

Multi-period Stochastic Network Design for Combined Natural Gas and Hydrogen Distribution

Citation for published version (APA):

Hastürk, U., Schrottenboer, A., Roodbergen, K. J., & Ursavas, E. (2023). *Multi-period Stochastic Network Design for Combined Natural Gas and Hydrogen Distribution*. (pp. 1-33). arXiv.org.
<https://doi.org/10.48550/arXiv.2312.13388>

Document license:

CC BY

DOI:

[10.48550/arXiv.2312.13388](https://doi.org/10.48550/arXiv.2312.13388)

Document status and date:

Published: 20/12/2023

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Multi-period Stochastic Network Design for Combined Natural Gas and Hydrogen Distribution

Umur Hasturk^a, Albert H. Schrottenboer^b, Kees Jan Roodbergen^a, Evrim Ursavas^a

^a*Department of Operations, Faculty of Economics and Business, University of Groningen, Nettelbosje 2, Groningen, 9747 AE, The Netherlands*

^b*Operations, Planning, Accounting and Control Group, School of Industrial Engineering, Eindhoven University of Technology, De Zaale, Eindhoven, 5600 MB, The Netherlands*

Abstract

Hydrogen is produced from water using renewable electricity. Unlike electricity, hydrogen can be stored in large quantities for long periods. This storage ability acts as a *green battery*, allowing solar and wind energy to be generated and used at different times. As a result, green hydrogen plays a central role in facilitating a climate-neutral economy. However, the logistics for hydrogen are complex. As new hydrogen pipelines are developed for hydrogen, there is a trend toward repurposing the natural gas network for hydrogen, due to its economic and environmental benefits. Yet, a rapid conversion could disrupt the balance of natural gas supply and demand. Furthermore, technical and economic developments surrounding the transition contribute additional complexity, which introduces uncertainty in future supply and demand levels for both commodities. To address these challenges, we introduce a multi-period stochastic network design problem for the transition of a natural gas pipeline network into a green hydrogen pipeline network. We develop a progressive hedging based matheuristic to solve the problem. Results demonstrate our matheuristic is efficient, both in computation time and in solution quality. We show that factoring in uncertainty avoids premature expansion and ensures the development of an adequate pipeline network meeting long-term needs. In a case study in the Northern Netherlands for *Hydrogen Energy Applications in Valley Environments for Northern Netherlands* initiative, we focus on two key scenarios: local production and importation, exploring their impacts on performance indicators. Our case insights exemplify the solid foundation for strategic decision-making in energy transitions through our approach.

Keywords: Transportation, Network Design, Green Hydrogen, OR in energy, Progressive Hedging

Acknowledgments

This project has received funding from the Fuel Cells and Hydrogen 2 Joint Undertaking (now Clean Hydrogen Partnership) under Grant Agreement No 875090. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme, Hydrogen Europe and Hydrogen Europe Research. Albert H. Schrottenboer has received support from the Dutch Science Foundation (NWO) through grant VI.Veni.211E.043. We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Hábrók high performance computing cluster.

1. Introduction

The European Union has set a goal to achieve carbon neutrality by 2050 (European Union, 2012). This necessitates an increased reliance on renewable energy sources for electricity generation. However, the supply from renewable energy sources, such as wind and solar, is intermittent and uncertain (Hodge et al., 2018). Typically, natural gas-fired power plants bridge the gap when demand surpasses renewable supply (Safari et al., 2019). Such an approach, while practical, contradicts the carbon-neutral directive and increases the dependency on countries rich in natural gas. An emergent and sustainable alternative is *green hydrogen*, which is produced carbon-free through electrolysis powered by renewable sources (Megía et al., 2021). It allows the storage of excess energy as hydrogen, which can be converted back to power or directly used in transportation, residential heating, and as an industrial feedstock, thereby highlighting its potential as a multifaceted energy solution (Staffell et al., 2019).

Today’s energy landscape mainly centers around networks that consist of natural gas producers, customers, intermediate nodes, and storage units, linked by pipelines that facilitate the natural gas flow. In the coming decades, these networks are envisioned to gradually transition from natural gas to hydrogen-based infrastructure, a vision already taking shape with projects like *Hydrogen Energy Applications in Valley Environments for Northern Netherlands* (HEAVENN) in the Netherlands, where Europe’s first hydrogen valley is being built. This shift involves not only the construction of new hydrogen pipelines but also the retrofitting of existing natural gas pipelines for hydrogen distribution. Compared to building new pipelines, repurposing existing natural gas pipelines for hydrogen is both cost-effective and more environmentally friendly (Gordon et al., 2023). However, this process of repurposing is complex, as it requires a strategic approach to balance current natural gas supply and demand with the transition towards hydrogen to avoid potential disruptions. Rapid conversion could disrupt this balance, while delayed conversion might hinder the establishment of the hydrogen economy. Moreover, it is uncertain how fast hydrogen demand will emerge in society (Espegren et al., 2021). Consequently, it is crucial to strategically plan the new pipeline construction and the gradual conversion of existing pipelines for the coming decades.

In this study, we address a novel multi-period multi-commodity stochastic network design problem focused on the transition of gas-based pipelines into hydrogen-based pipelines over a given time horizon. At the beginning of the horizon, the pipeline network balances the existing supply and demand for all commodities (e.g., gas and hydrogen). As time progresses, we encounter a stochastic transition between these commodities, reflected in the changes in supply and demand at associated nodes in the network. Inventories are kept at each location, allowing for the storage of the commodities for the demand of the subsequent periods. We choose periods aligning with the seasonality in the energy markets, which helps us understand the interplay between inventory levels across different seasons. We employ a two-stage recourse model representation of this multi-period problem. The first stage involves making strategic decisions about the construction of new pipelines and retrofitting existing ones, while the second stage addresses uncertainty in demand and supply by adopting a scenario-based reformulation. Within each scenario, we determine the operational decisions relating to commodity flow on pipelines and inventory management at network nodes. In cases where the demand is not satisfied, we incur a penalty, mirroring

real-world scenarios where trucks might be utilized temporarily to transport excess demand.

To solve this problem, we design an efficient matheuristic based on progressive hedging. We augment this matheuristic with several acceleration techniques. On well-calibrated benchmark instances, we show that our matheuristic outperforms a compact mixed-integer programming (MIP) formulation that is solved by Gurobi. Using the matheuristic, we study the structure of our optimal solutions. Among others, considering the stochastic nature of supply and demand leads to (i) an increased deployment of hydrogen infrastructure at the end of the planning horizon, reflecting a strategic investment that enhances the network’s capacity and efficiency, (ii) a later adoption of hydrogen infrastructure, thereby reducing the risk of overcapacity and underutilization through better-informed planning, and (iii), a higher need for installation capacity in the peak adaptation periods as more pipelines need to be built and retrofitted in a shorter amount of time.

Our practice-based study makes several theoretical contributions to the theory of network design. Our problem relates to the literature on so-called stochastic fixed-charge capacitated multi-commodity network design problems (Hewitt et al., 2021). Authors focus on balancing the uncertain supply and demand of multiple commodities by determining optimal arc construction —pipelines in our context— and flow distribution (Paraskevopoulos et al., 2016; Sarayloo et al., 2021). Within our paper, we offer several contributions to this literature. Firstly, we introduce a novel *arc-sharing* setting within the stochastic fixed-charge capacitated multi-commodity network design problem, ensuring each pipeline can be used for any commodity, but is designated to one specific commodity and operates within its own distinct network in any particular period (see also Tsiklios et al., 2022). Secondly, we contribute to the area of incremental network design, where existing models address the multi-period recovery of partially disrupted networks (Engel et al., 2017; Kalinowski et al., 2015). While their multi-period strategic focus aligns with ours, their scope does not consider the multi-commodity aspect associated with the necessary transition between commodities over time. Finally, we present a case study in the Northern Netherlands associated with the HEAVENN initiative, pioneering the hydrogen economies (New Energy Coalition, 2020). Our study focuses on two main scenarios with local production versus hydrogen import, and analyses how the transition evolves from natural gas to hydrogen. In summary, by evaluating how to effectively implement a transition from one commodity network to another over a time horizon, utilizing existing infrastructure, and constructing new pipelines when necessary, we provide valuable insights into the dynamics of the energy transition process, a perspective previously unexplored in the literature.

The remainder of the paper is organized as follows. In Section 2, we review the relevant literature. In Section 3, we define the problem and the corresponding mathematical model. In Section 4, we propose a progressive-hedging-based matheuristic to solve our problem. In Section 5, we present and discuss the key insights and findings derived from our computational experiments. In Section 6, we tackle a case within the Northern Netherlands (HEAVENN, 2022). In Section 7, we conclude our paper and present an outlook for future research.

2. Literature Review

This section discusses several areas of the literature that relate to our work. In Section 2.1, we consider studies with various aspects of hydrogen network design, encompassing supply chain planning, pipeline network configurations, and discussions on optimal pipeline diameter selection. Section 2.2 discusses the literature on stochastic fixed-charge capacitated multi-commodity network design problems. We end this section by discussing how our work relates to existing studies on incremental network design.

2.1. Hydrogen Network Design

In the energy literature, several works focus on practical use cases within supply chain optimization. While effective in their specific applications, these models often provide limited theoretical insights from an Operations Research perspective. We highlight a selection of such works. Almansoori and Shah (2012) present a multi-stage stochastic model for hydrogen supply chain network design. They encompass strategic decisions related to fueling stations, production plants, and storage facilities. Seeking to enhance the efficiency of such complex models, Nunes et al. (2015) employ sample average approximation and validate their approach with Great Britain’s liquid hydrogen infrastructure. We refer the interested reader to Li et al. (2019) for an optimization-oriented review of hydrogen supply chain network design.

Research on hydrogen pipeline network design has grown in recent years. For example, André et al. (2013) focus on constructing pipelines and their capacities for a case study in France. They incorporate nonlinear construction costs that vary depending on the determined diameters of the pipes. Reuß et al. (2019) assess the impact of linearized versus nonlinear diameter cost models on solution quality. A hydrogen production and transmission model is introduced by Johnson and Ogden (2012). Applied to the Southwestern United States, they provide insights into supply location selection and the optimal dimensions of pipelines, showing its potential in effectively utilizing existing infrastructure. André et al. (2014) look at how hydrogen transportation might change over time. Using a so-called backward heuristic, they predict pipeline growth in Northern France. Their work shows that trucks might be relevant for the short run, but pipelines become a better option when the hydrogen economy constitutes at least 10% of the fueling market. This supports the idea in our paper of using a penalizing scheme, where trucks fill in when pipelines cannot meet demand for short-term imbalances.

We enrich the aforementioned literature stream in several ways. To the best of our knowledge, previous studies have not explicitly examined the multi-commodity setting that considers the potential conversion of existing pipelines from natural gas to hydrogen. Furthermore, these studies consider pipeline capacities as a continuous decision variable, which does not reflect the industry practice of selecting from a limited set of standard pipeline diameters. Our approach incorporates a realistic approach by selecting pipeline diameters from these regional standard options, thereby ensuring the feasibility of implementation while reflecting real-world construction limitations. While some aspects of our problem are touched upon in the above studies, our research is the first to combine these elements.

2.2. Stochastic Fixed-Charge Capacitated Multi-commodity Network Design

The fixed-charge capacitated multi-commodity network design problem is an optimization problem that is faced in various domains including less-than-truckload (LTL) transportation, energy distribution,

and telecommunications. In this design, the network distributes multiple commodities between their origin and destination via the selected capacitated arcs that incur a fixed cost (i.e., charge) and a variable cost per unit for commodity distribution. Gray (1971) introduces an exact method for a direct delivery case using decomposition and traditional branch-and-bound techniques. Ghamlouche et al. (2003) later address the general case, proposing a tabu search approach with a cycle-based neighborhood. However, real-world applications often lack complete upfront information —such as demand and costs—, leading to the study of this problem under uncertainty.

The literature concerning network design under uncertainty mostly considers two-stage recourse models to deal with demand uncertainty. The first stage determines arc capacity decisions, and the second stage decides the flow across chosen arcs, accounting for demand uncertainty via finitely many scenarios. Notable solution techniques include a metaheuristic framework based on the progressive hedging algorithm (Crainic et al., 2011), an accelerated Bender’s decomposition framework (Rahmaniani et al., 2018), and partial Benders decomposition (Crainic et al., 2021) for the case where arc capacity is also uncertain. The literature also presents various extensions expanding upon the core concepts of this problem. Hu et al. (2019) solve the problem with a multi-echelon setting, utilizing a multi-stage recourse model to address the problem. Considering the extreme scenarios, Müller et al. (2021) propose fixing the first-stage decision variables with a high penalty incurred in the second stage. On a distinct note, Thapalia et al. (2012) delve into a variant of the problem where the arc capacity for a single commodity can be utilized in both directions of a directed network.

