Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

INCI YUEKSEL-ERGUEN, THORSTEN KOCH, JANINA ZITTEL

Mathematical optimization based flow scenario generation for operational analysis of European gas transport networks based on open data

ZIB Report 24-03 (March 2024)

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30 84185-0
Telefax: +49 30 84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Mathematical optimization based flow scenario generation for operational analysis of European gas transport networks based on open data

Inci Yueksel-Erguen[1], Thorsten Koch[1,2], and Janina Zittel[1]

[1] *Applied Algorithmic Intelligence Methods Department, Zuse Institute Berlin, Takustraße 7, 14195 Berlin, Germany; yueksel-erguen@zib.de, koch@zib.de, zittel@zib.de*
[2] *Chair of Software and Algorithms for Discrete Optimization, Technische Universität Berlin, Straße des 17. Juni 135, Berlin,10623, Germany*

March 27, 2024

Abstract

The decarbonization of the European energy system demands a rapid and comprehensive transformation while securing energy supplies at all times. Still, natural gas plays a crucial role in this process. Recent unexpected events forced drastic changes in gas routes throughout Europe. Therefore, operational-level analysis of the gas transport networks and technical capacities to cope with these transitions using unconventional scenarios has become essential.

Unfortunately, data limitations often hinder such analyses. To overcome this challenge, we propose a mathematical model-based scenario generator that enables operational analysis of the European gas network using open data. Our approach focuses on the consistent analysis of specific partitions of the gas transport network, whose network topology data is readily available. We generate reproducible and consistent node-based gas in/out-flow scenarios for these defined network partitions to enable feasibility analysis and data quality assessment.

Our proposed method is demonstrated through several applications that address the feasibility analysis and data quality assessment of the German gas transport network. By using open data and a mathematical modeling approach, our method allows for a more comprehensive understanding of the gas transport network's behavior and assists in decision-making during the transition to decarbonization.

Keywords— linear programming, network optimization, scenario generation for European gas transport network, open data, real-world data consistency

1 Introduction

Natural gas transport networks play a fundamental role in achieving a smooth and effective energy transition in Europe. There are several reasons for that. First, in Europe, natural gas is diversely utilized by households, particularly for heating [Fra23], industry, and power generation [MTZD23]. It is critical to replace the utilization of natural gas with sustainable energy sources on the path to decarbonization while guaranteeing the security of the energy supply.

Furthermore, the gas transport network provides flexibility to the electricity network. This flexibility is necessary to balance the intermittency caused by the stochastic nature of renewable energy sources (RES), particularly during the energy transition. Here, the gas-powered plants, which are activated with low set-up time, are in the focus. They provide stable energy to the power grid whenever RES cannot produce sufficient energy due to weather conditions. Besides, power-to-gas (P2G) can contribute to decarbonization by converting surplus renewable electricity into hydrogen that can be injected into gas pipelines. Therefore, the gas network can store excess energy from renewable sources to provide it when necessary, balancing the grid and ensuring energy security. Besides, re-purposing some of the existing gas pipelines to transport pure hydrogen is another focus during the energy transition [The21, FNB21].

The criticality of natural gas during the energy transition manifests itself in the economy by reflecting any uncertainty in gas supply to gas prices. For instance, the gas supply issues with the political crisis in 2022 led to a significant increase in gas prices, resulting in higher electricity prices and inflation. As can be revealed by the unprecedented gas flow directions since 2022, such a crisis impacts the operation of the gas networks. Similarly, emerging technologies like power-to-gas facilitating the European energy transition to reach decarbonization targets have impacts on the conventional gas flow scenarios. A thorough impact analysis requires detailed operational-level models to understand whether the underlying physical network is capable of transporting the gas given the new circumstances. Thus, the operational analysis of gas networks has become more important than ever to analyze novel situations led by the impact of such events.

Mathematical modeling has been used extensively in the operational analysis of gas networks in the literature. The available models rely on proprietary data belonging to organizations [CBB+14] or companies [GMSS18]. Or, they are tested by simplified or distorted data sets that do not accurately reflect real-world conditions [HS19]. The resulting models, therefore, have limited utility for researchers, who rely on open data, for exploring solutions to intricate real-world problems.

Operational-level decision-making for gas networks requires employing detailed and mathematically precise models based on the physical flow of the gas in the pipelines. These models aim to find feasible operational settings of the network elements that allow the supplied gas entering the network from entry nodes to reach the exit nodes where it is demanded [KHPS15, PGH+15]. Hence, node-based gas in-/out-flow scenarios are required by operational-level analysis. Each scenario consists of the amount of gas entering the network from entry nodes and exiting the network from the exit nodes.

However, this endeavor is hindered by data quality since pan-European gas transport networks are not exempt from data limitations [PMD+22]. The available open pan-European gas transport network data are not sufficiently detailed for operational analysis. Moreover, the historical flow data is no longer valid in some cases because of the above-mentioned disruptive events. Hence, there is a pressing necessity for methods that can operate effectively on the open, accessible data to incorporate pan-European gas transport constraints into operational gas network analysis. The development of such methods is essential to support informed decision-making and ensure the robustness and resilience of gas transport networks, particularly in the face of unexpected events.

Despite the plethora of research on the European energy transition planning [Eur18, CBvAO+21], data limitations restrain the operational analysis of the pan-European gas transport network as a whole. One limitation is due to the insufficiency of network topology data quality. The detailed data is confidential due to security and its commercial value. The open pan-European network data sets are inadequate for models with the detail and precision required for operational-level network analysis. Additionally, the multiple-owner property structure of the pan-European gas transport network, which belongs to more than 55 transmission system operators (TSOs), introduces a challenge for consolidating data. The consolidated data by organizations like the European Network of Transmission System Operators for Gas (ENTSOG) is only at the level of interconnection points (IPs) [ENT18d], further limiting analysis. Disruptive events, such as changes in supply sources or gas properties, render historical flow data invalid for making data-driven inferences about gas distribution in Europe. Therefore, methods are required to facilitate operational analysis of pan-European gas transport networks to overcome the data limitations.

This research presents a mathematical modeling-based scenario generator for gas transport that utilizes open data to analyze pan-European gas transport networks. It is comprised of two linear programming (LP) models interconnected at the data exchange level to construct a hierarchical modeling framework. This framework employs open data to generate node-based gas in-/out-flow scenarios for a restricted region in Europe, for which a detailed enough data set exists for operational analysis. In other words, the framework allows the integration of pan-European gas transport constraints to a more precise mathematical model of a region in Europe. In this way, it facilitates feasibility analysis as well as assessment of the robustness of the underlying physical network by generating reproducible scenarios with open data. Moreover, the framework permits analyzing the consistency of the exchanged data sets that address the same physical entities with varying detail.

In light of its various utilization domains, the proposed scenario generator provides a deeper understanding of the gas transport network dynamics within the pan-European energy system. It, therefore, supports the development of effective policies for the European energy system transition to reach the decarbonization targets. Overall, this study provides a valuable contribution to gas transport network optimization, highlighting the potential of open data to facilitate more reliable decision-making processes during the European energy transition.

In this paper, we first provide a brief introduction to the operational-level gas network optimization models and their data requirements in Section 2. In this section, we also present the open data landscape for pan-European gas transport networks and related work in the literature on gas network scenario generation. Before going into details of the proposed scenario generator, in Section 3, we explain the mathematical models and the modeling framework that constitutes the scenario generator. We present the scenario generation process using the scenario generator in Section 4 with examples from the German gas transport network. We give examples from the applications in which the scenario generator is used in Section 5 and make our concluding remarks in Section 6.

2 Operational-level Decision Making with Open Data

In this section, we provide a concise overview of an operational-level gas network optimization model (GNO), outlining its essential characteristics and data requirements. Subsequently, we explore the open data landscape, offering insights into the existing open data sources, and relevant literature, identifying gaps between the GNO's requirements and the available data.

2.1 Gas network optimization and data requirements

Gas flows in pipes according to the thermodynamic rules, i.e., from a high-pressure point to a low-pressure one, and while flowing, its pressure drops. There are network elements that regulate the gas pressure and hence the flow direction of the gas. For instance, compressor stations increase the pressure while control valves decrease when active. These network components, with valves, also change the direction of the gas by decoupling the adjacent network nodes when they are closed. Furthermore, the gas compression ability and allowed gas direction by each compressor station change according to the selected configuration of its sub-components. The node pressures, states of the network elements, and selected configuration of each compressor station constitute the network state at a particular point in time.

An operational-level mathematical model of a gas network, as presented in detail in [KHPS15], is detailed and precise enough to use the properties and operational dynamics of single network elements. Hence, these mathematical models are suitable for finding a feasible network state that enables the gas flow as defined by a gas in-/out-flow scenario [HS20, HKS+18]. Such an analysis requires a detailed and consistent network topology data set, an accurate compressor station data set, and consistent node-based gas in-/out-flow scenarios [Zus18, SAB+17]. In this study, we focus on node-based scenarios.

In the literature, benchmark scenarios [Zus18, SAB+17] provide node-based scenarios for operational-level gas transport network analysis. However, these scenarios are distorted data sets from real-world cases, due to confidentiality of data. Again, datasets used by publications on real-world gas transport networks are either from TSOs [HAHB+21] and they are not open, or very limited in terms of geographical span [DWS00]. In the presence of high-resolution historical demand/supply data from the TSOs, node-based scenarios are generated by forecast [PCG+22] or adversarial nomination generating heuristics [HHH+15]. To the best of our knowledge, there is no tool using open data and open models for node-based scenario generation to be employed in operational-level analysis for the pan-European gas transport network.

2.2 Available open data for European gas transport

The European gas transport network is a highly connected network that is operated by more than 55 TSOs [ENT18d]. The network spans the continent and Great Britain by transmission pipelines of more than 200,000km long. There are more than 200 IPs on this network connecting pipelines belonging to the different TSOs [ENT18c]. Besides, there are more than 170 underground storage facilities (UGSs) connected to it [ENT18d, Gas21b]. Regrettably, comprehensive open data for analyzing the entire European network using GNO is currently unavailable. The existing data is dispersed across various sources and lacks uniform aggregation, posing a considerable challenge for consolidation into a cohesive dataset.

