SPONSORED BY THE

Federal Ministry of Education and Research

MODAL FORSCHUNGSCAMPUS

FORSCHUNGS CAMPUS öffentlich-private Partnerschaft für Innovationen

Data Preprocessing and Data Quality Assessment for Energy System Optimization

Inci Yüksel-Ergün, Thorsten Koch, Janina Zittel

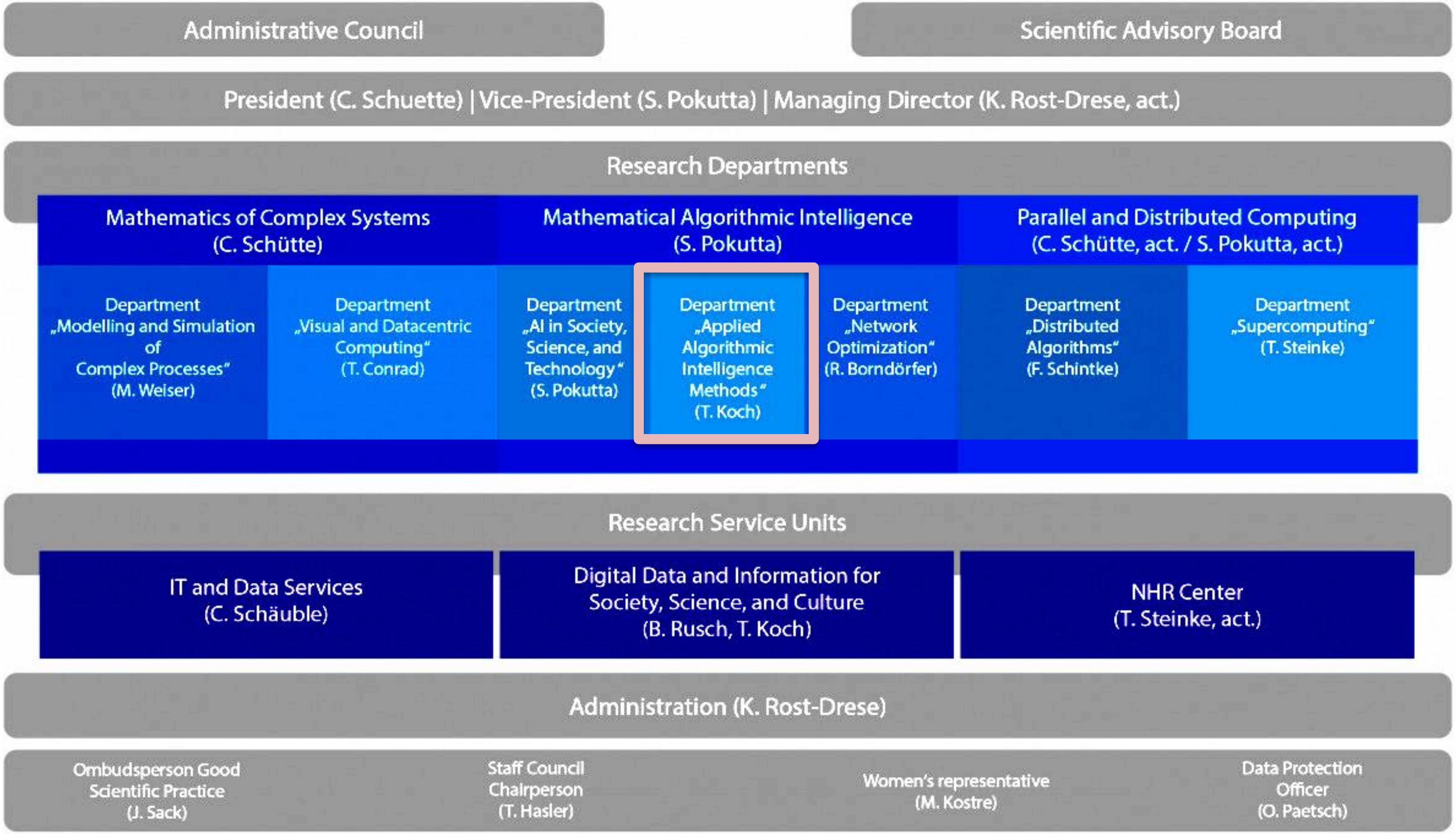
CO@WORK 2024
25.09.2024

Chair of
Software and Algorithms for
Discrete Optimization

Department:
Applied Algorithmic
Intelligence Methods

ZUSE INSTITUTE BERLIN

Applied Algorithmic Intelligence Methods (A^2IM) Department



Applied Algorithmic Intelligence Methods (A²IM) Department

Applied Algorithmic Intelligence Methods (A²IM)

Aims at computing **smart** decisions

Applies advanced **algorithmic intelligence methods** from mathematical optimization and machine learning to explore algorithmic solutions for **real-world problems**