The chosen set of scenarios in the second stage profoundly influences both the solution’s quality and computational efficiency. In general, the more scenarios are considered, the closer the model is to reality. However, research has indicated that reducing the number of scenarios can significantly expedite solution times without severely compromising the solution quality. In this context, Crainic et al. (2014) introduce an innovative approach to grouping scenarios based on their similarities, thereby enhancing computational performance. Advancing the field, Jiang et al. (2021) propose a soft clustering method, which, unlike traditional methods, allows scenarios to have probabilistic membership to multiple bundles.

Two-stage recourse models are also often selected to address the multi-period version of this design problem. Within the first stage, the decision centers on which arcs should be selected throughout the entire planning horizon. Thereafter, the second stage concentrates on optimizing the commodity flow through these arcs for the planning horizon. Adopting this paradigm, Hewitt et al. (2019) and Wang et al. (2019) partition an infinite time horizon into distinct, manageable periods, such as weeks, emphasizing periodic and cyclic scheduling. Taking a slightly different approach, Jiang et al. (2021) and Boland et al. (2017) find solutions within a finite time horizon by constructing a time-space network. Notably, Fragkos et al. (2021) presume that as the horizon progresses, arc construction becomes more economical, and present a solution to the deterministic version of the problem.

Our contributions to this literature stream are as follows. Firstly, in contrast to most of the existing literature where each commodity is often from a singular origin and a single destination, our study focuses on a small number of commodities, dispersed across the network with multiple supply, demand, and storage nodes. Secondly, contrary to the common approach in LTL settings, where commodities

can share arcs, our model uniquely assigns each arc to a specific commodity, ensuring that they operate independently within their own designated networks. Thus, one arc is only used for one commodity at a time. Lastly, our model addresses a transitional phase where the volume of at least one commodity diminishes over time. We leverage the existing network infrastructure of this diminishing commodity to benefit others, facilitating a dynamic conversion of arcs from one commodity to another throughout the planning horizon.

2.3. Incremental Network Design

The incremental network design problem focuses on the restoration of networks impaired due to unforeseen events like natural disasters. Researchers aim to strategically rebuild or repair arcs over several periods. This concept is similar to our setting, where decisions on arc construction are made incrementally across a multi-period horizon. Constraints often include set budgets per period, repairs requiring more than one period, or a fixed number of arcs that may be built in each period.

Kalinowski et al. (2015) focus on the incremental network design, with an emphasis on maximizing the cumulative flow over the planning horizon by adding arcs in each period. Similar to our approach where additional arcs can be constructed every period, they operate with a single commodity from a single origin to a destination, aiming to optimize the flow of the commodity. On the other hand, Baxter et al. (2014) study a problem related to shortest paths in incremental network design. While they also consider the addition of arcs across periods, they restrict it to one arc per period, emphasizing the minimization of the total length of flow prior to the complete repair of the network.

Our research offers new perspectives on this existing literature. While we apply a multi-commodity framework to a multi-period network design problem—diverging from traditional incremental network design—we also introduce the arc-sharing concept. In our context, arc-sharing ensures that pipelines are designated for a single commodity, with the flexibility to switch commodities between periods, hence preventing cross-commodity interference and optimizing the flow within segregated networks. Our approach, specifically tailored for the energy transition network design problem, therefore enhances the scope of incremental network design literature, especially in scenarios where diverse types of resources are considered as multiple commodities.

3. Problem Definition

We first provide an overview of our system and detail the dynamics within. Afterward, we introduce the MIP that models our energy transition network design problem.

3.1. System Description

We consider an energy transition network design problem on a finite planning horizon with periods $t \in \mathcal{T} := \{0, 1, \dots, T - 1\}$. We consider a directed network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, where \mathcal{N} is the node set and \mathcal{A} the arc set.

At period $t = 0$, we assume the network is constructed so that it enables matching demand and supply in the existing energy market. For this, we consider multiple commodities $k \in \mathcal{K} := \{1, 2, \dots, K\}$ that

model the different energy types. Each node $i \in \mathcal{N} := \{1, 2, \dots, N\}$ has a net supply quantity for each commodity, which is negative for demand nodes and zero for intermediate nodes. For some commodities, there may be no market initially, which means that their net supply quantities are zero for each node. This also means that an intermediate node for a commodity could become a supply or a demand node in later periods of the time horizon, for example when an electrolysis facility for hydrogen is built at that node. We assume each of these nodes is included in the initial set of nodes \mathcal{N} , and thus the node set does not change over periods of the horizon. One node may be shared by multiple commodities, for example, a supplier node for one commodity may be a demand node for another commodity.

The arc set $\mathcal{A} := \{1, 2, \dots, A\}$ consists of all initial arcs and all candidate arcs to be constructed. There may be multiple arcs between a pair of nodes, for example, each representing a pipeline with a different standard capacity. $\mathcal{A}_k \subseteq \mathcal{A}$ is the set of arcs that are in use at $t = 0$ for commodity $k \in \mathcal{K}$, where $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$. \mathcal{A}_0 represents the candidate arcs that are not constructed at $t = 0$ and thus might be constructed in future periods. The cardinality of the set of arcs \mathcal{A} , thus, does not change over periods of the horizon. An arc can only be assigned to a single commodity k at any given time, but this assignment can switch to a different commodity during the planning horizon. We enable the switch between commodities via decision variables, detailed below, and thus the set \mathcal{A}_k does not change over time. This conversion process is less costly compared to constructing a new arc. We impose that each arc can be converted at most once during the planning horizon. Furthermore, we assume construction and conversion occur in between periods, thus once constructed, an arc is available on all the following periods.

The net supply at each node for each period is assumed to be stochastic. Besides assuming periodical variations (e.g., to reflect seasonal effects) leading to variations in quantities, it is also uncertain when demand for commodities will emerge or decrease. For example, we assume that it is not exactly known which locations would be the supplier nodes within an emergent hydrogen economy. This also results in uncertainty of the inventory capacity of network nodes, since the capacity depends on the type of the location. For instance, supplier nodes are typically connected to larger storage units suitable for extended periods, whereas demand nodes, such as those in residential areas or refueling stations, generally have smaller storage capacities suitable for shorter periods. Under these assumptions, creating a robust network for each and every scenario might be economically infeasible. Thus, we assume the unsatisfied demand is penalized, simulating, for example, transport by truck. We will model the uncertainty using a finite set of scenarios.

The resulting model is a two-stage stochastic mixed integer program with continuous recourse. In the first stage, we make decisions on which arcs to construct for the entire planning horizon, and simultaneously assign a specific commodity to each newly constructed arc, as well as decide on converting existing arcs from one commodity to another. In the second stage, we optimize for each scenario the distribution of each commodity to ensure supply and demand are balanced. Excess supply is allowed to be stored and carries over to the next period. The goal is to find a network design that minimizes the construction cost and the expected distribution cost.

3.2. Mixed Integer Programming Formulation

Let y_{at}^k be a binary decision variable, equaling 1 if pipeline arc $a \in \mathcal{A}$ is constructed at time period t for commodity k , which has a fixed construction cost f_{at}^k . Let β_{at}^k be a binary decision variable, equaling 1 if pipeline arc a is converted for commodity k at the beginning of time period t , which has a fixed conversion cost g_{at}^k . Each arc a has a fixed flow capacity of U_a . Let z_{at}^k be a binary decision variable, having a value of 1 if the pipeline arc a is assigned to be used at time period t for commodity k .

We model uncertainty with finitely many scenarios $\omega \in \Omega := \{1, 2, \dots, W\}$. Each scenario ω has a probability of occurrence of p_ω . Under scenario ω , $s_{it}^{k\omega}$ represents the net supply of node i for commodity k in period t . Let $x_{at}^{k\omega}$ be the flow amount of commodity k on arc a on period t under scenario ω , which costs c_{at}^k per unit, which we set linearly proportional to the arc length. Let $I_{it}^{k\omega}$ be the inventory level of node i for commodity k at the beginning of period $t \in \mathcal{T} \cup \{T\}$ under scenario ω , which cannot exceed the inventory capacity of $C_{it}^{k\omega}$. Let $e_{it}^{k\omega}$ be the amount of unsatisfied net demand in node i for commodity k in period t under scenario ω , which is penalized with a cost of h per unit. Let the sets $\delta^+(i)$ and $\delta^-(i)$ represent the arc sets that enter and leave node i , respectively.

The two-stage stochastic mixed integer program with continuous recourse is formulated as:

$$\min \sum_{a \in \mathcal{A}} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} (f_{at}^k y_{at}^k + g_{at}^k \beta_{at}^k) + \sum_{\omega \in \Omega} p_\omega \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \left(\sum_{a \in \mathcal{A}} c_{at}^k x_{at}^{k\omega} + \sum_{i \in \mathcal{N}} h e_{it}^{k\omega} \right) \quad (1)$$

$$\text{s.t. } I_{i,t+1}^{k\omega} \leq I_{it}^{k\omega} + s_{it}^{k\omega} + e_{it}^{k\omega} - \sum_{a \in \delta^-(i)} x_{at}^{k\omega} + \sum_{a \in \delta^+(i)} x_{at}^{k\omega} \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (2)$$

$$I_{it}^{k\omega} \leq C_{it}^{k\omega} \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T} \cup \{T\}, \quad (3)$$

$$x_{at}^{k\omega} \leq U_a z_{at}^k \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (4)$$

$$\sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} y_{at}^k \leq 1 \quad \forall a \in \mathcal{A}, \quad (5)$$

$$y_{at}^k \leq z_{at}^k \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (6)$$

$$\sum_{k \in \mathcal{K}} \sum_{t'=0}^t y_{at'}^k = \sum_{k \in \mathcal{K}} z_{at}^k \quad \forall a \in \mathcal{A}, t \in \mathcal{T}, \quad (7)$$

$$z_{at}^k + \sum_{k' \in \mathcal{K} \setminus \{k\}} z_{a,t-1}^{k'} \leq 1 + \beta_{at}^k \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, t \in \mathcal{T} - \{0\}, \quad (8)$$

$$\sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \beta_{at}^k \leq 1 \quad \forall a \in \mathcal{A}, \quad (9)$$

$$y_{a0}^k = 1 \quad \forall k \in \mathcal{K}, a \in \mathcal{A}_k, \quad (10)$$

$$I_{i0}^{k\omega} = 0 \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, \quad (11)$$

$$x_{at}^{k\omega} \geq 0 \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (12)$$

$$I_{it}^{k\omega} \geq 0 \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T} \cup \{T\}, \quad (13)$$

$$e_{it}^{k\omega} \geq 0 \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (14)$$

$$y_{at}^k \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (15)$$

$$z_{at}^k \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (16)$$

$$\beta_{at}^k \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, t \in \mathcal{T}. \quad (17)$$

The objective function minimizes the total cost of arc construction, arc conversion, the expected cost of arc flow, and the expected penalty cost. Constraint (2) ensures the inventory levels are balanced between periods according to the net supply, as well as the inflow and outflow of the commodity. Constraint (3) ensures that the inventory respects the capacity. Constraint (4) ensures that the flow capacity of an arc is not exceeded, and is only available for the assigned commodity. Constraint (5) ensures each arc can be built once in the planning horizon. Once constructed, the arc is assigned to the corresponding commodity by Constraint (6) for that period. For other periods, Constraint (7) satisfies that each constructed arc is assigned to one of the commodities. We link the variable z with the conversion variable β by Constraint (8), ensuring the conversion cost is reflected in the objective function value. Each arc is restricted to be converted at most once during the planning horizon by Constraint (9). The initial networks for each commodity are created by Constraint (10). Finally, with Constraint (11) we make sure that the initial inventories are assigned. The remaining constraints model the decision variables' domains.

4. Progressive Hedging Based Matheuristic

The progressive hedging algorithm (PHA) is a method that uses an iterative process to solve stochastic optimization problems, such as a two-stage recourse model. By taking the Lagrangian relaxation on the non-anticipativity constraints, the problem can be decomposed among each scenario. Then, the PHA progresses iteratively by modifying the objective function of the single scenario deterministic optimization problems. The modification is done in such a way that the solutions to the deterministic optimization problems become closer to each other as the algorithm progresses through these iterations. The algorithm continues until the solutions of each single scenario problem share the same first-stage solution variables.