In light of these limitations, a more manageable approach is to analyze smaller regions within Europe while maintaining a connection to the broader European network. It is plausible that network topology data for individual countries or data specific to TSOs may be obtainable. However, this approach necessitates the formulation of node-based scenarios tailored to these smaller regions, ensuring consistency with pan-European gas transport constraints. In addition, in open datasets, demand distribution is often provided per postal code or NUTS3 region. In this subsection, we present the available open data that is useful for generating node-based flow scenarios for regions in Europe.

The European high-level gas transport network data is provided by ENTSOG in several datasets. The first dataset is provided by ENTSOG via its transparency platform (TP) [ENT18d] which we call the ENSTOG IP dataset throughout the paper. The data set includes the relevant IPs of the European gas transport network, their capacities, and the historical physical flow via these IPs. The IPs represent the interconnection between different gas infrastructures such as two TSOs, or a TSO and a UGS/LNG facility. The IP data with hourly physical flow and capacities are published by the ENTSOG TP. However, since the data only includes the relevant points, it does not completely represent the distribution of the gas through Europe. For example, this data set lacks historical gas consumption data for some countries such as Austria, Switzerland, Czechia, Sweden, Denmark, and Finland. Again, since the relevant nodes include only aggregated final consumer nodes for Germany, the IP data in Germany is not complete. So, the data cannot be used directly to represent the entire historical gas flow in the European gas transport network, even with a high-level representation with IPs. Besides, data specific to the UGSs such as withdrawal and injection capacity are not provided in this data set.

Therefore, we require other data sources that complement the ENTSOG IP data set for compiling the input data for the European-level entry-exit model. These are reports and publications of ENTSOG and Gas Infrastructure Europe (GIE) TP. The historical gas consumption data per country is also published by ENTSOG with higher temporal granularity in their security of supply reports [ENT17c]. Again, data sets including gas demand and supply forecast are published by ENTSOG for countries yearly in the ten-year network development plans (TYNDP)[ENT18b]. GIE TP complements the ENTSOG data by providing data on UGSs [Gas21b]. In these cases, the data has to be temporally and geographically disaggregated consistently.

On the other hand, the ScigridGas data, stemming from a BMWK-funded project, offers an open dataset containing the pipeline network across Europe [Sci18]. Nevertheless, to facilitate operational-level analyses, more detailed information is essential than the network topology presented by pipelines. This includes access to physical properties for individual network components, covering both active components and node-based demand series. Moreover, it includes demand data at the subregion level. A parallel dataset for Germany is available through the LKD-EU project [LE18, KKS$^+$17, KWH$^+$17], maintaining congruent aggregation levels for demand time series and topological details. To establish consistent node-based scenarios using these datasets, methods for disaggregation are crucial, ensuring alignment with the broader pan-European network's demand and supply structures.

3 Scenario Generation Tool

Potentially, sufficiently precise network topology data sets are accessible for gas transport networks of smaller regions in Europe. To exemplify, accessing the data set of a network belonging to a country, as shown here for the case of Germany, or a TSO is more likely than a detailed consolidated data set at the pan-European scale.

Using the historical data to characterize the gas in and out-flow to such reduced networks is possible using statistical and predictive methods based on fully accessible, accurate, and timely historical data [HHL$^+$15]. However, this is not the case when the historical data is collected only from open data sources, or analysis of unconventional scenarios is required. In the former case, open data sources are not complete to generate a historical flow scenario for operational-level analysis in Europe. For the latter, historical data becomes obsolete in some cases when we face unexpected events or when we aim to analyze the impacts of unconventional what-if scenarios in the future. In both cases, methods to explore the gas in and out-flow to reduced networks in Europe conforming to the constraints of the pan-European gas transport are necessary. Such methods should generate gas in and out-flow scenarios, which will be called gas flow scenarios in the text from this point forward, for the reduced network such that the scenarios

- are reproducible using open data and open models,
- conform to the limitations of the network infrastructure given the pan-European gas supply,
- robust against data uncertainties, and
- are able to reflect the future uncertainties of gas transport so that what-if scenarios can be easily integrated from top to bottom.

In this study, we propose a mathematical modeling-based scenario generator to generate meaningful realistic scenarios. The proposed scenario generator includes two LP models as outlined in Figure 1.

The first LP (M_1) models the high-level pan-European gas network. It aims to find the gas in-/out-flow of a region in Europe given the gas supply and cumulative gas demand of the European gas transport network. M_1 results in cumulative gas in-/out-flow of the selected region based on the gas infrastructure defined in ENTSOG TP data.

The second LP (M_2) models the regional gas transport network as a linear model, i.e., without the nonlinear thermodynamic gas properties or combinatorial configurations. M_2

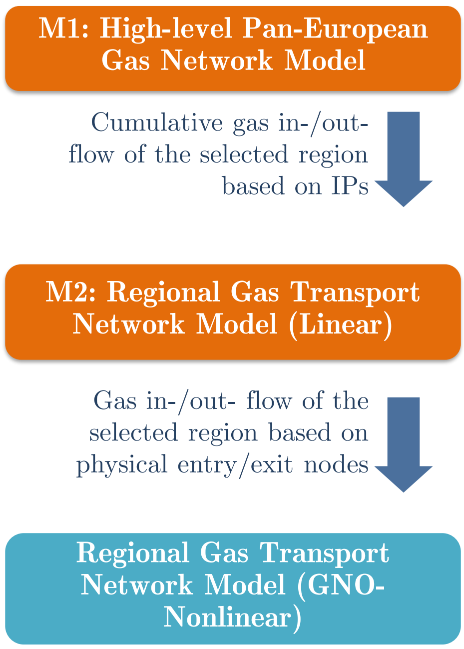


Figure 1: Scenario Generator

disaggregates the gas infrastructure-based gas in-/out-flow of the specified region to its gas network entry/exit nodes by ensuring the correct system/node associations, meeting subregion demand distribution of the specified region, and feasibility given approximate maximum pipeline gas capacity. The results of M_2 are the node-based scenarios required for the operational-level analysis of the regional gas transport network, i.e., by GNO.

The details for M_1 and M_2, which use open data sources as input and are integrated at the data exchange level, are presented in the ensuing subsections. We then present methods to enhance the models by exploiting the properties of LP to effectively explore the feasible region of the problems.

3.1 European Level Entry-Exit Model - M_1

We propose modeling the European-level entry-exit network as an LP using the available open data sources. Our objective is to explore feasible gas distributions in the European gas transport network with minimum gas curtailment given the existing gas infrastructure and possible gas demand and supply. We model the network as a capacitated minimum-cost network flow problem. The required inputs for this model are network topology, capacities of the edges, amount of gas supply from source nodes, and amount of gas demand from sink nodes. Consequently, we seek the answer to the question of how the gas supplied to the EU is distributed to the countries so that the fulfilled demand is maximized.

Input:

- The high-level network topology data provided by ENTSOG
- Gas supply to Europe via imports, LNG facilities, UGSs, and production
- Gas demand of countries or balancing zones, gas injected to UGSs, exports, LNG
- UGS working capacity, injection and withdrawal capacities based on gas volume in the UGSs

The selection of gas demand and supply data sources depends on the aimed analysis. For example, if the analysis of a future time frame is required, then the supply and demand forecast is used. Again, if the integrated analysis of electricity and gas grids is aimed, the input data is prepared by merging the gas demand and supply profiles resulting from the electricity grid model and the applicable demand and supply data from open data sources. To exemplify, the supply from P2G facilities is added to the yearly supply forecast data from ENTSOG TYNDP scenarios to analyze the effect of installed P2G facilities. In this case, physical flow data can serve as a meta-distribution for the spatial distribution of the yearly forecast values, if applicable.

Output:

- Minimum possible demand curtailment given the supplied gas
- Gas distribution to the high-level gas network, i.e., utilization of the pipeline capacities and UGSs, cross-border exchange limitations, imbalances in the network due to cross-border exchange capacities: curtailed demand vs. stored gas
- Geographical disaggregation of the supplied gas to Europe

Mathematical model:
ENTSOG TP data represents the European network based on the IPs. Each IP is composed of two gas infrastructures exchanging gas in one direction. So, we define a graph using these gas infrastructures exchanging gas in the IPs as nodes. Again, the ENTSOG IPs constitute the directed and capacitated arcs of the graph.

Table 1: Notation of the European entry-exit mathematical model

Sets	
V	Entry-exit nodes in the high-level European gas network
A	Arcs between the nodes in V
GS	Gas Systems
IB	Internal Bottlenecks
IC	International Connections
St	UGSs
StI	UGS Interconnection Nodes
LN	LNG Facilities
B	Balancing Zones
Sp	Suppliers
Subsets and Indexed sets	
$V_{Set} \subset V$	Nodes denoting the infrastructure given by the *Set* definition
$V_B^b \subset V$	Nodes in the balancing zone $b \in B$
Parameters	
c_{ij}	Capacity of arc $(i, j), (i, j) \in A$
U_i	Capacity of a UGS, $i \in V_{St}$
d_i	Demand of node I, $i \in V$
s_i	Supply of node I, $i \in V$
ϵ_1	UGS gas volume adjustment coefficient
Variables	
x_{ij}	Gas flow on the arc $(i, j) \in A$
u_i	Amount of gas injected to the UGS $i, i \in V_{St}$
y_i	Demand curtailment of node $i \in V$

The model is defined on a directed graph $G_E = (V, A)$ where V denotes nodes consisting of entry-exit system components in the open data set published by ENTSOG [ENT19, ENT17c]. These components include infrastructure types such as gas systems, UGSs and

LNG facilities, suppliers, and international connections. An arc $(i, j) \in A$ represents the capacity of the IPs between the nodes in V.

