieve a realistic energy demand fuelled by hydrogen. Important to mention is that this replacement factor will be different for every unique company in reality. Some companies will maybe replace 80% of energy use by hydrogen, others only 20%. This study assumes a percentage of 50% for every company, taking the average percentage.
The dataset from CBS-Statline (2025) provides energy demand for a detailed set of sectors, which is more detailed than the sectors used in figure 5.1. Therefore, every company is assigned to a specific sub-sector, so that appropriate demand data is assigned to every company. The assignment of sub-sectors has been done in a similar way as the assignment of sectors, but this time it's mostly done by own insights. Figure 5.2 shows the distribution of energy demand for all companies used in this study. After excluding locations with no natural gas demand, the final set of locations consists of 291 unique locations.

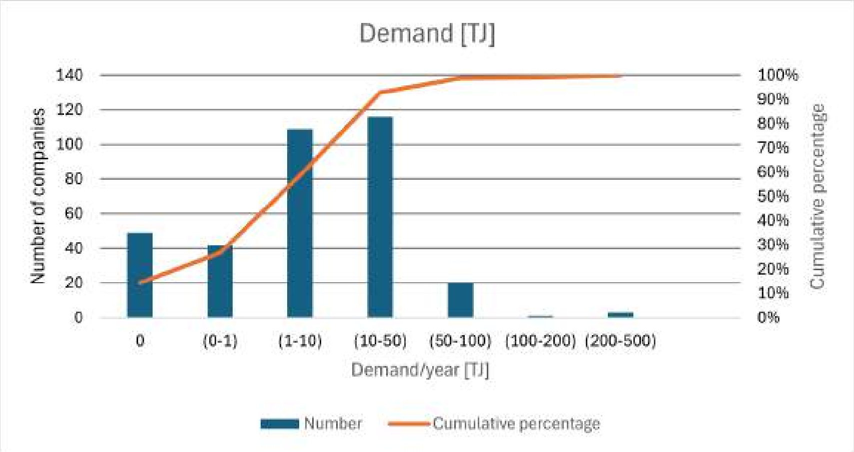


**Figure 5.2:** Energy demand for hydrogen

Now that demand for every industry location is known, total demand for every sector can also be calculated. Figure 5.3 plotted the demand for every sector, combined with the number of companies for that sector. Both are projected as a percentage of total number of companies or demand, which provides insights in the actual impact of the sectors. Despite the large amount of companies in the food sector, it isn't the largest sector with regard to total demand. However, the chemical sector, which includes less companies within its sector, has the highest total demand. Another interesting example is the waste treatment and recycling sector. This sector consists of 8% of the total number of companies, but only 1% of total demand.

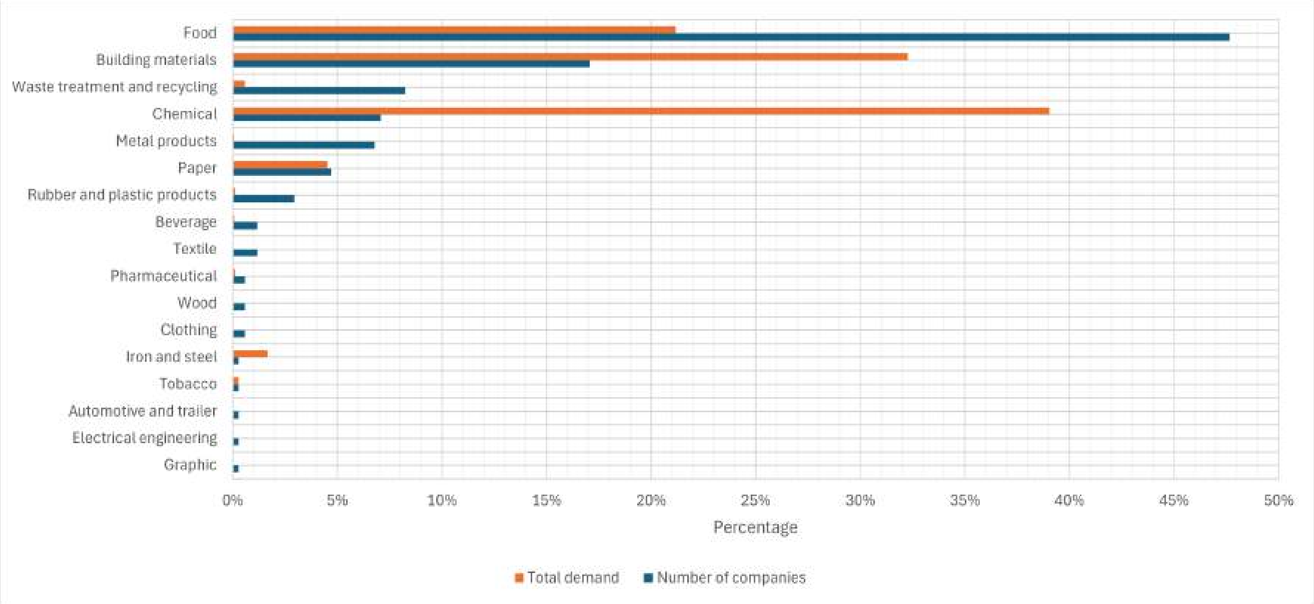


**Figure 5.3:** Demand & number of companies for every sector

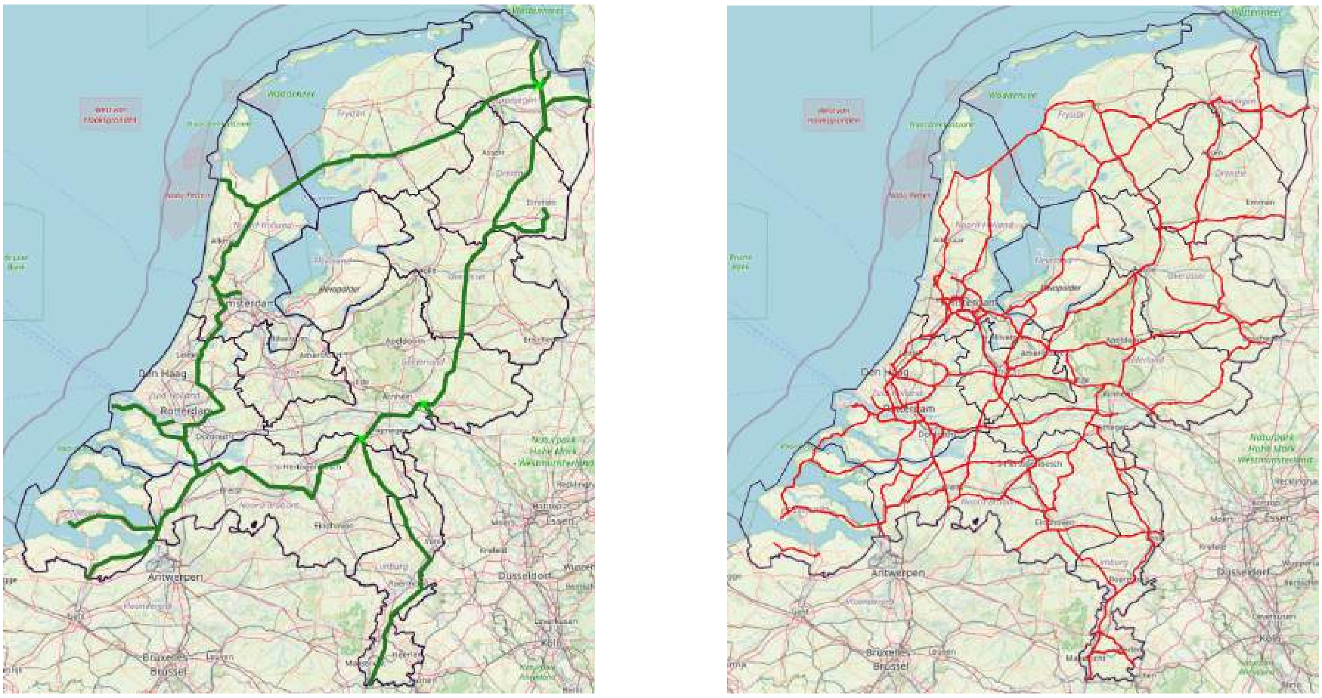
### 5.1.2 Distance to transport networks

This section will analyse the three chosen transport networks. This analysis has been done by using the geographical information program QGIS for visualizing data and Python coding for calculating distances between networks and industry locations.

As mentioned earlier, the Netherlands are developing a hydrogen pipeline network which connects the five main industry clusters. Besides that, it connects to neighbouring countries, so that the export and import of hydrogen can be done. This network is expected to be fully operational in 2033 (Hynetwork, 2024). For this study, it is assumed that the Dutch pipeline network is completely finished and can be used to supply regional industries as well. The network has been drawn in QGIS by hand, because no geographical datasets of the network are available yet, except maps of the network. Figure 5.4a shows the drawn network. The network is mostly located along the border of the Netherlands, leaving an area in the centre of the country uncovered by the pipeline backbone. It also includes some connections to neighbouring countries for import and export.

Second, the Dutch road network is retrieved from [PDOK] (2025b). This dataset contains all roads in the Netherlands, and is therefore a large dataset. In order to save computation time, this dataset is is filtered using Python, so that highways only are left, which are shown in figure 5.4b. This filtered network will be used in further calculations and models. A high density is seen in the western part of the country, connecting the high number of cities located in that area. Highways are less present in the northern part of the Netherlands.

Last, the Dutch water network is retrieved from [PDOK] (2025a) and is also filtered. Because not all waterways in the Netherlands are suited for inland vessels, only the main transportation waterways are selected (Eichhorn, 2021). Figure 5.4c shows the final waterways selection, which is mostly covered by the largest Dutch rivers, located in the middle part of the country. The provinces Groningen, Drenthe, Overijssel and Noord-Brabant are less covered by waterways, as is the area in the middle of the Netherlands.



**(a)** Pipeline network

**(b)** Road network

**(c)** Water network

**Figure 5.4:** Relevant networks in the Netherlands

**Analysis of transport networks**

Finally, the distances between the different networks and every industry location is calculated. This is executed by a Python script, which calculates the shortest euclidean distance between a location and every network. First, average distances and standard deviations resulting from the calculations are shown in table 5.1 to analyse the calculations.

| Transport option | Average distance (km) | Standard deviation (km) |
|---|---|---|
| Pipeline backbone | 11.60 | 10.04 |
| Waterway | 8.53 | 9.71 |
| Highways | 3.14 | 3.24 |

**Table 5.1:** Statistics of distances for different networks

This shows clear differences between highway network distances, and pipeline backbone and waterway network distances. Highways are relatively close to every industry location. This is logical, because the density of the highway network is high, compared to pipeline backbone and waterway network. Besides that, this network which has been used for these calculations is an already filtered selection. In real-life, the road network is even more dense and distances between roads and industry locations will probably be close to zero. On the other hand, the average distances for pipeline backbone and waterway networks are more than three times larger than highway network distances. This can be explained by the lower density of the networks, shown in figure 5.4. Standard deviation data of the waterway distances is relatively large, which shows that distances to a waterway vary a lot. Figure 5.5 visualises the distribution of the data more detailed. The cumulative percentage lines show that already 77% of all 291 companies are within five kilometres of the highway network, and all industry locations are within 25 kilometres of the highway network. In comparison: the share of companies within five kilometres is 32% for the pipeline backbone network and 53% for waterway network. For both options, all industry locations are located within 50 kilometres of the networks.

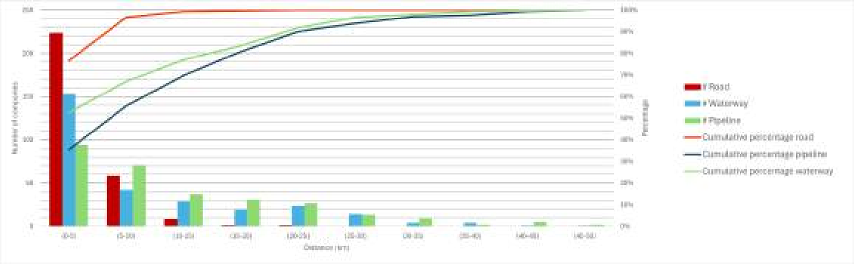


**Figure 5.5:** Distance distribution for transport networks

## 5.2 Conclusion

This chapter analysed the characteristics of Cluster 6 companies, which showed that there is a variety in all characteristics. First, the allocation of sectors showed a large differentiation in types of sectors, existing in Cluster 6. A large share of companies is located in the food sector, followed by building materials and waste treatment and recycling facilities. Not all locations selected at first could be allocated to a useful sector for this study, which led to the exclusion of 34 industry locations. The final set of locations now consists of 291 locations. Next, demand was allocated to every industry location, based on their specific sub-sector. This revealed large differences between sectors; the two sectors with highest demand are responsible for 71% of all demand. This confirms that the type of sectors influences the amount of demand for hydrogen, and therefore influence the cost-optimal network.
The three most relevant transport options for this study (pipeline, vessels over water and trucks) has some differences. The density of the highway network resulted in short distances between the network and industry locations, even for this filtered set of roads. The (future) pipeline backbone and waterways network showed three times larger distances between industry locations and the networks. Besides that, the standard deviation for the data resulted in a three times higher value as well, which showed the wider spread of values for the pipeline backbone and waterway network. Therefore, it can be concluded that the optimal transport option differs for every area, depending on which network is closest to industry locations. This shows the importance of modelling for different type of areas, where different transport networks are most prominent available.

# 6

# Categorisation of regional industry

Now that the relevant characteristics for all industry locations are known, the locations will be analysed for similarities in their characteristics. This will be done in two different ways, both with use of clustering methods. The first clustering will decrease the number of locations which will be put into the final model, so that computation time is diminished. This will be done by clustering the companies based on their location, using DBSCAN clustering method. This will result in the merge of locations close by each other.

The second clustering will support in understanding the type of companies present in Cluster 6. By using K-means clustering, companies with similar characteristics (demand, distance to networks) will be put in the same cluster. This way, the type of companies can be described in a such a way, that their generic for a set of industry locations. This will help determine the research areas and support the conclusions of this study for specific clusters within the Netherlands.

## 6.1 Clustering on distance

All company coordinates are gathered in previous sections. Figure 6.1 displays all locations used in this study, coloured by industry sector. The size of a location corresponds to the size of their hydrogen demand. This figure has been made with the use of Python, based on the database created for chapter 5. Some locations are far apart from each other, but some are relatively close to each other. Two examples are provided in figure 6.2, which shows industries in Veendam and Meppel, two villages in the Netherlands. If companies are this close to each other, they will probably be supplied by the same transport means, because there distance to transport networks is similar. Therefore, it is more convenient for optimal network analysis to consider locations this close to each other as one location. Besides that, the clustering of locations close to each other will save computation time for the final model.

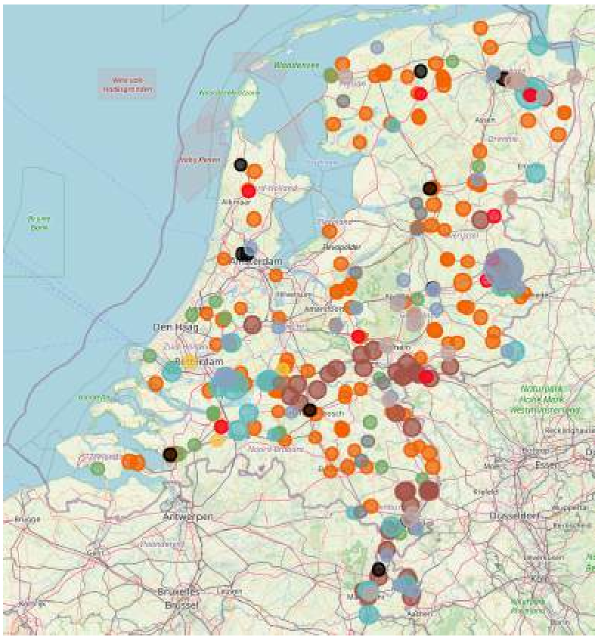
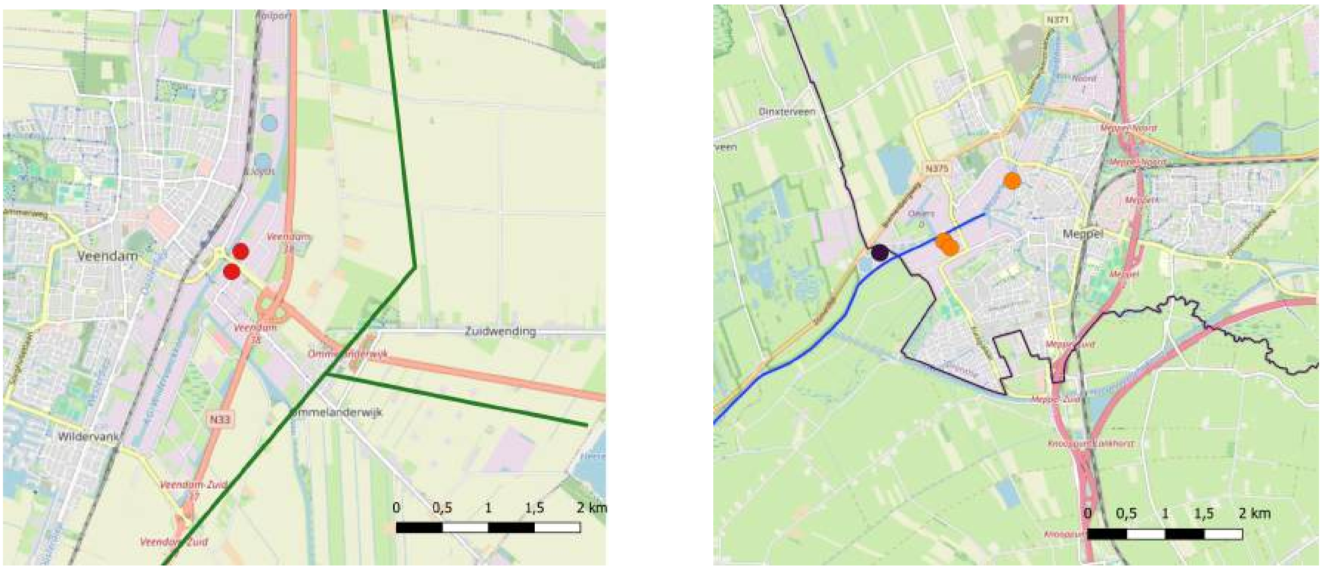


**Figure 6.1:** Industry locations (coloured to matching sector and sized to demand)

In order to apply this clustering method to the dataset used in this study, the right parameters needed to be established. As mentioned earlier, outliers/points with no neighbours are accepted. The size of the neighbourhood is chosen to be 2000 meters, after considering different distances. Lower values than 2000 meters didn't always add data-points near to each other in the same cluster, where a higher value added data-points which were far away from each other to the same cluster. As can be seen in figure 6.2, the distance between companies within the same village, is approximately 2000 meters.



**(a)** Industry locations Veendam

**(b)** Industry locations Meppel

**Figure 6.2:** Example of industry locations close to each other

One disadvantage of clustering points together, is that they 'lose' their unique characteristics, such as hydrogen demand and location, which are needed for the final network optimisation. However, because only locations close to each other are clustered, their main characteristics (distance to various supply networks) is almost equal. The only characteristics that can be significantly different is hydrogen demand. In order to take care of demand, the following has been done. After the DBSCAN clustering was completed, the sum of demand for all data-points in the same cluster is taken. This extra step makes sure, the total amount of demand for that cluster is still taken into account. This increases realistic behaviour of the network optimisation, because now it takes into account that demand of a neighbourhood can be higher than a single data point in that neighbourhood. Otherwise, it might connect all data-points separately to a transport network, where it could be financially beneficial to provide one major connection to supply all data-points at once.

Secondly, the unique location of a data-point is lost when clustering points. For this study, the location of the new data-point representing the clustered points, is set to be the geometric centre of all clustered points. An example can be found in figure 6.3. The black point in the middle represents the new clustered data-point, the other nodes represent the original data-points. The black point is located at the geometrical centre. The labels show demand for every point, where the sum of all orange points is equal to the value of the black point.

Demand and location are taken care of, but the specific industry sector label for every data-point is now lost. Because the industry sector is used specifically for insight in demand, the specific industry cluster is no longer relevant for further clustering and modelling.

**Figure 6.3:** Example DBSCAN clustering (Taken from QGIS, labels representing demand)

The DBSCAN clustering (including centering locations and summing demand) has been executed for all company locations in the Netherlands. After clustering, 217 points remained, reducing the initial set from 291 points. These points will be used in the network optimisation, further explained in chapter 8.