To apply the PHA to our model, we first create a reformulation of model (1) - (17). This reformulation duplicates each first-stage variable for each scenario, and then adds non-anticipativity constraints to make sure these variables are equal to each other. Explicitly, we create a copy for each first-stage variable for each scenario. We denote these variables by $y_{at}^{k\omega}$, $z_{at}^{k\omega}$, and $b_{at}^{k\omega}$. Furthermore, we add the non-anticipativity Constraints (35) - (37). This results in the following reformulation:

$$\min \sum_{\omega \in \Omega} p_{\omega} \left[\sum_{a \in \mathcal{A}} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} (f_{at}^k y_{at}^{k\omega} + g_{at}^k \beta_{at}^{k\omega}) + \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \left(\sum_{a \in \mathcal{A}} c_{at}^k x_{at}^{k\omega} + \sum_{i \in \mathcal{N}} h e_{it}^{k\omega} \right) \right] \quad (18)$$

$$\text{s.t. } I_{i,t+1}^{k\omega} \leq I_{it}^{k\omega} + s_{it}^{k\omega} + e_{it}^{k\omega} - \sum_{a \in \delta^-(i)} x_{at}^{k\omega} + \sum_{a \in \delta^+(i)} x_{at}^{k\omega} \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (19)$$

$$I_{it}^{k\omega} \leq C_{it}^{k\omega} \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T} \cup \{T\}, \quad (20)$$

$$x_{at}^{k\omega} \leq U_a z_{at}^{k\omega} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (21)$$

$$\sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} y_{at}^{k\omega} \leq 1 \quad \forall a \in \mathcal{A}, \omega \in \Omega, \quad (22)$$

$$y_{at}^{k\omega} \leq z_{at}^{k\omega} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (23)$$

$$\sum_{k \in \mathcal{K}} \sum_{t'=0}^t y_{at'}^{k\omega} = \sum_{k \in \mathcal{K}} z_{at}^{k\omega} \quad \forall a \in \mathcal{A}, \omega \in \Omega, t \in \mathcal{T}, \quad (24)$$

$$z_{at}^{k\omega} + \sum_{k' \in \mathcal{K} \setminus \{k\}} z_{a,t-1}^{k'\omega} \leq 1 + \beta_{at}^{k\omega} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T} - \{0\}, \quad (25)$$

$$\sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \beta_{at}^{k\omega} \leq 1 \quad \forall a \in \mathcal{A}, \omega \in \Omega, \quad (26)$$

$$y_{a0}^{k\omega} = 1 \quad \forall k \in \mathcal{K}, a \in \mathcal{A}_k, \omega \in \Omega, \quad (27)$$

$$I_{i0}^{k\omega} = 0 \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, \quad (28)$$

$$x_{at}^{k\omega} \geq 0 \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (29)$$

$$I_{it}^{k\omega} \geq 0 \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T} \cup \{T\}, \quad (30)$$

$$e_{it}^{k\omega} \geq 0 \quad \forall i \in \mathcal{N}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (31)$$

$$y_{at}^{k\omega} \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (32)$$

$$z_{at}^{k\omega} \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (33)$$

$$\beta_{at}^{k\omega} \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (34)$$

$$y_{at}^{k\omega} = y_{at}^k \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (35)$$

$$z_{at}^{k\omega} = z_{at}^k \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (36)$$

$$\beta_{at}^{k\omega} = \beta_{at}^k \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (37)$$

$$y_{at}^k \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (38)$$

$$z_{at}^k \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}, \quad (39)$$

$$\beta_{at}^k \in \{0, 1\} \quad \forall a \in \mathcal{A}, k \in \mathcal{K}, \omega \in \Omega, t \in \mathcal{T}. \quad (40)$$

As proposed by Rockafellar and Wets (1991), we relax the non-anticipativity constraints by removing Constraints (35) - (40). This removal enables us to separate the model for each $\omega \in \Omega$ and solve them individually. Upon this, we modify the objective for scenario ω with the corresponding Lagrangian coefficients as in the following for each ω :

$$\begin{aligned} \min \quad & \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}} \sum_{k \in \mathcal{K}} (f_{at}^k + \lambda_{at}^{k\omega}) y_{at}^{k\omega} + (g_{at}^k + \delta_{at}^{k\omega}) \beta_{at}^{k\omega} + \pi_{at}^{k\omega} z_{at}^{k\omega} \\ & + \frac{\rho}{2} \left((y_{at}^{k\omega} - \bar{y}_{at}^k)^2 + (\beta_{at}^{k\omega} - \bar{\beta}_{at}^k)^2 + (z_{at}^{k\omega} - \bar{z}_{at}^k)^2 \right) \\ & + \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \left(\sum_{a \in \mathcal{A}} c_{at}^k x_{at}^{k\omega} + \sum_{i \in \mathcal{N}} h e_{it}^{k\omega} \right). \end{aligned} \quad (41)$$

Here, λ , δ , and π are Lagrangian multipliers for decision variables y , β , and z respectively. ρ is the pre-assigned penalty ratio of the progressive hedging algorithm. The average values of decision variables over scenarios, \bar{y} , $\bar{\beta}$, and \bar{z} , are calculated from the previous iteration of the algorithm. Since all the first-stage variables are binary in our model, the quadratic objective function (41) would be equal to the following linearized objective function for all binary solutions:

$$\begin{aligned} \min \quad & \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}} \sum_{k \in \mathcal{K}} (f_{at}^k + \lambda_{at}^{k\omega}) y_{at}^{k\omega} + (g_{at}^k + \delta_{at}^{k\omega}) \beta_{at}^{k\omega} + \pi_{at}^{k\omega} z_{at}^{k\omega} \\ & + \left(\frac{\rho}{2} - \rho \bar{y}_{at}^k \right) y_{at}^{k\omega} + \left(\frac{\rho}{2} - \rho \bar{\beta}_{at}^k \right) \beta_{at}^{k\omega} + \left(\frac{\rho}{2} - \rho \bar{z}_{at}^k \right) z_{at}^{k\omega} \\ & + \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \left(\sum_{a \in \mathcal{A}} c_{at}^k x_{at}^{k\omega} + \sum_{i \in \mathcal{N}} h e_{it}^{k\omega} \right). \end{aligned} \quad (42)$$

Using this reformulation and relaxed model, we describe the steps of the PHA in Algorithm 1.

Algorithm 1 Progressive Hedging.

Input: $\lambda := 0, \delta := 0, \pi := 0, \rho > 0$.

Output: A feasible solution to the model (1) - (17).

- 1: **Initialization:** For each $\omega \in \Omega$, solve (18) - (34) by assuming $\Omega := \{\omega\}$ and $p_\omega = 1$.
 - 2: **while** convergence on first-stage variables is not satisfied **do**
 - 3: **Average:** $\bar{y}_{at}^k := \sum_{\omega \in \Omega} p_\omega y_{at}^{k\omega}, \bar{\beta}_{at}^k := \sum_{\omega \in \Omega} p_\omega \beta_{at}^{k\omega}, \bar{z}_{at}^k := \sum_{\omega \in \Omega} p_\omega z_{at}^{k\omega}$.
 - 4: **Price:** $\lambda_{at}^{k\omega} := \lambda_{at}^{k\omega} + \rho(y_{at}^{k\omega} - \bar{y}_{at}^k), \delta_{at}^{k\omega} := \delta_{at}^{k\omega} + \rho(\beta_{at}^{k\omega} - \bar{\beta}_{at}^k), \pi_{at}^{k\omega} := \pi_{at}^{k\omega} + \rho(z_{at}^{k\omega} - \bar{z}_{at}^k)$.
 - 5: **Solution:** For each $\omega \in \Omega$, solve (19) - (34) with the objective (42) by assuming $\Omega := \{\omega\}$ and $p_\omega = 1$.
 - 6: **end while**
-

We apply four accelerating techniques to the base PHA in our solution method. The first acceleration technique is to consider bundles of scenarios instead of single-scenario subproblems. This results in solving two-stage recourse models (i.e., subject to the scenarios in the bundle) in the PHA, instead of solving deterministic single scenario problems. It has been shown that for well-calibrated bundle sizes, this reduces computation time without affecting the quality too much (Crainic et al., 2014). We use a set of bundles Ξ where each scenario $\omega \in \Omega$ is in exactly one bundle $\xi \in \Xi$. The probability of a bundle is assigned as $p_\xi = \sum_{\omega \in \xi} p_\omega$. The objective function of the two-stage recourse model for bundle ξ is given below in Equation (43).

$$\min \sum_{a \in \mathcal{A}} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \left(f_{at}^k y_{at}^{k\xi} + g_{at}^k \beta_{at}^{k\xi} \right) + \sum_{\omega \in \xi} \frac{p_\omega}{p_\xi} \left[\sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \left(\sum_{a \in \mathcal{A}} c_{at}^k x_{at}^{k\omega} + \sum_{i \in \mathcal{N}} h e_{it}^{k\omega} \right) \right]. \quad (43)$$

Secondly, we observe that a majority of first-stage decision variables converge to a common value at early iterations of PHA. In this case, these decision variables may be fixed to their converged value for all scenarios for the remaining iterations of the algorithm. This is called *slamming* in Watson and Woodruff (2011) and has shown to be an effective way of accelerating the algorithm without affecting the solution quality substantially. Authors fix variables if they are equal for all scenarios for a given number of consecutive iterations. This given number of iterations is arranged in such a way that it decreases over time to enhance convergence.

We use a similar acceleration technique with two main differences. Firstly, we enable fixing variables even for the cases where they are not equal for all scenarios but for the majority of scenarios. If the total probability of scenarios where the decision variable shares a common value exceeds an input probability parameter p_H , we fix the corresponding decision variables. Secondly, we observe that even if some decision variables do not share the same value, we may see similarities between the values of decision variables in terms of our problem context. Thus, we implement a problem-specific convergence rule that does not depend on individual decision variables as in Watson and Woodruff (2011). Instead, the convergence condition depends on subsets of decision variables. For example, consider a case where one arc is built for the majority of scenarios from node i_1 to i_2 for commodity k . However, in each scenario, a different arc with a different capacity may be built, or the same arc may be built in different time periods. In this

case, the individual values of $y_{at}^{k\omega}$ suggest the algorithm is far from being converged, but the following would be satisfied for the majority of scenarios:

$$\sum_{a \in (i_1, i_2)} \sum_{t \in \mathcal{T}} y_{at}^{k\omega} \geq 1. \quad (44)$$

This means that for the majority of scenarios, the network for k requires an arc in the flow (i_1, i_2) . This is frequently seen, for example, when (i_1, i_2) is one of the few candidate paths connecting some local supply node to a demand point of k .

To implement this logic of convergence, we create the acceleration technique given in Algorithm 2. This algorithm checks for Constraint (44) which is defined for decision variable y . We observe in preliminary experiments that the complexity of our model is mainly driven by the y variables. As such, there is no need to consider the same logic for the variables z and β . Indeed, they can be found rather easily for a fixed set of y variables. This procedure is run once at each iteration of PHA.

Algorithm 2 addHeuristic

Input: A probability level of p_H .

- 1: **for each** pair of nodes (i_1, i_2) **do**
 - 2: **for each** commodity k **do**
 - 3: **if** total probability of bundles satisfying Constraint (44) is at least p_H **then**
 - 4: Add Constraint (44) to all bundles.
 - 5: **end if**
 - 6: **end for**
 - 7: **end for**
-

The third acceleration technique enables early termination of the PHA. As mentioned above, the majority of first-stage decision variables converge rather fast in the PHA. Often, the algorithm spends the majority of the iterations for some small discrepancies in the last decision variables. In order to accelerate, we terminate the algorithm before it is fully converged, and solve the rest of the problem with a MIP solver. One way of doing this would be fixing a subset of first-stage decision variables, and then solving the MIP for the rest. However, we implement a similar idea to the heuristic constraints mentioned above and create heuristic constraints without fixing individual decision variables. It has the same heuristic constraint structure of Algorithm 2, but for early convergence, we also check the cases where Constraint (44) sums to zero. At each iteration, Algorithm 3 is executed in our method. Once the number of these heuristic constraints reaches the target stopping condition, we solve a MIP for the base model including these constraints.

Finally, since the major complexity of the problem is due to decision variables y , as the fourth acceleration technique, we implement the progressive hedging algorithm only on decision variables y . This means, we eliminate calculating the average values of $\bar{\beta}_{at}^k$ and \bar{z}_{at}^k , and remove the corresponding Lagrangian multipliers, δ and π . Therefore, the convergence is checked only on decision variables y . This

Algorithm 3 earlyConvergence

Input: A probability level of p_E , $i_t := 0$, $i_c := 0$.