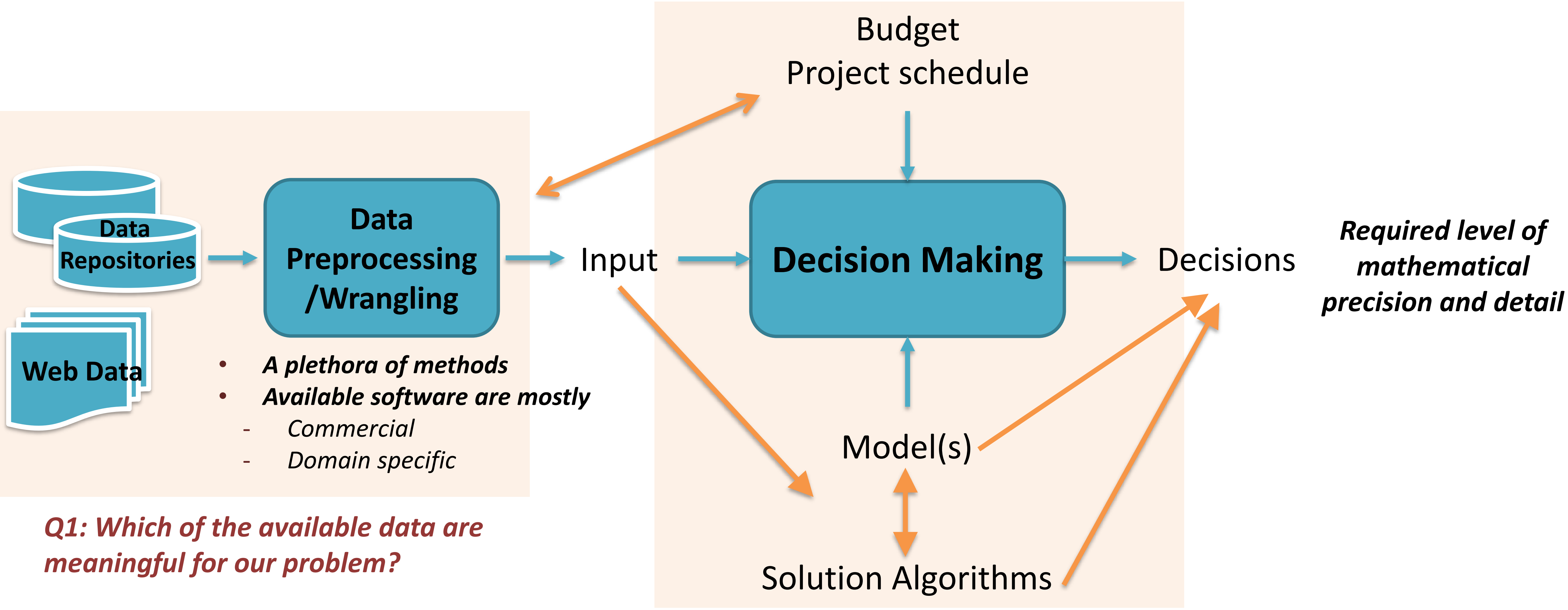
Addresses better planning, extension, and control of **vital and complex infrastructure networks**

MODAL FORSCHUNGSCAMPUS

Stadt-Land-Energie

ForNe | TRR 154 | Mathematische Modellierung, Simulation und Optimierung am Beispiel von Gasnetzwerken | UNSEEN | plan4res | motif

Data Preprocessing vs Decision Making



- ***A plethora of methods***
- ***Available software are mostly***
  - *Commercial*
  - *Domain specific*

Q1: Which of the available data are meaningful for our problem?

European Energy System Transition - Challenges

Energy system transition to decarbonization

Decarbonization of energy system vs. Secure energy supply

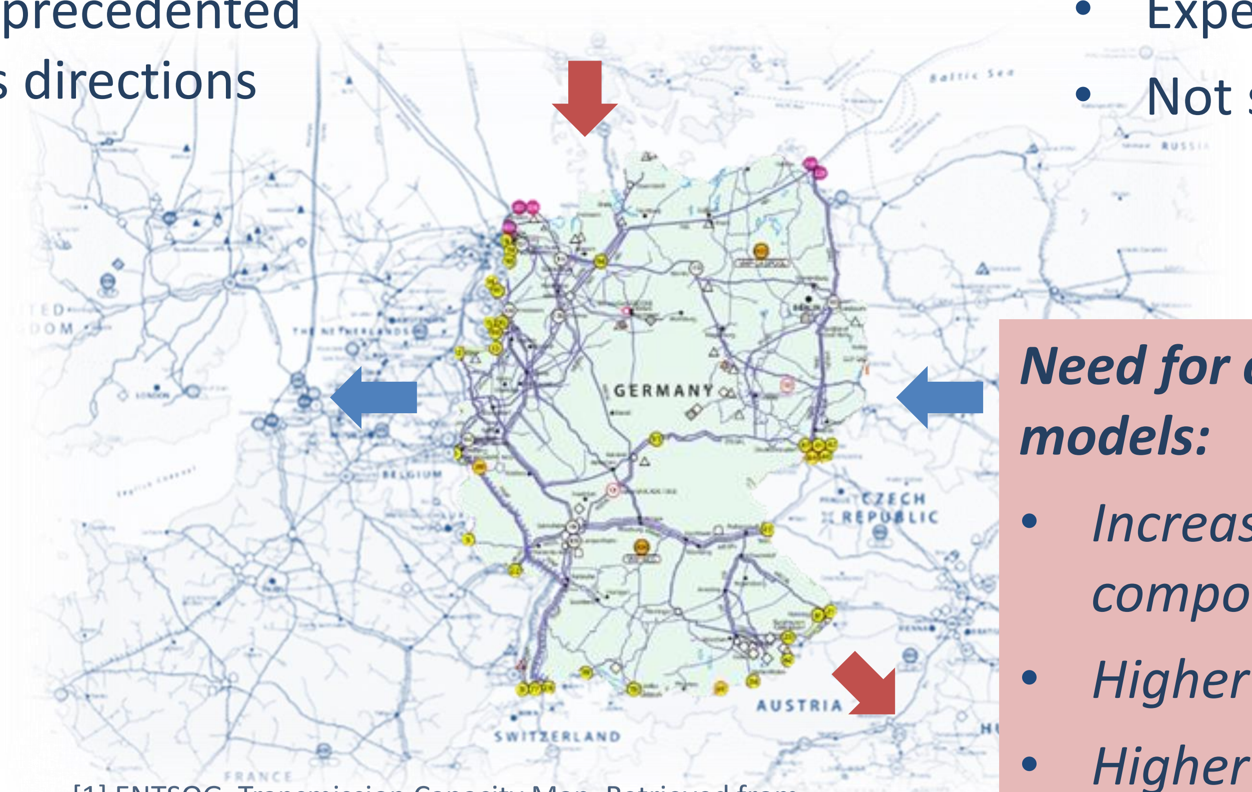
Renewable Energy Vectors by Vecteezy, https://www.vecteezy.com/free-vector/renewable-energy