## 6.2 Clustering on characteristics

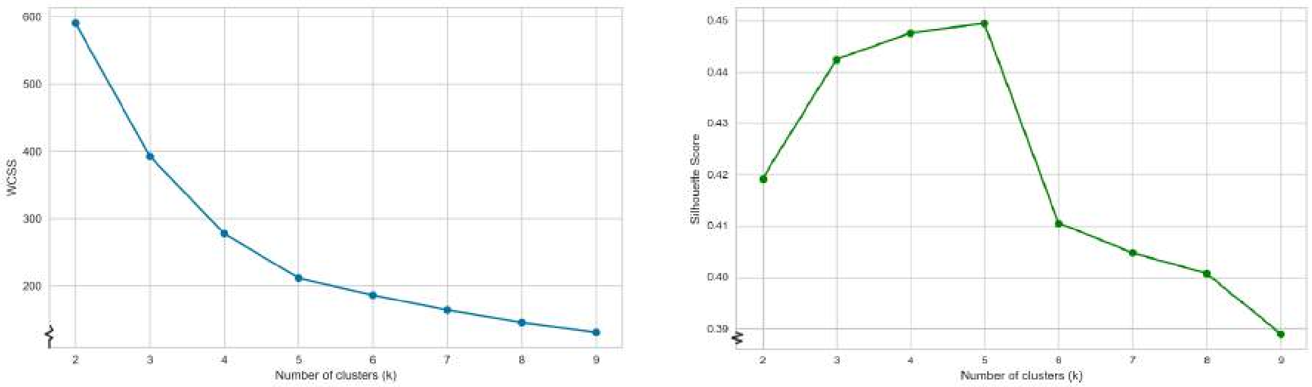
The total set of industry locations consists of many different type of companies, as concluded in chapter 4. Clustering those companies into specific groups, which can all be described by their common characteristics, will support drawing conclusions for a specific area. For example, a government would like to investigate a specific area for the development of hydrogen transport to regional industries. If it's known which type of companies is most present, it will be possible to calculate the type of hydrogen transport needed for this specific area. Therefore, clustering based on company characteristics will support in calculating what the optimal solution might be for a specific area.

### 6.2.1 Execution of K-means clustering

For this specific study, the k-means clustering has been executed by following a few steps. The full described version can read in appendix A, where every step has been extensively described. This section will summarize the results of every step, concluding by the final optimal set of clusters.

- Step 1 - Data selection and preparation
- Step 2 - Plot data for visual inspection
- Step 3 - Remove outliers for optimal clustering
- Step 4 - Pre-analyse data for optimal k-values
- Step 5 - Run and analyse K-means clustering for optimal k-values
- Step 6 - Determine final k-value

These steps support in systematically selecting the correct number of clusters, based on multiple performance metrics. After the execution of step 1-3, it was clear that the data needed some 'cleaning'. Distance data for the highway network and industry locations is not used in this analysis, because every location is close to a highway, and every location in reality has a direct connection to a road. Secondly, data is standardized using the StandardScaler method to be able to compare different scales of data for distances and demand Besides that, a few outliers with relative large demand were excluded from the clustering, thereby increasing the quality of the final clusters. Because outliers can impact heavily the averages that K-means uses in order to select centroids, the final set of clusters might not be representative of the actual data. This is why the outliers are removed.
After selecting the final dataset and parameters for this clustering method, step 4 showed how different values of k performed on the elbow method and silhouette score. Figure 6.4 shows the performance for k-values 2-9 for both metrics.



**(a)** Elbow plot

**(b)** Silhouette score plot

**Figure 6.4:** Analysis plots for k-values

The elbow method plots the 'Within Cluster Sum of Squares' (WCSS) for a range of k-values. WCSS represents the total variance within each cluster. This metric is plotted against the range of k-values. The lower the WCSS value, the better the clustering is done (Kavlakoglu & Winland, 2024). By searching for the 'elbow point', a well-performing number of clusters can be determined. The consideration is that a higher number of clusters might improve the WCSS, but might decrease the ability to describe clusters well. The silhouette score calculates how well a specific data-point fits to its appointed cluster, compared to other clusters. This score varies between -1 and 1, where a higher value means that clusters are more dense and separated from other clusters (Kumar, 2020). It is important to mention that the differences between the number of clusters seems to be large, but the vertical axis has a relatively small range (0.39-0.44) It is therefore harder to draw conclusions from this graph. K-values 3-5 do have the highest values in this graph, but this needs to be put in perspective, as mentioned. After analysing the different k-values in detail (full description in appendix A), the k-values 4-7 showed to be best performing for the two metrics. A detailed overview of performance on silhouette score is shown in table 6.1.

| Number of clusters (k) | Below average silhouette scores | Size fluctuations | Negative silhouette scores |
|---|---|---|---|
| 4 | **+/-** | **-** | **-** |
| 5 | **+** | **+/-** | **-** |
| 6 | **+** | **+** | **-** |
| 7 | **+** | **-** | **-** |

**Table 6.1:** Overview performance k-values on silhouette score

The final two steps evaluated those 4 k-values on their ability to describe different clusters. This analysis showed that k-values 4 and 7 were excluded at first, due to their lack/abundance of describing all relevant types of industries. Combining the results from the elbow method and (detailed) silhouette score, k=6 seems to be the optimal k-value for this data set. The different clusters all describe a unique group of data points. Next to that, k=6 performs best on silhouette scores (table A.1). Therefore, this study will use 6 different types of companies within Cluster 6. The final clusters are described in table 6.2 and shown in figure 6.5, where every coloured node represents a different cluster.

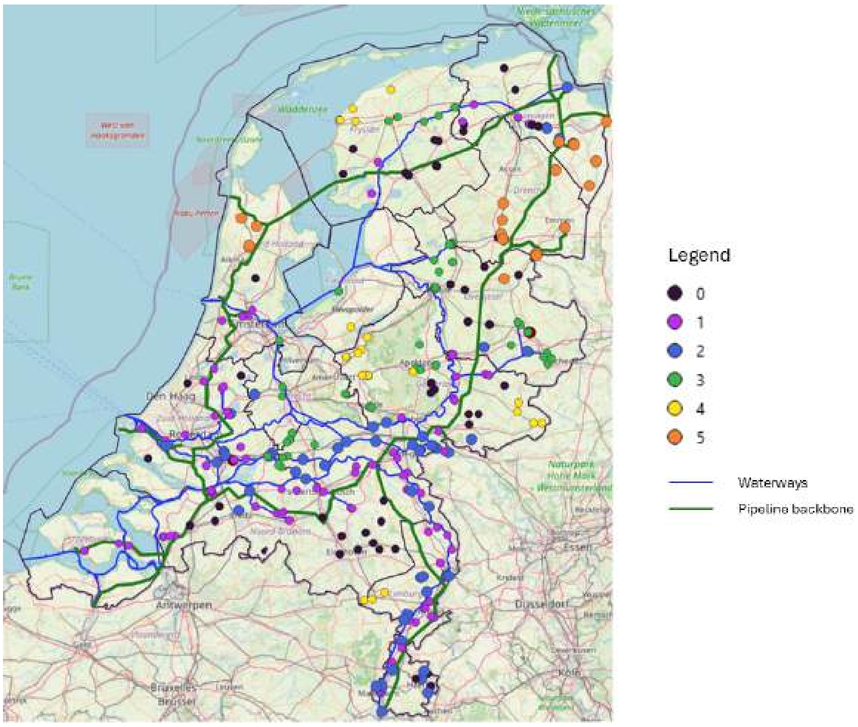


**Figure 6.5:** Cluster overview

| Cluster | Hydrogen demand | Distance to hydrogen backbone | Distance to waterways network | Number of industries |
|---|---|---|---|---|
| 0 | Low | Low | Moderate-high | 52 |
| 1 | Low | Low | Low | 89 |
| 2 | High | Low | Low | 54 |
| 3 | Low | High | Low | 47 |
| 4 | Low | High | High | 22 |
| 5 | Moderate | Low | High | 23 |

**Table 6.2:** Characteristics of final clusters

## 6.3 Conclusion

In this chapter, the set of industry locations has been categorised in two different ways. First, DBSCAN clustering was used to merge nearby (within 2000m) companies into single locations. This step reduces computation time for the optimisation model, while still taking into account the total demand of clustered companies, ensuring sufficient supply is transported to the cluster. This approach does remove the unique coordinates of individual company locations which are clustered, which could decreases the realism of the final model slightly.Secondly, K-means clustering was applied to group companies with similar characteristics. This clustering method supported in understanding what type of companies are part of Cluster 6. By using hydrogen demand and the distance to the pipeline and waterway networks, the final set of clusters identified six distinct types of companies. These clusters give insight into the variation within the research area and make it possible to generalise results for different groups of companies, rather than for every individual location. Overall, these two clustering methods prepare the dataset for the optimisation model. On the one hand, they make the model computationally feasible, and on the other hand, they provide meaningful groups of companies that can be used to interpret and compare results in later chapters. In the next chapter, the clustering on characteristics will play a relevant role in determining the research areas for this study.

# 7

# Selection of research areas

In order to save computation time of the network optimisation model, and analyse specific industry clusters, a number of interesting areas will be selected. The following sections describe how these areas are chosen and what their characteristics are.

## 7.1 Analysing grid-sections

By creating a grid over the Netherlands, the country is classified into different tiles. This will simplify selecting different areas. Those tiles/areas will each be analysed on what type of industry locations is present, based on their cluster number, established in the previous chapter.

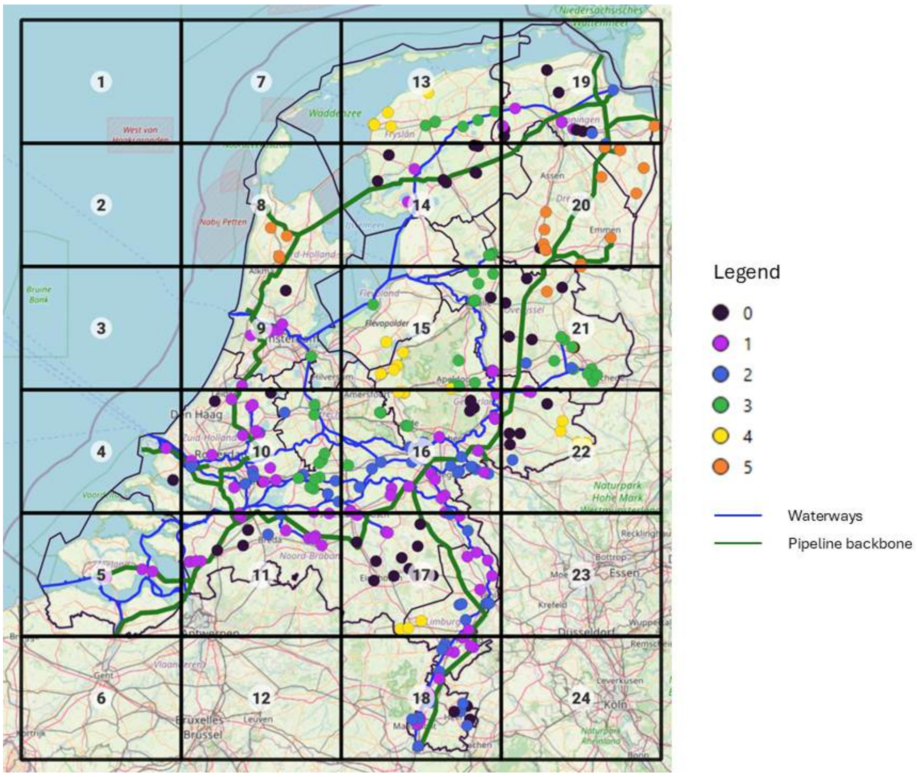


**Figure 7.1:** Grid for selecting research areas

The size of the tiles have been established by visually evaluating different measurements. The tiles shouldn't be too small, to prevent loosing sight into the context of an area. The tiles also shouldn't be too large, to prevent large computation times for large areas. Figure 7.1 shows the grid of four by six tiles, including all nodes, pipeline backbone and waterway network. The highway network is slightly visible, but not highlighted because clustering didn't account for the highway network. The nodes have been coloured based on the cluster they belong to.
The grid has been made using QGIS. With use of QGIS-functions and Excel, the number of companies for every tile has been established. More specifically, the number of companies per cluster has been calculated, for every tile. This will support in selecting areas which include a large number of companies, belonging to the same cluster. Such an area will provide the possibility to derive conclusions for the optimal hydrogen-transport, for the specific type of companies.

Table 7.1 shows the number of companies per cluster, for every tile. Tiles which don't consist any companies are left out of the table. The full table is colour-graded, making the highest number dark-green and the lowest number white. Everything in between is coloured accordingly. The colours help to quickly understand which clusters are most present in a tile.

| | | **Tiles** | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | T-4 | T-5 | T-8 | T-9 | T-10 | T-11 | T-13 | T-14 | T-15 | T-16 | T-17 | T-18 | T-19 | T-20 | T-21 | T-22 |
| **Cluster** | C-0 | 1 | 0 | 0 | 1 | 2 | 4 | 0 | 8 | 1 | 4 | 10 | 3 | 7 | 1 | 6 | 4 |
| | C-1 | 1 | 3 | 0 | 5 | 14 | 15 | 0 | 2 | 4 | 16 | 11 | 11 | 5 | 0 | 0 | 2 |
| | C-2 | 0 | 0 | 0 | 0 | 11 | 2 | 0 | 0 | 0 | 14 | 8 | 13 | 2 | 1 | 2 | 1 |
| | C-3 | 0 | 0 | 0 | 1 | 10 | 0 | 6 | 4 | 12 | 5 | 0 | 0 | 0 | 0 | 9 | 0 |
| | C-4 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 7 | 3 | 3 | 0 | 0 | 0 | 0 | 4 |
| | C-5 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 16 | 2 | 0 |

**Table 7.1:** Number of cluster nodes for each tile (n>0)

The goal of this analysis is to find tiles/areas which have a sufficient number of companies and have one or two clusters with a large share of the total number of companies of the tile. Ideally, there would be a separate research area for each cluster, where that cluster is most present. However, setting up six experiments is time-consuming and therefore the three most interesting research areas are chosen.

The first research area which is interesting, is one where the distance to pipeline backbone is large. This can be interesting, because the main transport option throughout the country is the pipeline backbone. It is important to analyse an area which doesn't have easy access to the pipeline backbone. This is represented by cluster 3 or 4, for example. Both clusters are most present in tile 15. By analysing this area, companies with large distance to pipeline backbone (cluster 3) and companies far from both the pipeline backbone and the waterway network (cluster 4) can be analysed.

Secondly, it is interesting to analyse an area where there is high demand for hydrogen. This can influence the type of transport option strongly. Cluster 2 represents companies with high demand, but are close to each transport network. Cluster 2 is most present in tile 16. However, this tile exists of many other companies from other clusters. Tile 18 contains one company from cluster 2 less than tile 16, but doesn't contain that many other nodes. Tile 18 is therefore the second area which will be investigated. The benefit of this tile is that there companies of cluster 0 are also present, which therefore can be analysed as well.

Last, cluster 5 is interesting to investigate, due to its large distance to the waterway network, combined with low hydrogen demand. This might create opportunities for road transport. Cluster 5 is most present in tile 20, which only has two other companies which belong to another cluster. This will provide the possibility to draw conclusions tailored to cluster 5.

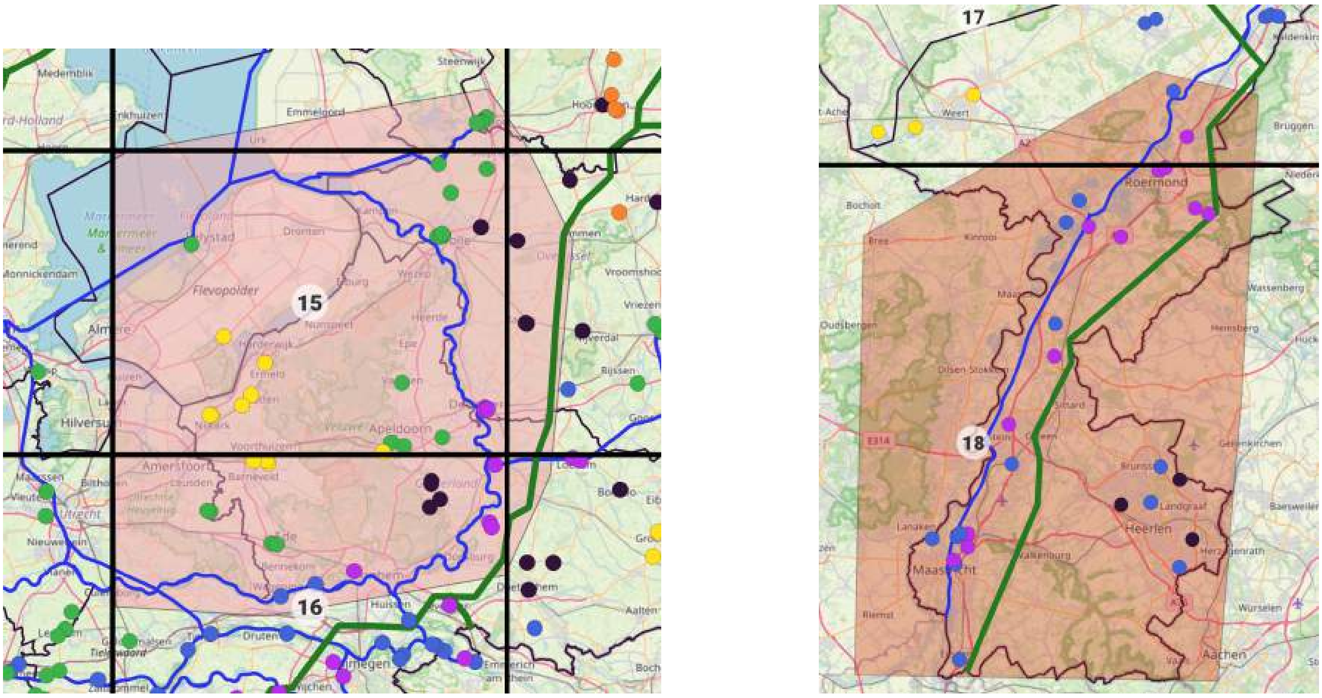
## 7.2 Description of selected research areas

In the following subsections, the chosen research areas will be described shortly. The final research areas which will be modelled, are slightly expanded with respect to the exact area of the chosen tiles. This is done to model areas where context is taken into account, maybe influencing the optimal solution. Sometimes, other companies or part of the network are just outside the chosen tile, but can strongly influence the optimal network. The boundaries of the research areas are therefore chosen by visual inspection in QGIS.

### 7.2.1 Research area 1 - T15/16

The first area is centred around nature reserve 'De Veluwe', showed in figure 7.2a. The area has been extended into tile 16 mostly, taking other companies located within the 'waterway circle', which is located around the 'De Veluwe'. On the right side, the area is extended to include a segment of the pipeline backbone. This has been done to include the possibility to connect to the national pipeline backbone. The type of company that is mainly investigated in this area is a company far away from the pipeline backbone and in some cases from waterways as well. The demand of the company is relatively low. The final research area includes the following list of companies for each cluster:

- Cluster 0 - 8 companies
- Cluster 1 - 8 companies
- Cluster 2 - 2 companies
- Cluster 3 - 20 companies
- Cluster 4 - 10 companies
- Cluster 5 - 0 companies



**(a)** Research area 1

**(b)** Research area 2

**Figure 7.2:** Research area 1 & 2

### 7.2.2 Research area 2 - T17/18

The second research area is located around the boundaries of the province Limburg (figure 7.2b. A small part of tile 17 has been included, because a few locations were located closely to the border of tile 18. This locations in this area are mostly close by the different transport networks, except for a few companies in the bottom right corner. Those companies are also high demanding companies, which makes this area more interesting. Research area 2 consists of the following companies for every cluster:

- Cluster 0 - 3 companies
- Cluster 1 - 12 companies
- Cluster 2 - 14 companies
- Cluster 3 - 0 companies
- Cluster 4 - 0 companies
- Cluster 5 - 0 companies

### 7.2.3 Research area 3 - T19/20

The final research area has been located in the North-east of the Netherlands. Besides tile 20, a large part of tile 19 has been included, due to locations close to the boundary of tile 20. Because the number of companies present in tile 20 is not extremely high, it has been decided to include the major part of tile 19, so that the context of the area is complete. The type of company investigated in this area is a company with large distance to waterways, but close to pipelines. The demand of the companies is moderate. The third research area consists of the following companies per cluster.

- Cluster 0 - 6 companies
- Cluster 1 - 3 companies
- Cluster 2 - 3 companies
- Cluster 3 - 0 companies
- Cluster 4 - 0 companies
- Cluster 5 - 18 companies

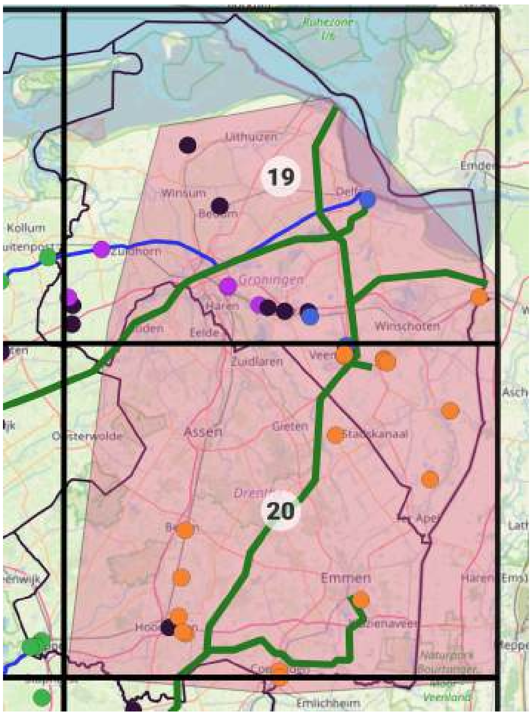


**Figure 7.3:** Research area 3

# 8

# Model setup

After selecting the research areas, the actual network modelling will be executed. This will be done using the Optimal Network Layout Tool (ONLT) developed by Heijnen (2025a). The model is adjusted to fit this study. These changes will be mentioned in the first section of this chapter. Following the explanation of the tool, the implementation of new transport options and associated cost functions are described in detail. After this, all assumptions used in the tool are described.