Output: The stopping condition flag (Y/N), and the set of heuristic constraints.

```
1: for each pair of nodes  $(i_1, i_2)$  do
2:   for each commodity  $k$  do
3:      $i_t := i_t + 1$ .
4:     if  $\sum_{a \in (i_1, i_2)} \sum_{t \in \mathcal{T}} y_{at}^{kw} = 0$  for all scenarios then
5:        $i_c := i_c + 1$ .
6:       Record heuristic constraint as  $\sum_{a \in (i_1, i_2)} \sum_{t \in \mathcal{T}} y_{at}^{kw} = 0$ .
7:     end if
8:     if  $\sum_{a \in (i_1, i_2)} \sum_{t \in \mathcal{T}} y_{at}^{kw} \geq 1$  for all scenarios then
9:        $i_c := i_c + 1$ .
10:      Record heuristic constraint as  $\sum_{a \in (i_1, i_2)} \sum_{t \in \mathcal{T}} y_{at}^{kw} \geq 1$ .
11:    end if
12:  end for
13: end for
14: if  $i_c/i_t \geq p_E$  then
15:   Return true with the set of heuristic constraints.
16: else
17:   Return false.
18: end if
```

would yield the Lagrangian relaxed objective function in Equation (45).

$$\begin{aligned} \min \quad & \sum_{a \in \mathcal{A}} \sum_{t \in \mathcal{T}} \sum_{k \in \mathcal{K}} \left(f_{at}^k + \lambda_{at}^{k\xi} + \frac{\rho}{2} - \rho \bar{y}_{at}^k \right) y_{at}^{k\xi} + g_{at}^k \beta_{at}^{k\xi} \\ & + \sum_{\omega \in \xi} \frac{p_\omega}{p_\xi} \left[\sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \left(\sum_{a \in \mathcal{A}} c_{at}^k x_{at}^{k\omega} + \sum_{i \in \mathcal{N}} h e_{it}^{k\omega} \right) \right]. \end{aligned} \quad (45)$$

The above-mentioned acceleration techniques are implemented in order to create our progressive-hedging-based matheuristic algorithm given in Algorithm 4.

5. Computational Experiments

This section tests the computational efficiency of our progressive-hedging-based matheuristic. Specifically, we are interested in the added value of obtaining a solution stochastically, expressed as the difference between the two-stage stochastic program solution and the solution in which we simply take the expected values of stochastic parameters. This measures the value of taking uncertainty into account in the decision-making processes. In Section 5.1, we introduce the benchmark instances and computational parameters. In Section 5.2, we give an overview of the efficiency of our developed methods. In Section 5.3, we test the impact of the bundle size on our matheuristic. In Section 5.4, we discuss the difference in

Algorithm 4 Progressive Hedging Based Matheuristic

Input: $\lambda := 0, \rho > 0, p_H, p_E, \Xi$.

Output: A feasible solution to the model (1) - (17).

- 1: **Initialization:** For each $\xi \in \Xi$, solve (2) - (17) with the objective (43) and $\Omega := \xi$.
 - 2: **while** stopping condition flag is false **do**
 - 3: **Average:** $\bar{y}_{at}^k := \sum_{\xi \in \Xi} p_\xi y_{at}^{k\xi}$.
 - 4: **Price:** $\lambda_{at}^{k\xi} := \lambda_{at}^{k\xi} + \rho(y_{at}^{k\xi} - \bar{y}_{at}^k)$.
 - 5: addHeuristic(p_H)
 - 6: **Update:** For each $\xi \in \Xi$, solve (2) - (17) with the objective (45) and $\Omega := \xi$.
 - 7: earlyConvergence(p_E)
 - 8: **end while**
 - 9: **Solution:** Solve (1) - (17) by the set of heuristic constraints.
-

solution structure when accounting for uncertainty compared to taking the expected net supply. Notice that the case study is presented in Section 6.

5.1. Benchmark Instances and Computational Parameters

All the experiments are performed using an AMD 7763 CPU (with 16 cores @ 2.45 GHz) with 32GB RAM. All algorithms are executed in C++20 in combination with Gurobi 10.0.1. We generate 20 instances for some combinations of $N \in \{8, 10, 12, 15, 20\}$, $T \in \{8, 10, 15\}$, and $W \in \{60, 120, 240\}$. Each of these instances is solved by our progressive-hedging-based matheuristic (PHM) and Gurobi (GRB). We also obtain the expected value solution (EEV) by solving a single-scenario expected value problem and then simulating the performance of this solution over all scenarios in Ω . For PHM, we apply a MIP gap of 10% for the *Initialization* and *Update* steps of Algorithm 4. We further apply a time limit of 1 hour and a MIP gap of 1% for the *Solution* step. We set p_H and p_E to 0.2 and 1, respectively. We multiply these probabilities by 0.97 in each iteration of the algorithm for earlier convergence. We select the bundle size of PHM as 6 for the instances with $W = 60$, and 8 for $W = \{120, 240\}$ according to the preliminary experiments. In Section 5.3 we provide further insights on the bundle size. We solve the base model (1)-(17) via GRB. The EEV solution is obtained by taking the average of all stochastic parameters to create an average scenario for each instance. Both GRB and EEV have a 3-hour time limit. The GRB MIP gap is set as 1%. We believe this is a competitive setting to test for GRB, since without these limits it either takes too much time to solve the problem or the chances of going out of memory are too high.

For further details on instance generation and scenario construction, please consult Appendix Appendix A.

5.2. Performance of the Matheuristic

In this section, we compare our progressive-hedging-based matheuristic (denoted as PHM) with the MIP solution of the base model (denoted as GRB), and the expected value solution (denoted by EEV). A summary of the average performance over all instances is provided in Table 1.

Table 1: Results the relative performance of PHM and EEV compared to GRB

N	A	T	W	$\#_{\text{GRB}}$	$\#_{\text{PHM}}$	PHM				EEV				
						Gap (%)		Time (%)		Gap (%)		Time (%)		
						min	avg	min	avg	min	avg	min	avg	
8	79	8	60	20	20	-0.31	1.34	0.56	3.21	4.28	12.35	0.13	0.72	
			120	20	20	-0.33	1.61	0.40	1.59	4.50	12.15	0.05	0.34	
			10	60	20	20	0.16	1.24	0.60	2.32	4.27	12.32	0.09	1.19
			120	14	20	0.01	1.46	0.83	3.09	4.06	10.46	0.08	0.78	
10	104	8	60	20	20	0.70	2.52	0.44	5.65	6.17	16.91	0.02	0.93	
			120	16	20	0.10	1.83	0.67	5.38	5.28	16.29	0.03	0.26	
			10	60	19	20	0.51	1.70	0.64	7.11	6.67	17.36	0.05	1.39
			120	10	20	-10.87	0.00	2.09	8.50	5.24	15.63	0.05	0.77	
12	134	8	60	20	20	0.33	1.41	0.45	6.72	8.00	20.23	0.05	0.76	
			120	9	20	-0.44	1.81	1.13	14.08	6.63	18.46	0.04	1.14	
			10	60	20	20	0.35	1.91	0.72	11.14	10.13	21.16	0.19	2.03
			120	10	20	-8.49	0.27	1.82	23.44	6.70	17.42	0.16	8.10	
15	179	10	120	16	19	-17.06	-8.98	9.22	28.94	8.19	18.26	1.39	31.58	
			240	0	20	-	-	13.69	36.88	-53.25	11.57	0.80	27.70	
20	258	10	240	0	6	-	-	10.78	35.43	6.17	9.26	1.09	71.95	
			15	240	0	1	-	-	47.50	47.50	10.38	10.38	1.63	91.73

Each row in Table 1 denotes the performance over 20 instances. Column A denotes the average number of arcs over the 20 instances, which is not a constant as potential arcs are generated based on the customer locations in each individual instance, which vary. The columns $\#_{\text{GRB}}$ and $\#_{\text{PHM}}$ are the number of instances that are solved without hitting the memory limit. We do not report the number of solutions for the EEV, since it stays within the memory limit for all instances. The following columns are relative comparisons. First, we compare PHM relative to GRB both in objective value (Gap (%)) and time (Time (%)). These values are derived from the instances that are solved within the memory limits both with GRB and PHM. In case $\#_{\text{GRB}} = 0$, we compare the solution time of PHM with the limit of 3 hours. The same information is provided for the EEV solution, however here, we compared the objective to the best-found solution by either GRB or PHM.

We observe that our algorithm can find on average solutions that are within 1 – 2% of optimality by only spending 5–6% of the MIP solution time for smaller instances. Moreover, our algorithm outperforms GRB as shown by the negative value for “Gap” for some of the instances. It also has fewer memory issues. This is significant for larger-sized instances, as most of these instances cannot be solved by GRB. The time ratio on average increases with larger-sized instances, but this is mostly because GRB has a fixed 3-hour time limit. This partly results in an average 8.98% improvement by PHM compared to GRB for the instances with $N = 15$, $T = 10$, and $W = 120$, while PHM spends on average only 28.94% of the time GRB takes on these instances.

We also make comparisons with the EEV. In the column “Gap” of EEV, the EEV solution is compared with the best solution found via GRB and PHM. Note for the cases that GRB cannot solve any instance, this is always the solution provided by PHM. The solution time of EEV is the percentage of GRB time, except for the cases $\#_{\text{GRB}} = 0$ where we compare the EEV time by 3 hours. While EEV works fast on small instances, we observe an overall gap of 10 – 20% in its solution quality. Moreover, there is a significant increase in its time compared to GRB on larger instances, since GRB starts hitting the 3-hour limit more and more. PHM has similar time ratios on the same larger instances while having better results for the GAP results.

It is important to note that for the instances with $N = 20$, PHM improves the EEV solution by about 10% while spending half of the EEV solution time. This is significant since while the EEV is a 1-scenario problem, the PHM deals with thirty 8-scenario problems on $W = 240$. The algorithm’s success can be attributed not only to the high MIP gap but also to the variable fixing through heuristic constraints, which enhances efficiency. Each iteration is solved as fast as possible, and at the end of each iteration, we fix decision variables as much as possible while keeping overall solution quality high. Only in one extreme case on $N = 15$, $T = 10$, and $W = 240$, we observe that the EEV is significantly better (due to PHM not converging quickly enough) as it provides a solution with a -53.25% gap. However, even when this is included, on average these instances are improved by 11.57% in terms of the solution quality by PHM.

5.3. Impact of Scenario Count per Bundle on Solution

This subsection analyzes how varying the number of scenarios per bundle affects the solution quality of the PHM. We randomize various instances by using the same parameter settings as before. We test various numbers of scenarios per bundle, i.e. $|\xi| \in \{1, 2, 3, 4, 6, 8, 12, 24\}$ to randomly assign the scenarios into bundles. We report their solution in terms of the quality and the solution time in Figure 1, where Figure 1a and Figure 1b provide the individual and the average behavior, respectively.

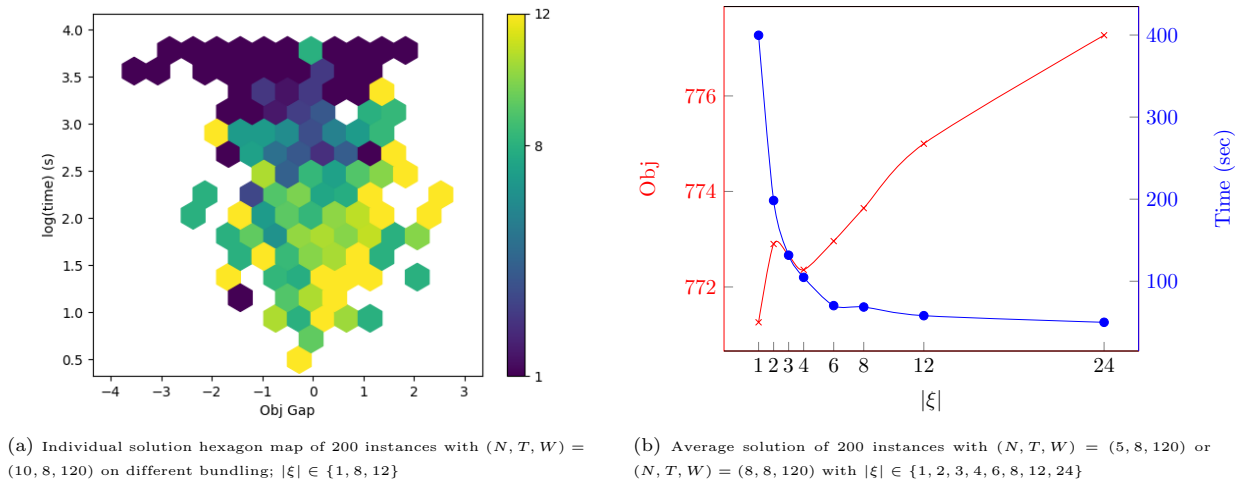


Figure 1: The effects of $|\xi|$ in solution quality and computational time

We provide a hexagon scatter plot in Figure 1a for a subset of the number of scenarios considered. We solve a total of 200 instances with $(N, T, W) = (10, 8, 120)$ for each value $|\xi| \in \{1, 8, 12\}$ by grouping

scenarios randomly. The colors represent the composition of data included in the hexagon. The objective gap is normalized; it is the gap to the average objective taken from the same instances for all number of scenarios in $\{1, 2, 3, 4, 6, 8, 12, 24\}$. The time is reported in a logarithmic scale, for example, a score of 1.5 means it takes $10^{1.5}$ seconds to solve. The figure shows the trade-off between the solution quality and the solution time. While increasing the number of scenarios per bundle decreases the solution quality in terms of the objective value, it also decreases the solution time on average. We like to stress that the differences in objective value are small.