The vertex set V is partitioned into the following sets regarding the definitions in the ENTSOG documents [ENT17c, ENT17b, ENT18a, ENT17a]:

- V_{GS} denote demand attached nodes having installed gas infrastructure, i.e., transmission and distribution, except for those including UGSs and LNG facilities
- V_{IC} denote inter-connector pipelines having a special regime connected to the EU entry-exit system
- V_{StI} have interconnection from UGSs to other gas systems.
- V_{St} denote the UGSs.
- V_{LN} denote the LNG facilities.
- V_{IB} are determined by ENSTOG and can depict seasonal variability of gas demand or supply, as in France.
- V_{Sp} denote gas suppliers to EU countries.

The vertices in the above-listed partitions are clustered in special topologies called balancing zones. A balancing zone $b \in B$ consists of at least one demand-attached node. The total gas entering a balancing zone b is equal to the gas leaving it. Nodes belonging to a particular country may constitute a single balancing zone or a country may have more than one balancing zone. We denote a balancing zone as a set of nodes such that $V_B^b = \{i : i \in V\}, b = 1, ..., |B|$. The structure of a balancing zone is illustrated in Figure 2.

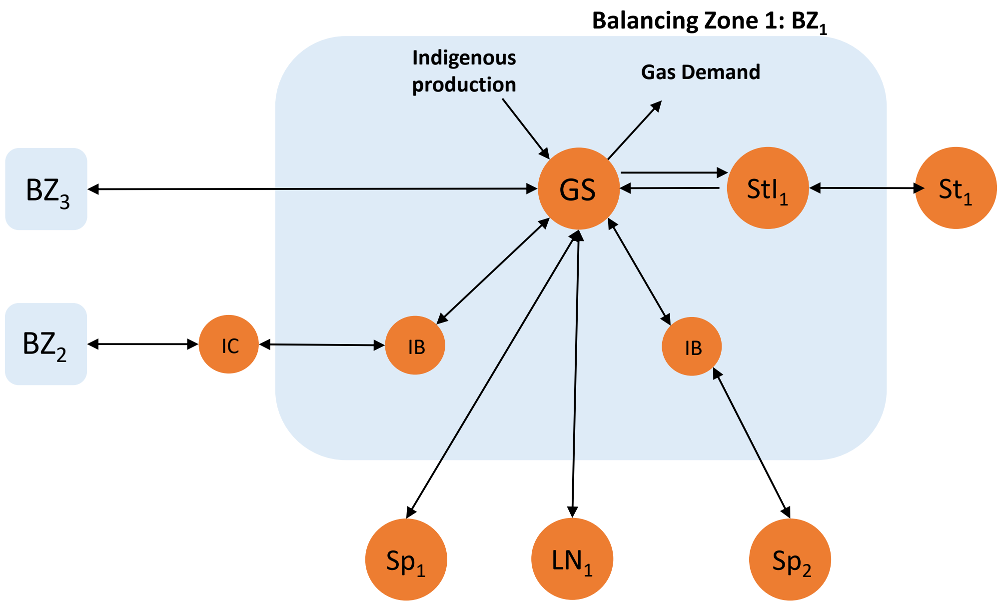


Figure 2: Illustration of a balancing zone structure

There are different types of arcs in the topology in terms of how we assign the capacity values $c_{ij}, (i, j) \in A$:

- UGS arcs denote injection of gas to storage, i.e., $a = (i, j)$ such that $i \in V_{StI}$ and $j \in V_{St}$, or withdrawal of gas from storage, i.e., $a = (i, j)$ such that $i \in V_{St}$ and $j \in V_{StI}$. The capacity of these arcs is the injection and withdrawal capacity of the UGSs, respectively, which are determined dynamically according to the gas deliverability curves of the UGSs as provided by GIE [Gas21a].
- LNG arcs denote injection of gas to LNG facility, i.e., $a = (i, j)$ such that $i \in V_{GS}$ and $j \in V_{LN}$, and withdrawal of gas from LNG facility, i.e., $a = (i, j)$ such that

$i \in V_{GS}$ and $j \in V_{LN}$. The capacity of these arcs is the injection and withdrawal capacity of the LNG facilities [ENT17c], respectively.

- the rest of the arcs $a = (i, j), a \in A$, such that $i, j \in V_{GS} \cup V_{IC} \cup V_{StI} \cup V_{LN} \cup V_{IB} \cup V_{Sp}$ are the connection arcs. The capacity of these arcs is the capacity of the associated IPs listed in ENTSOG data [ENT17c].

There are four sources of supply to the European Union (EU) countries. These are the existing amount of gas in UGSs, LNG facilities, imported gas from supplier countries, and national indigenous production. They are denoted by the parameter s_i such that

- if $i \in V_{Sp}$, s_i denotes the imported amount of gas from the supplier i,
- if $i \in V_{StI}$, s_i denotes the working gas volume of the UGS i,
- if $i \in V_{LN}$, s_i denotes the amount of gas that can be injected into the network from the LNG facility i,
- if $i \in V_{GS}$, s_i denotes the indigenous gas production of the gas system i,
- otherwise, $s_i = 0$.

The demand for gas is denoted by d_i and attached to the gas systems and internal bottlenecks, i.e., $d_i = 0$ if $i \notin V_{GS} \cup V_{IB}$. The demand attached to EU countries represents the gas demand of final consumers in these countries. Whereas, the demand of non-EU denotes the exported gas to these countries from the EU countries.

The capacity of a UGS $i \in V_{St}$ is denoted by U_i where $U_i \geq 0$ if $i \in V_{St}$. This parameter implies the upper bound on the gas injection to the UGS i.

The variable x_{ij} is the flow of gas from node i to node j such that $i, j \in V$. If the gas demand of a particular node $i \in V_{GS} \cup V_{IB}$ cannot be met by the supply, then the demand is curtailed. The demand curtailment variable for each node is denoted by y_i such that $y_i \geq 0$ for all $i \in V_{GS} \cup V_{IB}$ and $y_i = 0$ otherwise. Another variable used in the model is the amount of gas injected into the UGSs, u_i where $0 \leq u_i \leq U_i$ if $i \in V_{ST}$ and $u_i = 0$ otherwise.

The European-level entry-exit network as modeled in M_1 is illustrated in Figure 3 and its mathematical model is presented below using the notation presented in Table 1.

$$min \sum_{i \in V} y_i - \epsilon_1 \sum_{i \in V_{St}} u_i \tag{1}$$

subject to

$$\sum_{\substack{j \in V \\ (j,i) \in A}} x_{ji} - \sum_{\substack{j \in V \\ (i,j) \in A}} x_{ij} + y_i - u_i = d_i - s_i, \qquad \forall i \in V \tag{2}$$

$$\sum_{i \in V_B^b} \Big(\sum_{\substack{j \notin V_B^b \\ (j,i) \in A}} x_{ji} - \sum_{\substack{j \notin V_B^b \\ (i,j) \in A}} x_{ij} + y_i - u_i \Big) = \sum_{i \in V_B^b} (d_i - s_i), \qquad \forall b \in B \tag{3}$$

$$y_i \geq 0, \qquad \forall i \in V_{GS} \cap V_{IB} \tag{4}$$

$$y_i = 0, \quad \forall i \in V \setminus V_{GS} \cap V_{IB} \tag{5}$$

$$u_i = 0, \qquad \forall i \in V \setminus V_{St} \tag{6}$$

$$0 \leq x_{ij} \leq c_{ij}, \forall (i,j) \in A; 0 \leq u_i \leq U_i, \qquad \forall i \in V_{St} \tag{7}$$

The objective is to minimize the total demand curtailment while keeping the gas in the UGSs at an acceptable level. In this paper, we added the total gas amount injected into the UGSs to the objective function (1) with a coefficient $\epsilon_1 < 0$ to adjust a relevant level of stored gas.

Constraints (2) are flow conservation constraints. Constraints (3) are balancing zone constraints to ensure the amount of flow entering each balancing zone is equal to the amount of flow leaving the balancing zone. Constraints (4) depict non-negativity for

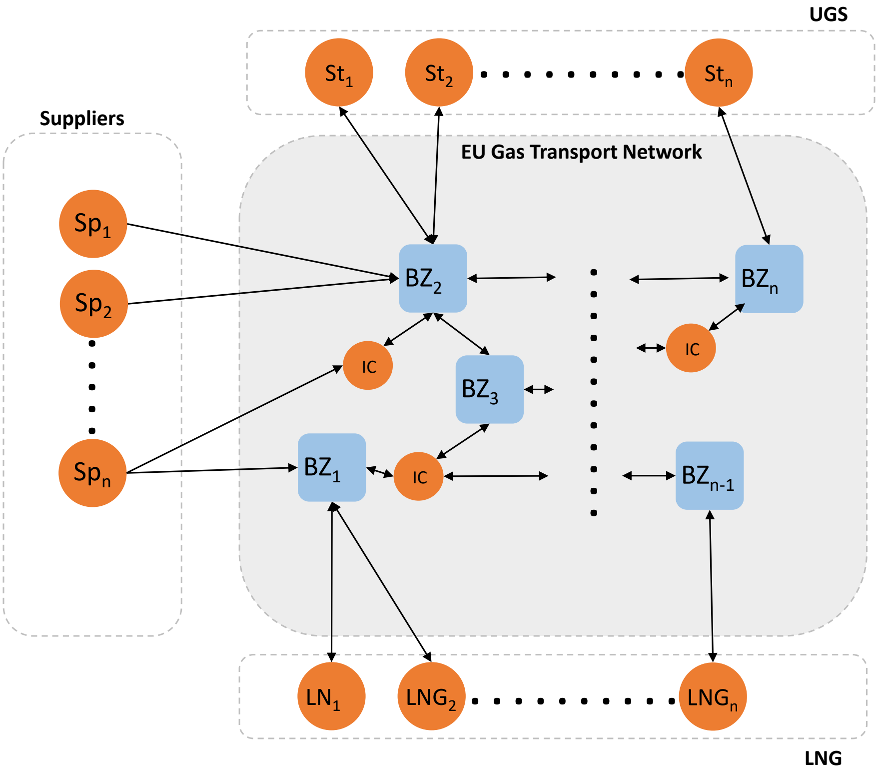


Figure 3: The illustration of the European Level Entry-Exit Model

demand curtailment for demand-attached nodes while constraints (5) guarantee that the demand curtailment is zero for others. Constraints (6) ensure that u_i denoting the amount of gas injected into UGSs are non-zero only for relevant nodes. Constraints (7) are capacity constraints for edge flows and the amount of gas that can be injected into the UGSs, respectively.