## 8.1 Adjustments to ONLT

For this study, the ONLT model has been adjusted to fit desired capabilities. The new model has been build and tested on an area in the Netherlands. This test-area has not been chosen by hard criteria, but based on first insights on interesting areas. By testing and developing the model first, the major errors are accounted for, so the modelling of the final research areas will execute relatively easily. The most significant changes are done in the cost function of the tool, which is described in section 8.3. Besides that, the file that delivers the input for the model has been adjusted to fit this study. The original file used simple coordinate notation, for example (40,8), which indicates the vertical and horizontal position. This has been modified so that the tool can read and include UTM coordinates, which provides the opportunity to include real-life coordinates. Finally, the implementation of new transport options and associated cost functions are the largest changes to the tool, but those are explained in detail in the following sections.

## 8.2 Implementation of new transport options

The biggest change to the existing model is the extra transport options that the model needs to account for. Most changes are implemented in the cost function, but some other changes are done as well. First, every edge is labelled with a category/transport mode. In the previous model, only one transport mode was possible, so this wasn't needed. This label is given to existing connections beforehand, and to new connections after the costs are calculated, where the cheapest category is assigned to an edge. For every transport option, a different edge colour is assigned to visually distinguish the transport options.

## 8.3 Cost function

The ONLT's standard cost-function isn't suitable for this specific study, because it focusses only on pipelines. In this study, three different transport options are used, which means that there are three different cost-functions needed. Besides that, the original cost function results in unit-less costs, where this study uses real costs in euros. The python code of the complete edge cost function is added in appendix B. Every cost function consist of capital costs (CAPEX) and operational costs (OPEX) and calculates the costs for one year, because demand data is used for one year. Below, the CAPEX and OPEX functions for every transport option are explained.

### 8.3.1 Pipeline costs

As mentioned in section 8.4, it is assumed that the pipeline backbone is already built. Because of that, the model doesn't consider capital costs for using existing pipeline connections. Only operational costs are taken into account when using existing pipeline connections. When new pipeline connections (Connections beside the pipeline backbone) are needed to be built to deliver hydrogen, the model does account for capital and operational costs.

For pipeline capital costs, a cost function is used from Hammond et al. (2024), already applied to hydrogen infrastructure networks. Different costs are taken into account, such as labour, material and other various costs which are needed to install a pipeline. Lafeber (2024) has converted the original function into a million euros, because the original function used pounds.

$$C^{\text{pipe}} = C^{\text{pipe}}_{\text{capex}} + C^{\text{pipe}}_{\text{opex}} \tag{8.1}$$

$$C^{\text{pipe}}_{\text{capex}} = 50833 \cdot L \cdot e^{0.0697} \cdot D + 297 \cdot D^2 + 71800 \cdot D + 54658 \tag{8.2}$$

$$C^{\text{pipe}}_{\text{opex}} = 0.04 \cdot C^{\text{pipe}}_{\text{capex}} \tag{8.3}$$

Equation 8.2 shows that capital costs depend on the length and diameter of a pipeline. The length of a pipeline is taken from the ONLT, but the diameter is not an output of the ONLT. For this study, it is assumed that the diameter of every pipeline is set to 0.508 meters, based on R. Yang et al. (2025). Yang mentions that hydrogen pipelines are commonly smaller than natural gas pipelines, which vary between 0.1 meters and 1.22 meters. A common size for hydrogen pipelines is 20 inches, or 0.508 meters. Because most companies have a relative low hydrogen demand, compared to the five main clusters in the Netherlands, this size should be sufficient to deliver hydrogen to all companies. Because the capital cost function takes pipeline diameters into account and the diameter is a fixed value, demand doesn't influence the costs of hydrogen pipelines. The costs of a pipeline can therefore only change by varying the length of an edge. In real-life situations, the diameter of a pipeline would be tailored to fit hydrogen demand, but to reduce complexity of the model, the diameter is set to a fixed value in this study.

| Symbol | Description | Value |
|---|---|---|
| $D$ | Pipeline diameter | $0.508$ m |
| $L_p$ | Pipeline lifetime | $40$ years |
| $f^{\text{pipe}}_{\text{opex}}$ | Annual OPEX factor | $0.04$ |

**Table 8.1:** Pipeline parameters

The operational costs (maintenance for example) are set to 4% of the capital costs per year, which is often used in literature (Hammond et al., 2024; Parolin et al., 2022). If an existing pipeline is used, the capital costs are also calculated to determine the operational costs. The capital costs are not considered in final costs in this case.

### 8.3.2 Road/waterway costs

Road and waterway costs are build up differently than pipeline costs. Because the waterway costs structure is almost equal to road costs structure (Busch et al., 2023; Parolin et al., 2022), this section will provide the formulas which are used for both transport options. Both transport options costs are based on the number of vehicles used and the length of transportation. Road or waterway specific values or formulas are mentioned separately. The vehicles used for road transport are trucks, and for waterways these are vessels. Both terms will be used further in this section for explanation.

The total costs of roads and waterways also consists of capital and operational costs, but aren't different for new or existing connections. For roads, it is assumed that no new roads have to be built, due to the fact that every industry location is already connected to a road in real life. This means that for every road connection used in the model, the costs of this connection can be calculated as if it is an existing connection. Therefore, the costs for new and existing connections are assumed to be equal. For waterways, only existing edges are considered, because it is quite complex to built new waterways. Therefore, it is assumed no new waterways are built in this model. In order to calculate capital and operational costs, a few parameters are used, which are provided below for trucks. The parameters for vessels will be mentioned later.

| **Symbol** | **Description** | **Value** |
|---|---|---|
| $Q_t$ | Truck capacity | $4324.9$ kg $\mathrm{H_2}$/truck |
| $c_t^{\text{var,exist}}$ | Variable costs (existing edge) | $0.0008029$ €/m |
| $c_t^{\text{var,new}}$ | Variable costs (new edge) | $0.0009973$ €/m |
| $L_t$ | Truck lifetime | $11$ years |
| $C_t^{\text{inv}}$ | Investment cost per truck | $860{,}000$ € |

**Table 8.2:** Truck parameters

In order to achieve the capacity of one truck in kilograms of hydrogen, the density of liquid hydrogen (70.9 kg/$m_3$ (Demaco Holland B.V., 2025)) is multiplied by the capacity of one truck (61 $m_3$, used in Busch et al. (2023). In order to calculate the variable costs for transporting hydrogen by truck, the variable costs are used from Busch et al. (2023). Those costs are based on fuel costs, fuel consumption, operational hours and kilometres driven per operational hour. A distinction is made between existing connections, which are highways, and new connections, which are regional ways. A higher driving speed on highways is assumed, compared to regional ways, increasing the total amount of distance driven in one operational hour, which finally leads to lower variable costs for existing connections, compared to new connections.

Now that the input parameters are clear, the equations for capital and operational costs are described in equation 8.4, 8.5 and 8.10. The operational costs consist of a fixed and variable part. The fixed operational costs are 2% of capital costs, taken from Parolin et al. (2022).
It is important to distinguish between existing and new edges in the network when calculating transport costs for roads. For existing edges, the total costs consist only of operational costs. The capital costs of vehicles are not included per edge, because a single truck or vessel can travel along multiple existing edges without requiring additional vehicles. Including capital costs for every existing edge would overestimate the total cost, as it would imply that a new vehicle is needed for each edge. In contrast, for new edges, the total costs include both operational and capital costs to account for the investment required to provide sufficient transport capacity along routes that do not yet exist. This approach ensures that the model correctly represents vehicle utilization and the incremental costs associated with expanding the transport network.

$$C^{\text{rw}} = C^{\text{rw}}_{\text{capex}} + C^{\text{rw}}_{\text{opex}} \tag{8.4}$$

$$C^{\text{rw}}_{\text{opex}} = C^{\text{rw}}_{\text{fix}} + C^{\text{rw}}_{\text{var}} \tag{8.5}$$

$$C^{\text{rw}}_{\text{fix}} = 0.02 \cdot C^{\text{rw}}_{\text{capex}} \tag{8.6}$$

Because trucks, vessels and pipelines all have a different lifetime, this needs to be accounted for. Because pipeline has the longest lifetime of 40 years (Lipiäinen et al., 2023), this lifetime is set as the calculation period. This means, enough trucks or vessels have to be bought to achieve 40 years of lifetime in total. The number of trucks or vessels that have to be bought is used in the capital and operational costs calculation (equation 8.7.

In order to know how many trucks or vessels are needed, the demand needed in a specific edge is divided by the capacity of a truck or vessel (equation 8.8). This number of loads is rounded up, because in case half the capacity of a truck is needed, still one complete truck needs to be available. Next, the amount of trucks/vessels needed to transport the demand is calculated (equation 8.9, assuming a truck/vessel can be on the road/water one time a day, for every day of the year.

$$N_{\text{rep}} = \frac{T}{L_v} \tag{8.7}$$

$$N_{\text{load}} = \frac{Q_d}{Q_v} \tag{8.8}$$

$$N_v = \frac{N_{\text{load}}}{365} \tag{8.9}$$

$$C^{\text{rw}}_{\text{capex}} = \frac{N_{\text{rep}} \cdot N_v \cdot C^{\text{inv}}}{T} \tag{8.10}$$

Now that the capital costs are calculated, the operational costs can be calculated as well. The formula for fixed operational costs is mentioned in equation 8.6. The variable operational costs are mainly determined by the length of transportation for a specific edge and the amount of loads needed to supply hydrogen demand. Equation 8.11 shows that the length of and edge is multiplied by two, because it is assumed a truck or vessel needs to return to its starting point, therefore travelling the distance twice. The specific variable costs for trucks per kilometre are dependent on if the edge is new or existing, as can be seen in the mentioned parameters for trucks.

$$C^{\text{rw}}_{\text{var}} = 2 \cdot L \cdot c^{\text{var}} \cdot N_{\text{load}} \tag{8.11}$$

**Waterway/vessel costs**

Waterways have some different parameters, compared to road transport. The parameters are showed below and are used in the equations described earlier in this section. The capacity of a vessel is determined by using the same mass density for liquid hydrogen as for trucks, which is multiplied by the volume of a vessel, which is 5000 $m_3$ (Busch et al., 2023). The parameters show that vessels have 82 times the capacity of a truck. Combined with a longer lifetime, it is expected that a much smaller amount of vessels is needed to deliver the same amount of hydrogen, compared to trucks. However, both variable costs per kilometre (20 times higher) and investment costs (4.8 times higher) are much higher compared to trucks.

| **Symbol** | **Description** | **Value** |
|---|---|---|
| $Q_v$ | Vessel capacity | $354{,}500$ kg $H_2$/vessel |
| $c_v^{\text{var}}$ | Variable costs per meter | $0.01599$ €/m |
| $L_v$ | Vessel lifetime | $25$ years |
| $C_v^{\text{inv}}$ | Investment cost per vessel | $4{,}100{,}000$ € |

**Table 8.3:** Vessel parameters

## 8.4 Model assumptions

In modelling, the goal is to get as close as possible to real-life situations. Still, a model will never fully reflect everything that happens in reality. To make the model work, certain assumptions had to be made. In this section these assumptions are described, so it is clear under which conditions the results of the model should be seen. The generic assumptions are mentioned in Heijnen (2025b), the assumptions mentioned here are applicable for this specific study. Cost assumptions are provided in section 8.3.

### Transport options assumptions

1. All transport networks are already built and fully available to use, including the future pipeline network.
2. New edges can only be build from existing road or pipeline backbone edges. Waterway edges can only be existing, not new.
3. Liquid hydrogen is assumed for trucks and vessels, gaseous hydrogen for pipelines. No transformation costs are calculated to suit these different sorts of consumers/producers
4. The hydrogen backbone (existing connections) can exist of new pipelines and reused natural gas pipelines, as mentioned in chapter 1.3. The model considers the complete hydrogen backbone as existing connections, without differences between natural gas pipelines or new pipelines in the hydrogen backbone. If new pipeline connections are needed in the network, besides the existing hydrogen backbone, it is assumed that these are new pipelines and not existing natural gas pipelines.
5. One truck or boat can supply multiple industry locations at once, if total demand of those locations is smaller than or equal to the capacity of the vehicle. If the capacity needed is larger than the capacity of one truck, a new truck/boat is needed to invest in.

### Network assumptions

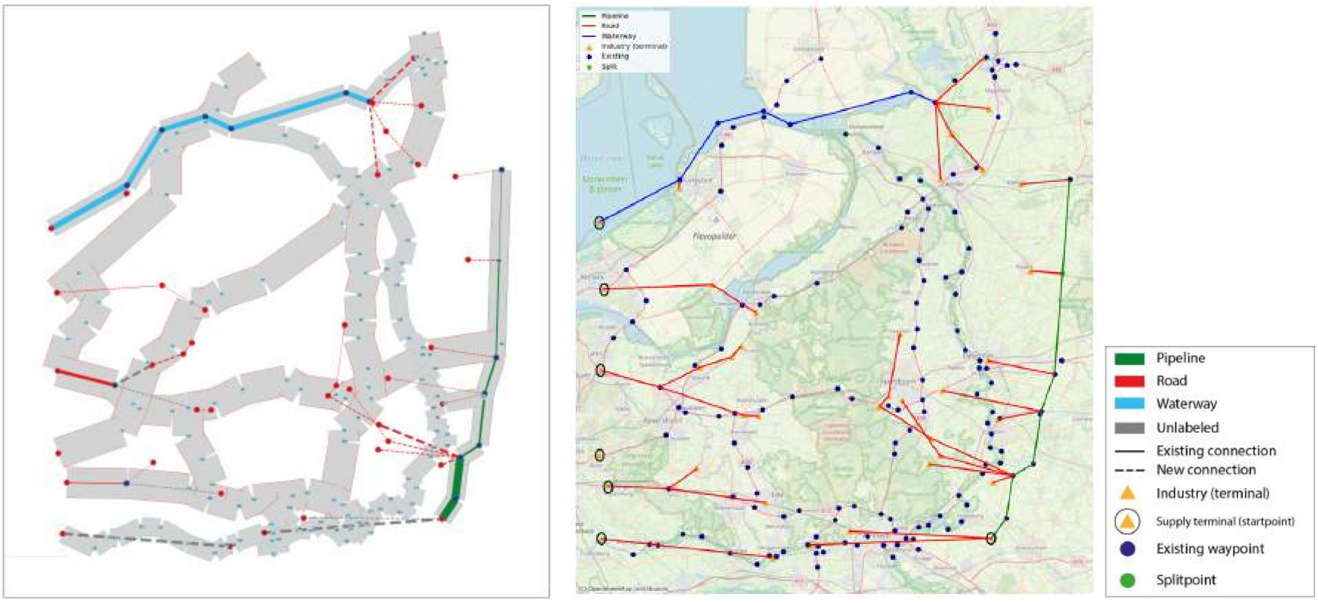
1. The model assumes all edges are connected to the exact location of a consumer/industry location.
2. Every consumer can receive both forms of hydrogen, gaseous and liquid.
3. There is no limit to hydrogen supply of any transport option. Supply can be scaled up if needed.
4. The model simulates a yearly demand in one time step, assuming all demand for one year is delivered at once. More specifications on this assumption can be found in the cost function explanation of trucks and vessels.
5. The model draws edges direct to the consuming node, no obstacles of any form are taken into account. Obstacles are identified as a significant part of network modelling in the theoretical background (chapter 2), but do to time limits this is left out of scope.
6. For every transport option present in a research area, one or more supply nodes are chosen. Every supply node is located at the border of a research area. It is assumed that supply can be provided from different locations:
   (a) Waterways: Harbour of Rotterdam, IJmuiden, or Eemshaven. If research areas don't include one of those harbours, the waterway coming from that direction is chosen as the supply node.
   (b) Roads: Supplied from direction of the same harbours, or roads leading from neighbouring countries (Simulating import of hydrogen)
   (c) Pipelines: Same assumptions as for roads, meaning harbours and neighbouring countries are chosen as supply points.
7. New connections can be built from any place on an existing connection. Split points for the optimal location of these new connections are determined by the model.
8. The ONLT has the option to include routing networks, which are networks/nodes which are obligatory to use. No routing networks are used in this study.

# 9
# Results

This chapter highlights the ONLT results for every research area. The results will show how every transport option is used in different environments. This is supported by the visualisation of every network and data on total costs and length used for every transport option. Appendix C shows all resulting networks for every step, for every research area. The discussion chapter will explain how these results should be interpreted. Figure 9.1c shows the legend for all maps used in this chapter. Important to mention is that some connections might be coloured grey, being unlabelled. However, the model did label these connections as road connections, but in some way the final visualisations didn't get this label. The grey connections are road connections, which should be taken into account when analysing the figures.

## 9.1 Research area 1 - Low demanding industries with large distances to supply networks

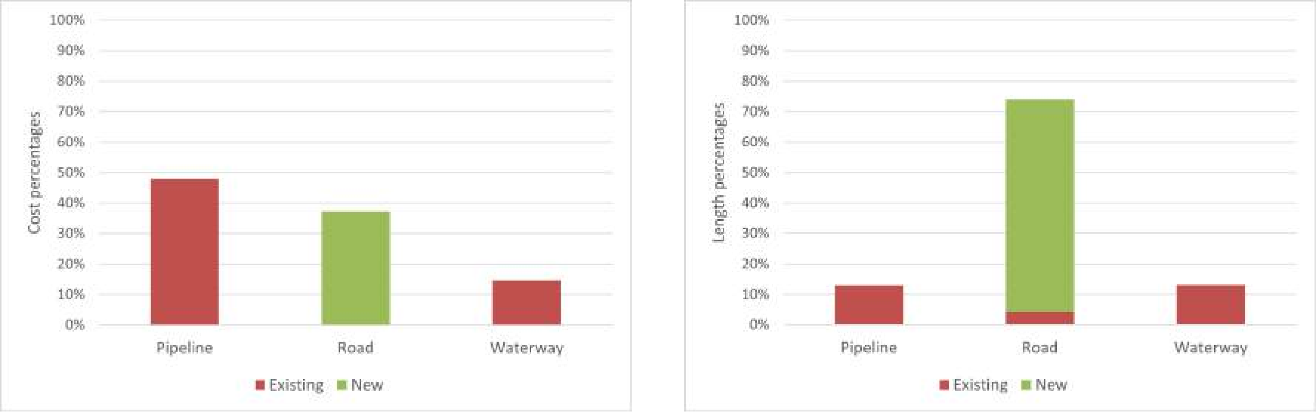
Research area 1 is a relatively large area, when looking at the number of industrial locations included in the model. This area focusses on the industries which are far away from both waterways and the pipeline backbone, which are represented in cluster 3 and 4. Figure 9.1 visualises the final networks created by the ONLT, including a network drawn on Openstreetmap for better interpreting the results in the actual area. As can be seen in figure 9.1b, the main supply routes are far away from the industries. This is probably due to the nature area 'De Veluwe' in the middle, which is visualised through a darker green colour in figure 9.1b. Industries are locates at the borders of this area mostly.



**(a)** Final network from ONLT **(b)** Final network with Openstreetmap **(c)** Legend

**Figure 9.1:** Results Research Area 1

The final network uses pipelines and waterways as main supply routes for the upper and right part of the area. Because pipelines and waterways aren't present at the left part of this area, multiple road supply routes are used in this area. However, when final distance delivery transport modes are analysed, trucks are used to arrive at the actual industry locations. The network therefore shows one main branch for waterway and pipeline, and multiple branches for road transport. This confirms that waterways and pipeline have larger capacities and are used to transport large amounts of hydrogen. When smaller amounts are transported for company specific delivery, trucks are cheaper to use. Interesting to see is that new road connections are modelled relatively close to each other, where it might be expected that these connections are combined, to reduce the amount of trucks needed and therefore reducing costs. An example of such a situation is visible in the bottom-right part of figure 9.1a. The discussion chapter will dive deeper into the behaviour of the model for such situations. Figure 9.2 visualises this as well, where it's clear that pipeline and waterways only use existing connections, and roads are mostly new connections. This means that the roads outside of highways are needed in order to supply hydrogen for the final distances to companies. Besides that, figure 9.2 shows the relation between share of costs and length for every transport option. Interesting to see is that pipelines are responsible for almost 50% of all costs, but only responsible for 13% of total length of used connections, resulting in final costs/km of €52.68. Road costs on the other way, are much cheaper (7.13 €/km), as are waterways (15.71 €/km). The average costs/km for this area are 14.12 €/km.