A similar behavior is seen in Figure 1b. The figure provides the average objective value and corresponding solution times of a total of 200 instances with either $(N, T, W) = (5, 8, 120)$ or $(N, T, W) = (8, 8, 120)$. Each of these instances is solved for various randomized bundles with $|\xi| \in \{1, 2, 3, 4, 6, 8, 12, 24\}$. We observe that the trade-off observed in Figure 1a persists, except for $|\xi| \in \{2, 3\}$. This trade-off is due to using our soft variable fixing approach, which incorporates heuristic constraints when a certain percentage of instances share common pipeline arcs. As $|\xi|$ increases, even a few bundles sharing the same arc can trigger the heuristic constraints. In the highest value of $|\xi| = 24$, heuristic constraints are introduced even if only one bundle contains a particular arc. Because when $W = 120$, $p_H = 20\%$ of the solutions consist of exactly one bundle with $|\xi| = 24$ scenarios. Thus, with higher values of $|\xi|$, the incorporation of heuristic constraints becomes more frequent. While this restriction on the solution space indeed tends to increase the objective value, it also leads to a decrease in the solution time.

For $|\xi| = \{1, 4, 6, 8, 12, 24\}$, the decision-maker may select the trade-off between solution quality and solution time. Regardless of the chosen trade-off, our method consistently exhibits small objective gaps across all values of $|\xi|$, underscoring its robustness and efficacy. We select either $|\xi| = 6$ or $|\xi| = 8$ for the computational experiment instances. This choice is influenced by the observation that the average objective gap appears to be relatively insensitive to the number of scenarios in our particular setting, and for 6 or more values the solution time decreases sharply. However, the objective gap range highly depends on the instance parameters selected. For instance, scenarios characterized by higher uncertainty in supply/demand and high penalty costs may exhibit more substantial differences in objective value. In this case, it might be sensible to decrease $|\xi|$ to find a better balance between the objective function value and the solution time. Similarly, for larger instances requiring extended computational time, it may be advantageous to increase the number of scenarios $|\xi|$ even further, aiming to find some feasible solutions in reasonable computational time. This is particularly important because as computational time increases, the likelihood of encountering memory issues also rises.

5.4. *Timing of the Transition in the Hydrogen Economy*

In this section, we investigate the dynamics of transitioning from a natural gas network to one primarily based on hydrogen. We conduct 100 instances on each of the parameter sets $(N, T, W) = (8, 10, 60)$, representing a smaller network with fewer nodes and potential connections, and $(N, T, W) = (12, 10, 60)$, illustrating a larger, more complex network structure. We compare the average proportion of network arcs allocated to hydrogen transportation in each period. All these instances are solved via PHM and EEV, with $|\xi| = 6$ for PHM. Figure 2 illustrates the evolution of hydrogen implementation

over time, where the y-axis represents the fraction of arcs being designated for hydrogen transport.

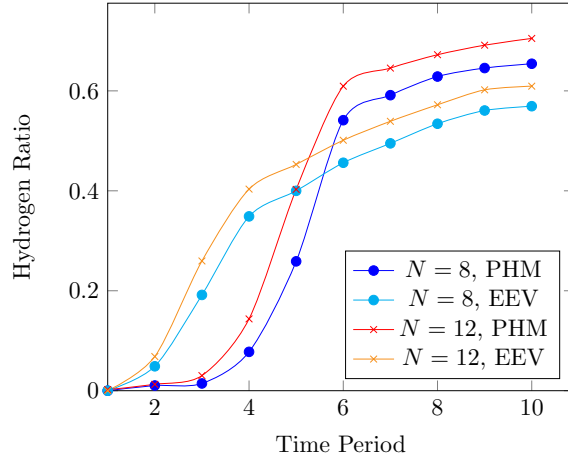


Figure 2: Hydrogen Implementation Over Time

As depicted in Figure 2, by the tenth period, about 65 – 70% of network arcs should be designated for hydrogen transport according to the PHM solution. Nevertheless, roughly 30% of the natural gas infrastructure endures even though we assume that the natural gas economy mostly phases out by $t = 10$ in the instances. This is due to the inherent incompatibilities between certain natural gas arcs and the evolving hydrogen transport routes. These incompatibilities arise from various factors. Firstly, some natural gas pipelines may require new counterparts to serve new demands arising in the future hydrogen economy, for example with pipelines at a different capacity. Secondly, certain natural gas supplier locations at $t = 0$ may not serve as hydrogen suppliers in the future. In such cases, the pipelines originating from these locations are less likely to be converted for hydrogen transport, leading to the existence of pipelines that are no longer in use at $t = 10$. This also partly explains that the end ratio of hydrogen pipelines in the case of $N = 12$ is about 5% higher than in the case of $N = 8$. With more locations in the problem instance, there are increased opportunities for retrofitting existing natural gas infrastructure for hydrogen transport due to the denser network with more supply and demand locations, resulting in a more effective transition.

It is also noteworthy that there is a difference in the average behaviors between the two solution approaches, PHM and EEV. The EEV suggests converting pipelines earlier in the transition process, with approximately 40% of arcs allocated to hydrogen by $t = 4$, whereas PHM only has about 10% at the same time in our instances. Conversely, the EEV solutions result in about 8% less conversion compared to the PHM at $t = 10$. These disparities arise primarily due to the way the uncertainty is handled. PHM adopts a multi-scenario solution approach, which inherently accounts for a range of potential scenarios, including extreme cases. In contrast, EEV employs a deterministic equivalent of the problem, assuming a single-scenario (mean) perspective, which leads to a more optimistic outlook. With this mean-value perspective, EEV solutions tend to initiate pipeline construction earlier, assuming a degree of certainty in decision-making. Meanwhile, PHM focuses on optimizing the timing of pipeline construction to minimize overall costs, often favoring delayed pipeline construction—which is cheaper due to the assumption that costs decrease over time—while paying the penalty of unsatisfied demand for

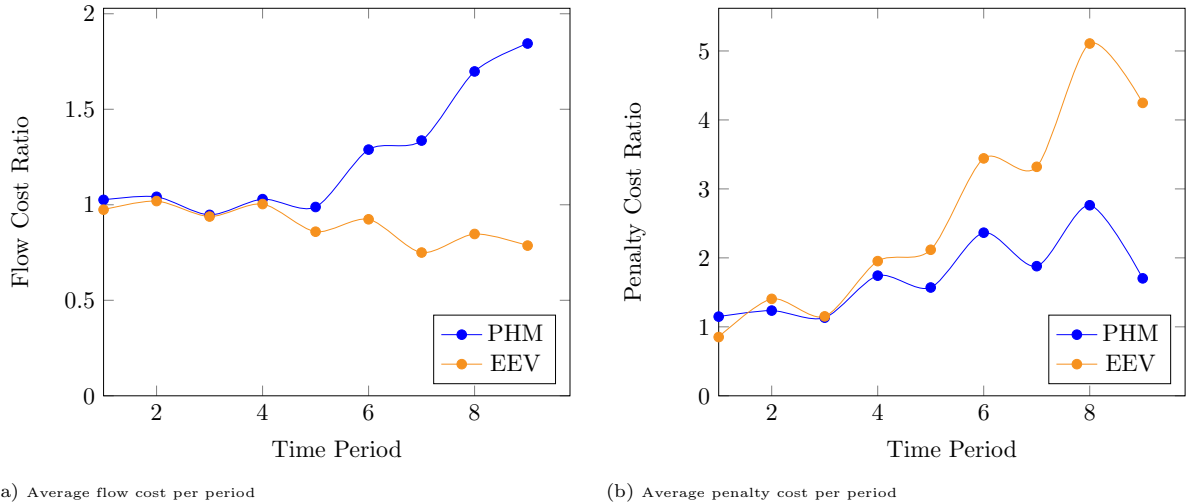


Figure 3: Expected flow and penalty cost per period of PHM and EEV

the extreme cases at early periods. A similar pattern is seen, for example, in other investment decisions under uncertainty. Deterministic approaches might lean towards early investments due to anticipated benefits, while stochastic methods delay them to hedge against potential risks (see, e.g., Dixit and Pindyck, 1994).

Nevertheless, due to the additional uncertainty information considered, PHM’s solution proves to be more cost-efficient in the long run. Figure 3a and 3b reflect the average per period cost among two sets of instances, $(N, T, W) = (8, 10, 60)$ and $(N, T, W) = (12, 10, 60)$, where EEV values are from simulations with suggested construction policies. The costs are provided as ratios to the average cost of PHM and EEV at $t = 1$. Figure 3a demonstrates that PHM increasingly succeeds in satisfying the demand through flow costs as the planning horizon progresses. The widening gap in flow costs indicates a better covering of the demand of the network, which results in PHM incurring lower penalty costs compared to EEV, especially evident in later periods as shown in Figure 3b. We also observe a periodic pattern in costs due to seasonal factors, which is further explained in Appendix Appendix A. The observed differences are most pronounced towards the end of the time horizon, as period $t = 10$ mimics sustained long-term use beyond the planning horizon, with significant penalty costs for any unmet demand at $t = 10$.

In summary, the interplay between infrastructure expansion and emerging supply demands adds depth to our understanding of the gradual transition from natural gas to hydrogen within the network. We specifically observe a series of periods, mainly occurring in the middle of the planning horizon, marked by a rapid and pronounced transition. Planning and budgeting for these middle periods will be crucial, as they require more resources. This aligns with current energy transition plans in the HEAVENN project as well, where a small adoption of hydrogen until 2040, followed by a rapid growth phase from 2040 to 2050 is expected. Our results mirror this trend.

6. Case Study and Discussion: Transitioning Natural Gas Network to Hydrogen in the HEAVENN Region

In this section, we discuss the implications of transitioning the extensive natural gas network within the Northern Netherlands region into a green hydrogen distribution infrastructure. This is a critical step towards achieving a climate-neutral economy as is also a part of the project *Hydrogen Energy Applications in Valley Environments for Northern Netherlands* (HEAVENN, 2022), which is Europe’s first hydrogen valley initiative. While new hydrogen pipelines are being developed, there is a parallel effort to repurpose the existing natural gas network for hydrogen transport, a cost-effective and environmentally friendly alternative. Based on discussions with our project partners, we focus on a set of locations representing the main critical areas in the network, as shown in Figure 4. These locations encompass five distinct types: suppliers, hydrogen refueling stations or analogous demand units for natural gas, residential areas with heating gas demand, industrial zones requiring gas as a feedstock, and intermediate/storage points facilitating pipeline connections.

We investigate two variants of the network structure to capture different supply strategies that may emerge in the region. The first variant relies on a single supplier, the Eemshaven port in our specific case, representing the import of hydrogen from abroad, a possible scenario foreseen for regions without sufficient local hydrogen production facilities. In the second variant, we study a situation where several smaller suppliers are spread across the Northern Netherlands region, producing hydrogen locally. For clarity, we refer to the first variant as *Eemshaven import*, and the second variant as *decentralized production*. While we anticipate a blend of both import and local production, our focus on these two settings aims to offer distinct insights into their potential consequences. The *Eemshaven import* variant is characterized by the certainty at the supply location, unlike the *decentralized production* variant which has variability in supplier locations across scenarios. Both variants start with the initial natural gas network structure represented in Figure 4. On this starting network, we consider a subset of major pipelines in the region. Capacities for these pipelines are categorized into three levels: S, M, and L, representing small, medium, and large capacities, respectively.

We assign parameter values based on consultations with our industry partners in the Northern Netherlands region and from prior studies conducted within the HEAVENN project (Bakker, 2023; Pol, 2023). Key parameters include information on pipeline transition costs, possible pipeline capacities, initial network of natural gas, seasonality on demand, and the expected pace of the transition from natural gas to hydrogen. The majority of the parameters align with those used in the computational experiments (see also Appendix Appendix A). Distinctively, we adjust pipeline conversion cost as 30% of the fixed cost of construction of the same pipeline. For the progressive hedging algorithm, we employ the same parameters as in the computational experiments. Our instance generation method is detailed in Appendix Appendix A, and it is implemented with 60 scenarios.