3.2 Linear Gas Network Optimization Model - M_2

The input of an operational model requires the amount of gas entering the region through the entry nodes of the regional gas transport network and leaving the network from the exit nodes. For a stationary gas optimization problem, this scenario has to be balanced. In other words, we need a physical-node-based gas in/out-flow scenario that induces an equal amount of gas entering and leaving the region. However, such gas in/out-flow scenarios are not directly available from open data. They should be derived by integrating different data sources.

The final consumer nodes are, by legislation, not relevant (see Section 2.2 for details). Therefore, their physical flows are not published by open data sources such as TSO and TSO organizations' TPs. Instead, gas demand distribution to subregions in countries is accessible from open data. In the literature, the distribution data is often provided per postal code region or NUTS3 region. Here, care should be taken since the demand distribution to subregions possesses uncertainty. Besides, the distribution estimations available for countries, such as the one provided by [KKS+17] for Germany, reflect the average given conventional scenarios or historical data. Hence, the model should provide some flexibility to relax the constraints imposed by the demand distribution to effectively search the feasible scenario space.

We use an LP to dispatch the cumulative gas supply and demand of a region in Europe to physical network entry-exit nodes. With this LP, we explore gas in/out-flow scenarios that meet the regional demand and balance the gas scenarios. Besides, we introduce applicable side constraints specific to the region and type of the performed analysis to model the quality-of-service of the emerging technologies. We model the regional gas transport network as a capacitated minimum-cost network flow problem. We transform the network to generate feasible solutions with minimum deviation from the input total gas supply and demand, and demand distribution. Furthermore, the physical network topology is modeled as a graph and given to the model as an input with the edge capacities denoting the maximal gas flow of the physical pipelines.

Input:

- Network topology data of the region including physical entry, exit and inner nodes, and pipelines with predefined capacity upper bounds
- Amount of gas demand and supply of the region as a result of M_1
- Physical boundary node association to the high-level data set given in M_1
- Gas demand distribution to subregions, i.e., NUTS-3/postal code regions

Output:

- A valid dispatch of the gas demand and supply results in M_1 to the boundary nodes of the regional network
- Amount of gas reduction in supply/demand by source
- Amount of gas exchange between subregional demand

Mathematical model:

The physical gas transport network of the selected region is modeled as a bi-directional graph $G_P = (V_P, E_P)$ where V is the set of vertices and E_P is the set of edges such that $i, j \in E_P, i, j \in V_P$.

The boundary nodes where the gas enters and leaves the gas network are called entry and exit nodes and are denoted by V_P^+ and V_P^-, respectively. The junction points between the pipelines are the transshipment nodes and are denoted by V_P^0. Hence, the set of vertices is $V_P = V_P^+ \cup V_P^- \cup V_P^0$. The set of arcs E_P represents the pipelines between the physical nodes of the network.

The gas leaving from and entering the G_P via the boundary nodes is the output of the European entry-exit model M_1. However, the output of this model is cumulative, i.e., there is not necessarily a one-to-one relationship between the output flow amounts of the European-level model M_1 and the physical nodes of the regional model M_2. So, we have to define the interface between G_E and G_P by associating the output flow from the European model and boundary nodes of the physical region. We, therefore, transform the G_P, so that the resulting augmented graph dispatches the cumulative output of M_1 to the boundary nodes of the G_P. We define the set of transformations so that the resulting graph

- ensures that resulting node-based scenarios are feasible in terms of the European-level constraints and the demand distribution needs for the selected region,
- eliminates the scenarios that are already infeasible given the relaxed linear gas network optimization model, and
- allows exploring the feasible scenario space by changing the objective function values of the network flow models and distribution of gas in the physical network.

To model this interface between the graphs, we define artificial vertices and arcs, denoted by V_{Ar} and A_{Ar}, respectively. Then we augment the regional physical graph with these artificial sets to have a mixed graph representing the regional gas network $G_R = (V_R, E_P, A_{Ar})$, where $V_R = V_P \cup V_{Ar}$. We apply the following network transformations (NT) to integrate the graphs G_E and G_R by using the coarse-to-fine (C2F) approach, which is explained in detail in [YEKZ23].

- NT_1 **- matching boundary nodes:** We add edges in the M_1 solution associated with the selected region as a node in the augmented graph. We add artificially directed links to the associated nodes to each link. These associations are a priori-defined accounting for the infrastructure types of the nodes in G_E and the physical boundary nodes, V_P^+ and V_P^-. We add the amount of flow on the links in the M_1 solution to these nodes as supply or demand according to the infrastructure type.
- NT_2 **- geographical dispatch:** For the final consumers, the association is not direct such as the exit nodes associated with the other infrastructures. The demand is based on subregions and there can be gas exchange between them. So, we model the subregions and their association with the physical nodes by a set of artificial nodes, V_{Ar}^{SX}. Adding the artificial arcs emanating from nodes in $V_P^{-,S}(r)$ to the artificial node set $A_{Ar}^{P,SX}$ enables us to determine the resulting gas flow from the final consumer nodes associated with the regions. Thus, defining two non-negative variables π_r^+ and π_r^- for deviation from the given distribution per subregion $r \in S$ and limiting the deviation from total subregion demand by another non-negative variable, θ, we deal with the uncertainty in demand distribution data. These variables and the transformation also allow us to comment on the required gas redispatch between subregions given the results of M_1. This is especially important for integrated electricity-gas networks since this exchange means a redispatch requirement for the electricity grid when the subregions include gas-powered plants.
- NT_3 **- exploring artificial feasible solution:** We generate feasible solutions to M_2 by adding an artificial supply and an artificial demand node, t and $k \in V_{Ar}$, and an artificial arc $(t,k) \in A_{Ar}$. The directed artificial arcs between these nodes and artificial nodes that we added in NT_1 allow us to explore alternative solutions in the neighborhood of the M_1 solutions. Here, the amount of flow on arc (t,k) shows the amount of gas that cannot be routed in the gas network. The amount of flow of the artificial arcs between the nodes t, k and other artificial nodes added in NT_1 show the break-down of this *unrouted flow* on arc (t,k) to the infrastructure and gas systems in G_E. Hence, in case of infeasibility, we can understand how close the solution is to feasibility by checking the flow on (t,k) and its root cause by checking the flows on the arcs emanating from (going into) k (t).

After these transformations, M_2 can be modeled as a capacitated minimum cost network flow defined on $G_R = (V_R, E_P, A_{Ar})$ as follows using the notation presented in Table 2.

$$min \quad f_{t,k} + \epsilon_2 \sum_{r \in S} \pi_r^+ + \pi_r^- \tag{8}$$

subject to

$$\sum_{\substack{j \in V_R, \\ (j,i) \in A_{Ar} \vee \{i,j\} \in E_P}} f_{ji} - \sum_{\substack{j \in V_R \\ (i,j) \in A_{Ar} \vee \{i,j\} \in E_P}} f_{ij} = d_i - s_i, \qquad \forall i \in V_R \tag{9}$$

$$\sum_{\substack{i \in V_P^{-,S}(r), j \in V_{Ar}^{SX} \\ (i,j) \in A_{Ar}^{P,SX}}} f_{ij} + \pi_r^+ - \pi_r^- = d^r \qquad \forall r \in S \tag{10}$$

$$\sum_{r \in S} \pi_r^+ - \pi_r^- \leq \theta \tag{11}$$

$$0 \leq f_{ij} \leq c_{ij}, \forall (i,j) \in A_{Ar}; -c_{ji} \leq f_{ij} \leq c_{ij}, \quad \forall \{i,j\} \in E_P \tag{12}$$

$$0 \leq \pi_r^+, 0 \leq \pi_r^-, \qquad \forall r \in S \tag{13}$$

The model aims to find a scenario minimally deviating from the demand distribution given the total final consumer demand while maximizing the amount of gas routed in the

Table 2: Notation of the regional gas network model

Sets	
V_R	Vertices in the regional network
E_P	Physical gas network edges
A_{Ar}	Artificial arcs to dispatch exchanged gas to individual nodes in $V_P^+ \cap V_P^-$
S	Set of subregions
Subsets and Indexed sets	
$V_P \subset V_R$	Physical gas network nodes of the selected region in EU
$V_P^0 \subset V_P$	Transshipment nodes
$V_P^+ \subset V_P$	Entry nodes
$V_P^- \subset V_P$	Exit nodes
$V_P^{-,S}(r) \subset V_P^-$	Admissible exit nodes serving to subregion $r, r \in S$
$V_{Ar} \subset V_R$	Artificial nodes representing the gas exchange of the physical gas network with EU network
$V_{Ar}^{SX} \subset V_{Ar}$	Subregion-admissable exit node association
$A_{Ar}^{P,SX} \subset A_{Ar}$	Artificial arcs between $V_P^{-,S}(r)$, $\forall\ r \in\ S$ and V_{Ar}^{SX} denoting the flow from final consumer nodes to subregions
Parameters	
c_{ij}	Maximum allowable flow on the connection between nodes i and j: $i, j \in V_R$ and $(i, j) \in A_{Ar}$, or $i, j \in E_P$ and $i, j \in V_P$
d_i	Demand of node i, $i \in V_{Ar}$
d^r	Demand of subregion r, $r \in S$
s_i	Supply of node i, $i \in V_{Ar}$
ϵ_2	Adjustment coefficient for demand distribution among subregions
θ	Allowable deviation from total final consumer demand
Variables	
f_{ij}	Gas flow on the arc (i, j) if $(i, j) \in A_{Ar}$ or on the edge i, j if $i, j \in E_P$
π_r^+	Non-negative increase in demand for subregion $r, r \in S$
π_r^-	Non-negative decrease in demand for subregion $r, r \in S$

gas network. So, the objective function (8) minimizes the total deviation of demand of subregions and the amount of flow on the artificial arc (t, k).