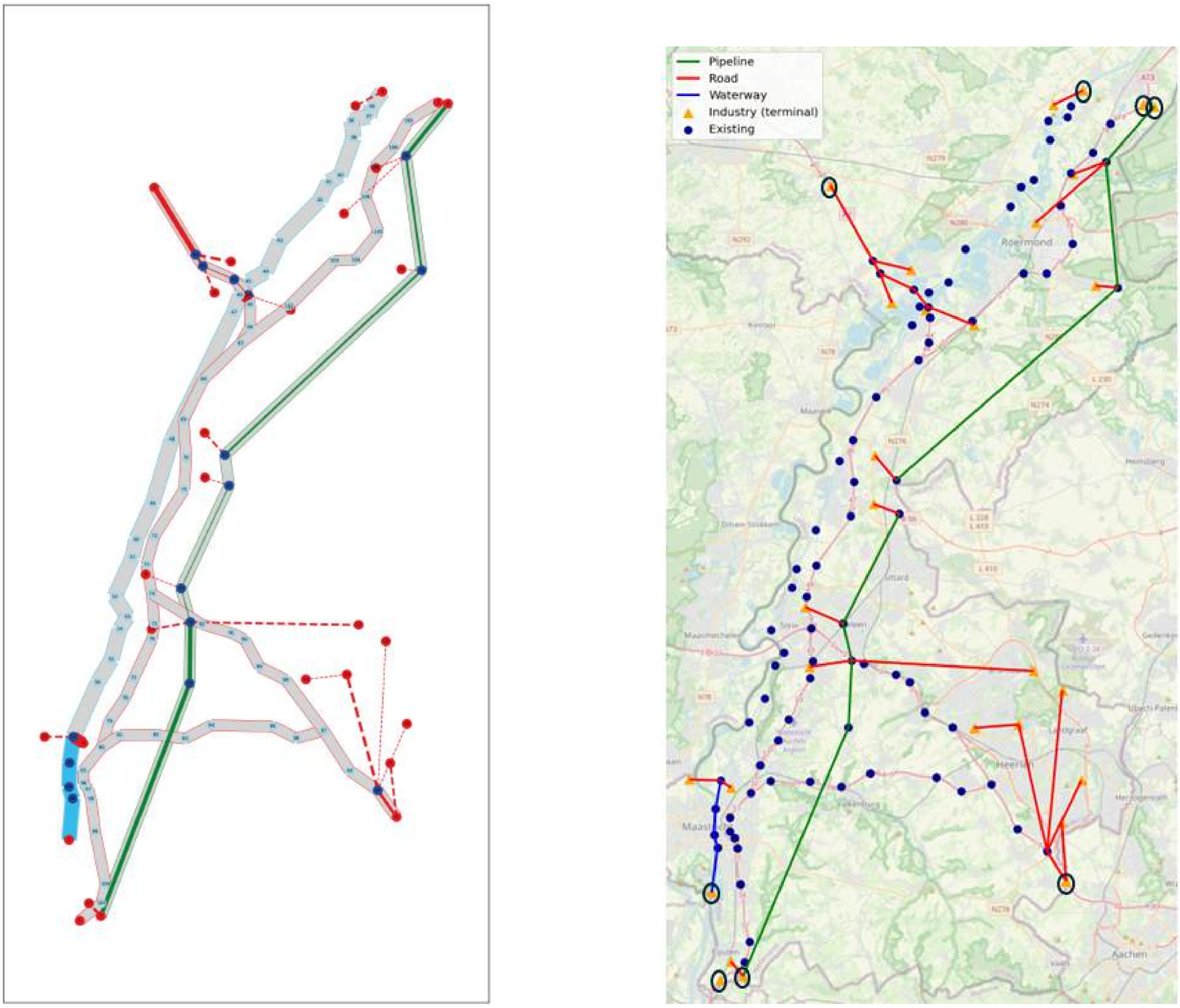
**(a)** Cost percentages

**(b)** Length percentages

**Figure 9.2:** Share of length and cost for transport options - Research area 1

## 9.2 Research area 2 - High demanding industries close to supply networks

The second research area focusses on high demanding companies, close to all supply networks. Figure 9.3a shows that all existing connections from different transport options are highly present throughout the whole research area. The ONLT therefore uses significantly more existing connections of the pipeline backbone, in order to transport large volumes of hydrogen. Again, road connections are mostly used outside of highways (new ONLT-connections, existing roads in real life), therefore delivering the hydrogen for the final kilometres to companies.

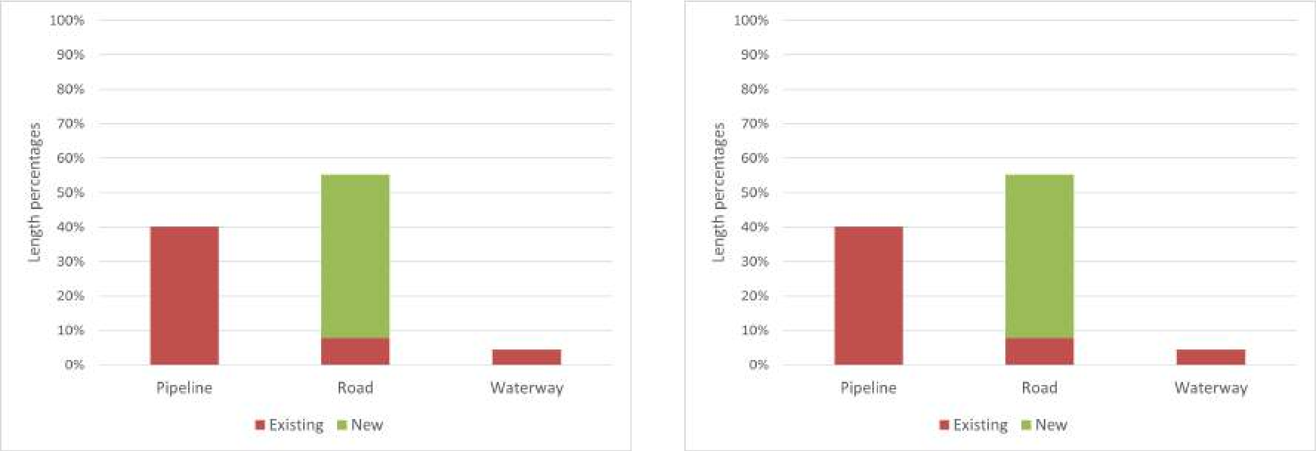


**(a)** Final network from ONLT

**(b)** Final network with Openstreetmap background

**Figure 9.3:** Results Research Area 2

Figure 9.4 shows that the share of pipeline-length is larger compared to research area 1, where the share of costs didn't increase as much as length. Costs per kilometre almost didn't change for pipelines however (52.67 €/km), whereas it did change for roads and waterways. The road costs/km have risen to 19.89 €/km, which can be explained by the increasing amount of trucks that's needed to supply larger amounts of hydrogen. Waterways costs have increased even more, resulting in 90.38 €/km, showing that waterways have transported large amounts of hydrogen as well. The average costs/km for this area are 36.28 €/km.



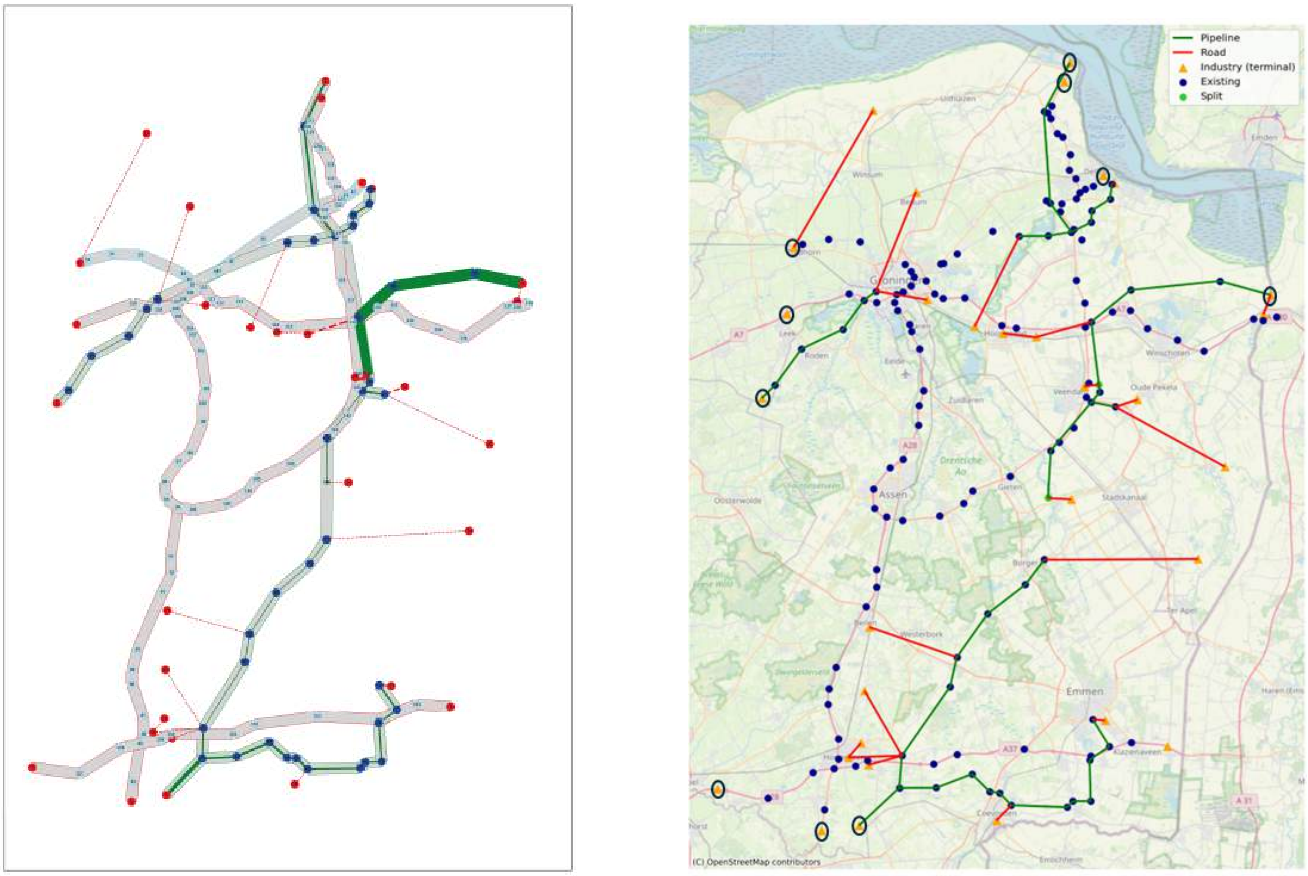
**(a)** Cost percentages

**(b)** Length percentages

**Figure 9.4:** Share of length and cost for transport options - Research area 2

## 9.3 Research area 3 - Medium demanding industries close to pipelines, far from waterways

The third research area consists of industries close to the pipeline backbone, but far away from any waterway. Therefore, the pipeline backbone is the main supply network for large amounts of hydrogen. Demand of the companies present is average.



**(a)** Final network from ONLT

**(b)** Final network with Openstreetmap background

**Figure 9.5:** Results Research Area 3

The resulting networks figure 9.5 show that this area imports a large amount of hydrogen from Germany, in the top right corner of this area. This is done by pipelines, as is the supply for the southern part of this area. The existing pipelines supply a large amount of hydrogen, which may be the case due to the lack of waterways, excluding the possibility to transport large amounts of hydrogen. Because pipeline is used for more than 50% of total edge length, the pipeline costs account for 85% of total costs (figure 9.6), resulting in similar average costs per kilometre for this research area (€35.83/km) as for research area 2. Trucks are used again for final connections to the companies, resulting in the use of only new road connections and no existing road connections. Pipeline costs/km result in €54.15 and road transport result in €12.25.

**(a)** Cost percentages **(b)** Length percentages

**Figure 9.6:** Share of length and cost for transport options - Research area 3

# 10

# Discussion

This chapter will discuss the results of this study by interpreting the results of the final model results. Interpreting these results will clarify why certain choices were made by the model, which is determined by the input that is made in earlier steps in this study. Therefore, the results from previous steps such as clustering and selection of research areas will be taken into account. After that, the limitations of the study are discussed.

## 10.1 Interpretation of results

**Use of transport options**

One goal of this study is to implement different transport options in the ONLT and evaluate how these options are used, compared to only using pipelines. The results in chapter 9 showed that only road transport is used, if new connections are needed to be made. This applies for every research area. This means that pipeline transport is too expensive to build a new connection and waterways aren't available as an option to build a new connection. In this case, pipelines are extremely expensive compared to the use of trucks, especially for the small amounts of hydrogen which are transported in these cases. This is because the regional industry are small companies on average, resulting in small amounts of hydrogen demand (section 5.1.1). Besides that, trucks on roads are very flexible, due to the high density of the road network, as discussed in 5.1.2.These results line up with expectations, because the cost functions defined in section 8.3 shows that trucks are less expensive than pipelines.

However, for the transportation of large amounts of hydrogen, the benefits of pipeline and vessels are becoming clear. The large capacity of vessels and pipelines provide cheaper opportunities for transporting hydrogen in large amounts, if it is assumed the pipeline backbone is already in place and only operational costs are calculated for. This means, that those two options are used as a 'backbone' for hydrogen transporting. This results for all research areas in the use of existing waterway and pipeline connections, which are combined with road branches, in order to distribute the final distance to the actual company. Therefore, roads are mainly used as a distribution network, whereas pipeline and waterways are used as transmission networks.

**Meaning of different costs/km**

The results showed different costs/km for every transport option in the research areas. This has been calculated by dividing the total costs for a specific transport option by the total length of the transport option. For waterways and roads, the costs/km can vary due to the increase or decrease of the amount of vehicles that is needed, or the transportation length. For example, if large amounts of hydrogen are needed to be transported over a short distance, the costs/km are relatively high. It is important to keep this in mind, because it doesn't necessary mean that options with high costs/km are a bad transport option. This metric therefore needs to be combined with the amount of hydrogen that is transported for that distance as well to gain an honest view of the results.

On the other hand, this metric can be used to compare different research areas, by using the average costs/km for all costs and all distances of transportation in one research area. To draw conclusions for the different clusters which were present in the research areas, the average costs/km can be compared to gain insights in which type of companies cause higher/lower costs/km, which will be done in section 11.1.5. This will support to predict which areas will probably be more or less expensive to provide regional hydrogen supply for.

**Influence of research area**

The boundaries of research areas have been chosen based on visual checks. These checks evaluated if possible companies close to the border of the tiles were included as well. This has been done to ensure that companies which might influence routing of the network are included in the model. The same goes for including parts of a transport network, to make sure the model can use all three transport options in some way. Otherwise, the model would make decisions which wouldn't be realistic. However, there will always be limitations in the amount of context that can be taken into account, when choosing such research areas. It is therefore important to mention that choosing boundaries of research areas differently, might lead to different results. Including specific data-points or not, can determine if a specific connection is made, therefore possibly influencing total distance for a specific transport option, leading to change of costs. Ideally, the Netherlands would be modelled completely to ensure full context in the model.

**Meaning of industry-clusters**

The results from chapter 6 showed that six different clusters were made for regional industry. By identifying these clusters, suitable transport options can be assigned to every specific company, increasing the possibility to implement these solutions earlier. This has been concluded, based on the current dataset of companies, which has been used for this study. This dataset might, and probably will, change over the coming years. Companies might disappear, new companies can appear, and therefore this set of clusters isn't the standard for the future necessarily. This means that the clusters are needed to be revised occasionally, to assure that hydrogen supply is matched properly to the companies present.

**Use of DB-SCAN clustering**

In order to save computation time and reduce complexity, the industry locations have been clustered on location, combining locations within a range of 2000 meters. The results from this clustering showed a reduction from 291 locations to 217 locations, 'removing' 74 locations (25%) of the original set of locations. This seems like a significant reduction, but when zoomed into the research areas, the reduction is smaller due to smaller areas than the complete country.
It is however expected that the ONLT provides similar results when not using the DB-SCAN clustering. Because industry locations close to each other are combined, those locations will probably be in the same cluster, established by K-means clustering for distance and demand. The only relevant difference between DB-SCAN clustered locations, can be their hydrogen demand. However, because the demand of combined locations is summed up for DB-SCAN clustering, the demand is still accounted for, or even optimized compared to not using DB-SCAN clustering. DB-SCAN will increase the probability that a transport option is chosen which suits the surrounding area of a location best, by combining demand of nearby locations. Now, the model considers the combined demand of an area and can supply this by a transport option suitable for larger amounts of hydrogen, instead of multiple smaller supply connections, probably increasing costs (The more connections made, the higher the costs).
It can be concluded that DB-SCAN might not decrease the complexity and computation time of the model with large numbers, but it doesn't impacts the results of the model in a negative way as well. Industry locations combined by DB-SCAN clustering are likely to be in the same K-means cluster, due to similar distance characteristics, therefore not undermining the final statements that will be done for the different K-means clusters.

## 10.2 Limitations of this study

The model and results presented in this study are subject to several limitations. These are mainly related to the transport options that were included, the way the cost function was designed, and the way the network is represented. It is important to keep these in mind when interpreting the results.

**Demand data**

The demand data used in this study is based on the natural gas consumption of companies in 2024. This demand was converted into hydrogen demand by applying an energy equivalence factor and assuming that 50% of the energy demand would be fulfilled by hydrogen, based on expectations in literature (Parolin et al., 2024). However, this dataset is probably not fully representative for the year in which the hydrogen network is assumed to be operational (around 2033). Because this study models for that year, 2033, both the total demand and the share of hydrogen in the energy mix may be significantly different from the assumptions applied here. This creates uncertainty in the model outcomes, as the actual demand for hydrogen in 2033 may diverge from the demand patterns that were used in this study. This could influence the capacity needed to transport to specific industries, thereby maybe changing the optimal network layout. If smaller amounts of hydrogen are needed, more road transport might be used compared to waterways and pipelines and vice versa.

Besides that, the replacement percentage of 50% is an average of all companies. In reality, every company might have a different replacement percentage, altering the actual demand of the companies. If this would be taken into account in this study, demand changes will influence the use of the different transport options. Some company nodes which have a high demand right now, might have a low demand if accounted for actual replacement percentages and vice versa. This will change optimal network solutions.

**Available transport options**

The second limitation is related to the available transport options. Existing natural gas pipelines were not included in this study. Although their potential reuse is frequently mentioned in literature and the Dutch hydrogen backbone project (Gasunie, 2024), the technical and financial uncertainties are too large at this stage, and including them would have required a much more detailed analysis. Datasets of natural gas pipelines are relatively large and are therefore time-consuming to use in the right way. Secondly, the pipeline backbone that is used in the model is only a schematic representation, because not all locations are exactly known, because the pipeline backbone is still in development. This means that it shows general connections, but not the exact location of pipelines in reality.

**Cost function**

The cost function includes some simplifications. For example, no $CO_2$ taxation was included, which would have increased costs for road transport in particular. However, it is expected that this wouldn't have impacted the choice of transport options largely, because the differences between transport options in costs are relatively large. It may reroute the use of roads slightly, also depending on the share of fossil-fuelled trucks in use. It can be expected that the share of eco-friendly trucks is increased in the future, reducing the $CO_2$-emissions and thereby the effect of $CO_2$ taxations.

Secondly, the pipeline diameter was fixed, while in practice smaller diameters could be applied for smaller flows, which would reduce costs. Furthermore, the model does not include transformation costs between liquid and gaseous hydrogen, or when switching between transport modes. This could make the modelled costs lower than they would be in reality, especially in cases where different modes are combined. This is often the case, when pipelines or waterways are used to transmission large quantities of hydrogen, followed by the distribution by trucks, requiring a change of hydrogen form.

Because this original cost-function was replaced, it is possible that the construction and placement of splitting points in step 3 of the ONLT (section 3.5) isn't optimal. The original function uses a factor $\beta$, which denotes the scale-factor for the costs of pipelines and is default set at 0.6. The original function is $C = l \cdot q^{\beta}$. The use of $\beta$ causes that costs will not increase linearly if capacity increases linearly, but costs will increase less than proportionally (with diminishing returns). This affects the location of splitting points, because these are chosen based on the original cost function. Now that the cost functions for all transport options have changed, and $\beta$ isn't included any more, it's not sure if the locations of splitting points are still optimal. The results showed, however, that splitting points are almost not used, except for a few in research area 1 and 3. Therefore, this limitation will not significantly influence the outcomes of this study, but can be further investigated.

### Network design

Finally, there are some limitations in the way the network is designed. The model is designed to follow existing connections where possible. However, in some cases it still creates new direct connections that run parallel to existing ones, instead of using them. Due to the assumption that new connections can be build anywhere, new connections can be of a shorter distance than following the existing connection. An example of such a situation is displayed in figure 10.1. The blue dots represent the waypoints of existing highways. The bright, red edges are the connections drawn by the model. As can be seen, it runs parallel to the waypoints of existing highways. This leads to less realistic network layouts. In real-life, the road-network that is already present would be used, leading to longer distances, meaning higher costs for road connections.
10.1.