6.1. Results

In our analysis, we address two distinct network supply settings. As previously outlined, the *Eemshaven import* variant involves a single supplier, situated at the Eemshaven, representing the import

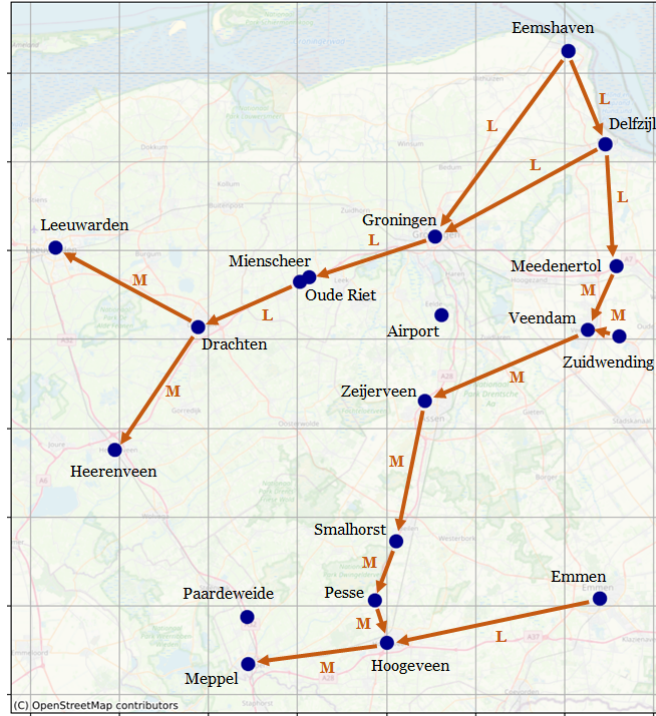


Figure 4: Initial natural gas network

of hydrogen from external sources. In contrast, the *decentralized production* variant explores a scenario where multiple local suppliers are dispersed across the Northern Netherlands region, with a specific focus on uncertainty regarding their locations. The candidate locations for these suppliers are Eemshaven, Delfzijl, Groningen, Zeijerveen, Zuidwending, and Eelde Airport. We randomly assign these locations as supply nodes or intermediary nodes in each of the 60 scenarios. Additionally, this allocation process also establishes whether these nodes can function as storage facilities.

The pipeline connections for hydrogen at the latest period ($t = T$) are given in Figure 5 for both variants. As a comparison, we also show the deterministic (EEV) solution for both variants to highlight the need to include uncertainty in decision-making for a future hydrogen economy. For clarity in representation, when at least two pipelines exist between a given pair of locations, we denote the capacity label of the largest capacity among the pipelines. Pipelines transporting green hydrogen are green and those for natural gas are orange. Note that despite the phase-out of the natural gas economy at the latest period, the graph still includes pipelines belonging to the old natural gas network. Those are the pipelines where repurposing for hydrogen use was not deemed necessary during the planning horizon. For a more detailed discussion on this behavior, see Section 5.4.

In Figure 5a, we observe a pipeline network that primarily extends from Eemshaven (most upper-right node) to various destinations, mostly optimized for the shortest paths to their respective endpoints. In contrast, our analysis of the decentralized production reveals a more crowded network in Figure 5b, with a higher number of pipelines near potential supply locations. The total installation capacity remains relatively consistent between both variants. The Eemshaven variant picks for pipelines with greater capacity, resulting in fewer pipelines overall. Conversely, the decentralized production variant

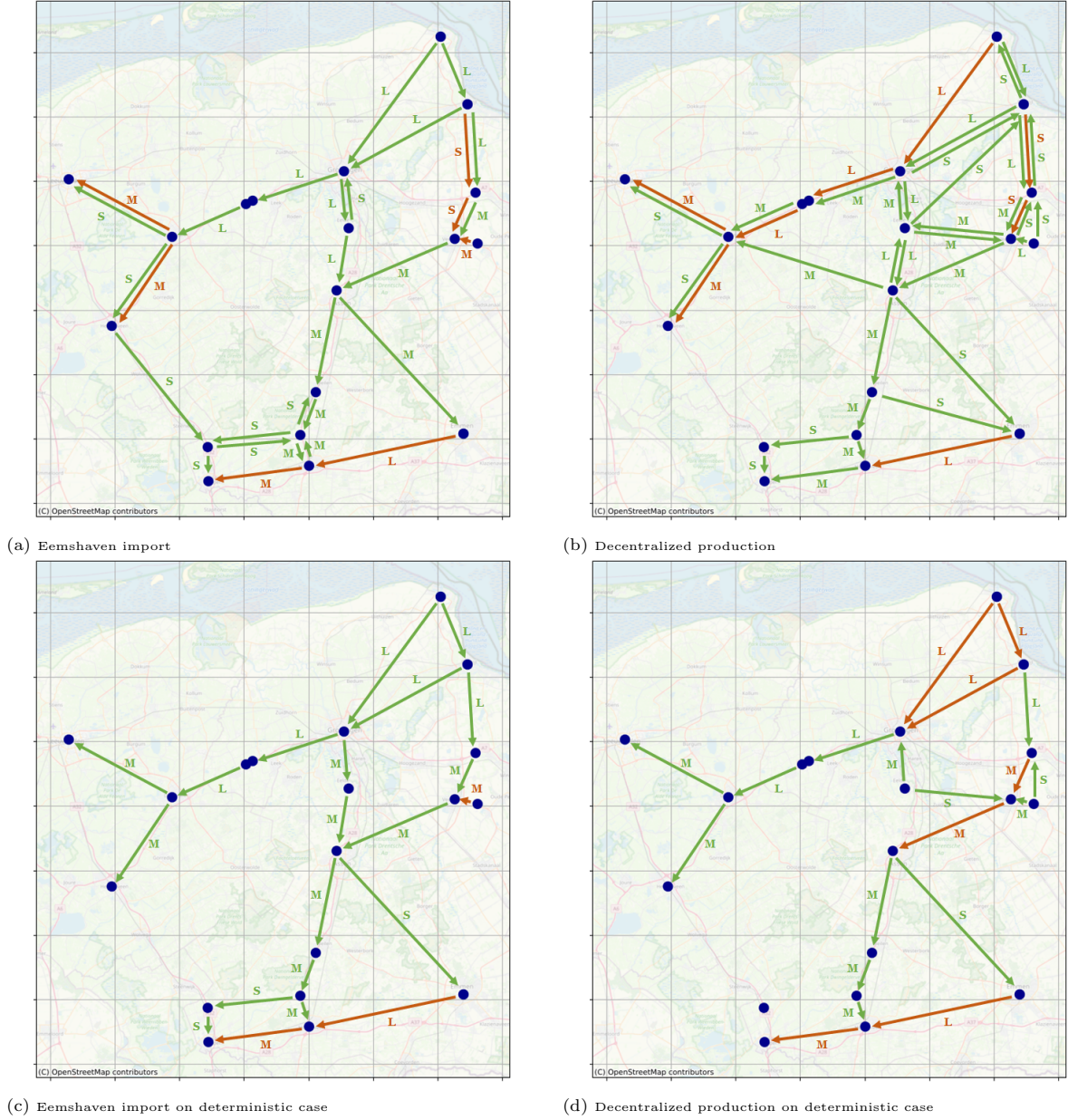


Figure 5: Hydrogen networks selected at $t = T$

features a larger number of pipelines, each with smaller capacities. In Appendix Appendix B, we provide ten figures detailing the evolution of the suggested energy network over each time period for decentralized production. The majority of the network transition is observed around the center of the horizon, specifically at $t = 4$.

The Eemshaven variant repurposes more of its initial natural gas pipeline network for hydrogen as compared to the decentralized variant. Specifically, the ratio of conversion to construction costs is 70% higher in Eemshaven variant than in the decentralized variant. Since the conversion costs are lower compared to the new construction costs, this leads to 17.5% less total infrastructure costs on the Eemshaven variant than the decentralized variant. However, the Eemshaven variant also experiences approximately 50% more penalty costs than the decentralized variant due to the greater need for truck transportation in times of peak demand. This difference arises because the extensive network of the

decentralized variant, even with lower capacities, provides a more reliable setting for unexpected increases in demand.

We additionally examine the cases where we use the mean values and solve the deterministic version of the problem. For the deterministic case of the Eemshaven, the recommended policy in Figure 5c spends 60% of the total infrastructure costs observed in its stochastic equivalent. However, due to the high penalty costs incurred in the deterministic policies, the suggested pipeline network translates to an expected overall cost increase of 20%. The disparity becomes more pronounced in the decentralized production variant. In the deterministic case of the decentralized production, because only 22% of the total infrastructure cost of its stochastic equivalent is allocated as in Figure 5d, many extreme scenarios are overlooked, resulting in a 115% rise in overall expected costs. This significant increase is partly attributed to certain locations being completely omitted from green hydrogen pipeline connections due to their low likelihood of being essential in the long-term perspective, reflecting the inherent uncertainties surrounding locations. As these uncertainties increase, deterministic solutions become more inadequate, often omitting pipeline connections altogether.

Our discussions of our findings with project partners have highlighted some additional points of network supply strategies. Ports like Eemshaven, for instance, are ideal locations for hydrogen production due to their abundant access to resources essential, such as land and water. The findings from our case study indicate that centralized supply already leads to lower total infrastructure costs. However, it has a higher risk of incurring penalty costs to cover peak demand. This is due to its efficient but less dense infrastructure network as compared to the decentralized production. Thus, if decision-makers proceed with the suggested option of centralized production, we recommend they concentrate on strategic plans where the effects of uncertainties can be decreased, such as by investing in sufficient storage locations.

In both variants, our analysis assumes sufficient supply to meet demand. Therefore, unsatisfied demand via pipelines mainly results from the consideration that it is more cost-effective to incur penalty costs than to invest in new pipeline construction, along with the associated flow costs. In situations where supply is limited, our analysis indicates that optimal solutions prioritize meeting demand at locations closer to supply points, which is evidently more cost-efficient due to reduced construction and the flow costs for closer customers. The outcome could differ, however, if the decision maker assumes varying levels of importance to customers. For instance, higher penalty costs imposed on significant customers may influence the optimal solution, potentially leading to different pipeline structures to meet their demand.

6.2. Summary of Case Study

In summary,

- o We analyzed two distinct network supply strategies: one with a single supplier at Eemshaven and the other involving multiple local suppliers adding location uncertainty.
- o The Eemshaven variant displayed a simpler pattern with pipelines radiating mostly directly from Eemshaven, while the decentralized production variant featured a more crowded network with numerous smaller-capacity pipelines near potential supply locations.

- o The Eemshaven variant exhibits a higher level of integration of the initial natural gas network, driven by the optimistic assumption of location certainty and the alignment with the existing pipeline structure’s flow direction. However, this led to 50% more penalty costs for the need of trucks compared to the decentralized production variant due to partially neglecting peak demand scenarios.
- o While the Eemshaven variant reduces total infrastructure costs by 17.5%, the single supplier location offers economies of scale and easy access to resources like land and water at the port, which is important for strategic energy transition decisions. Conversely, decentralized production ensures energy independence and extensive pipeline coverage for peak demand scenarios.
- o Finally, as uncertainty increases, deterministic policies lead to insufficient pipeline infrastructures. This results in sharp increases in penalty costs, leading to an increased total expected cost of the network design.

7. Conclusions

In this paper, we address the energy transition network design problem over a finite planning horizon. The problem focuses on strategic decision-making associated with the construction and conversion of natural gas pipelines into green hydrogen pipelines, and on balancing supply and demand in a transitioning network under uncertain conditions. We provide a novel two-stage stochastic mixed integer programming model with continuous recourse to formulate our stochastic network design problem on energy transition. In the first stage, decisions regarding the construction and conversion of pipelines are made. The second stage, including uncertainties represented by a finite number of scenarios, determines the optimal energy flow decisions and derives corresponding inventory storage policies.

Our algorithm consistently yields solutions that are near-optimal, outperforming benchmark methods in the balance of solution quality and computational time. We also use our algorithm on a case study in the Northern Netherlands, as part of the *Hydrogen Energy Applications in Valley Environments for Northern Netherlands* (HEAVENN) initiative. This real-world application not only emphasizes the practical relevance of our model but also highlights its adaptability, particularly in scenarios with high uncertainties. We observe that uncertainty in future supplier locations is a key factor for network design decisions.

For future studies, several suggestions arise. Firstly, exploring multi-stage recourse models for this problem could offer a more nuanced approach to decisions across different periods. Second, considering heterogeneous customers might be practical, leading to a different penalty cost per customer. Third, while our study focuses on pipeline network transition, it captures uncertainties regarding changes in locations over time. An extension of this study might also decide on optimal infrastructure placements, like suppliers and storage units. Lastly, integrating principles of land-use planning or urban development might offer insights into how the strategic positioning of hydrogen refueling stations or households influences energy costs in the network.