Impact of Disruptive Events

Russian war on Ukraine

Impacts on gas transport

- Unprecedented gas directions



[1] ENTSOG. Transmission Capacity Map. Retrieved from https://www.entsog.eu/maps#transmission-capacity-map-2021. Accessed on 31.10.2022

Impacts on decision-making models

- Experience from past becomes obsolete
- Not sufficient historical data for the new cases

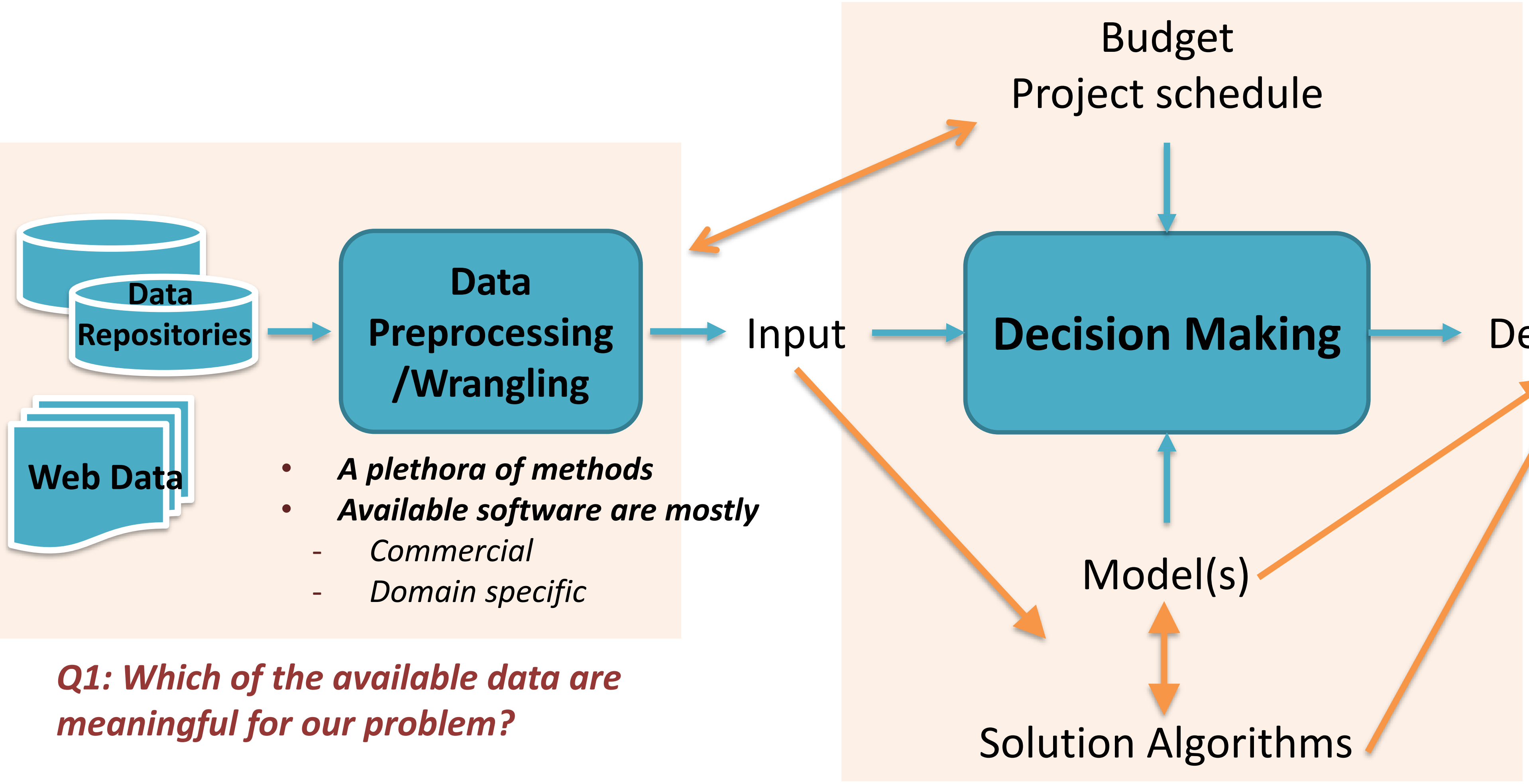
Need for operational-level decision making models:

- *Increased modeling detail → individual network components, such as compressor stations*
- *Higher spatio-temporal resolution*
- *Higher mathematical precision*

Example: Gas Network Planning Problems

Example Planning Problem	Modeling Detail	Computational Size		Modeling Precision	Typical models used in literature:
		Geographical	Temporal		
Investment, Capacity expansion	Flow between countries / important gas systems	Span: Continent Res: Countries / Gas systems	Span: Decades / years	LP	High-level gas network optimization model
	Gas flow on a linearized gas pipeline network	Span: Country Res: Physical nodes	Span: A year/months	LP	Linear gas network optimization model
Nomination Validation	Steady-state physical flow based gas flow on a gas network including active components; operational state of the active components	Span: Country Res: Physical nodes	Span: A day	MINLP	Stationary gas network optimization model
Grid operation, Live dispatch	Time dependent physical flow based gas flow on a gas network including active components; operational state of the active components	Span: Sub-country Res: Physical nodes	Span: Hours Res: Minutes	MINLP	Transient gas network optimization model

Data Preprocessing vs Decision Making



At operational-level required model:

- *is highly detailed,*
- *has high spatio-temporal resolution*
- *has high precision*

Q1: Which of the available data are meaningful for our problem?

Q2: Is the input data set of the sufficient quality?