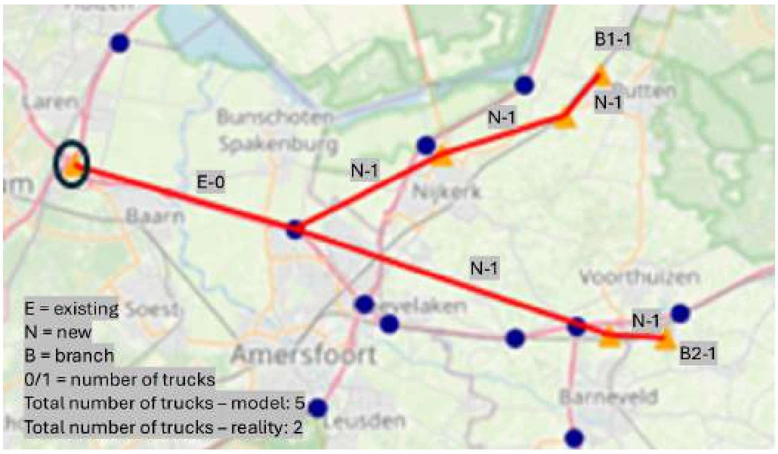


**Figure 10.1:** Example of number of trucks limitation

Another issue is the number of trucks that is counted for new road connections. In the current model, every new road edge is assumed to need a new truck, while in reality the same truck could serve multiple new edges if they are connected. As a result, the number of trucks (and therefore costs) can be overestimated. Although some improvements were made to avoid the worst cases of overestimation, this part of the model is still a simplification. Figure 10.1 is used for this limitation as well. Every edge-label shows if it's an existing or new connection, and the number of trucks needed to transport the assigned capacity for this specific edge. The representation of number of trucks is presented as the model calculates these numbers at the moment. As mentioned and can be seen in the figure, every new connection uses a separate truck, increasing costs for every connection by including investment costs for a new truck, for every new edge. In this example, this means that the investment costs for five separate trucks are taken into account. The ideal situation, however, calculates the amount of trucks needed for one connected branch, of which are two in this situation (B1 and B2). The number of trucks needed for these separate branches is one truck for each branch, and a total of two (calculations for number of trucks are given in section 8.3). This shows that the actual number of trucks needed in this case is two instead of five. The costs in the model are therefore significantly higher than it would be in real life. This is also the case for waterways, where every connection requires a new vessel. This is unrealistic, because all waterway edges are connected to each other, and therefore it can be assumed that the same vessel will travel along those edges, instead of a separate vessel for every edge. A small benefit to this limitation is the fact that it applies to both roads and waterways. If this was fixed, costs of both transportation options would decrease, preserving the cost ratio between the road and waterway transport options. However, there would be cost-differences compared to pipelines, because the pipeline cost stay the same. This could lead to a higher share of waterway use in some areas.

In addition, no restrictions were applied for obstacles in the landscape such as nature reserves, residential areas or other land-use constraints. As a result, some connections may appear more straightforward in the model than they would be in practice.

Last, because obstacles are not accounted for in this study, all new road and pipeline connections drawn by the model are constructed in a straight line, from one point to the other. This is due to the assumption made in section 8.4, that new edges are directly connected. In reality, this wouldn't be the case. Therefore, the model isn't completely fair with regard to the length of new road and pipeline edges. It is expected that in real life, the new edges would be longer, thereby increasing costs for these transport options. For this study, it means that length of edges aren't totally represented, but could increase by a factor between 1.2 and 1.4, which is mostly used in literature (Weißenburger et al., 2025). This would increase costs for both road transport and pipeline transport, making it more expensive compared to waterway transport.

## 10.3 Verification and validation

To increase the robustness and reliability of the ONLT results, several verification and validation steps could be performed in future work:

**Pipeline diameters**
In this study, one fixed pipeline diameter was assumed. For validation, the model could be tested with different diameters, to better reflect the economies of scale in pipeline construction. This would also allow for comparison with the original cost function (including the $\beta$ parameter), providing insight into the sensitivity of the network design to this assumption.

**Variation in truck and vessel parameters**
Truck and vessel capacities, as well as operational costs, could be varied to test the influence on the choice of transport modes. Additionally, including $CO_2$ taxation on fossil-fuelled vehicles would allow for a more realistic estimation of future transport costs. These sensitivity tests would demonstrate whether road transport remains dominant under different boundary conditions.

**Clustering approach**
The effect of DB-SCAN clustering on the model outcomes could be validated by running the model without clustering and comparing results. This would verify whether the clustering step indeed reduces computation time, but still deliver the same results, as argued in this study.

**Research area boundaries**
Finally, the sensitivity of the results to different research area boundaries could be checked. By slightly shifting or enlarging the areas, it can be verified whether the outcomes are robust to small contextual changes.

These steps would provide confidence that the model results are not overly dependent on specific assumptions or simplifications, and would improve the validity of the conclusions drawn from this study.

# 11

# Conclusions and recommendations

The goal of this thesis was to optimise hydrogen supply networks for regional industry in the Netherlands. By identifying the regional industry, categorising them into similar groups, and finally modelling the optimal network layout for three different research areas, this study aimed to answer all research questions. This final chapter concludes the thesis by answering all subquestions, leading to the answer of the main research question. This is followed by a discussion of the scientific relevance, describing how this study contributes to the existing literature. In addition, the social relevance is addressed. Finally, recommendations for future work are presented.

## 11.1 Answering research questions

### 11.1.1 What are relevant characteristics of regional industries with regard to transport optimisation?

This study aims to optimise hydrogen supply for regional industries, thereby excluding large industrial sites such as those in Chemelot, Rotterdam and IJmuiden. All regional industries are grouped under the name 'Cluster 6'. Due to the large number of companies in this cluster, it is crucial to know what types of companies this study focuses on, and more specifically how their characteristics affect the cost optimisation of hydrogen supply. Therefore, the first subquestion focuses on identifying relevant characteristics that influence the costs of a hydrogen supply network. This was done by evaluating multiple academic literature sources and specific Cluster 6 reports.

From the literature review, it is concluded that demand and distance to any supply network are mentioned in almost every evaluated report. The larger the hydrogen demand, the greater the capacity of the transport option needs to be, thereby increasing costs. Distance to supply affects the length of transport that must be executed, which influences the amount of fuel needed or the investment required for pipeline length. Besides demand and distance to supply, several sources also highlighted the impact of obstacles, different transport options, and the type of industrial processes. These factors can increase realism (Hammond et al., 2025), provide flexibility in transport (Busch et al., 2023), or determine the level of hydrogen demand (Namazifard et al., 2024).

Cluster 6 reports confirm the range in demand and distance to transport networks as relevant factors that influence the feasibility of providing sufficient hydrogen supply. These are also the challenges Cluster 6 faces regarding the reduction of fossil fuel use and the switch to green energy sources. Sufficient infrastructure is more difficult to achieve due to larger distances to hydrogen supply networks and the wide geographical spread of industries. The literature review showed that hydrogen demand and transport distance are the most commonly used factors when optimising hydrogen transport. These factors were therefore used in the following steps.

### 11.1.2 How can the relevant characteristics be identified and prepared for clustering?

After determining the relevant characteristics, the dataset for all industry locations was completed by identifying the specific location and sector for every company, leading to demand and distance to different transport networks. Analysis of the different sectors in Cluster 6 showed that the food and building materials sectors are the largest in terms of company numbers. In total, 17 different sectors are present in Cluster 6, which demonstrates the wide variety of companies.

Matching demand to these sectors showed that a large number of companies does not necessarily imply a high demand. Smaller sectors can account for a large share of hydrogen demand. For example, the chemical sector consists of 7% of all companies, but accounts for 39% of hydrogen demand. The demand data therefore supports the identification of high-demand locations. However, the data is based on 2024 natural gas demand and could change in the future, depending on various factors. This uncertainty is also mentioned in section 10.2. Secondly, the distance to transport networks was determined. Research showed that different transport options are used in literature, leading to the selection of three transport options. First, pipeline networks were included, being the commonly used option for large distances and quantities of gaseous hydrogen. The average distance to the hydrogen network for all identified company locations in the Netherlands is the highest (11.60 kilometres). Secondly, waterways were used to transport liquid hydrogen over water, providing an alternative option for large quantities. The average distance to waterways is slightly lower, at 8.53 kilometres. Finally, the Dutch road network (highways only) was included for transporting liquid hydrogen. This option provides high flexibility, being able to deliver (small amounts of) hydrogen to every company on already existing roads. The average distance of 3.14 kilometres shows that the highway network is close to all industries, not even taking regional roads into account. Therefore, it is assumed that roads are available to every industry location.

Variety in both the number of companies, demand, and distance to transport networks can be concluded from this subquestion. A dataset was prepared and is ready to be categorised into similar clusters.

### 11.1.3 What characteristics-based clusters of regional industry can be identified?

By clustering the company locations in two ways, the data was prepared for both the final model and the ability to analyse specific types of companies with regard to hydrogen supply transport. First, DBSCAN clustering showed that various locations could be grouped if they were within 2000 metres of each other. By summing up demand and taking the geographical average location, a new data-point representing the clustered locations was created. This way, the total demand was still accounted for, while reducing computational complexity and model run time. Secondly, using a six-step approach, the industry locations were clustered into six different groups, applying K-means clustering. This was done based on hydrogen demand, distance to the pipeline network, and distance to the waterway network. By identifying and excluding outliers, the dataset was well-prepared for clustering. The final number of clusters was determined by evaluating the elbow plot, silhouette score, and the ability to describe each cluster. The evaluation showed that six clusters performed best for these criteria, providing good performance on both the elbow plot and silhouette score. Furthermore, all six clusters can be clearly described and differ sufficiently from each other. Setting the number of clusters lower or higher would have reduced performance and the ability to meaningfully distinguish different types of companies.

### 11.1.4 What are suitable research areas, where a specific regional industry cluster is strongly represented?

By selecting three different research areas, this subquestion showed which areas are most relevant for modelling a hydrogen supply network, based on the presence of specific categories of industry. Using a grid-analysis, three areas were selected based on the presence of specific clusters. The first area allows the analysis of companies far away from both the pipeline and waterway networks, or only far from pipelines. The demand of these companies is low. This area is interesting because it provides insights into situations where supply options with large capacity are limited, but demand is low. In such a situation, road transport is expected to dominate. The second area includes companies with characteristics opposite to research area 1. Companies are close to all supply options and have either high or low demand for hydrogen. This area will show which transport option is used most when all options are easily accessible. The final research area consists mainly of cluster 5 companies, with some cluster 0 companies. Cluster 5 companies have a moderate hydrogen demand, are close to the pipeline network, and far away from waterways. This area therefore analyses how the model performs when pipelines are the only available transport option with large capacity, and how this is reflected in final costs.

This research question provided insights into relevant research areas, based on actual characteristics in the Netherlands, thereby increasing realism for the final model. By choosing these areas, specific types of companies can be evaluated regarding their need for hydrogen supply. Narrowing the scope to three smaller areas instead of the whole country also reduced complexity and computation time.

### 11.1.5 How are different transport options used in a cost-efficient network model for different research areas?

The results from the ONLT show that the cost-optimal networks are designed differently for the specific conditions of each research area. One clear pattern is visible in all research areas: waterways and pipelines are mainly used as backbone connections for transporting large amounts of hydrogen, while road transport is used primarily for last-mile connections, distributing hydrogen from larger supply connections such as waterways and pipelines.

In research area 1, where industries are far from both waterways and pipelines, truck transport dominates. This highlights the flexibility and low costs of small-scale connections. This is reflected in the average costs/km for this area, which are €14.12/km. Research area 2 shows a larger share of pipelines as transport option, due to higher-demand companies located close to all transport networks. The short distances to all options reduce the amount of road transport needed, because large parts of the hydrogen supply can be transported using existing pipelines. Road transport is again used for the last-mile deliveries. Due to the increased use of pipelines, the average costs/km rose compared to research area 1, reaching €36.28/km. Research area 3 clearly lacks waterways, thereby increasing the use of pipelines for transporting large volumes of hydrogen. Road transport is again used for final deliveries only. The average costs for this area are similar to research area 2, at €35.83/km.

Overall, the model demonstrates clearly that the use of different transport options depends strongly on demand size and distance to supply networks. Pipelines and waterways transport large quantities of hydrogen but increase average costs. Road transport, on the other hand, is highly flexible thanks to the dense road network, enabling low-cost deliveries. When road transport is the dominant option, average costs are significantly lower compared to scenarios relying on pipelines and waterways. The differences between the three research areas show that no single transport option is universally dominant; instead, the optimal mix is determined by the spatial and demand-related context of each area.

### 11.1.6 How can hydrogen supply transport for Dutch regional industry be cost-optimized?

This section answers the main research question, as formulated in the title of this section. The results show that the starting point for optimising hydrogen supply transport lies in understanding what type of industries are present, and how their characteristics influence the design of a cost-optimal network. By identifying and categorising industries based on their hydrogen demand and distance to existing supply networks, a structured overview of the diversity within regional industries could be created. This categorisation makes it possible to distinguish between different clusters of industries, which together represent the wide variety of companies present across the Netherlands.

Based on this categorisation, research areas can be selected that represent different types of regional industries. These areas provide insight into how the network model performs for specific companies and how transport options are used depending on demand levels and spatial context. The model results confirm a clear division of roles between transport options. Pipelines and waterways act as backbone connections for transporting large quantities of hydrogen over longer distances, whereas road transport provides essential flexibility for last-mile deliveries. In areas with low demand and limited access to backbone networks, road transport dominates, keeping average costs relatively low. In contrast, areas with high demand or without access to multiple transport modes rely heavily on pipelines, resulting in higher average costs per kilometre.

To conclude, the results demonstrate that the optimisation of hydrogen supply transport for regional industry requires a different approach than network design for large industrial clusters alone. The diversity within regional industries is much greater, which means that significant preparatory work in identifying and categorising industries is needed before meaningful network optimisation can be carried out. Only by linking the characteristics of regional industries to the performance of different transport modes can realistic and cost-efficient hydrogen transport networks for regional industry in the Netherlands be developed.

## 11.2 Scientific & social relevance

The research questions constructed for this study are based on the theoretical background from chapter 2. This chapter concluded a knowledge gap in current literature. There exist a large amount of research into hydrogen networks nowadays, mostly focussing on large geographical areas, such as entire countries or continents(Bique & Zondervan, 2018; Busch et al., 2023; Kotek et al., 2023). This results in global outlines of networks, due to the broad scope. Most studies use mathematical models such as MILP (Timalsina et al., 2025), but graph theory models, such as the ONLT, aren't used often. Besides that, transport options are mostly limited to pipelines, reducing flexibility in the network. This is probably related to the large scales at which networks are studied, where high demand levels are naturally more suited to pipeline transport.
However, research is missing into the hydrogen supply for small-scale industries by the use of multiple transport options. This study adds to current literature by developing a method which is able to identify characteristics of regional industry, and applying a model which optimises the supply transport for multiple transport options. By diving deep into the characteristics of demanding companies, and categorise them as well, this study is able to model hydrogen supply transport for on a high level of detail. Modelling for detailed research areas, and using actual transport networks available in that area, makes this study a valuable addition to the current field of hydrogen network research.

Besides the relevance within the scientific field, this study is relevant for society as well. As mentioned in chapter 1.1, one of the main challenges for Cluster 6 companies to increase the use of sustainable fuels, is to have access to supply transport. Geluk and Koopman (2020) concluded that smaller industries are dependent on regional distribution grids for hydrogen in order to connect to hydrogen supply networks. This study provides a method which can be applied to all companies, even on a relatively small scale. By using the method developed in this study, every company in need of green hydrogen can be identified, categorised and added to the ONLT. This will support in designing transport networks for regional scale as well, while keeping costs as low as possible, which is in favour of small companies. The National Cluster Energy Strategy for Cluster 6 could apply this method to improve the quality of industry data, thereby increasing the realism of the model and enabling the development of more concrete and actionable designs in a shorter timeframe. Besides that, further looking into the availability of road and waterway transport might be a solution to use in the coming period. Building the pipeline network is expected to be finished in 2033, but before that, other modes of transport could be a solution. This research showed that pipelines aren't the only transport option, especially for small-scale regional industry from Cluster 6.

## 11.3 Recommendations for future work

While this study has provided insights into the cost-optimisation for hydrogen supply transport for regional industry, there always exist limitations to a study, as described in chapter 10.2. This can be due to time limitations or limited access to data for example. This section presents recommendations for further research, so that this method and model can be further optimised and developed.
First, some aspects regarding the ONLT can be further developed. Including obstacles such as nature areas or villages would increase the realism of this model certainly, and add even more to the knowledge gap found in literature. Besides that, behaviour of road transport could be optimised, through a more realistic cost function, including $CO_2$-taxes, or further dive into the use of highways versus regional ways, investigating congestion for example.
Another interesting aspect is to analyse the use of existing natural gas pipelines, which could decrease costs for the overall network quite certain. This is a challenging subject, due to the high uncertainty in availability, quality of pipelines and the combination with new pipelines. Analysing the the use of those pipelines could be a study on its own, but could definitely add as an existing transport network to the ONLT. Now that the tool is suited for multiple transport options, another one could relatively easy be added. This does increase complexity and computational time of the model.
This study analysed parts of the country, based on the location of relevant clusters. It would be very interesting to analyse the optimal network for the Netherlands completely, to obtain full context of all regional industry. By developing a method which could analyse large areas like the Netherlands

at once, but still optimise on this level of detail would be a valuable addition to the method.
Lastly, it is recommended to analyse the hydrogen demand for regional industry. Identifying actual demand for every company would increase realism of this model as well, as would distinguishing demand for liquid and for gaseous hydrogen. If there would be companies which only need liquid hydrogen for example, this would highly influence the choices that the current model takes.
To conclude, recommendations for future work are mostly focussed on increasing realism of the current model. It is suggested to remain or even increase the level of detail from this study, because it supports in achieving results which could actually be applied in real-life scenarios, instead of remaining purely theoretical.

# Bibliography

Beagle, E., Lewis, M., Pecora, B., Rhodes, J., Webber, M., & Hebner, R. (2023). Model to inform the expansion of hydrogen distribution infrastructure. *International Journal of Hydrogen Energy*, *49*, 105–113. https://doi.org/10.1016/j.ijhydene.2023.07.017

Bique, A. O., & Zondervan, E. (2018). An outlook towards hydrogen supply chain networks in 2050 — design of novel fuel infrastructures in germany. *Process Safety and Environmental Protection*, *134*, 90–103. https://doi.org/10.1016/j.cherd.2018.03.037

Busch, T., Groß, T., Linßen, J., & Stolten, D. (2023). The role of liquid hydrogen in integrated energy systems–a case study for germany. *International Journal of Hydrogen Energy*, *48*(99), 39408–39424. https://doi.org/10.1016/j.ijhydene.2023.05.308

CBS Statline. (2025, March 17). *Energiebalans; aanbod en verbruik per sector*. Retrieved April 22, 2025, from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83989NED/table?dl=223E7

CBS-Statline. (2025, March 17). *Energiebalans; aanbod en verbruik per sector*. Retrieved April 22, 2025, from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83989NED/table?dl=223E7

Centraal Bureau voor de Statistiek. (2025). Meer broeikasgassen uitgestoten in eerste kwartaal van 2025. https://www.cbs.nl/nl-nl/nieuws/2025/24/meer-broeikasgassen-uitgestoten-in-eerste-kwartaal-van-2025

Cheng, W., & Cheng, Y. F. (2023). A techno-economic study of the strategy for hydrogen transport by pipelines in canada. *Journal of Pipeline Science and Engineering*, *3*(3), 100112. https://doi.org/10.1016/j.jpse.2023.100112

Cluster 6. (2025). *Landelijke ces – cluster 6*.

Demaco Holland B.V. (2025, January 24). *De energiedichtheid van waterstof: Een unieke eigenschap*. https://demaco-cryogenics.com/nl/blog/de-energiedichtheid-van-waterstof-een-unieke-eigenschap/

DEP. (2024). *Hoe krijgt het zesde (regionale) industrie cluster toegang tot waterstof? - duurzaam energie perspectief - dep*. https://www.dep.nl/artikelen/hoe-krijgt-het-zesde-regionale-industrie-cluster-toegang-tot-waterstof/#:~:text=Om%20de%20geproduceerde%20waterstof%20beschikbaar,tot%20grote%20volumes%20aan%20waterstof

Eichhorn, E. (2021, May 11). *Vaarwegenoverzicht nederland*. https://www.binnenvaartkennis.nl/2021/02/vaarwegenoverzicht-nederland/

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Knowledge Discovery and Data Mining*, 226–231. https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf

European Hydrogen Backbone. (n.d.). *The european hydrogen backbone (ehb) initiative | ehb european hydrogen backbone*. Retrieved August 4, 2025, from https://ehb.eu/

Franco, A., & Giovannini, C. (2023). Routes for hydrogen introduction in the industrial hard-to-abate sectors for promoting energy transition. *Energies*, *16*(16), 6098. https://doi.org/10.3390/en16166098

Gasunie. (2024, December 10). *Waterstof door gasleiding: Veilig en duurzaam*. https://www.gasunie.nl/expertise/waterstof/waterstof-door-gasleiding-veilig-en-duurzaam#:~:text=Waterstof%20kan%2C%20net%20als%20aardgas,noodzakelijk%20is%20in%20de%20energietransitie.