(9) are the flow conservation constraints. Subregion demand constraints (10) ensure that the demand of each subregion is met with a deviation amount of $\pi_r^+ - \pi_r^-$. Subregion demand deviation constraint, (11), guarantees that the total subregion demand is met with a small perturbation amount denoted by θ. The smaller θ, the smaller the deviation from the final consumer demand. We note that we can impose other rules by adding constraints to the π_r^+ and π_r^- in the model to limit the deviation of the generated scenarios from the given demand distribution (for details please see Section 4). (12) are the capacity constraints for arcs A_{Ar} and edges E_P, respectively. Finally, (13) guarantee nonnegativity for π_r^+ and π_r^-.

4 Scenario Generation Using LP

Employing M_1 and M_2 integrated by the C2F approach, the scenario generator produces the balanced node-based gas flow scenarios to be analyzed by operational-level analysis of a regional gas transport network in Europe. In this context, using LP-based models brings about both advantages and disadvantages.

One of the main advantages is the ability to effectively eliminate infeasible scenarios by considering higher-level constraints induced by M_1 and M_2. Acknowledging the computationally expensive high-precision models, scenario filtering by relaxed models saves us computation time. In addition, in this way, we can compartmentalize the reasons for infeasibility during the analysis. Furthermore, we can more effectively evaluate the root causes whenever a generated scenario is infeasible at the operational level.

Additionally, M_1 and M_2 offer the advantage of allowing the incorporation of side constraints, which enables even finer exploration for feasible scenarios complying with customized model constraints. Thus, the resulting scenarios are aligned with the decision requirements of the aimed decision analysis. For instance, if the scenarios are employed for an integrated electricity and gas grid analysis, the scenario generation tool should also serve in the capacity of merging the gas profiles from the electricity grid and non-electricity gas consumption (see Section 5). Again, M_1 and M_2 have parameters to provide flexibility to incorporate uncertainty in the resulting scenarios.

On the other hand, there are some drawbacks to using an LP-based scenario generator to consider. For instance, an LP solver searches the feasible space in the direction of the objective function. Again, the optimum solution is an extreme point of the feasible region. As M_1 and M_2 are minimum-cost network flow problems, their extreme solutions are even more restrictive for feasible scenario exploration.

In this section, we propose methods to eliminate the drawbacks of LP-based scenario generation and exploit the advantages to effectively explore the feasible scenario space for operational-level analysis. Besides, we demonstrate the implementation results with an example process flow on the scenario generator employing a case study that generates scenarios based on pan-European historical flow data. Thus, in the ensuing subsections, we first introduce the details of the data used for the demonstration, and then we introduce our methods for effective scenario generation using the LP-based scenario generation using the data.

4.1 Case study settings

For demonstration, we select Germany as the region for generating the node-based gas flow scenarios. We preprocessed the available open gas transport network topology data set for Germany provided by the LKD-EU project [LE18, KKS$^+$17, KWH$^+$17] and generated a sufficiently high-quality data set for operational-level analysis. Moreover, the average demand distribution data in the same data sets based on the average demand of NUTS3 regions in Germany is employed.

For the high-level pan-European gas transport network topology, we use the balancing zone information in the ENTSOG-IP database [ENT18d]. For maintaining consistency, the capacities for the links in the topology are taken from the ENSTOG Security of Supply report data [ENT17c] that also constitutes the demand and supply flow data set. The data set includes the monthly historical demand data as well as the indigenous production data. We demonstrate our results based on November 2017 data provided in the report. In the data set, the European Union (EU) countries have demand including the consumption of final consumers, collected from the TSOs. This demand data is aggregated at the balancing zone level. This is important, especially for countries having more than one balancing zone in November 2017 including Germany. The amount of gas exported to the non-EU countries from the EU countries is not explicitly listed in the report. It is stated in the report that, historically, the total export of the EU countries is 5% of their winter demand, 2% of which is sent to Ukraine. We add the associated gas amount as demand to the listed gas systems of the non-EU countries. This dataset further includes the average imports to EU countries from non-EU countries including Russia, Norway, Algeria, and Libya.

We use the storage data including working volume, and injection/withdrawal rates based on the existing gas from GIE [Gas21a] aggregated by the balancing zones as given in the ENTSOG Security of Supply report [ENT17c].

We match the boundary nodes of the improved Germany gas transport network dataset with the IPs in the ENTSOG IP data set [YEZW+20] to use the NT_1 of the C2F approach for integrating M_1 and M_2. In addition, for each NUTS3 region, we find the nodes within a predefined distance to the epicenter of the NUTS3 region. When defining the distance, we make sure all NUTS3 regions are served by at least one exit node in V_P^-. Then, we mark at most the k closest exit nodes to the epicenter as admissible nodes to serve this particular NUTS3 to meet the final consumer demand.

4.2 Effective scenario generation

Enhancing objective function: We aim to explore alternative ways to route gas throughout Europe. Thus, we eliminated cost parameters from the models. The models generate optimal solutions to maximize the amount of gas in the network while meeting the capacity constraints. To change the search direction as needed, we define the following two parameters in the models.

- In M_1, ϵ_1 does not only serve as a scaling parameter but also provides flexibility to the user to control the amount of gas injected into the UGSs. If the epsilon is set to zero, then M_1 uses gas only to fulfill the demand. Yet, this is not realistic in the presence of the UGSs. The supplied gas can be injected into the UGSs, especially in summer, when the demand is less than the supply. Likewise, UGSs act as a gas supply when the supply from imports or LNG is low. Hence, ϵ_1 enables us to model the seasonality effect or impact of gas prices on the gas transport to explore the potential gas transport scenarios.
- In M_2, the objective is to route as much gas as possible given the capacities of the gas transport network to meet the demand distributed to the subregions. Slightly perturbing the demand distribution allows us to explore scenarios for robustness analysis of the gas transport network to the demand uncertainties. In some cases, this also is necessary to cope with the data uncertainty associated with the demand distribution. Hence, we use a non-negative ϵ_2 to perturb the total deviation of demand per region from the demand distribution, which is $\pi_r^+ + \pi_r^-$. Thus, we let the model to redispatch some amount of gas between subregions to allow more gas to be routed in the network.

Generating artificial feasible solutions We aim to generate node-based scenarios that are feasible both at the pan-European level and the selected region subject to the linearized regional gas network capacities. Besides, we also aim to know how much of the available supply or demanded gas could be routed in the network. We enable this by introducing slack variables, i.e., the demand curtailment variable, y_i in M_1, and artificial arc (t, k) connected to the artificial nodes denoting the infrastructure types in M_2. Also enabled by our C2F approach integrating the two models, we can comment on the feasibility of the scenarios at each modeling layer.

More specifically, the resulting feasible scenarios from M_2 are still feasible in M_1 after applying just NT_1 and NT_2. However, when we apply NT_3, the scenarios are feasible for both M_1 and M_2 if the amount of flow on $(t, k) \in A_{Ar}$ is zero, independent of the exchanged gas amounts between the exit nodes. Otherwise, the feasibility of the scenario in M_1 depends on the gas that cannot be routed by M_2. This should be investigated for individual cases by checking the non-negative flows on the arcs (j, t) and (k, j) in M_2 solution for all $j \in V_{Ar}$.

Using this method, we can explore the feasible scenarios in the neighborhood of a base scenario, as well as get insight into the infeasibility cases. Hence, we can find the nearest possible solution to the feasible solution in the case of infeasibility.

Incorporating side constraints We utilize side constraints to model unprecedented events or emerging technologies, deal with data uncertainties, and, last but not least, effectively generate alternative solutions to the models by changing the basic feasible solution (BFS) structure.

To exemplify, we can explore the effects of potential future regulations such as the injection of H_2 in the gas network no more than x% of the natural gas on a single pipe by adding side constraints to the M_2 for the entry nodes at a region where a P2G facility is built [MWSYE21]. Again, in M_2, the θ parameter provides flexibility to change the demand distribution of the region by π_r^- and π_r^+ variables. If there is room for total deviation from the final consumer demand, then there can be better solutions that conform to the demand distribution. We can also add side constraints on π_r^- and π_r^+ variables for particular regions, to make gas exchange admissible or inadmissible between different regions, i.e., depending on the distance or the regions. Similarly, in M_1, we can set flow variables denoting the cross-border gas flow between some gas systems or countries to zero, if there are any inadmissible flow constraints that we would like to explore.

On the other hand, we deal with the limitations on the generated scenarios dictated by the structure of the BFSs by adding side constraints. The reason is that an optimal solution to an LP is an extreme point with non-basic variables at the upper bound or at the lower bound. This has a restrictive implication for the solutions. For instance, in M_2, if a node in V_{Ar} is connected to multiple nodes in V_P^+, the solution may involve only the cases where some of the pipelines are utilized to their capacity while some of them are not utilized at all. To explore the feasible scenario space effectively, we add additional constraints to make sure that gas is shared among such pipelines.

Generating alternative solutions: We integrate these methods and models using the following workflow.