Dimensions of Data Quality

Accuracy is the ability to reflect reality, i.e., to be able to correctly describe real world objects

Completeness is existence of all required data in the data set, comprehensiveness

Validity is conformance to predetermined type, range and format

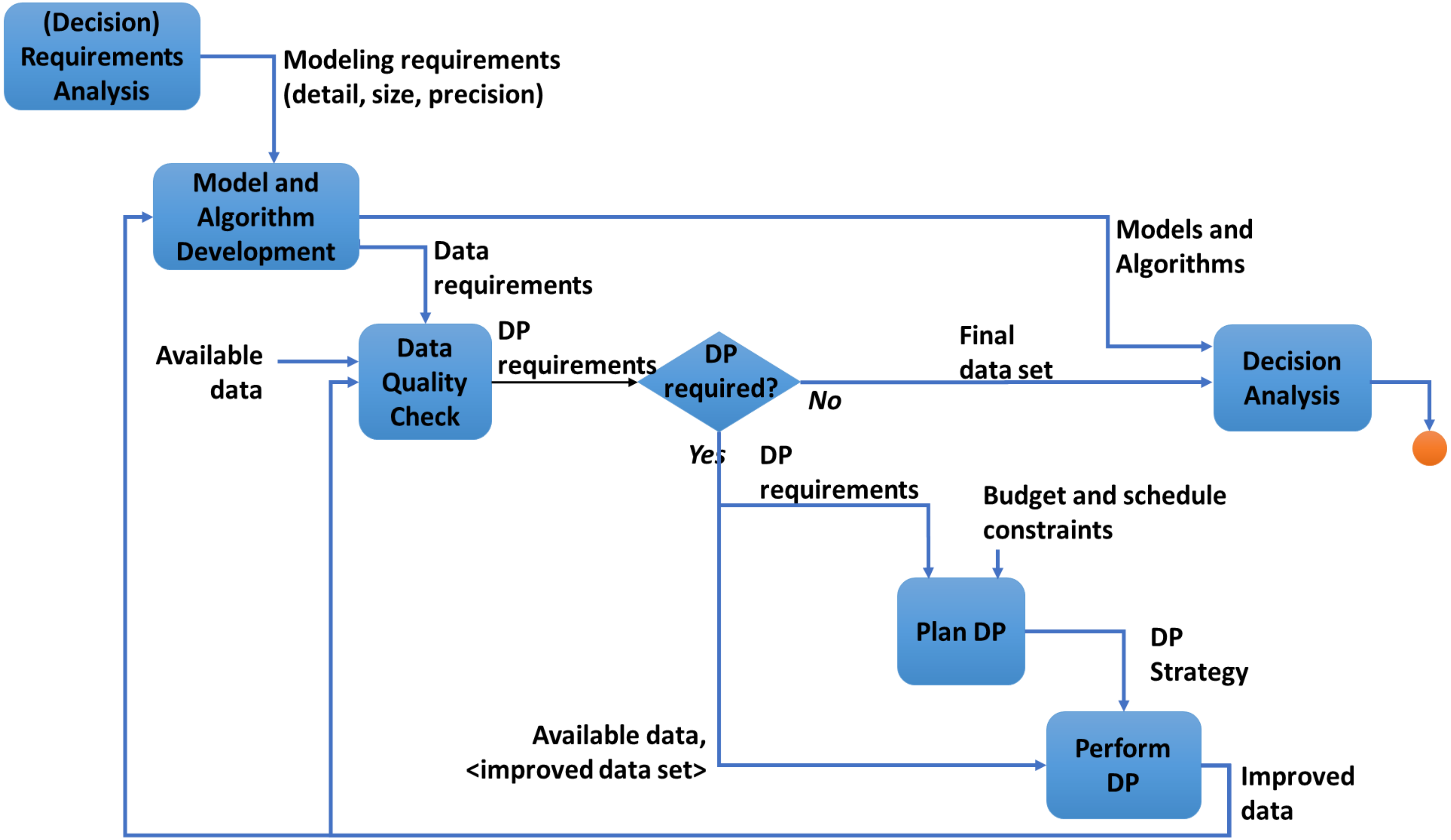
Consistency means conflict-free data, i.e., attributes of data entities addressing the same physical attribute should be consistent

Uniqueness means representing each real world object in the data set exactly once

Timeliness shows whether data is up-to-date data depending on the analysis context

[4] Askham et al. The six primary dimensions for data quality assessment. Technical Report, 2013, DAMA UK.

Selecting a Data Preprocessing Strategy



S1. Data cleaning - small interventions using generic tools and methods

S2. Data modeling - for (further) standardizing input data

S3. Data enrichment - using resources external to the data set

S4. Data error diagnosis and correction - by exploiting knowledge external to the data set

[5] Inci Yueksel-Erguen, J. Zittel, Y. Wang, F. Hennings, T. Koch. Lessons learned from gas network data preprocessing. Technical Report 20-13(v2). Zuse Institute Berlin, Takustr. 7, 14195 Berlin: ZIB, 2020.

Dimensions of Data Quality & Data Preprocessing

[4] Askham et al. The six primary dimensions for data quality assessment. Technical Report, 2013, DAMA UK.

Data Improvement Method	Related Data Quality Dimension	Error
Text mining, fuzzy queries	Uniqueness	Repetition of the objects using different names / ids
Schema validation	Uniqueness Validity	Format mismatch Nonconformance to bounds
Data augmentation	Completeness Accuracy	Insufficient detail Missing entries/attributes
Data generation	Completeness Accuracy	Insufficient detail Missing entries/attributes
Consistency check heuristics	Consistency Accuracy	Conflicts in the data set Wrong modeling assumptions Data preprocessing errors

Examples for consistency check heuristics: [5] Inci Yueksel-Erguen, J. Zittel, Y. Wang, F. Hennings, T. Koch. Lessons learned from gas network data preprocessing. Technical Report 20-13(v2). Zuse Institute Berlin, Takustr. 7, 14195 Berlin: ZIB, 2020.