References

- Almansoori, A., Shah, N., 2012. Design and operation of a stochastic hydrogen supply chain network under demand uncertainty. *International Journal of Hydrogen Energy* 37, 3965–3977.
- André, J., Auray, S., Brac, J., De Wolf, D., Maisonnier, G., Ould-Sidi, M.M., Simonnet, A., 2013. Design and dimensioning of hydrogen transmission pipeline networks. *European Journal of Operational Research* 229, 239–251.
- André, J., Auray, S., De Wolf, D., Memmah, M.M., Simonnet, A., 2014. Time development of new hydrogen transmission pipeline networks for France. *International Journal of Hydrogen Energy* 39, 10323–10337.
- Bakker, R., 2023. Roadmap towards the green hydrogen economy: design of a hydrogen energy system for the Northern Netherlands hydrogen valley. Master’s thesis. Rijksuniversiteit Groningen. the Netherlands.
- Baxter, M., Elgindy, T., Ernst, A.T., Kalinowski, T., Savelsbergh, M.W., 2014. Incremental network design with shortest paths. *European Journal of Operational Research* 238, 675–684.
- Boland, N., Hewitt, M., Marshall, L., Savelsbergh, M., 2017. The continuous-time service network design problem. *Operations Research* 65, 1303–1321.
- Crainic, T.G., Fu, X., Gendreau, M., Rei, W., Wallace, S.W., 2011. Progressive hedging-based meta-heuristics for stochastic network design. *Networks* 58, 114–124.
- Crainic, T.G., Hewitt, M., Maggioni, F., Rei, W., 2021. Partial Benders decomposition: General methodology and application to stochastic network design. *Transportation Science* 55, 414–435.
- Crainic, T.G., Hewitt, M., Rei, W., 2014. Scenario grouping in a progressive hedging-based meta-heuristic for stochastic network design. *Computers & Operations Research* 43, 90–99.
- Dixit, A.K., Pindyck, R.S., 1994. *Investment under uncertainty*. Princeton university press.
- Engel, K., Kalinowski, T., Savelsbergh, M.W., 2017. Incremental network design with minimum spanning trees. *Journal of Graph Algorithms and Applications* 21, 417–432.
- Espegren, K., Damman, S., Pisciella, P., Graabak, I., Tomasgard, A., 2021. The role of hydrogen in the transition from a petroleum economy to a low-carbon society. *International Journal of Hydrogen Energy* 46, 23125–23138.
- European Union, 2012. *Energy: Roadmap 2050*. Publications Office of the European Union. URL: <https://data.europa.eu/doi/10.2833/10759>, doi:10.2833/10759.
- Fragkos, I., Cordeau, J.F., Jans, R., 2021. Decomposition methods for large-scale network expansion problems. *Transportation Research Part B: Methodological* 144, 60–80.

- Ghamlouche, I., Crainic, T.G., Gendreau, M., 2003. Cycle-based neighbourhoods for fixed-charge capacitated multicommodity network design. *Operations Research* 51, 655–667.
- Gordon, J.A., Balta-Ozkan, N., Nabavi, S.A., 2023. Socio-technical barriers to domestic hydrogen futures: Repurposing pipelines, policies, and public perceptions. *Applied Energy* 336, 120850.
- Gray, P., 1971. Exact solution of the fixed-charge transportation problem. *Operations Research* 19, 1529–1538.
- HEAVENN, 2022. Hydrogen energy applications in valley environments for Northern Netherlands. URL: <https://heavenn.org>.
- Hewitt, M., Crainic, T.G., Nowak, M., Rei, W., 2019. Scheduled service network design with resource acquisition and management under uncertainty. *Transportation Research Part B: Methodological* 128, 324–343.
- Hewitt, M., Rei, W., Wallace, S.W., 2021. Stochastic network design. *Network Design with Applications to Transportation and Logistics* , 283–315.
- Hodge, B.M., Martinez-Anido, C.B., Wang, Q., Chartan, E., Florita, A., Kiviluoma, J., 2018. The combined value of wind and solar power forecasting improvements and electricity storage. *Applied Energy* 214, 1–15.
- Hu, S., Han, C., Dong, Z.S., Meng, L., 2019. A multi-stage stochastic programming model for relief distribution considering the state of road network. *Transportation Research Part B: Methodological* 123, 64–87.
- Jiang, X., Bai, R., Wallace, S.W., Kendall, G., Landa-Silva, D., 2021. Soft clustering-based scenario bundling for a progressive hedging heuristic in stochastic service network design. *Computers & Operations Research* 128, 105182.
- Johnson, N., Ogden, J., 2012. A spatially-explicit optimization model for long-term hydrogen pipeline planning. *International Journal of Hydrogen Energy* 37, 5421–5433.
- Kalinowski, T., Matsypura, D., Savelsbergh, M.W., 2015. Incremental network design with maximum flows. *European Journal of Operational Research* 242, 51–62.
- Komdeur, K.R., 2020. Modeling and optimization of hydrogen transmission pipeline networks. Master’s thesis. Rijksuniversiteit Groningen. the Netherlands.
- Li, L., Manier, H., Manier, M.A., 2019. Hydrogen supply chain network design: An optimization-oriented review. *Renewable and Sustainable Energy Reviews* 103, 342–360.
- Megía, P.J., Vizcaíno, A.J., Calles, J.A., Carrero, A., 2021. Hydrogen production technologies: From fossil fuels toward renewable sources. A mini review. *Energy & Fuels* 35, 16403–16415.

- Müller, J.P., Elbert, R., Emde, S., 2021. Intermodal service network design with stochastic demand and short-term schedule modifications. *Computers & Industrial Engineering* 159, 107514.
- New Energy Coalition, 2020. The Northern Netherlands hydrogen investment plan 2020: Expanding the Northern Netherlands hydrogen valley. Technical Report. URL: <https://www.newenergycoalition.org/custom/uploads/2020/10/investment-plan-hydrogen-northern-netherlands-2020-min.pdf>.
- Nunes, P., Oliveira, F., Hamacher, S., Almansoori, A., 2015. Design of a hydrogen supply chain with uncertainty. *International Journal of Hydrogen Energy* 40, 16408–16418.
- Paraskevopoulos, D.C., Bektaş, T., Crainic, T.G., Potts, C.N., 2016. A cycle-based evolutionary algorithm for the fixed-charge capacitated multi-commodity network design problem. *European Journal of Operational Research* 253, 265–279.
- Pol, J., 2023. Design of a global hydrogen import system to the Netherlands: A design science study. Master's thesis. Rijksuniversiteit Groningen. the Netherlands.
- Rahmaniani, R., Crainic, T.G., Gendreau, M., Rei, W., 2018. Accelerating the Benders decomposition method: Application to stochastic network design problems. *SIAM Journal on Optimization* 28, 875–903.
- Reuß, M., Welder, L., Thürauf, J., Linßen, J., Grube, T., Schewe, L., Schmidt, M., Stolten, D., Robinius, M., 2019. Modeling hydrogen networks for future energy systems: A comparison of linear and nonlinear approaches. *International Journal of Hydrogen Energy* 44, 32136–32150.
- Rockafellar, R.T., Wets, R.J.B., 1991. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research* 16, 119–147.
- Safari, A., Das, N., Langhelle, O., Roy, J., Assadi, M., 2019. Natural gas: A transition fuel for sustainable energy system transformation? *Energy Science & Engineering* 7, 1075–1094.
- Sarayloo, F., Crainic, T.G., Rei, W., 2021. A learning-based metaheuristic for stochastic multicommodity network design. *INFORMS Journal on Computing* 33, 643–656.
- Staffell, I., Scamman, D., Abad, A.V., Balcombe, P., Dodds, P.E., Ekins, P., Shah, N., Ward, K.R., 2019. The role of hydrogen and fuel cells in the global energy system. *Energy & Environmental Science* 12, 463–491.
- Thapalia, B.K., Wallace, S.W., Kaut, M., Crainic, T.G., 2012. Single source single-commodity stochastic network design. *Computational Management Science* 9, 139–160.
- Tsiklios, C., Hermesmann, M., Müller, T., 2022. Hydrogen transport in large-scale transmission pipeline networks: Thermodynamic and environmental assessment of repurposed and new pipeline configurations. *Applied Energy* 327, 120097.

Wang, X., Crainic, T.G., Wallace, S.W., 2019. Stochastic network design for planning scheduled transportation services: The value of deterministic solutions. *INFORMS Journal on Computing* 31, 153–170.

Watson, J.P., Woodruff, D.L., 2011. Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *Computational Management Science* 8, 355.

Appendix A. Instance Generation

In this section, we outline the procedure for generating instances for our study. Broadly, each instance is characterized by a **graph**, associated **costs**, and instance **parameters**. We discuss these components in sequence in this section. We determine costs and parameters on the expert recommendations of the HEAVENN project, and tailored them for our computational experiments, ensuring valuable insights for readers. Parameters associated with the solution methodology are elaborated in Section 5.2.

For the **graph**, nodes are uniformly randomized within a 10×10 square area. Subsequently, nodes are categorized based on their functional roles. A node can be designated as

1. a supply source,
2. a hydrogen refueling station or an analogous demand unit for natural gas,
3. a household region requiring gas for heating,
4. an industrial hub requiring gas as a feedstock,
5. and an intermediate node for pipeline connections.

Then, nodes are assigned these roles based on the following probabilities: $\{0.2, 0.3, 0.2, 0.1, 0.2\}$, respectively, different for each commodity. However, we allow for the possibility that certain locations may assume different roles depending on the scenario, introducing an element of uncertainty particularly for the type of location for supply (1) and intermediate nodes (5). Thus, a location selected as a supply node in one scenario of an instance might serve as an intermediate node in another scenario of the same instance. We do not, however, introduce this uncertainty for demand locations (2, 3, and 4), as the demand nodes are known beforehand with high certainty, e.g. household areas. Then, we establish potential pipeline connections between nodes. Specifically, given a predefined value of U^d and the distance d_{ij} between nodes i and j , if there exists at least one node ℓ such that $d_{ij} * U^d > d_{i\ell} + d_{\ell j}$, the candidate arcs between nodes i and j are all discarded. This reflects the practical observation that pipelines typically do not span long distances directly between two nodes. Instead, they traverse multiple intermediate nodes. For our computational experiments, we set U^d to 1.15. If no such ℓ node exists, then we add a total of 6 arcs to \mathcal{A} , three pipelines in each direction between i and j , each with diameters of $\{30, 75, 120\}$.

Following the graph generation, we define the associated **costs**. The fixed cost for pipeline construction is determined using the formula from André et al. (2013) as $(a_0 + a_1 u_{ij} + a_2 u_{ij}^2) \times d_{ij}$ where a parameters are coefficients and u_{ij} represents the diameter. Based on an area-specific study (Komdeur, 2020), we select the coefficients as $a_0 = 25/6$, $a_1 = 5/72$, and $a_2 = 1/5400$, by normalizing them to our distance scheme. We implement these values for all $f_{a_0}^k$. Motivated by considerations of opportunity

costs and economies of scale, the fixed costs for each pipeline linearly decrease to 50% at the end of the horizon. In other words, $f_{a,T-1}^k$ equals 50% of f_{a0}^k for every $a \in \mathcal{A}$ and $k \in \mathcal{K}$. The conversion cost, denoted as g_{at}^k , is set to half the fixed cost of construction, i.e., $g_{at}^k := 0.5f_{at}^k$ for every $a \in \mathcal{A}$, $k \in \mathcal{K}$, and $t \in \mathcal{T}$. As for the flow cost, c_{at}^k , we establish it as $c_{at}^k = 2.5 \times d_a$, with d_a being the distance of arc a . Notably, for the terminal period, the flow cost is adjusted to account for future flow costs in a fully hydrogen-transitioned network: $c_{a,T-1}^k$ is further multiplied by T . To deter penalties, we assign h a value of 1.

Lastly, we determine the problem **parameters**. The scenario probability, p_ω , is uniformly set to $1/\Omega$. The arc capacity, U_a , is adapted to our context as $U_a := (2/3)u_a d_a$, with u_a representing the diameters from the set $\{30, 75, 120\}$. Node capacities, $C_{it}^{k\omega}$, are categorized into three primary standards:

1. Region-specific high-capacity storage facilities (e.g., salt caverns for hydrogen),
2. Hub-specific medium capacity storage facilities (e.g., at industry hubs),
3. Short-term storage units (e.g., cylinder storage for compressed gases).

These standards correspond to capacities of $\{10000, 1000, 10\}$ units, respectively. Table A.2 illustrates the probability distribution of these capacities across node types for every $i \in \mathcal{N}$, $k \in \mathcal{K}$, and $\omega \in \Omega$. We assume that $C_{it}^{k\omega}$ remains the same for all $t \in \mathcal{T} \cup \{T\}$.

Table A.2: Probability of storage units per each location type

Capacity	Supplier	HRS	Household	Industry	Intermediate
10000	0.1	0.0	0.00	0.0	0.2
1000	0.9	0.3	0.05	0.5	0.1
10	0.0	0.7	0.95	0.5	0.7

We further generate demand samples, $s_{it}^{k\omega}$, by considering that $\mathcal{K} = 2$ with hydrogen and natural gas represented by indices $k = 0$ and $k = 1$, respectively. We introduce pr_t as the probability of observing $s_{it'}^{1\omega} = 0$ for $t' \in \{t+1, \dots, T\}$ for natural gas, and $s_{it'}^{0\omega} = 0$ for $t' \in \{0, \dots, t-1\}$ for hydrogen. Essentially, pr_t depicts the transition probability from natural gas to hydrogen at time t' . For each scenario ω , we select pr_t such that $pr_t \gg pr_{t-1}$ in order to represent the expected exponentiality of the transition from natural gas to hydrogen, where $pr_0 = 0$ and $\sum_{t \in \mathcal{T}} pr_t = 1$. Specifically, we select pr_t as:

$$pr_t = \frac{r^t - 1}{\sum_{t=0}^{T-1} (r^t - 1)}$$

where r is uniformly randomized in $[1.01, 1.4]$, differently per each scenario of each instance. Using the probability array pr_t , we randomly assign a transition time matrix, a_{ik} , indicating the period when commodity k transitions on node i . For example, if $s_{it'}^{k\omega} = 0$ for $t' \in \{t+1, \dots, T\}$ for natural gas, then $a_{ik} = t$, where $\mathbb{P}\{a_{ik} = t\} = pr_t$. For our base scenario, the net supply levels at the beginning are represented by s_{i0}^k , structured to account for two distinct seasons: winter and summer. Winter mostly has heightened demand but diminished supply, whereas summer experiences the opposite. Each period t alternates between these seasons. The initial net supply levels are given in the Table A.3.