Geluk, M., & Koopman, G. J. (2020). *Klimaattransitie door de nederlandse industrie - het 6e cluster*.

Hammond, J., Rosenberg, M., & Brown, S. (2024, January 1). *A genetic algorithm-based design for hydrogen pipeline infrastructure with real geographical constraints*. https://doi.org/10.1016/b978-0-443-28824-1.50106-x

Hammond, J., Rosenberg, M., & Brown, S. (2025). Understanding costs in hydrogen infrastructure networks: A multi-stage approach for spatially-aware pipeline design. *International Journal of Hydrogen Energy*, *102*, 430–443. https://doi.org/10.1016/j.ijhydene.2024.12.273

Heijnen, P. W. (2025a). *Optimal network layout tool (onlt)* (comp. software). https://gitlab.tudelft.nl/pheijnen/optimal-network-layout

Heijnen, P. W. (2025b, January 24). *Tutorial optimal network layout tool*.

Hynetwork. (n.d.). *Waterstofnetwerk nederland*. https://www.hynetwork.nl/over-hynetwork/waterstofnetwerk-nederland#

Hynetwork. (2024, November 10). *Conceptvoorstel aanpassing uitrolplan*.

Hynetwork. (2025, February 21). *Tweede kamer geïnformeerd over kostenraming waterstofnetwerk*. https://www.hynetwork.nl/kennisbank/artikel/tweede-kamer-geinformeerd-over-kostenraming-waterstofnetwerk

Jovicic, M., & West, K. (2023). *Future production scenarios and resulting hydrogen demand 2030-2050*. TNO. www.tno.nl

Kavlakoglu, E., & Winland, V. (2024, June 26). *K-means clustering*. Retrieved August 8, 2025, from https://www.ibm.com/think/topics/k-means-clustering#:~:text=K%2Dmeans%20is%20an%20iterative,the%20characteristics%20of%20the%20data.

Kim, C., Lee, Y., Lee, H., Lee, U., & Kim, K. (2023). Economic and environmental potential of green hydrogen carriers (ghcs) produced via reduction of amine-captured co2. *Energy Conversion and Management*, *291*, 117302. https://doi.org/10.1016/j.enconman.2023.117302

Kotek, P., Tóth, B. T., & Selei, A. (2023). Designing a future-proof gas and hydrogen infrastructure for europe – a modelling-based approach. *Energy Policy*, *180*, 113641. https://doi.org/10.1016/j.enpol.2023.113641

Kountouris, I., Bramstoft, R., Madsen, T., Gea-Bermúdez, J., Münster, M., & Keles, D. (2024). A unified european hydrogen infrastructure planning to support the rapid scale-up of hydrogen production. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-49867-w

Kumar, A. (2020, September 17). *Kmeans silhouette score explained with python example*. https://dzone.com/articles/kmeans-silhouette-score-explained-with-python-exam

Lafeber, M. (2024). *Hydrogen infrastructure planning under uncertainty in an industrial port cluster: A robust over time approach*. http://repository.tudelft.nl/

Lamboo, S., Weeda, M., & TNO Public. (2024, October 2). *Hydrogen use in dutch industry* (TNO 2024 R11702). TNO Public. https://publications.tno.nl/publication/34643187/tvsxleIr/TNO-2024-R11702.pdf

Lipiäinen, S., Sillman, J., Vakkilainen, E., Soukka, R., & Tuomaala, M. (2023). Hydrogen transport options for a large industrial user: Analysis on costs, efficiency, and ghg emissions in steel mills. *Sustainable Production and Consumption*, *44*, 1–13. https://doi.org/10.1016/j.spc.2023.11.021

Liu, H., Zhou, S., Gu, W., Zhuang, W., Gao, M., Chan, C., & Zhang, X. (2024). Coordinated planning model for multi-regional ammonia industries leveraging hydrogen supply chain and power grid integration: A case study of shandong. *Applied Energy*, *377*, 124456. https://doi.org/10.1016/j.apenergy.2024.124456

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 281–297. http://digitalassets.lib.berkeley.edu/math/ucb/text/math_s5_v1_article-17.pdf

McKinsey & Company. (2024, September 17). *What are scope 1, 2, and 3 emissions?* Retrieved July 25, 2025, from https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-are-scope-1-2-and-3-emissions

Mendler, F., Voglstätter, C., Müller, N., Smolinka, T., Holst, M., Hebling, C., & Koch, B. (2025). A newly developed spatially resolved modelling framework for hydrogen valleys: Methodology and functionality. *Advances in Applied Energy*, 100207. https://doi.org/10.1016/j.adapen.2025.100207

Namazifard, N., Vingerhoets, P., & Delarue, E. (2024). Long-term cost optimization of a national low-carbon hydrogen infrastructure for industrial decarbonization. *International Journal of Hydrogen Energy*, *64*, 583–598. https://doi.org/10.1016/j.ijhydene.2024.02.324

Neuwirth, M., Khanra, M., Fleiter, T., Jovicic, M., & Shinde, M. (2022). Future hydrogen demands from industry transition towards 2030 – a site-specific bottom-up assessment for north-western europe.

NU.nl. (2025, February 21). *Aanleg waterstofnet plots veel duurder: Kosten ruim verdubbeld naar 3,8 miljard*. https://www.nu.nl/klimaat/6346787/aanleg-waterstofnet-plots-veel-duurder-kosten-ruim-verdubbeld-naar-38-miljard.html

Oberle, S., & Klopstein, H. (2024). How feasible is the transition to a hydrogen distribution network? *2022 18th International Conference on the European Energy Market (EEM)*, 1–9. https://doi.org/10.1109/eem60825.2024.10608902

Parolin, F., Colbertaldo, P., & Campanari, S. (2022). Development of a multi-modality hydrogen delivery infrastructure: An optimization model for design and operation. *Energy Conversion and Management*, *266*, 115650. https://doi.org/10.1016/j.enconman.2022.115650

Parolin, F., Colbertaldo, P., & Campanari, S. (2024). Optimal design of hydrogen delivery infrastructure for multi-sector end uses at regional scale. *International Journal of Hydrogen Energy*, *79*, 839–849. https://doi.org/10.1016/j.ijhydene.2024.06.049

[PDOK], P. (2025a, June 6). *Nationaal wegen bestand (nwb) - vaarwegen*. https://www.pdok.nl/introductie/-/article/nationaal-wegen-bestand-nwb-vaarwegen

[PDOK], P. (2025b, June 6). *Nationaal wegen bestand (nwb) - wegen*. https://www.pdok.nl/introductie/-/article/nationaal-wegen-bestand-nwb-wegen

Pellegrini, L. A., Spatolisano, E., Restelli, F., De Guido, G., De Angelis, A. R., & Lainati, A. (2024, January 1). *Green h2 transport through lh2, nh3 and lohc*. https://doi.org/10.1007/978-3-031-66556-1

Pykes, K. (2024, August 21). *An introduction to graph theory*. https://www.datacamp.com/tutorial/introduction-to-graph-theory?dc_referrer=https%3A%2F%2Fwww.google.com%2F

SciKitLearn. (n.d.). *Standardscaler*. Retrieved August 8, 2025, from https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

TenneT. (2024, October 1). *Eerste grootschalige waterstoffabriek op het hoogspanningsnet*. https://www.tennet.eu/nl/nieuws/eerste-grootschalige-waterstoffabriek-op-het-hoogspanningsnet

Thailappan, D. (2024, October 16). *Understand the dbscan clustering algorithm!* https://www.analyticsvidhya.com/blog/2021/06/understand-the-dbscan-clustering-algorithm/

Timalsina, S., Zhang, C., Kumar, D., & Holubnyak, E. (2025). Optimizing hydrogen transportation infrastructure and supply chain under uncertainty: An integrated mixed integer linear programming and real options analysis. *International Journal of Hydrogen Energy*, *111*, 635–647. https://doi.org/10.1016/j.ijhydene.2025.02.336

Vergara, D., Fernández-Arias, P., Lampropoulos, G., & Antón-Sancho, Á. (2024). Hydrogen revolution in europe: Bibliometric review of industrial hydrogen applications for a sustainable future. *Energies*, *17*(15), 3658. https://doi.org/10.3390/en17153658

WaterEnergySolutions. (2022). *Ces - cluster 6, maart 2022*.

Weißenburger, B., Karkossa, L., Stephan, A., & McKenna, R. (2025, May 2). *Location matters: Exploring the effects of regional geographical and political characteristics on hydrogen pipeline costs globally*. https://arxiv.org/abs/2505.01124

Yang, M., Hunger, R., Berrettoni, S., Sprecher, B., & Wang, B. (2023). A review of hydrogen storage and transport technologies. *Clean Energy*, *7*(1), 190–216. https://doi.org/10.1093/ce/zkad021

Yang, R., Schell, C. A., Rayasam, D., & Groth, K. M. (2025). Hydrogen impact on transmission pipeline risk: Probabilistic analysis of failure causes. *Reliability Engineering & System Safety*, 110825. https://doi.org/10.1016/j.ress.2025.110825

# A

# K-means clustering steps

The K-means clusterings has been executed following six steps. Chapter 6.2 describes the results of this clustering method, this section describes the extended version of all steps.

**Step 1 - Data selection and preparation**
In the previous chapter, all data for every industry location has been gathered. For this clustering method, three variables will be the input for the model. The hydrogen demand data and the distance between industry locations and the pipeline and waterway networks are used. For this clustering, the distance between the industry locations and road network is not taken into consideration. Because the analysis of the data in section 5.4 showed that the average distance and standard deviation are relatively low, it's expected that this variable will not add to the clustering method. Every industry location is close to a road, even in the highway dataset which is used in this study. Therefore, there will be no relevant differentiation between different locations with regard to this parameter.
Secondly, the data that is used needs to be normalised, so that it can be compared to each other. This is needed because hydrogen demand and distances are displayed in different metrics. The StandardScaler method from the SciKitLearn (n.d.) package in Python is used to scale data to one generic metric. In the following step, the results of this step will be showed.

**Step 2 - Plot data for visual inspection**
Figure A.1 shows the distribution of all locations. Three subplots have been made to visualize the data. This way, every variable has been plotted against every other variable. In first sight, some outliers can be seen on the left and middle figure, where demand is displayed on the vertical axes. There is a high density of locations in the left-low corner, but some locations with relatively high demand are placed outside the main group of locations. The subplot on the right side shows the two distance variables, where the distribution of locations is more evenly spread, compared to the first two subplots. Besides these two-dimensional plots, an interactive three-dimensional plot has been created, which supported the visual inspection of the normalised data.

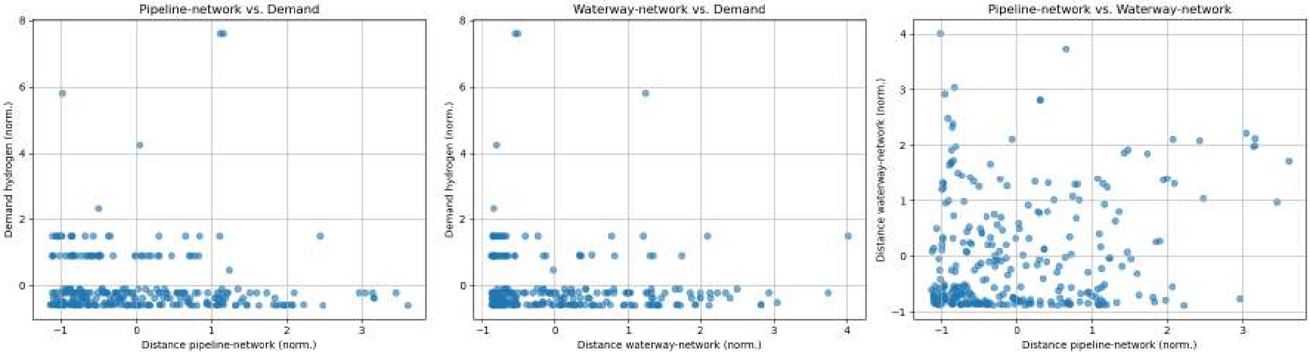


**Figure A.1:** Scatter-plot normalised data (complete dataset)

**Step 3 - Remove outliers for optimal clustering**
Outliers can influence the clustering by making relative distances between data-points smaller than they actually are. By removing outliers, the relative distances between all other data-points becomes more realistic, therefore creating better clusters. In order to select the actual outliers[1], which will be

---

[1] Relatively high values compared to average values

left out of the clustering, a few clustering rounds are executed. The k-value is varied between 2 and 9, so that it will be visible for different numbers of clusters which data-points are put into the 'outlier cluster'. For k=2, the expected outliers were put into the same cluster as data-points within the dense area, where all the other data-points are located. At k=3, the expected outliers were put into a separate cluster, as shown in figure A.2, where every cluster is represented with a separate colour, and a label at the centroid location. Cluster 3 consist of four data-points with high hydrogen demand.

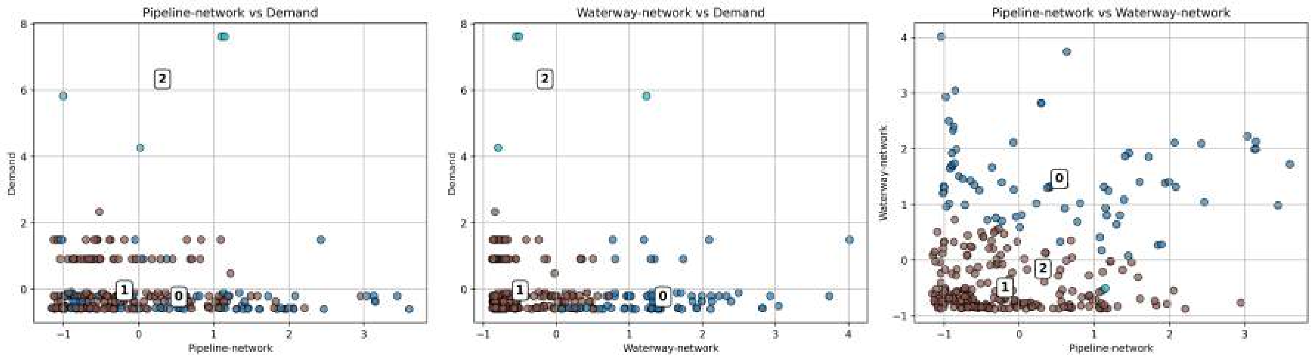


**Figure A.2:** Scatter-plot clusters (k=3)

However, it could be possible that other data-points might be an outlier as well, which aren't included in the same cluster for k=3. Therefore, the clustering has been done for k-values within the rang of 4-9. All clustering results for k within the range of 3-9, the same outlier cluster was constructed, which exists of the four data-points mentioned earlier in this section. Therefore, it's concluded that the four data-points are labelled as outliers. By removing the outliers, the clustering of the remaining data will be of higher quality, because the clustering algorithm isn't influenced by extreme values. Besides that, visual inspection will be made easier, because the scaling of the data will be adjusted.
Following this conclusion, the outliers have been removed from the dataset which will be used in the final clustering. The outliers will be part of the final network optimisation model.

**Step 4 - Pre-analyse data for optimal k-values**
Now that the final dataset has been chosen, the data has been scaled again. After that, the data is analysed, in order to determine the optimal k-values. This will be done by visually inspecting the data in the scatter-plot (figure A.3), but also by using the elbow method and silhouette score for a set of k-values, which will be explained later on.
Due to the removal of outliers, the data is now better interpretable through the scatter-plot in figure A.3. This figure is in some sort similar to figure 5.2 and 5.5, because it visualises the distribution of the company characteristics. The left and middle subplot in the scatter-plot show that demand is mostly relatively low, with part of the demand data being higher. In contrast to demand data, the distance metrics are less 'grouped' together, but spread throughout the graph, shown by the right subplot.

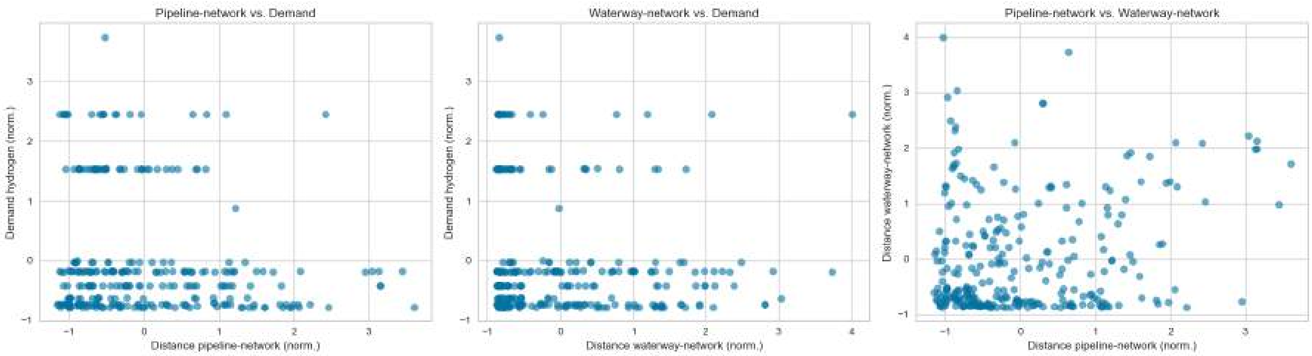
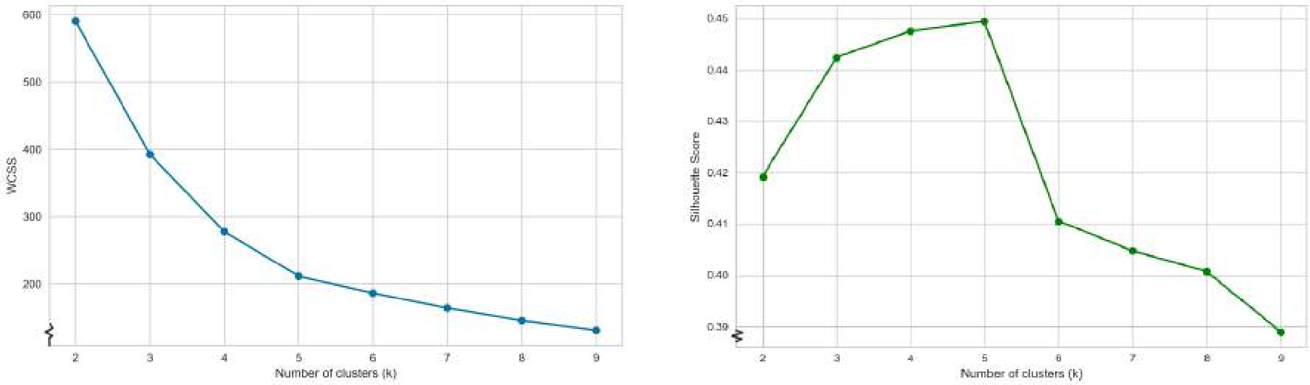


**Figure A.3:** Scatter-plot final dataset

Two other metrics which can be used for the analysis of K-means clustering are the elbow method and silhouette score. The elbow method shows the graph of the 'Within Cluster Sum of Squares' (WCSS) for a range of k-values. WCSS represents the total variance within each cluster. This metric is plotted against the range of k-values. The lower the WCSS value, the better the clustering is done (Kavlakoglu & Winland, 2024). By searching for the 'elbow point', the optimal number of clusters can be determined. However, not all WCSS plots have a clear 'elbow point', as is the case in figure A.4a. Therefore, no hard conclusions on the optimal number of clusters can be derived from this graph. It can still be concluded, that the slope of the WCSS graph decreases strongly from k=4/5, which means that adding more clusters, will increase the quality of clustering less, from these k-values on.



**(a)** Elbow plot

**(b)** Silhouette score plot

**Figure A.4:** Analysis plots for k-values

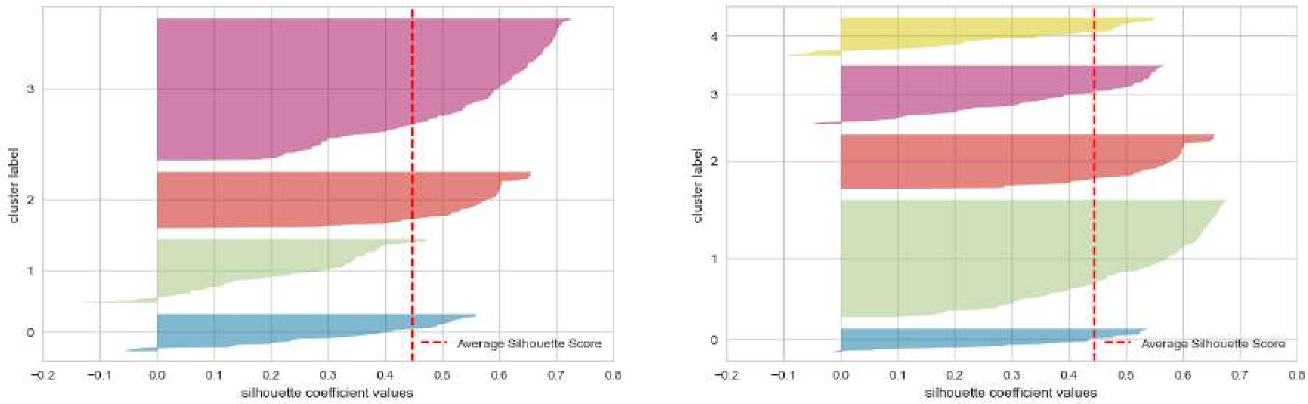
Besides the elbow method, the silhouette score is also calculated for the same range of k-values, so that the values of k can be judged on different metrics. The silhouette score calculates two distances for every data point. The first distance is the mean intra-cluster distance, which is the mean distance between a specific data point and all other points in the same cluster. The second distance is the mean nearest-cluster distance, which is the mean distance between the data point and all other data points of the next nearest cluster. These two values are used to calculate the the silhouette score (Kumar, 2020). The value of the silhouette score varies between -1 and 1. A score of 1 means that the clusters are dense and separated from other clusters. A score near 0 means overlapping clusters and a negative score indicates that data points have probably been assigned to wrong clusters.