Gas Network Optimization Model (GNO)

Nomination validation (NoVa): Is the given amounts of gas flow at entries and exits technically feasible? [2]

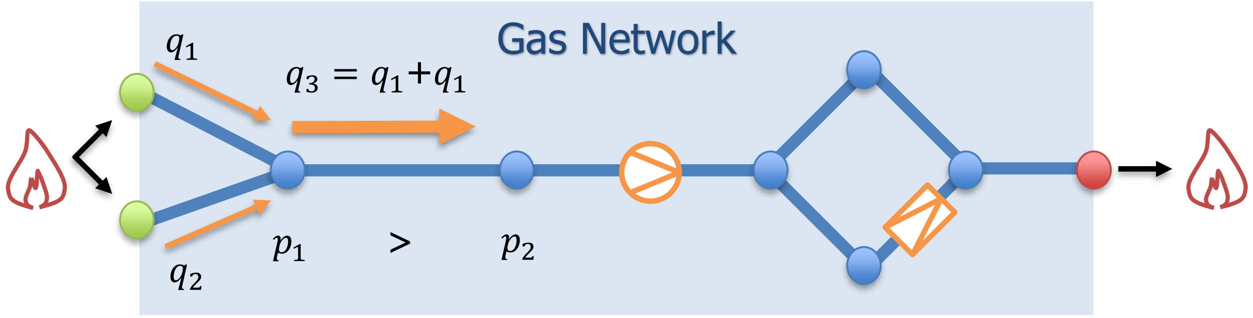
- A stationary gas network optimization model

Input:

- **A gas network**
  - Pipelines, nodes, exits & entries, active components
- **A Scenario**: Amounts of gas flow at entries and exits

Constraints

- Mass flow is conserved at nodes
- Gas moves according to thermodynamic laws
- Gas pressure drops as it flows through pipelines
  - Weymouth Equation → non-linear equations
- Gas pressure is regulated by active components
  - Valves, control valves, compressor stations → subnetworks
  - States: bypassed, closed, active → binary variables



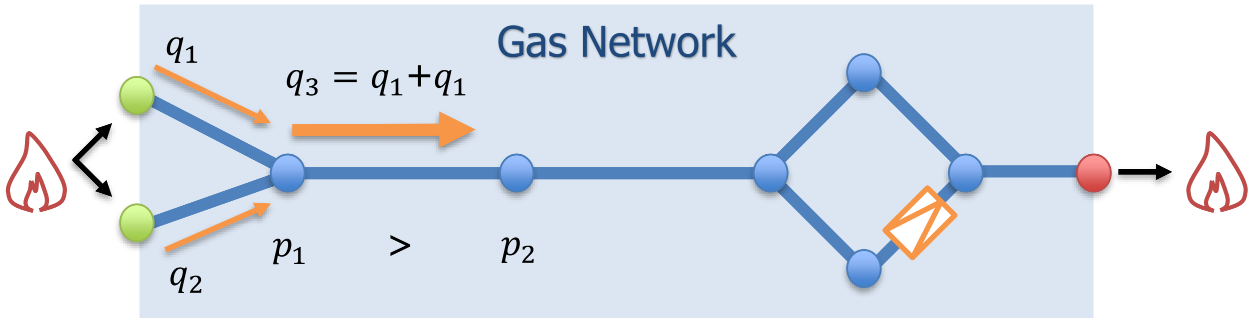
Weymouth Equation: $\alpha(p_1^2 - p_2^2) = \beta q_3 |q_3|$

[2] Koch et al. Evaluating Gas Network Capacities. Society for Industrial and Applied Mathematics, 2015, USA.

Gas Network Optimization Model (GNO)

Nomination validation (NoVa): Is the given amounts of gas flow at entries and exits technically feasible? [2]

- A stationary gas network optimization model



Weymouth Equation: $\alpha(p_1^2 - p_2^2) = \beta q_3 |q_3|$

Input:

- **A gas network**
  - Pipelines, nodes, exits & entries, active components
- **A Scenario**: Amounts of gas flow at entries and exits

Constraints

- Mass flow is conserved at nodes
- Gas moves according to thermodynamic laws
- Gas pressure drops as it flows through pipelines
  - Weymouth Equation → non-linear equations
- Gas pressure is regulated by active components
  - Valves, control valves, compressor stations → subnetworks
  - States: bypassed, closed, active → binary variables

[2] Koch et al. Evaluating Gas Network Capacities. Society for Industrial and Applied Mathematics, 2015, USA.

Gas Network Optimization Model (GNO)

Nomination validation (NoVa): Is the given amounts of gas flow at entries and exits technically feasible? [2]

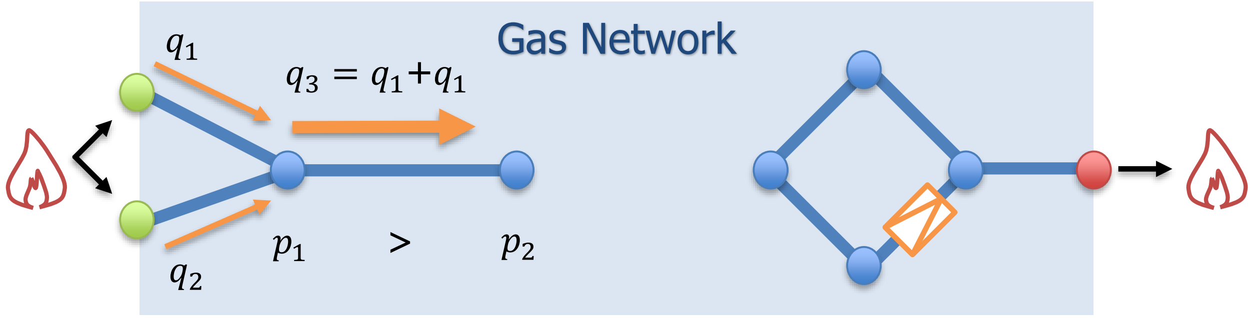
- A stationary gas network optimization model

Input:

- **A gas network**
  - Pipelines, nodes, exits & entries, active components
- **A Scenario**: Amounts of gas flow at entries and exits

Constraints

- Mass flow is conserved at nodes
- Gas moves according to thermodynamic laws
- Gas pressure drops as it flows through pipelines
  - Weymouth Equation → non-linear equations
- Gas pressure is regulated by active components
  - Valves, control valves, compressor stations → subnetworks
  - States: bypassed, closed, active → binary variables



Weymouth Equation: $\alpha(p_1^2 - p_2^2) = \beta q_3 |q_3|$

[2] Koch et al. Evaluating Gas Network Capacities. Society for Industrial and Applied Mathematics, 2015, USA.