The generation of $s_{it}^{k\omega}$ is further detailed in Algorithm 5. In this methodology, the growth factor G represents the expected demand increase ratio on each period over the planning horizon. We introduce

Table A.3: Initial net supply levels of locations per season

Season	Supplier	HRS	Household	Industry	Intermediate
summer	240	-30	-20	-50	0
winter	120	-40	-40	-40	0

an additional factor, H , specifically for hydrogen to capture the potentially elevated growth rates due to its emergence as a new energy source. Additionally, we introduce R to account for the uncertainty in short-term energy demand fluctuations. We pick values $[G_L, G_U] = [1.0, 1.1]$, $[H_L, H_U] = [1.3, 1.7]$, and $R = 0.2$ in our computational experiments.

Algorithm 5 Generation of $s_{it}^{k\omega}$

Input: $s_{i0}^k, [G_L, G_U], [H_L, H_U], R$.

Output: $s_{it}^{k\omega}$

- 1: **for each** scenario $w \in \mathcal{W}$ **do**
 - 2: Randomize a probability array pr_t for $t \in \mathcal{T}$.
 - 3: Assign transition time matrix; a_{ik} , according to pr_t .
 - 4: Randomize uniformly $G \in [G_L, G_U]$ and $H \in [H_L, H_U]$.
 - 5: For natural gas, assign $s_{it}^k = s_{i0}^k G^t$ for each node i for $t < a_{ik}$, and 0 for $t \geq a_{ik}$ by the corresponding season.
 - 6: For hydrogen, assign $s_{it}^k = s_{i0}^k G^t H$ for each node i for $t \geq a_{ik}$, and 0 for $t < a_{ik}$ by the corresponding season.
 - 7: For all node i , commodity k and period t , select $s_{it}^{k\omega}$ in $[(1 - R\frac{t}{T})s_{it}^k, (1 + R\frac{t}{T})s_{it}^k]$ uniformly random.
 - 8: **end for**
-

Appendix B. Distribution Production Case Timeline

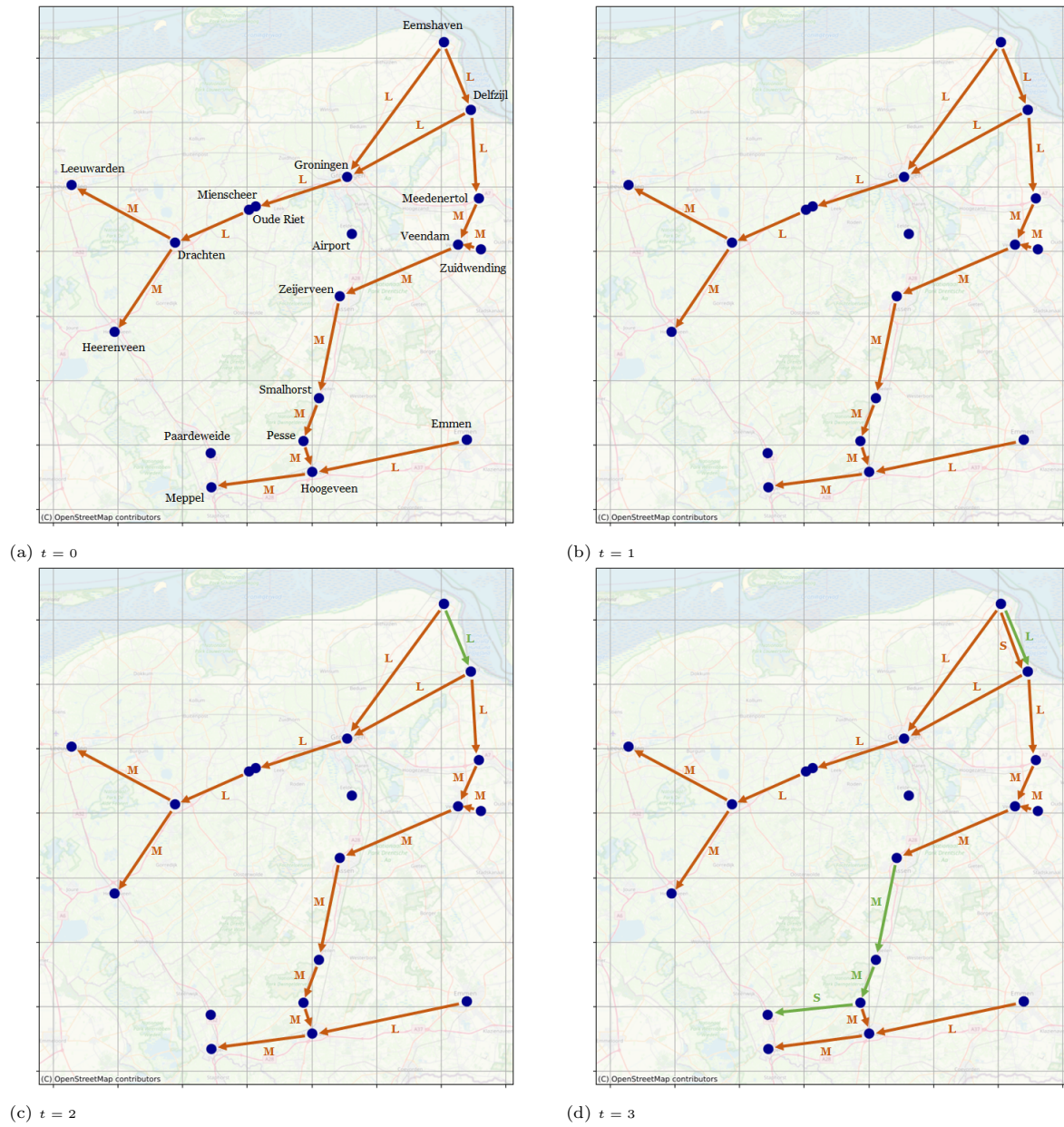
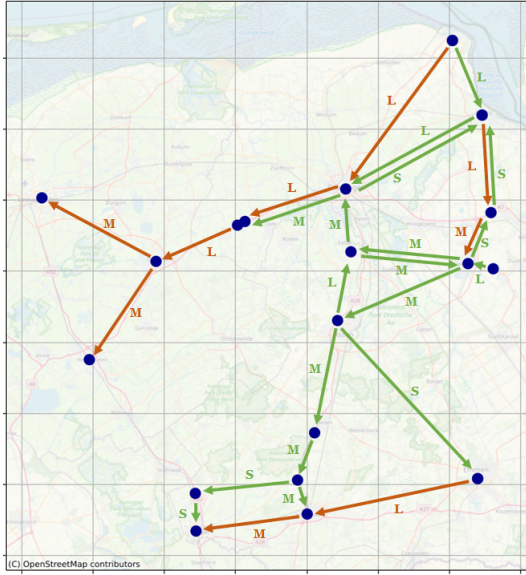
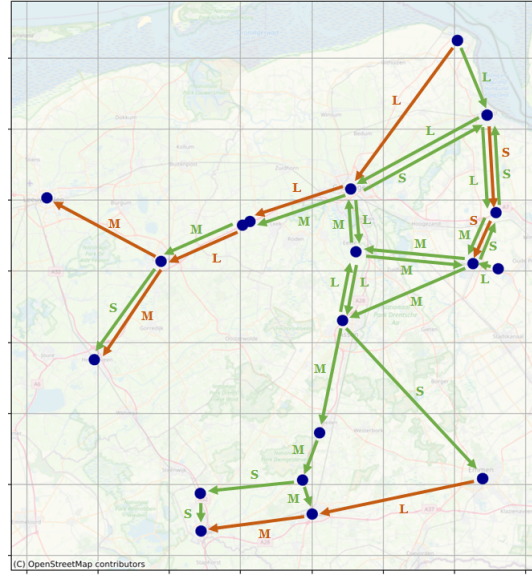


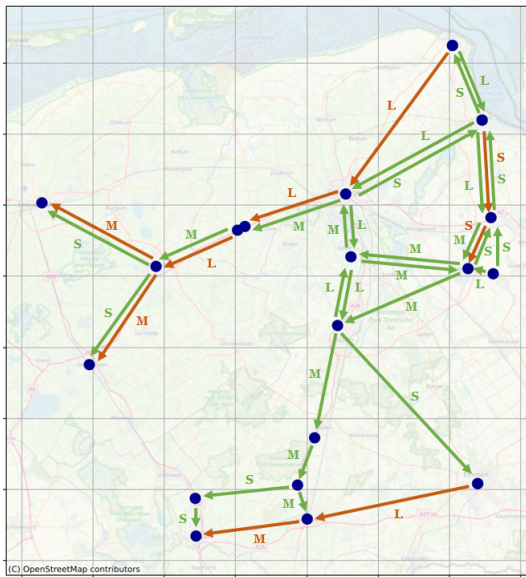
Figure B.6: The suggested pipeline network for decentralized production case at $t = 0$ to $t = 3$



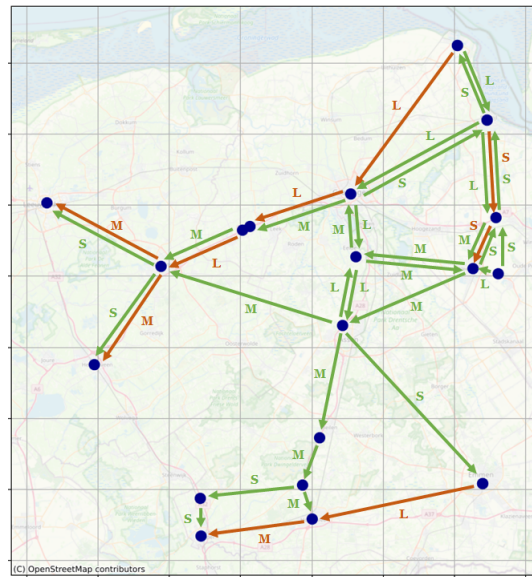
(a) $t = 4$



(b) $t = 5$

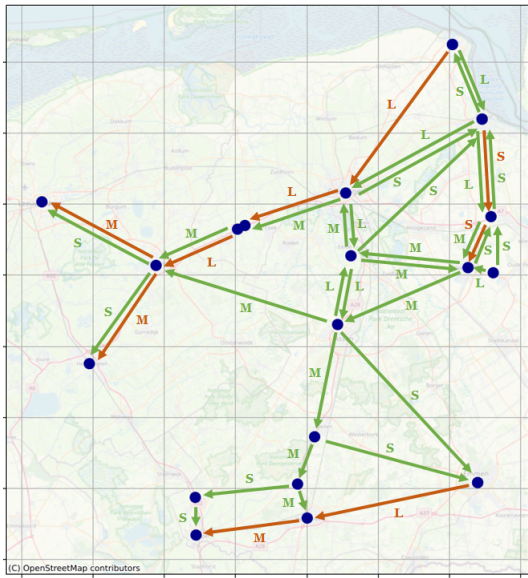


(c) $t = 6$

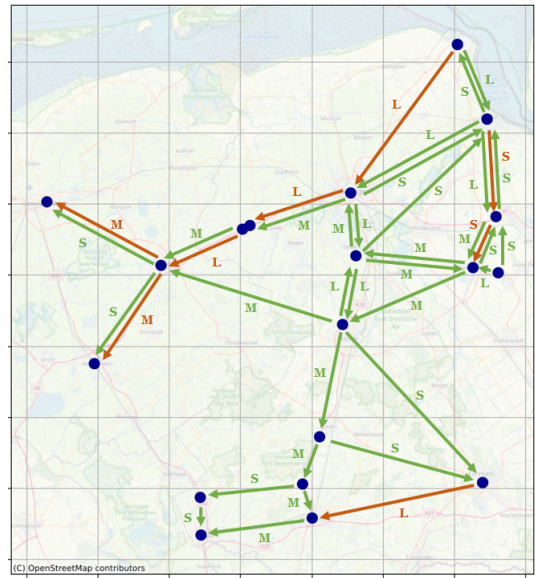


(d) $t = 7$

Figure B.7: The suggested pipeline network for decentralized production case at $t = 4$ to $t = 7$



(a) $t = 8$



(b) $t = 9$

Figure B.8: The suggested pipeline network for decentralized production case at $t = 8$ and $t = 9$