Figure A.4b plotted the silhouette scores for k-values between 2 and 9. It is important to mention that the differences between the number of clusters seems to be large, but the vertical axis has a relatively small range (0.39-0.44) It is therefore harder to draw conclusions from this graph. K-values 3-5 do have the highest values in this graph, but this needs to be put in perspective, as mentioned.

Figure A.5[2] provides the opportunity to dive deeper into the silhouette scores for the k-values 4-7. This range is chosen based on the elbow plot (figure A.4a). These plots provide the silhouette scores for every data point in every cluster, including the average silhouette score (same as in figure A.4b) with a red-dashed line. It is now possible to see if there are clusters which score below average, the fluctuations within a cluster and possible negative silhouette scores within a cluster.

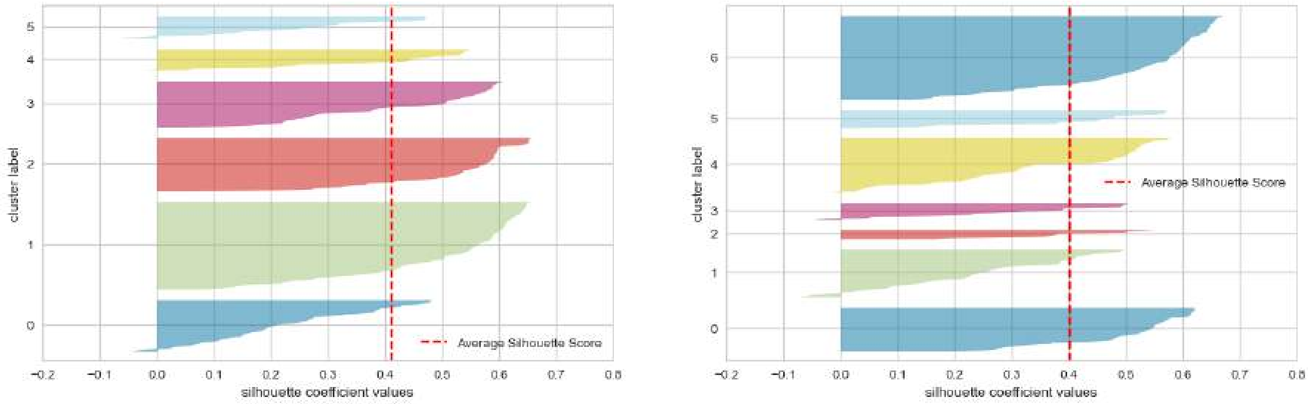
---

[2] Be aware: The numbering of clusters starts at 0 for clustering. Thus, if k=6, cluster 0-5 are created.



**(a)** Silhouette scores k=4

**(b)** Silhouette scores k=5



**(c)** Silhouette scores k=6

**(d)** Silhouette scores k=7

**Figure A.5:** Detailed silhouette scores for k (4-7)

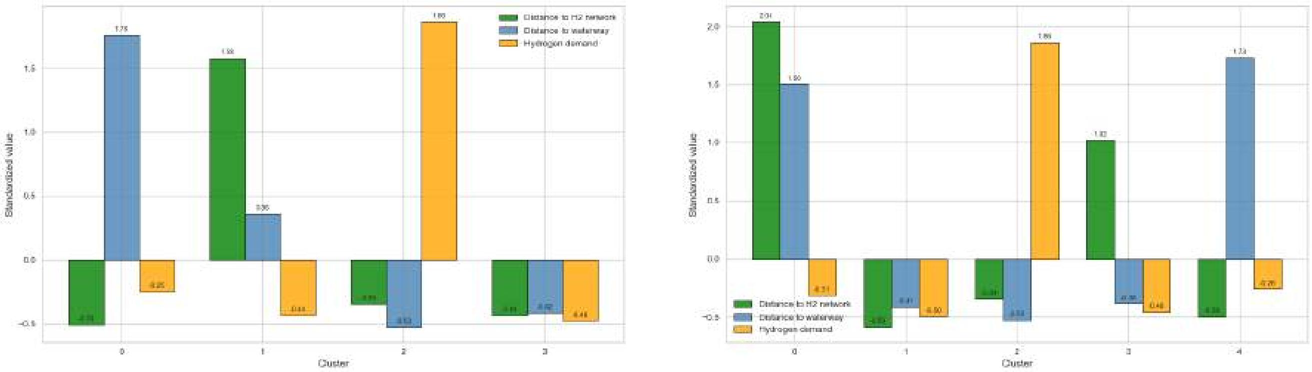
For k=4, there is a cluster which has below average silhouette scores for (almost) every data point. For k=5, k=6 and k=7, this is not the case. For k=4, k=5 and k=7, the fluctuations in cluster size are relatively large. The sizes of k=6 fluctuate also, but in smaller amounts. Lastly, negative values are noticed for every k-value. Table A.1 scores the k-values for the different checkpoints by providing - -,- ,+/-,+ or + +. The results shows that k=5 and k=6 perform relatively well on these performance metrics. This will support in choosing the final k-value.

| Number of clusters (k) | Below average silhouette scores | Size fluctuations | Negative silhouette scores |
|---|---|---|---|
| 4 | **+/-** | **-** | **-** |
| 5 | **+** | **+/-** | **-** |
| 6 | **+** | **+** | **-** |
| 7 | **+** | **-** | **-** |

**Table A.1:** Overview performance k-values on silhouette score

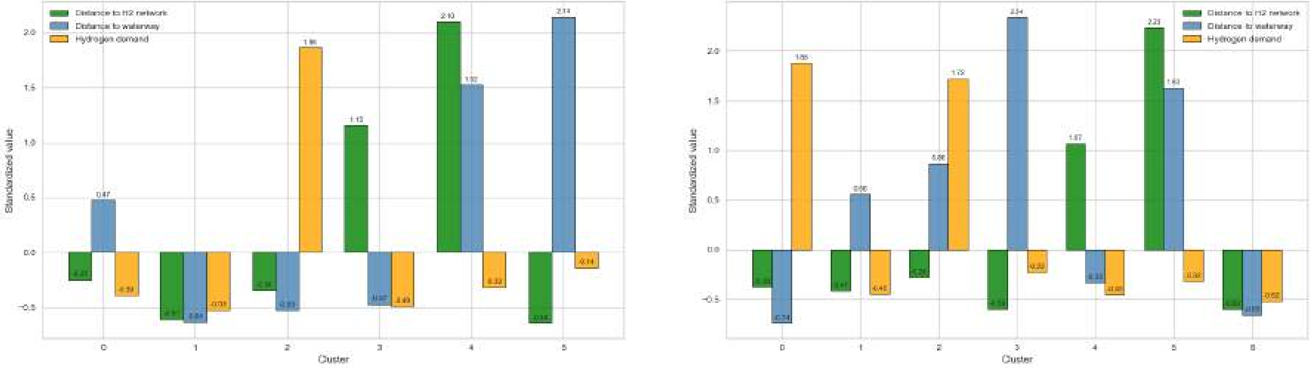
**Step 5 - Run and analyse K-means clustering for optimal k-values**

K=5 and k=6 performed well in the silhouette-specific metrics. K=7 performed well on the elbow method and k=4 performs best on the average silhouette score. Therefore, the final evaluation on clustering performance will be done for those 4 k-values. For this final analysis, the clustering will be evaluated on the description of a cluster. Figure A.6 shows the standardized, average parameter values for every cluster, for k-values 4-7.



**(a)** Cluster values k=4

**(b)** Cluster values k=5



**(c)** Cluster values k=6

**(d)** Cluster values k=7

**Figure A.6:** Standardized cluster values for final k-values

When clustering data, it’s important that every cluster describes an unique set of data, so that every cluster can be described differently. If one cluster is similar to another cluster, the number of clusters is too high. Besides that, it’s important to check the number of data points in a cluster. If this is too low, the cluster might not add to the evaluation of all data, because the characteristics of the cluster are almost never present.
The k=4 clusters can be classified as follows. For every parameter, there is a cluster in which one single parameter has a high average value, where the other two are low. These are cluster 0, 1 and 2. In cluster 3, every parameter has a low average value. The clusters are all different to each other, and the smallest cluster (0) consists of 36 data points.
If k is increased by one, the new cluster contains for both distance parameters a relatively high value, and a low value for demand. This can be an interesting cluster, because both transport options are far away, resulting in the road-network being closest. It might be interesting to see how the ONLT-model will provide a solution for these type of companies. This cluster (0) is the smallest cluster, consisting 23 data points.
Then, for k=6, cluster 0 is new, where distance to pipeline network and demand are low, and distance to waterway is a bit higher. However, as can be seen in figure A.6c, the distance to waterway average value isn’t as high, as it is in other clusters where distance to waterway is high. This new clusters does account for 52 data points, which shows that it might be a relevant group of data points.
Finally, when 7 clusters are formed, another cluster with high demand is introduced (cluster 1), combined with high distance to waterways and low distance to pipeline network. This cluster is relatively small (n=10) and has a below average silhouette score (as can be seen in figure A.5d), and is therefore not interesting.

**Step 6 - Determine final k-value**
To determine the optimal number of clusters for this dataset, all performance metrics described in this chapter are taken into account. Combining the results from the elbow method and (detailed) silhouette score, k=6 seems to be the optimal k-value for this data set. The different clusters all de-

scribe a unique group of data points. Next to that, k=6 performs best on silhouette scores (table A.1). Therefore, this study will use 6 different types of companies within Cluster 6. The final clusters are described in table A.2

| Cluster | Hydrogen demand | Distance to hydrogen backbone | Distance to waterways network | Number of industries |
|---|---|---|---|---|
| 0 | Low | Low | Moderate-high | 52 |
| 1 | Low | Low | Low | 89 |
| 2 | High | Low | Low | 54 |
| 3 | Low | High | Low | 47 |
| 4 | Low | High | High | 22 |
| 5 | Moderate | Low | High | 23 |

**Table A.2:** Characteristics of final clusters

# B

# Cost function

The cost function is described in section 8.3. This appendix provides the full code for edge costs, used in the ONLT-model, including all parameters used in equations.

### Parameters

```
transport_parameters = {
    ’Rekenperiode’:40,
    ’E_GH2’:0.000121,
    ’E_LH2’:0.000142,
    ’Pipeline’:{’Diameter’:0.508,
          ’Behoud energie’:0.9,
          ’Opex_fix’:0.04},
    ’Road’:{’Cap_truck’:4324.9,
               ’Cost_var’:0.0009973,
               ’Cost_var_snelweg’:0.0008029,
               ’Lifetime_truck’:11,
               ’Invest_truck’:860000,
               ’Opex_fix’:0.02,
               ’max_trucks_per_day’:25},
    ’Waterway’:{’Cap_boot’:354500,
               ’Cost_var’:0.01599,
               ’Lifetime_boot’:25,
               ’Invest_boot’:4100000,
               ’Opex_fix’:0.02}
    }

allowed_edge_transitions = [’Pipeline’, ’Road’]
```

### Definition edge costs

```
def edge_cost(G, beta, upc=0, cpc=1, transport_parameters=None,
    allowed_edge_transitions=None):
    # Maak een kopie van de graaf
    G1 = G.copy()
    rekenperiode = transport_parameters[’Rekenperiode’]
    E_GH2 = transport_parameters[’E_GH2’]
    E_LH2 = transport_parameters[’E_LH2’]
    for node1, node2 in G1.edges():
        weight = G1[node1][node2][’weight’]
        capacity = G1[node1][node2][’capacity’]
        category = G1[node1][node2].get(’category’, ’onbekend’)

        # Als het een nieuwe verbinding is
        if ’current’ not in G1[node1][node2]:
            if beta == 0 and capacity == 0:
                G1[node1][node2][’cost’] = 0
                G1[node1][node2][’category’] = ’onbekend’
                capex = 0
                opex = 0
            else:
```

```
best_cost = float('inf')
best_type ='onbekend'
# bereken voor elk edge-type de kosten
for edge_type in allowed_edge_transitions:
    if edge_type == 'Pipeline':
        # Parameters
        diameter = transport_parameters['Pipeline']['Diameter']
        behoud_energie = transport_parameters['Pipeline']['Behoud energie']
        opex_fix_pipeline = transport_parameters['Pipeline']['Opex_fix']

        #Formules
        gewicht_H2 = capacity/(E_GH2*behoud_energie)/365
        capex_full = 50833*weight*math.e**(0.0697*diameter)+297*(diameter**2)+71800*diameter+54658 #function from Hammond (2024), capital costs. Assuming for whole lifetime
        capex = capex_full/rekenperiode # Divide by calculation period to know costs/year
        opex = capex*opex_fix_pipeline #opex as percentage of capex

        cost = capex + opex
        temp_N_vehicle = None
    elif edge_type == 'Road':
        # Parameters
        cap_truck = transport_parameters['Road']['Cap_truck']
        cost_var_truck = transport_parameters['Road']['Cost_var']
        lifetime_truck = transport_parameters['Road']['Lifetime_truck']
        invest_truck = transport_parameters['Road']['Invest_truck']
        opex_fix_truck = transport_parameters['Road']['Opex_fix']
        max_trucks_day = transport_parameters['Road']['max_trucks_per_day']

        #Formules
        gewicht_H2 = capacity/E_LH2 #Amount of hydrogen needed in kg for 1 year for this specific edge
        N_truckloads = math.ceil(gewicht_H2/cap_truck) # Amount of truckloads needed for 1 year, ceiled.

        N_truck_purchase = math.ceil(N_truckloads/365) #Amount of trucks needed each year to supply demand, assuming a truck can drive every day once to location
        N_times_new_trucks = rekenperiode/lifetime_truck #Number of times trucks needs to be renewed within rekenperiode

        capex = (N_truck_purchase * N_times_new_trucks * invest_truck)/rekenperiode

        opex_fix = capex * opex_fix_truck
        opex_var = cost_var_truck*weight*2*N_truckloads #Variable opex, assuming a truck needs to drive back to start point as well.
        opex = opex_fix + opex_var

        cost = capex + opex
        temp_N_vehicle = N_truck_purchase
    else:
        cost = 9999999  # failsafe
        capex = 0
        opex = 0
```

```
                    if cost < best_cost:
                        best_cost = cost
                        best_type = edge_type

                G1[node1][node2]['cost'] = round(best_cost, 2)
                G1[node1][node2]['category'] = best_type
                G1[node1][node2]['N_vehicle'] = temp_N_vehicle

        # === Bestaande verbindingen ===
        else:

            if category == 'Pipeline':
                diameter = transport_parameters['Pipeline']['Diameter']
                behoud_energie = transport_parameters['Pipeline']['Behoud energie']
                opex_fix_pipeline = transport_parameters['Pipeline']['Opex_fix']

                #Formules
                #gewicht_H2 = capacity/(E_GH2*behoud_energie)/365
                capex_full = 50833*weight*math.e**(0.0697*diameter)+297*(diameter**2)
                    +71800*diameter+54658
                capex = capex_full/rekenperiode
                opex = capex*opex_fix_pipeline

                G1[node1][node2]['cost'] = round(opex, 2)
                G1[node1][node2]['N_vehicle'] = None
            elif category == 'Road':
                # Parameters
                cap_truck = transport_parameters['Road']['Cap_truck']
                cost_var_truck_snelweg = transport_parameters['Road']['Cost_var_snelweg
                    ']
                lifetime_truck = transport_parameters['Road']['Lifetime_truck']
                invest_truck = transport_parameters['Road']['Invest_truck']
                opex_fix_truck = transport_parameters['Road']['Opex_fix']
                max_trucks_day = transport_parameters['Road']['max_trucks_per_day']

                #Formules
                gewicht_H2 = capacity/E_LH2
                N_truckloads = math.ceil(gewicht_H2/cap_truck)

                N_truck_purchase = math.ceil(N_truckloads/365) #Amount of trucks needed
                     each year to supply demand
                N_times_new_trucks = rekenperiode/lifetime_truck #Number of times
                    trucks needs to be renewed within rekenperiode

                capex = (N_truck_purchase * N_times_new_trucks * invest_truck)/
                    rekenperiode

                opex_fix = capex * opex_fix_truck
                opex_var = cost_var_truck_snelweg*weight*2*N_truckloads
                opex = opex_fix + opex_var

                cost = opex #Only opex costs, capex costs are accounted for in new
                    edges, because we assume it is the same truck which drives on a new
                     edge and
                G1[node1][node2]['cost'] = round(cost, 2)
                G1[node1][node2]['N_vehicle'] = N_truck_purchase
            elif category == 'Waterway':
                # Parameters
                cap_boat = transport_parameters['Waterway']['Cap_boot']
```

```
                cost_var_boat = transport_parameters['Waterway']['Cost_var']
                lifetime_boat = transport_parameters['Waterway']['Lifetime_boot']
                invest_boat = transport_parameters['Waterway']['Invest_boot']
                opex_fix_boat = transport_parameters['Waterway']['Opex_fix']

                # Formules
                gewicht_H2 = capacity/E_LH2/365
                N_boatloads = math.ceil(gewicht_H2/cap_boat)

                N_boat_purchase = math.ceil(N_boatloads/365) #Amount of trucks needed each year to supply demand
                N_times_new_boats = rekenperiode/lifetime_boat #Number of times trucks needs to be renewed within rekenperiode

                capex = (N_boat_purchase * N_times_new_boats * invest_boat)/rekenperiode

                opex_fix = capex * opex_fix_boat
                opex_var = cost_var_boat*weight*2*N_boatloads
                opex = opex_fix + opex_var

                cost = capex + opex
                G1[node1][node2]['cost'] = round(cost, 2)
                G1[node1][node2]['N_vehicle'] = N_boat_purchase
            else:
                G1[node1][node2]['cost'] = 9999999  # failsafe
                capex = 0
                opex = 0

        G1[node1][node2]['capex'] = round(capex, 2)
        G1[node1][node2]['opex'] = round(opex, 2)
        G1[node1][node2]['new'] = 'New' if 'current' not in G1[node1][node2] else 'Existing'

    return G1
```

# C

# ONLT plots for every model step

The figures below visualize all steps calculated in the ONLT, for every research area. This shows the process of the model throughout the different steps. Figure C.1 shows the legend for all plots.