Gas Network Optimization Model (GNO)

Nomination validation (NoVa): Is the given amounts of gas flow at entries and exits technically feasible? [2]

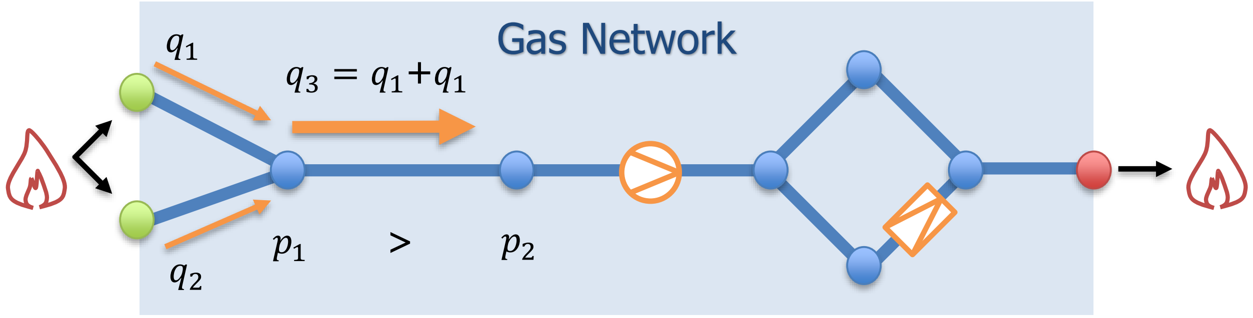
- A stationary gas network optimization model

Input:

- **A gas network**
  - Pipelines, nodes, exits & entries, active components
- **A Scenario**: Amounts of gas flow at entries and exits

Constraints

- Mass flow is conserved at nodes
- Gas moves according to thermodynamic laws
- Gas pressure drops as it flows through pipelines
  - Weymouth Equation → non-linear equations
- Gas pressure is regulated by active components
  - Valves, control valves, compressor stations → subnetworks
  - States: bypassed, closed, active → binary variables



Weymouth Equation: $\alpha(p_1^2 - p_2^2) = \beta q_3 |q_3|$

[2] Koch et al. Evaluating Gas Network Capacities. Society for Industrial and Applied Mathematics, 2015, USA.

Gas Network Optimization Model (GNO)

Nomination validation (NoVa): Is the given amounts of gas flow at entries and exits technically feasible? [2]

- A stationary gas network optimization model
- A mixed integer non-linear program (MINLP)

Input:

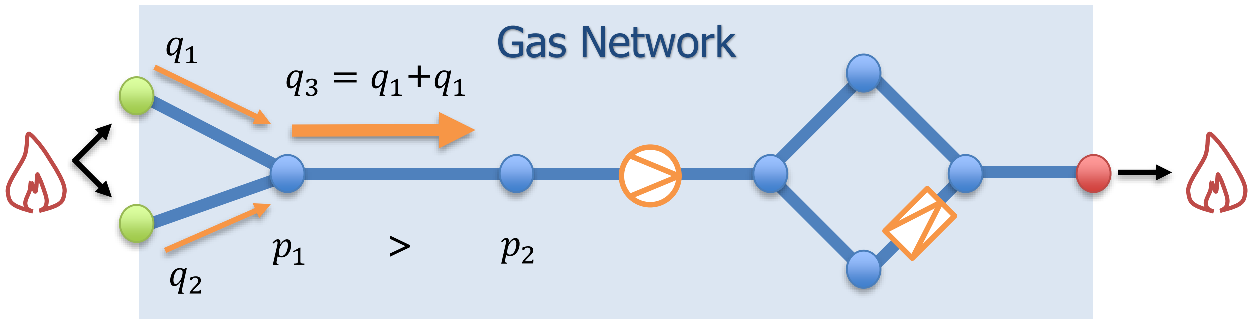
- **A gas network**
  - Pipelines, nodes, exits & entries, active components
- **A Scenario**: Amounts of gas flow at entries and exits

Constraints

- Mass flow is conserved at nodes
- Gas moves according to thermodynamic laws
- Gas pressure drops as it flows through pipelines
  - Weymouth Equation → non-linear equations
- Gas pressure is regulated by active components
  - Valves, control valves, compressor stations → subnetworks
  - States: bypassed, closed, active → binary variables

Output:

- Feasibility of the given scenario
- State of the network for a feasible scenario