- Generate feasible scenarios to the pan-European network: by varying ϵ_1 in M_1, we generate alternative feasible scenarios for the pan-European gas network. Since all resulting solutions are feasible to M_1, their convex combinations are also feasible to M_1, enabling us to explore the feasible scenario space for the pan-European network.
- Generate feasible solutions to the regional network: For each representative solution of M_1, we select an appropriate ϵ_2 and θ to find a feasible scenario that maximizes the amount of the gas routed by the regional network using the original M_2. Here, each ϵ_2 and θ pair represents a different distribution for the final consumer demand. Hence, to generate alternative scenarios for a particular demand distribution case, we use the side constraints to share the gas among pipelines given the resulting demand distribution and routed gas amount in the original model. So, we pursue a three-step procedure to generate alternative feasible node-based scenarios:
    1. given ϵ_2, θ, and M_1 solution, generate a feasible solution, Sol_1, using M_2 that maximizes the amount of routed gas in the regional network: this is solving the original M_2 as presented in subsection 3.2
    2. update M_2 to generate a feasible solution, Sol_2, that minimizes the maximum flow on the artificial arcs adjacent to a single artificial node in V_{Ar} given the total deviation of final consumer demand and flow amount on the arc (t, k): we drop the objective function in M_2, add constraints to the decision variables to bound them by the Sol_1 and add side constraints to manipulate the BFS structure of the original M_2. We also add a new objective function to force the model to share the gas among the alternative boundary nodes of the regional network.
    3. generate feasible node-based scenarios by convex combination of Sol_1 and Sol_2

4.3 Case Study Results

In the case study, by changing ϵ_1 in the interval $[-1.1]$, we explored the scenarios that minimized the gas demand that could not be met. As a result, we came up with five alternative feasible solutions to M_1 with the same amount of curtailed gas demand as the scenario with $\epsilon_1 = 0$, as presented in Figure 4. In the figure, zero corresponds to $\epsilon_1 = 0$, *max* (*min*) scenarios with $\epsilon_1 > 0$ ($\epsilon_1 < 0$) maximizing (minimizing) the gas injected to UGSs. These solutions involve three distinct levels of change in the gas levels of the UGSs, as well as supply from LNG facilities, while the cross-border exchange differs in all five scenarios. In Figure 5, the change in the cross-border exchange of Germany is presented in comparison with scenario *zero* whose $\epsilon_1 = 0$. Here, it is important to note that, since all these five scenarios are feasible solutions to the M_1, their convex combinations also constitute feasible solutions, hence they are valid scenarios for generating node-based scenarios of the selected region.

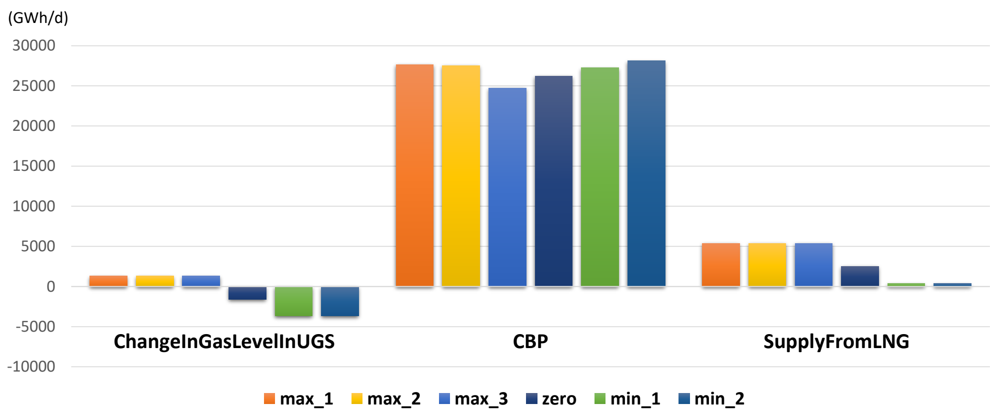


Figure 4: Gas Exchange Between Systems in Alternative Optimum Solutions

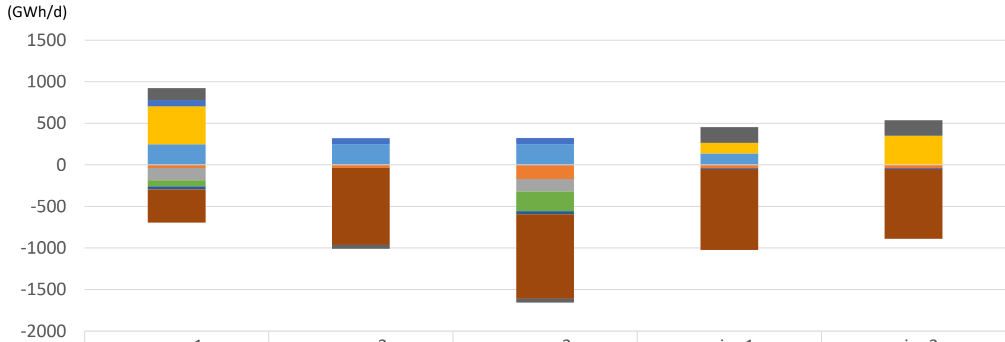
Then, using the pan-European gas transport scenario with $\epsilon_1 = 0.1$, we explored the feasible regional node-based scenarios for the German gas transport network. Here, we selected $\epsilon_2 = 1$ and $\theta = 0$ to explore the scenarios where the final consumer demand in the pan-European gas transport scenario is met with minimum deviation from the given demand distribution in the input dataset.

In Figure 6, we visually compare the generated node-based scenarios Sol_1 and Sol_2 by M_2 following the flow presented above. From this scenario visualization, we easily observe that the scenario generated by M_2 without adding additional constraints involves the injection of gas from a cumulative storage infrastructure through a single entry point. This particular example demonstrates the ability of the scenario generation tool to navigate between the alternative feasible node-based scenarios.

5 Implementations of Scenario Generator

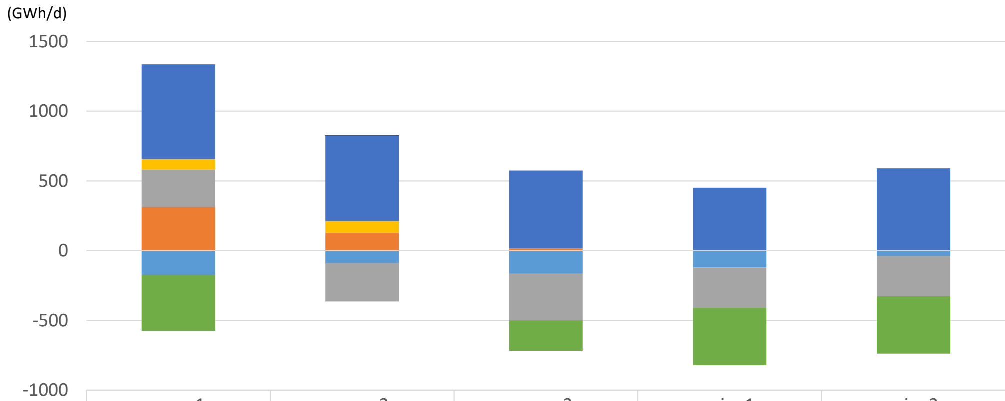
In our research, the scenario generator has been instrumental in creating consistent node-based scenarios for utilization in an operational-level gas network optimization model.

In the first application, derived from the Horizon 2020 funded project plan4res, we employed the scenario generator to integrate high-resolution gas demand/supply profiles from the electricity grid operation model and aggregated gas demand/supply forecast data for the pan-European gas transport network [DCM+20]. The former profiles were hourly time series per postal code region for the central western European countries including Germany, and per country for other EU countries, while the latter were yearly forecasts per country. Besides, the electricity grid operation data did not include the supplied gas



	max_1	max_2	max_3	min_1	min_2
PL	144	-49	-49	185	185
NL	-398	-921	-1013	-975	-836
LU	-37	0	-37	0	0
FR	-73	0	-238	0	0
DK	76	73	76	-10	-10
CZ	456	0	0	130	350
CH	-151	0	-151	0	0
BE	-35	-39	-169	-41	-41
AT	247	247	247	137	0

(a)



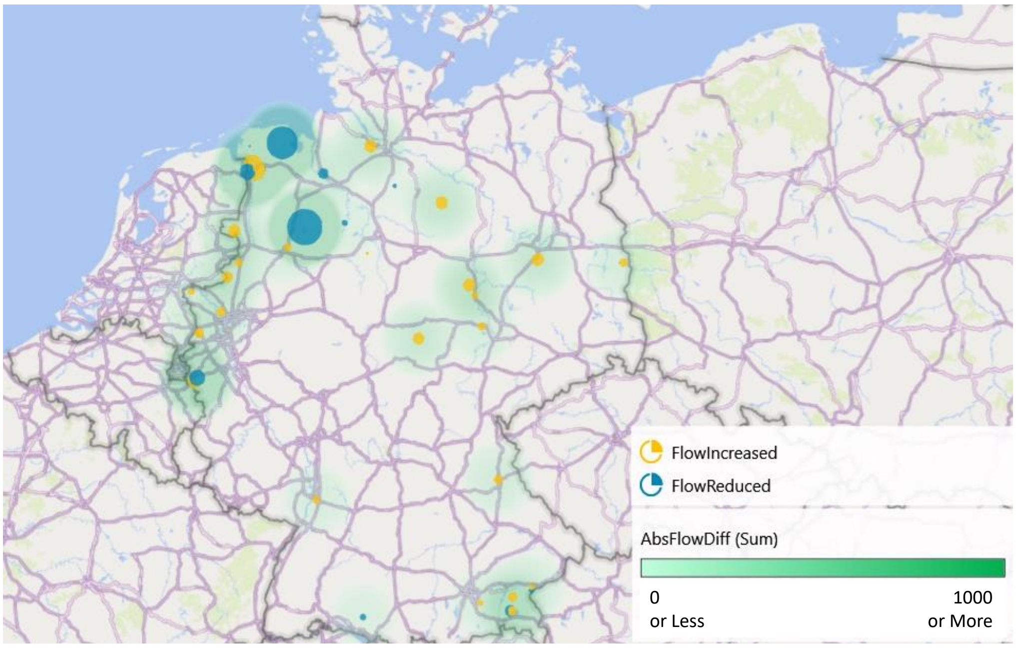
	max_1	max_2	max_3	min_1	min_2
PL	0	0	0	0	0
NO	-401	0	-218	-412	-412
NL	680	612	558	452	591
DK	76	83	0	0	0
CZ	267	-275	-334	-288	-288
BE	313	129	16	0	0
AT	-175	-90	-166	-121	-38

(b)