**Figure C.1:** Legend for maps

### Research Area 1

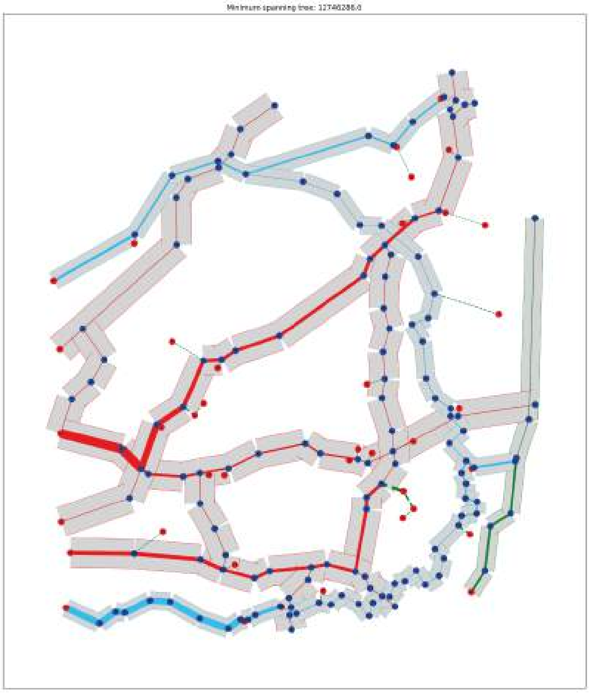


**Figure C.2:** Step 1 - RA-1

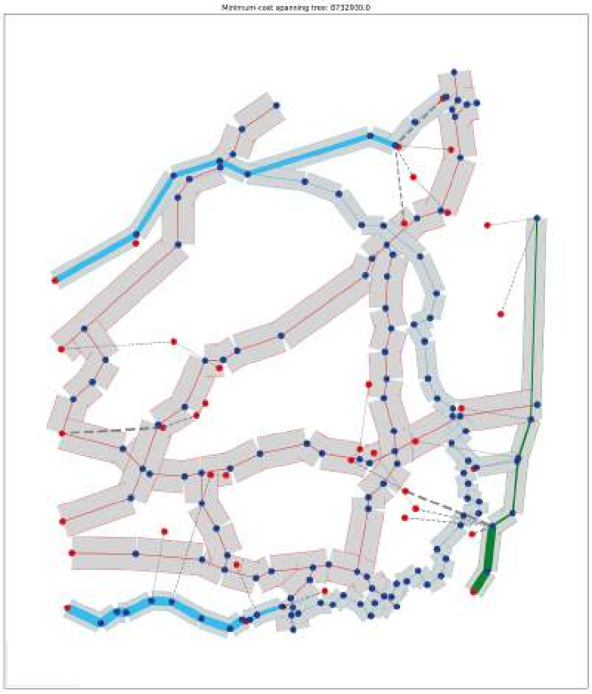


**Figure C.3:** Step 2 - RA-1

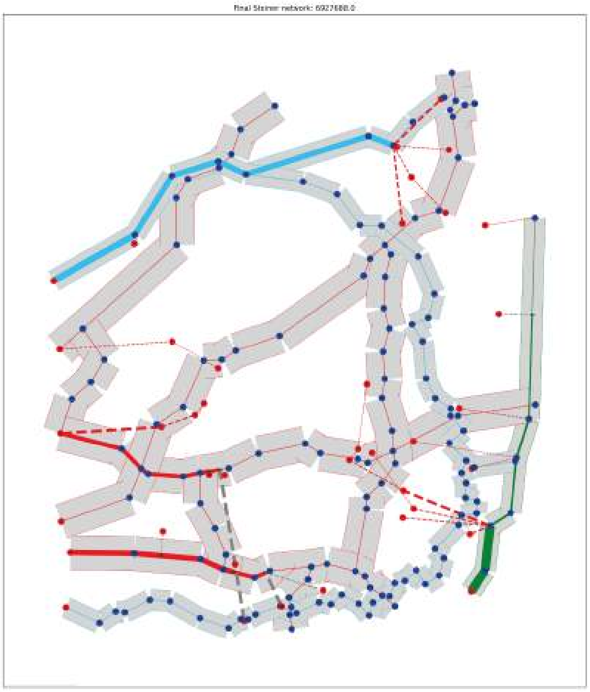


**Figure C.4:** Step 3 - RA-1

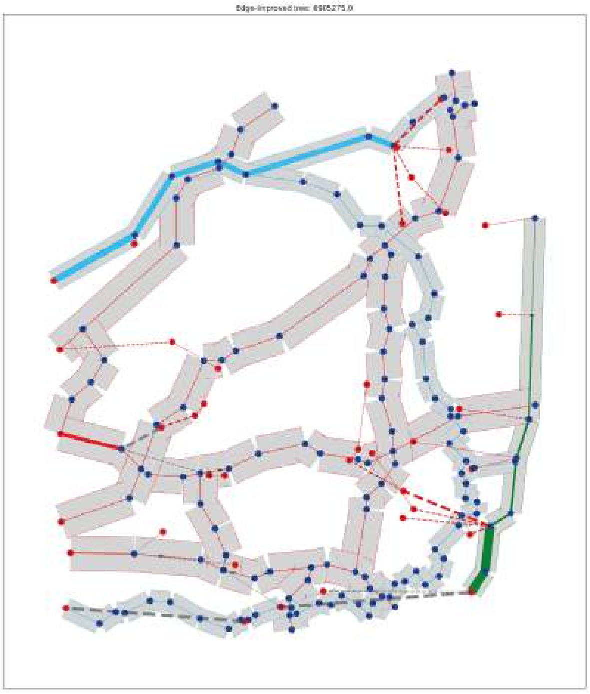


**Figure C.5:** Step 4 - edge improved tree - RA-1

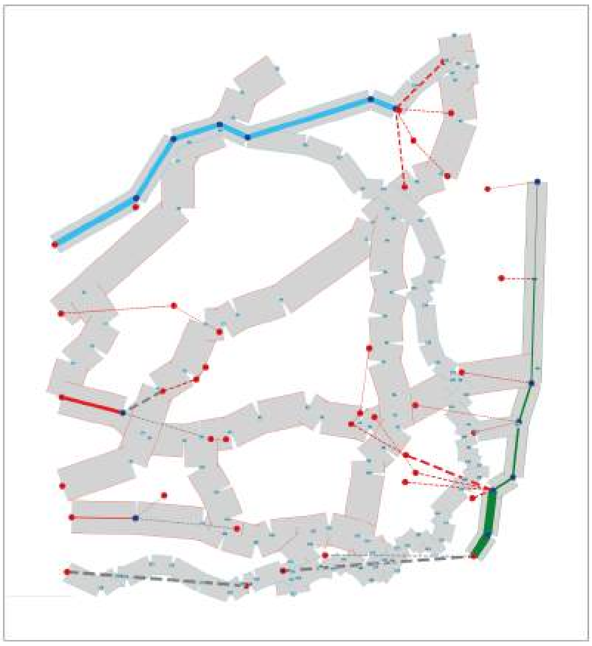


**Figure C.6:** Step 4 - final - RA-1

**Research area 2**

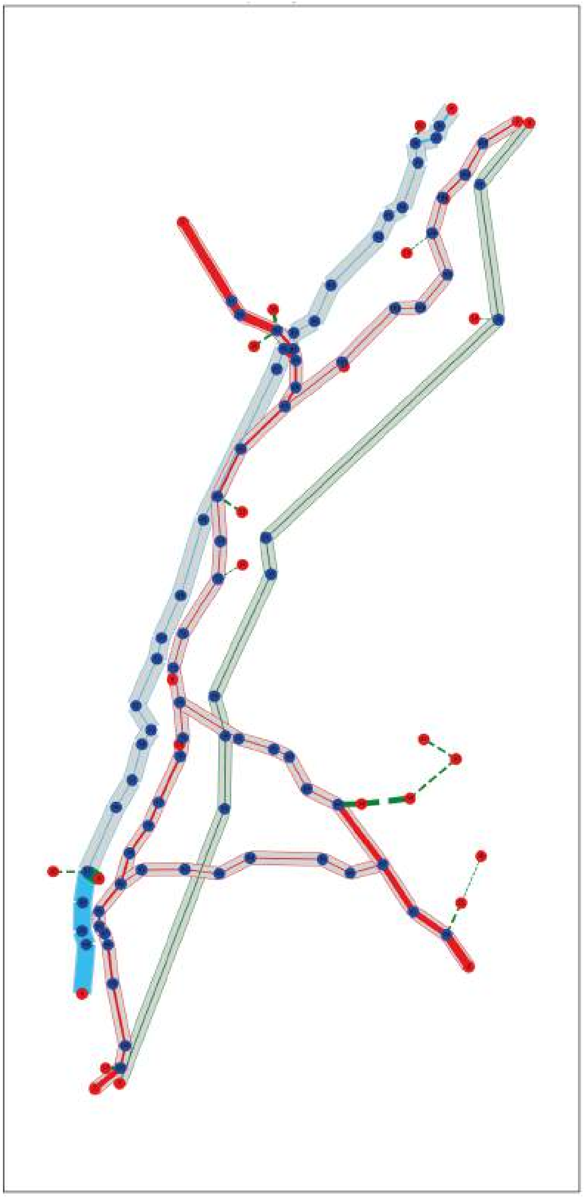


Figure C.7: Step 1 - RA-2

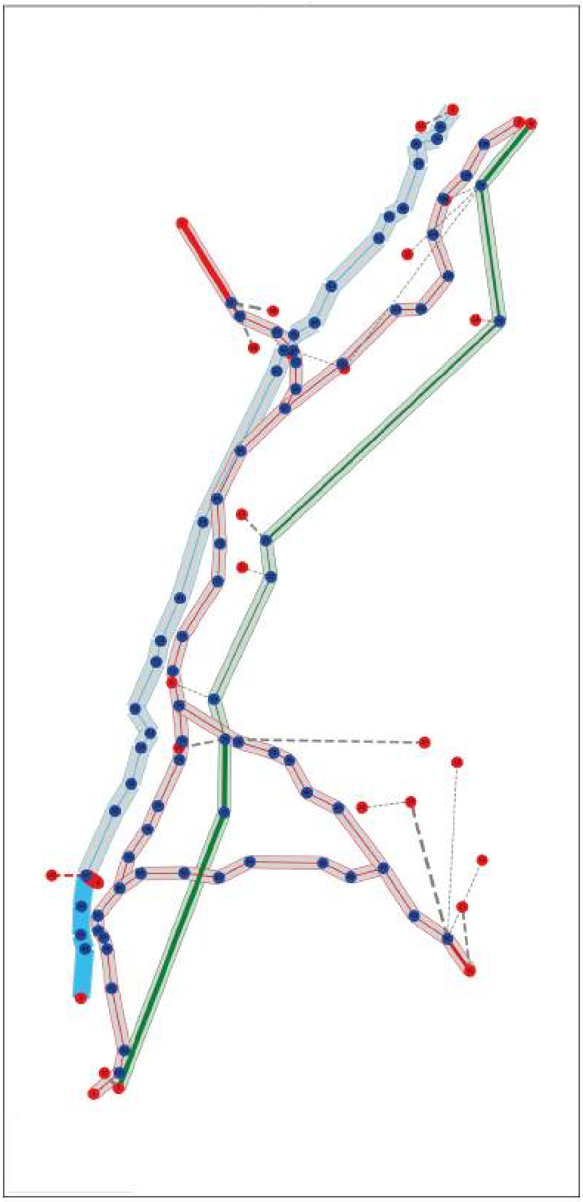


Figure C.8: Step 2 - RA-2

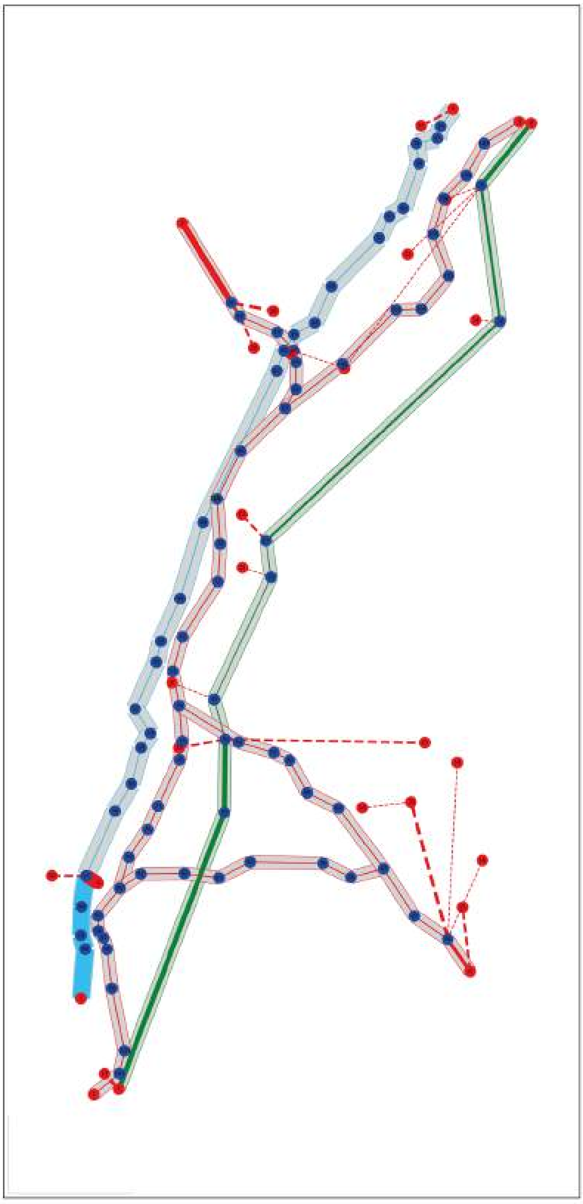


**Figure C.9:** Step 3 - RA-2

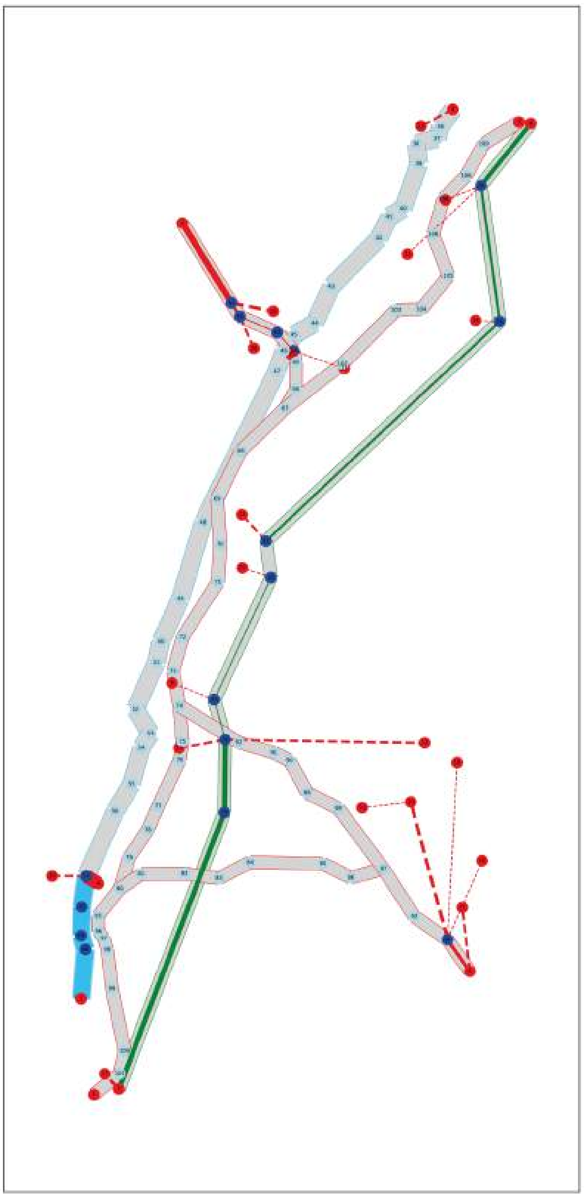
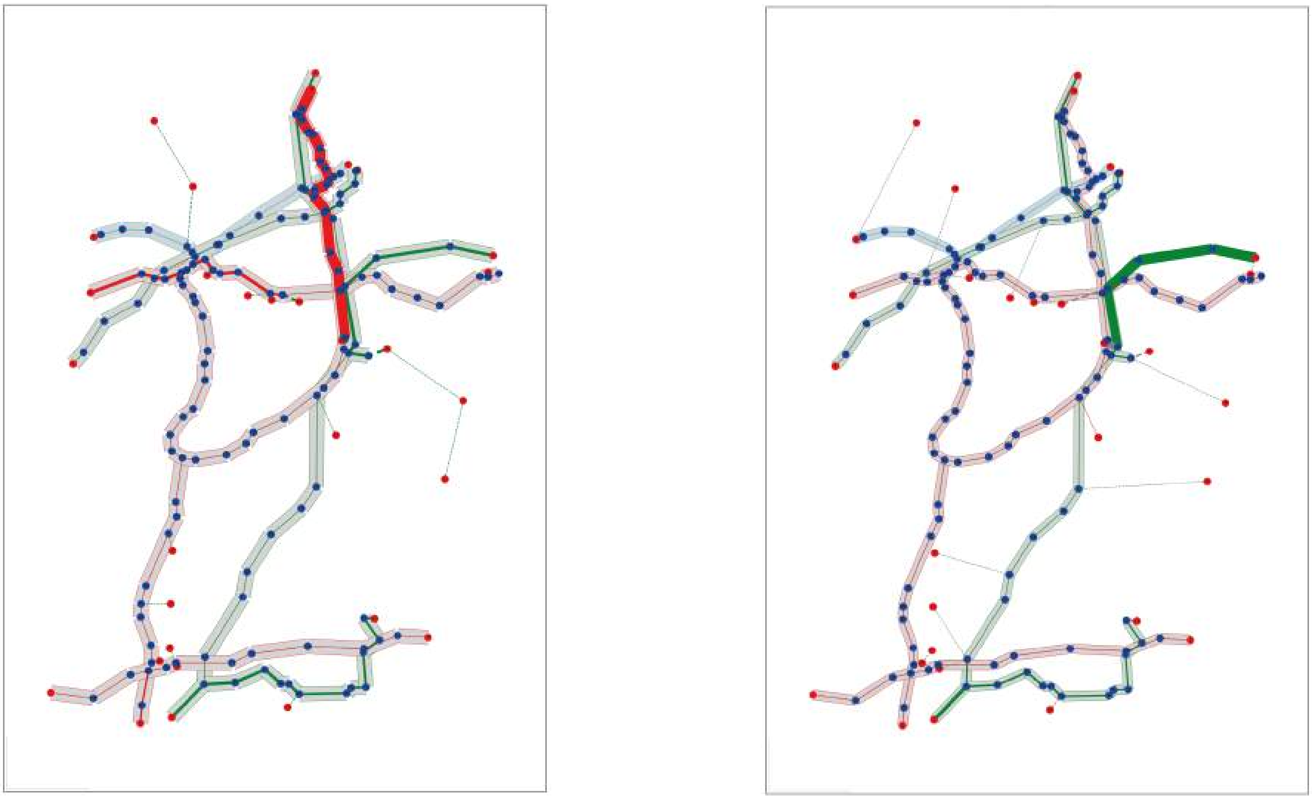


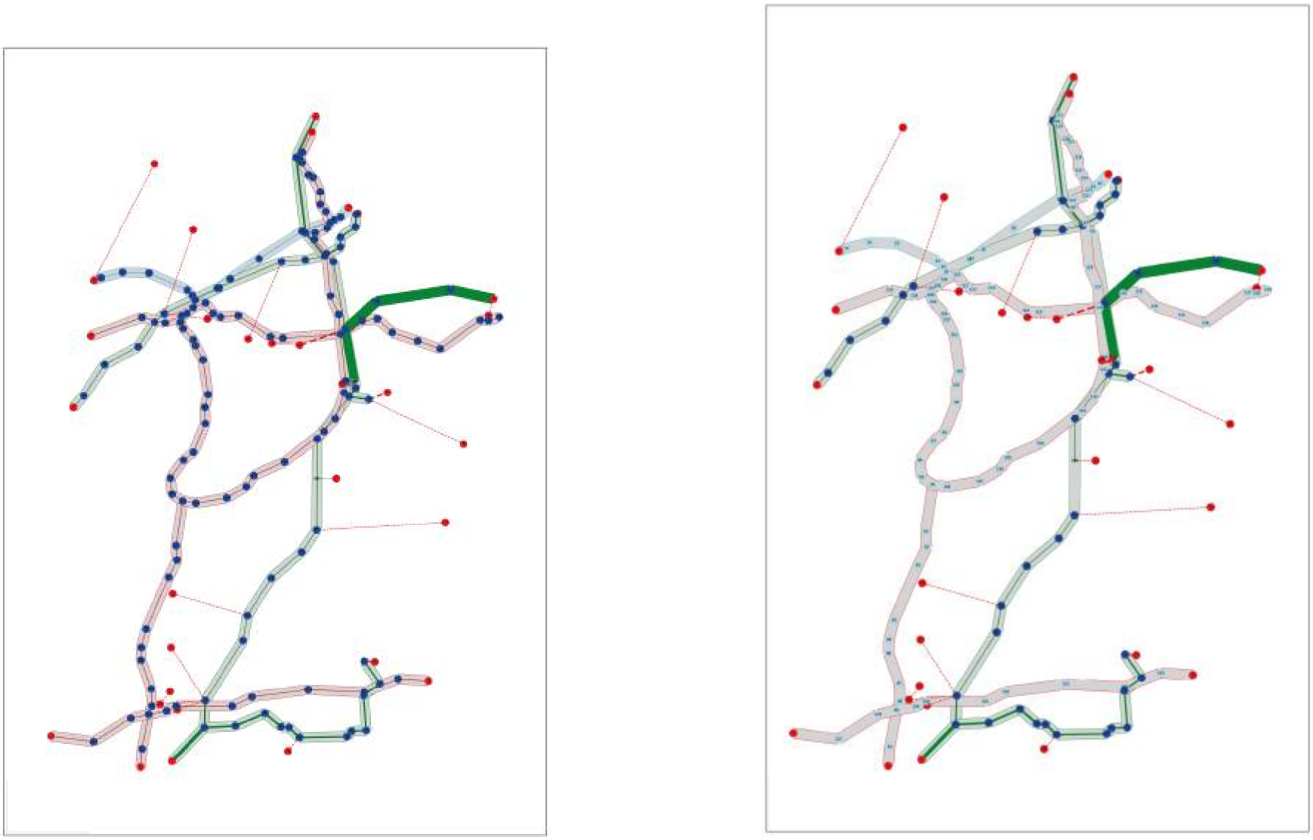
**Figure C.10:** Step 4 - RA-2

### Research Area 3



(a) Step 1

(b) Step 2



(c) Step 3

(d) Step 4

Figure C.11: All ONLT steps Research area 3