Weymouth Equation: $\alpha(p_1^2 - p_2^2) = \beta q_3 |q_3|$

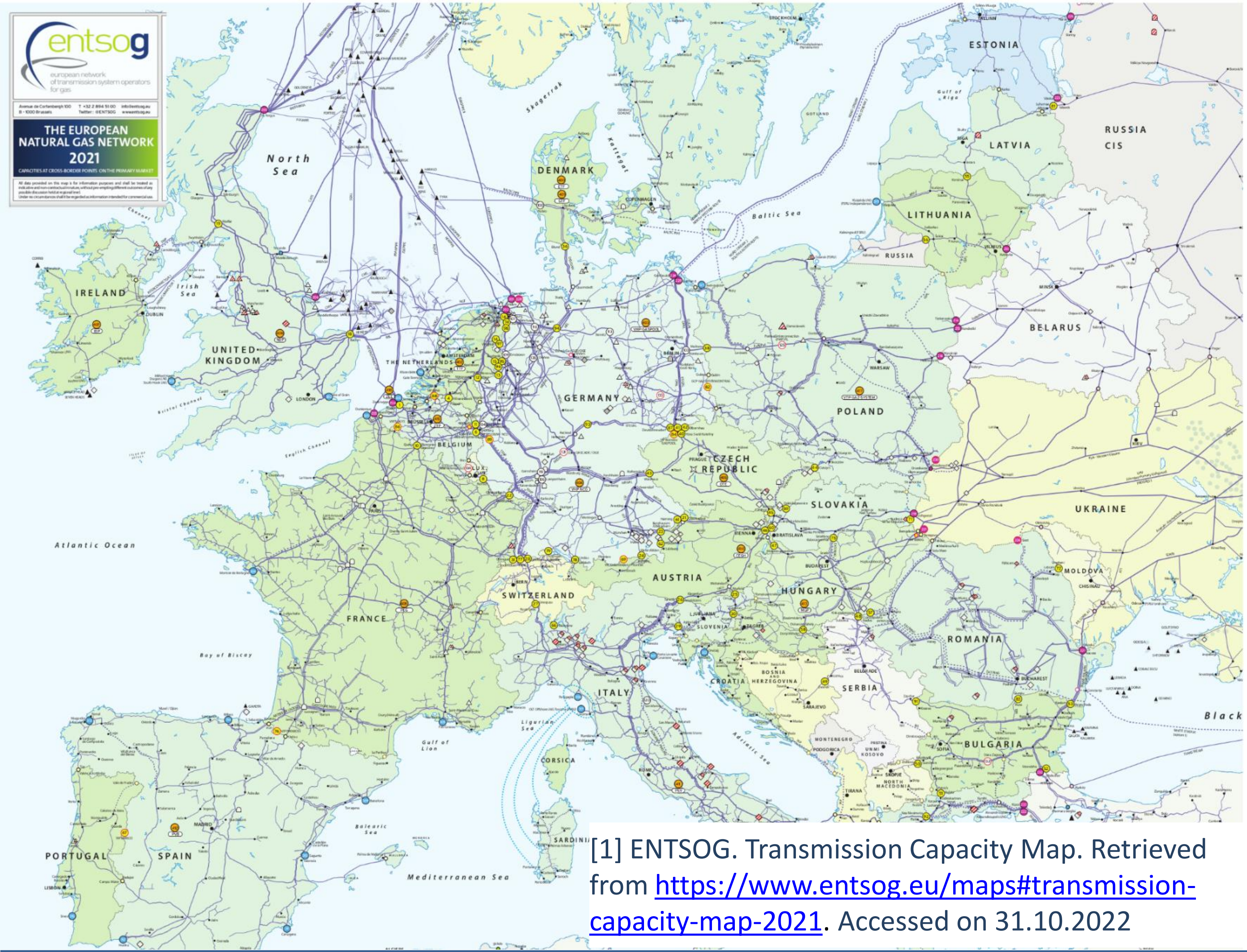
GNO data requirements – GasLib data model [3]:

- A node-based scenario
- Network topology and physical properties data for individual network components
- Compressor station data

[2] Koch et al. Evaluating Gas Network Capacities. Society for Industrial and Applied Mathematics, 2015, USA.

[3] Schmidt et al. GasLib - A Library of Gas Network Instances, Data 2, No. 4, article 40 (2017)

The European Gas Transport Network



[1] ENTSOG. Transmission Capacity Map. Retrieved from https://www.entsog.eu/maps#transmission-capacity-map-2021. Accessed on 31.10.2022

European gas transport network in numbers:

- 42 member, 10 associated partner, 2 observer TSOs
- > 200 interconnection points, > 170 storages
- ≈ 200,000 km transmission pipelines (EU+UK)

Data Limitations:

- Network topology and physical properties data for individual network components
- Compressor station data
- A node-based scenario

Decision-Making with Open Data - 1

Data Sets	Topology	Demand TS
ENTSOG IP [4]	High-Level (cumulative cap.)	Aggregated@IP (H)
ENTSOG SOS [5]		Aggregated@BZ (H)
ENTSOG TYNDP [6]		Aggregated@Cnt (F)
SciGrid Gas [7]	Pipeline Network (approximate cap.)	Subregion-Based
Germany LKD-EU [8]		
Network Topology Data + TSO data (node press.) Compressor station data	Pipeline Network + Active components (computed cap. + gas flow directions)	Node-Based

Operational (GNO):
A routing of gas in the network, resulting gas pressure at the nodes, and operational state of the complex facilities of the network

[6] ENTSOG TP, https://transparency.entsog.eu/, Last accessed 25.08.2023
[7] ENTSOG SOS, https://www.entsog.eu/security-of-supply-simulation#
[8] ENTSOG TYNDP, https://www.entsog.eu/tyndp#, , Last accessed 25.08.2023
[9] SciGRID Gas, https://www.gas.scigrid.de, Last accessed 06.07.2023
[10] Kunz, F. et al.: Reference Data Set (V1.0.0), https://doi.org/10.5281/zenodo.1044463, Last accessed 06.07.2023

Germany Gas Transport Network Data Set – Data Preprocessing

Text mining/ Fuzzy Queries:

- Associating the ENTSOG high-level topology with Germany network topology from LKD-EU data
- Associating the TSO node data with Germany network topology from LKD-EU data

Schema Validation:

- GasLib XML data model

Data augmentation:

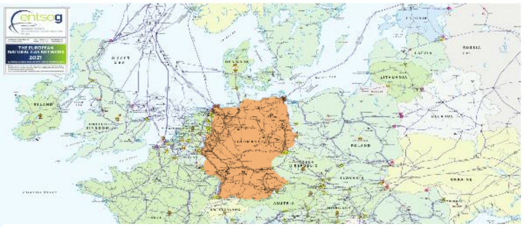
- Redefining entry and exit points using Transmission System Operator (TSO) data
- Adding height data to nodes via a geographical information system (GIS) software
- Associating known node pressures from TSO data for pipeline capacity computation
- Adding major pipelines built after the available LKD-EU data set was published