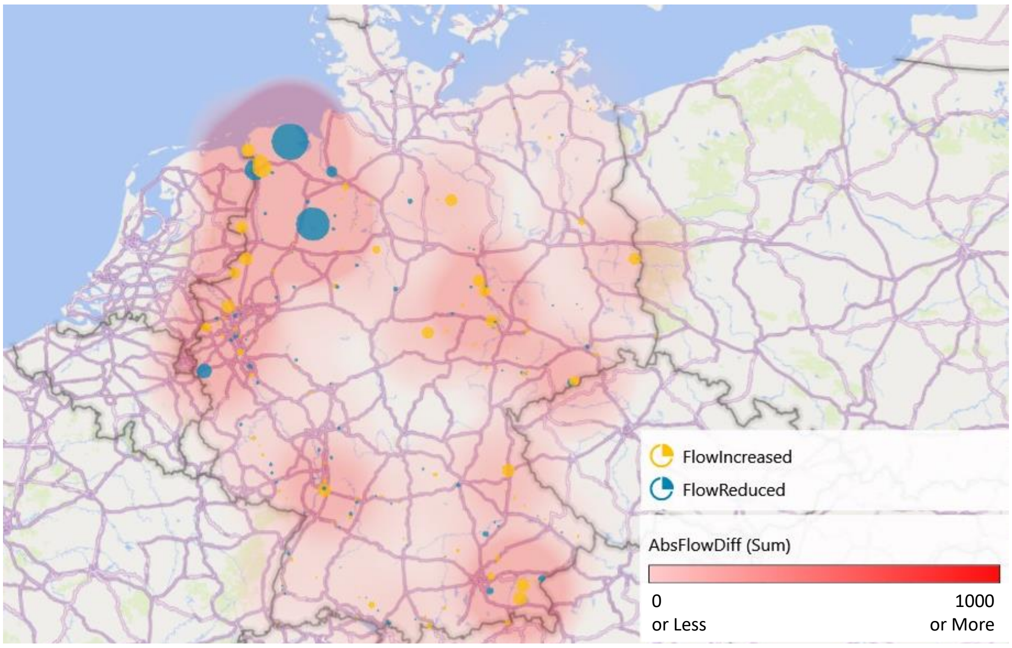
Figure 5: Change in Cross-Border Gas Exchange: (**a**: from Germany to its Neighbors) (**b**: to Germany from its Neighbors)

from non-EU countries as well as UGSs and LNGs. Thus, the scenario generator enabled us to generate the required node-based scenarios for operational-level gas network feasibility analysis of the German gas transport network by integrating the pan-European multi-energy systems data consistently. This allowed us to make the feasibility analysis of the strategic level techno-economic analysis of the pan-European energy system at the operational level by analysis of the integrated electricity and gas grids. Hence, we developed a single modeling framework to explore multi-modal pan-European energy concepts, considering sector coupling and CO2 emission reduction goals, and to assess their operational feasibility [YEMW$^+$23].

In the second application, we integrated two scenario generation tool models, M_1 and



(a)



(b)

Figure 6: Change in Gas Flow Direction and Amount Among Scenarios: (**a**: Entry nodes) (**b**: Exit nodes)

M_2, into the operational gas network optimization model at the data exchange level to assess and improve the quality of a gas network topology dataset [YEKHZ22]. M_1 and M_2 represent the logical pan-European gas network and the linearized German gas transport network, respectively. With the help of the scenario exploration ability of the scenario generator, we could generate an extensive number of node-based scenarios using open historical flow data. The generated scenarios facilitated the analysis of the gas transport network at varying levels of detail and precision, connecting three layers of models (see Figure 7). Notably, our study identified systematic errors including those related to pipeline capacity attributes in the linearized gas network data set, leading to a data improvement study and a reevaluation of assumptions in pipeline capacity calculations.

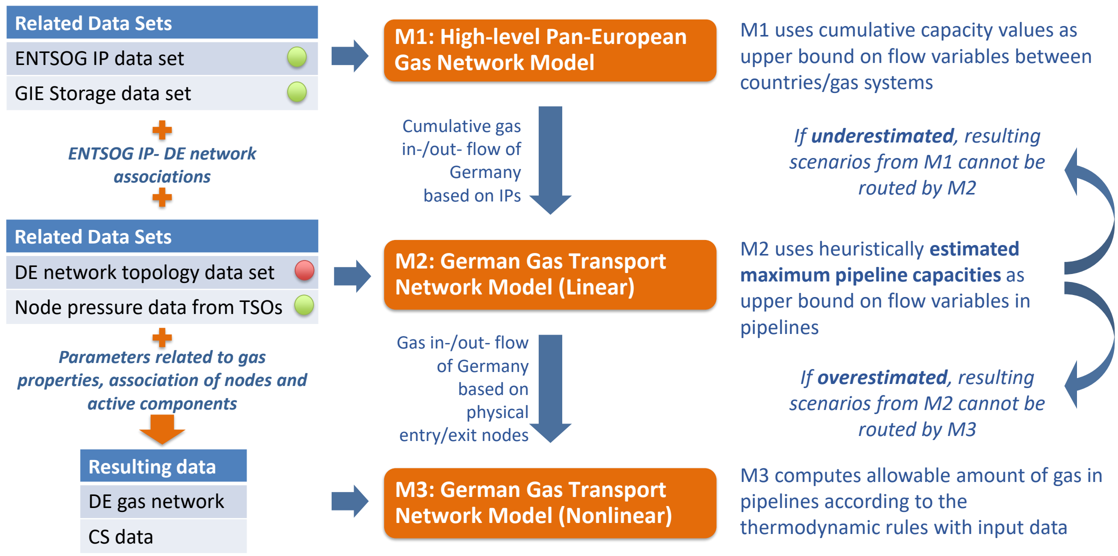


Figure 7: The hierarchical model set-up for data error detection

These successful implementations underscore the significance of the scenario generator in enhancing the accuracy and reliability of gas network optimization studies in diverse application fields.

6 Conclusion

In conclusion, our paper presents the development and implementation of a mathematical-modeling-based scenario generator designed to facilitate operational analysis of gas networks. This novel approach contributes to augmenting open data with mathematical modeling knowledge to improve the quality of data, particularly beneficial in multilayer network settings where open data addresses the aggregated upper layers. The case study showcased the scenario generator's capabilities, ensuring feasibility within European-level constraints and demand distribution needs for the chosen region, eliminating infeasible scenarios based on the relaxed linear gas network optimization model, and allowing exploration of the feasible scenario space through the definition of auxiliary variables. Our implementation in various applications demonstrated its value in both energy and open research data contexts.

Looking into the future, several enhancements and extensions for our scenario generator can be envisioned. Firstly, the current static nature of the tool, which generates a single snapshot for stationary analysis, could be transformed into a dynamic system capable of producing time series for a dynamic analysis. Secondly, while the tool is intentionally cost-agnostic, the incorporation of cost information could enable economic analyses, such as computing demand curtailment costs in M_1 and redispatching costs in M_2, expanding the tool's utility for economic assessments. Thirdly, originally designed for natural gas, the tool can be easily modified to accommodate hydrogen (H_2) networks in future applications. Lastly, the C2F method employed to integrate M_1 and M_2 shows promise for application in multilayer networks in various settings, particularly when coupled with NT_2 as a geographical disaggregation method. These prospects highlight the adaptability and potential broad impact of our scenario generator in diverse research and analysis domains.

Acknowledgements

The study received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 773897 and has been partly conducted in the Research Campus MODAL funded by the German Federal Ministry of Education and Research (BMBF) (fund numbers 05M14ZAM, 05M20ZBM).

References

[CBB+14] Rui Carvalho, Lubos Buzna, Flavio Bono, Marcelo Masera, David K Arrowsmith, and Dirk Helbing. Resilience of natural gas networks during conflicts, crises and disruptions. *PloS one*, 9(3):e90265, 2014.

[CBvAO+21] Sandrine Charousset-Brignol, Wim van Ackooij, Nadia Oudjane, Dominique Daniel, Slimane Noceir, Utz-Uwe Haus, Alfio Lazzaro, Antonio Frangioni, Rafael Lobato, Ali Ghezelsoflu, Niccolò Iardella, Laura Galli, Enrico Gorgone, Mauro dell'Amico, Spyros Giannelos, Alex Moreira, Goran Strbac, Stefan Borozan, Paula Falugi, Danny Pudjianto, Lothar Wyrwoll, Carlo Schmitt, Marco Franken, Daniel Beulertz, Henrik Schwaeppe, Dieter Most, Inci Yüksel-Ergün, Janina Zittel, and Thorsten Koch. Synergistic approach of multi-energy models for a european optimal energy system management tool. *The project repository journal*, 9:113 – 116, 2021.

[DCM+20] Mauro Dell'Amico, Sandrine Charousset, Dieter Most, Inci Yueksel Erguen, Daniel Beulertz, Spyros Gianellos, Dominique Daniel, Selini Hajidimitriou, Slimane Noceir, Nadia Oudjane, Danny Pudjianto, and Goran Strbac. Documentation of the plan4res public dataset - deliverable D4.5. Technical report, 2020. doi:10.5281/zenodo.3785010.

[DWS00] Daniel De Wolf and Yves Smeers. The gas transmission problem solved by an extension of the simplex algorithm. *Management Science*, 46(11):1454–1465, 2000.

[ENT17a] ENTSOG. Basic features of entsog network and market tool, 2017. Last accessed 10 October 2022. URL: https://www.entsog.eu/sites/default/files/2019-10/BasicfeaturesofENTSOGTYNDPtool.pdf.

[ENT17b] ENTSOG. ENTSOG Glossary of Existing Definitions, 2017. Last accessed 17 June 2022. URL: https://www.entsog.eu/sites/default/files/files-old-website/publications/Tariffs/2017/170421_ENTSOG_Glossaryofdefinitions.pdf.

[ENT17c] ENTSOG. Security of supply scenarios, 2017. Last accessed 10 October 2022. URL: https://www.entsog.eu/security-of-supply-simulation#union-wide-simulation-of-supply-and-infrastructure-disruption-scenarios.

[ENT18a] ENTSOG. ENTSOG Professional Data warehouse System Documentation Glossary and Definitions Contents, 2018. Last accessed 17 June 2022. URL: https://www.entsog.eu/sites/default/files/files-old-website/publications/TYNDP/2018/TYNDP062-18_Glossary_FINAL.pdf.

[ENT18b] ENTSOG. Ten year network development plans, 2018. Last accessed 10 October 2022. URL: https://www.entsog.eu/tyndp#.

[ENT18c] ENTSOG. Transmission Capacity and System Development Maps, 2018. Last accessed 17 June 2022. URL: https://www.entsog.eu/maps#system-development-map.

[ENT18d] ENTSOG. Transparency platform, 2018. Last accessed 10 October 2022. URL: https://transparency.entsog.eu/.

[ENT19] ENTSOG. ENTSOG TYNDP 2018 – Annex C - Capacities, 2019. Last accessed 17 June 2022. URL: https://www.entsog.eu/sites/default/files/2019-05/TYNDP2018-AnnexC-Capacities.xlsx.

[Eur18] European Commission. Horizon 2020 web site, 2018. Last accessed 10 October 2022. URL: https://ec.europa.eu/programmes/horizon2020/en.

[FNB21] FNB Gas. Hydrogen core network, 2021. Last accessed 12 March 2024. URL: https://fnb-gas.de/wasserstoffnetz-wasserstoff-kernnetz/.

[Fra23] Fraunhofer Institute for Systems and Innovation Research ISI - Press Release. New data set reveals the natural gas dependency of space heating in several eu countries, 2023. Last accessed 15 December 2023. URL: https://www.isi.fraunhofer.de/en/presse/2022/presseinfo-18-erdgas-abhaengigkeit-eu-raumwaerme-sektor-neue-daten.html.

[Gas21a] Gas Infrastructure Europe (GIE). Gie aggregated gas storage inventory - storage data, 2021. Last accessed 24 September 2021. URL: https://agsi.gie.eu/#/.

[Gas21b] Gas Infrastructure Europe (GIE). Gie web site, 2021. Last accessed 24 September 2021. URL: https://www.gie.eu/.

[GMSS18] Björn Geißler, Antonio Morsi, Lars Schewe, and Martin Schmidt. Solving highly detailed gas transport minlps: block separability and penalty alternating direction methods. *INFORMS Journal on Computing*, 30(2):309–323, 2018.

[HAHB+21] Felix Hennings, Lovis Anderson, Kai Hoppmann-Baum, Mark Turner, and Thorsten Koch. Controlling transient gas flow in real-world pipeline intersection areas. *Optimization and Engineering*, 22:687 – 734, 2021. doi:https://doi.org/10.1007/s11081-020-09559-y.

[HHH+15] Benjamin Hiller, Christine Hayn, Holger Heitsch, René Henrion, Hernan Leövey, Andris Möller, and Werner Römisch. Methods for verifying booked capacities. In Thorsten Koch, Benjamin Hiller, M. E. Pfetsch, and Lars Schewe, editors, *Evaluating Gas Network Capacities*, chapter 14, pages 291–315. SIAM-MOS series on Optimization, PA, Phyledelpia, 2015. doi:10.1137/1.9781611973693.ch14.

[HHL+15] Holger Heitsch, René Henrion, Hernan Leövey, Radoslava Mirkov, Andris Möller, Werner Römisch, and Isabel Wegner-Specht. Empirical observations and statistical analysis of gas demand data. In Thorsten Koch, Benjamin Hiller, M. E. Pfetsch, and Lars Schewe, editors, *Evaluating Gas Network Capacities*, chapter 13, pages 273–290. SIAM-MOS series on Optimization, PA, Phyledelpia, 2015.

[HKS+18] Benjamin Hiller, Thorsten Koch, Lars Schewe, Robert Schwarz, and Jonas Schweiger. A system to evaluate gas network capacities: Concepts and implementation. *European Journal of Operational Research*, 270(3):797–808, 2018. URL: https://www.sciencedirect.com/science/article/pii/S0377221718301681, doi:https://doi.org/10.1016/j.ejor.2018.02.035.

[HS19] Falk M. Hante and Martin Schmidt. Complementarity-based nonlinear programming techniques for optimal mixing in gas networks. *EURO Journal on Computational Optimization*, 7(3):299–323, 2019.

[HS20] Falk M. Hante and Martin Schmidt. Gas transport network optimization: Mixed-integer nonlinear models. In Panos M. Pardalos and Oleg A.

Prokopyev, editors, *Encyclopedia of Optimization*, pages 1–8. Springer International Publishing, Cham, 2020. doi:10.1007/978-3-030-54621-2_873-1.

[KHPS15] Thorsten Koch, Benjamin Hiller, M. E. Pfetsch, and Lars Schewe, editors. *Evaluating Gas Network Capacities.* SIAM-MOS Series on Optimization, PA, Phyledelpia, 2015.

[KKS+17] Friedrich Kunz, Mario Kendziorski, Wolf-Peter Schill, Jens Weibezahn, Jan Zepter, Christian von Hirschhausen, Philip Hauser, Matthias Zech, Dominik Moest, Sina Heidari, Björn Felten, and Christoph Weber. Data Documentation 92: Electricity, Heat, and Gas Sector Data for Modeling the German System, 2017. Last accessed 25 August 2022. URL: https://d-nb.info/1154864928/34.

[KWH+17] Friedrich Kunz, Jens Weibezahn, Philip Hauser, Sina Heidari, Wolf-Peter Schill, Björn Felten, and Christoph Weber. Reference Data Set: Electricity, Heat, and Gas Sector Data for Modeling the German System (Version 1.0.0), 2017. doi:10.5281/zenodo.1044463.

[LE18] LKD-EU. Lkd-eu project website, 2018. Last accessed 16 September 2021. URL: https://www.diw.de/de/diw_01.c.537097.de/projekte/langfristige_planung_und_kurzfristige_optimierung_des_elektrizitaetssystems_in_deutschland_im_europaeischen_kontext_lkd_eu.html.

[MTZD23] B. McWilliams, S. Tagliapietra, G. Zachmann, and T. Deschuyteneer. Preparing for the next winter: Europe's gas outlook for 2023. Technical Report Policy Contribution 01/2023, Bruegel, Brussels, Belgium, 2023. URL: https://www.bruegel.org/policy-brief/european-union-gas-survival-plan-2023.

[MWSYE21] Dieter Most, Lothar Wyrwoll, Carlo Schmitt, and Inci Yueksel-Erguen. plan4res D2.2 - Case Study 1 Report - Multimodal energy concept for achieving Europe's carbon reduction goals. Technical report, 2021. doi:10.5281/zenodo.5809338.

[PCG+22] Milena Petkovic, Ying Chen, Inken Gamrath, Uwe Gotzes, Natalia Selini Hadjidimitrou, Janina Zittel, Xiaofei Xu, and Thorsten Koch. A hybrid approach for high precision prediction of gas flows. *Energy Systems*, 13:383 – 408, 2022. doi:https://doi.org/10.1007/s12667-021-00466-4.

[PGH+15] Marc E Pfetsch, Ralf Gollmer, Benjamin Hiller, Jesco Humpola, Thorsten Koch, Thomas Lehmann, Alexander Martin, Antonio Morsi, Lars Schewe, Martin Schmidt, Robert Schwarz, Jonas Schweiger, Claudia Stangl, Marc C Steinbach, Stefan Vigerske, and Bernhard M Willert. Validation of nominations in gas network optimization: models, methods, and solutions. *Optimization Methods and Software*, 30(1):15–53, 2015.

[PMD+22] Adam Pluta, Wided Medjroubi, Jan C. Diettrich, Jan Dasenbrock, Hendrik-Pieter Tetens, Javier E. Sandoval, and Ontje Lünsdorf. Scigrid_gas - data model of the european gas transport network. In *2022 Open Source Modelling and Simulation of Energy Systems (OSMSES)*, pages 1–7, 2022. doi:10.1109/OSMSES54027.2022.9769122.

[SAB+17] Martin Schmidt, Denis Aßmann, Robert Burlacu, Jesco Humpola, Imke Joormann, Nikolaos Kanelakis, Thorsten Koch, Djamal Oucherif, Marc E. Pfetsch, Lars Schewe, Robert Schwarz, and Mathias Sirvent. Gaslib—a library of gas network instances. *Data*, 2(4), 2017. URL: https://www.mdpi.com/2306-5729/2/4/40, doi:10.3390/data2040040.

[Sci18] SciGrid Gas Project. Scigrid gas project website, 2018. Last accessed 10 October 2022. URL: https://www.gas.scigrid.de.

[The21] The European Hydrogen Backbone Initiative. The european hydrogen backbone (ehb) initiative website, 2021. Last accessed 14 July 2022. URL: https://ehb.eu/.

[YEKHZ22] Inci Yueksel-Erguen, Thorsten Koch, Felix Hennings, and Janina Zittel. Improving data quality in the presence of superhuman complexity in data errors. In Katsuki Fujisawa, Shizuo Kaji, Toru Ishihara, Masaaki Kondo, Yuji Shinano, Takuji Tanigawa, and Naoko Nakayama, editors, *Construction of Mathematical Basis for Realizing Data Rating Service*, volume 90, 2022. URL: https://joint.imi.kyushu-u.ac.jp/wp-content/uploads/2022/12/29d12d46a22d702596713780acde0689-3.pdf.

[YEKZ23] Inci Yueksel-Erguen, Thorsten Koch, and Janina Zittel. Consistent flow scenario generation based on open data for operational analysis of european gas transport networks. In *Operations Research Proceedings*, 2023. accepted for publication.

[YEMW+23] Inci Yueksel-Erguen, Dieter Most, Lothar Wyrwoll, Carlo Schmitt, and Janina Zittel. Modeling the transition of the multimodal pan-european energy system including an integrated analysis of electricity and gas transport. *Energy Systems*, pages 1–46, 2023.

[YEZW+20] Inci Yueksel-Erguen, Janina Zittel, Ying Wang, Felix Hennings, and Thorsten Koch. Lessons learned from gas network data preprocessing. Technical Report 20-13, ZIB, Takustr. 7, 14195 Berlin, 2020.

[Zus18] Zuse Institute Berlin. Gaslib - a library of gas network instances, 2018. Last accessed 10 October 2022. URL: https://gaslib.zib.de.