Data generation:

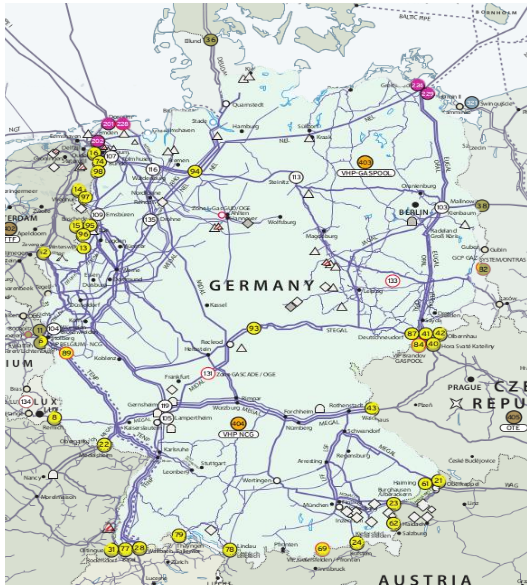
- Estimating compressor and driver data based on partially available public data and network topology using equations of appropriate mathematical models

Germany Gas Transport Network Data Set

European Gas Transport Network: Interconnections



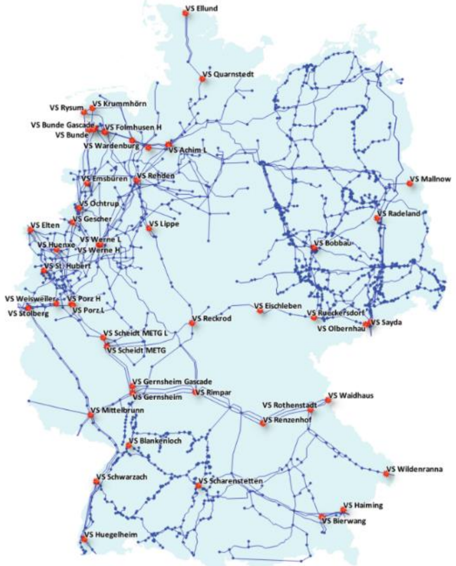
[1] ENTSOG. Transmission Capacity Map. Retrieved from https://www.entsog.eu/maps#transmission-capacity-map-2021. Accessed on 31.10.2022



[1] ENTSOG. Transmission Capacity Map. Retrieved from https://www.entsog.eu/maps#transmission-capacity-map-2021. Accessed on 31.10.2022

Germany Gas Transport Network: Topology & Complex facilities

- ≈70 entry & ≈900 exit points
- ≈1650 inner nodes
- ≈1770 pipes
- ≈95 control valves
- 58 compressor stations
  - 129 compressors & drivers
  - 200 valves



[11] Yueksel-Erguen et al. Modeling the transition of the multimodal pan-European energy system including an integrated analysis of electricity and gas transport. Energy Systems, 2023.
[10] Kunz et al. Reference Data Set: Electricity, Heat, and Gas Sector Data for Modeling the German System (Version 1.0.0), 2017. https://doi.org/10.5281/zenodo.1044463.

Gas Data Files: Technical properties

Network topology data: .net file (>34K lines; >28K data attributes)

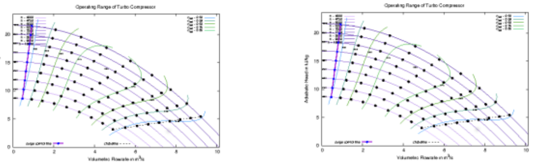
```
<source id="N1" x="-100" y="500">
  <height unit="m" value="200.0"/>
  <pressureMin unit="bar" value="30.0"/>
  <pressureMax unit="bar" value="70.0"/>
  <flowMin unit="1000m_cube_per_hour" value="50.0"/>
  <flowMax unit="1000m_cube_per_hour" value="750.0"/
  <gasTemperature unit="Celsius" value="10.0"/>
  <calorificValue  value="37" unit="MJ_per_m_cube"/>
  <normDensity value="0.785" unit="kg_per_m_cube"/>
  <coefficient-A-heatCapacity  value="32.23"/>
  <coefficient-B-heatCapacity  value="-0.01"/>
  <coefficient-C-heatCapacity  value="0"/>
  <molarMass  value="19.5" unit="kg_per_kmol"/>
  <pseudocriticalPressure  value="44.5" unit="bar"/>
  <pseudocriticalTemperature  value="190" unit="K"/>
</source>
```

```
<pipe from="N1" id="P12" to="N2">
  <flowMin unit="1000m_cube_per_hour" value="0.0"/>
  <flowMax unit="1000m_cube_per_hour" value="73500"/>
  <length unit="km" value="50"/>
  <diameter unit="m" value="1.1"/>
  <roughness unit="mm" value="0.01"/>
  <heatTransferCoefficient unit="W_per_m_square_per_K" value="2.0"/>
</pipe>
```

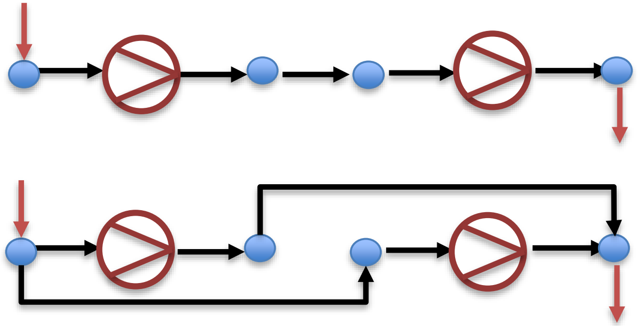
Compressor station data: .cs file (>54K lines)

```
<compressorStation id="CS3">
  <compressors>
    <turboCompressor drive="P_CS3_M1" id="T_CS3_M1">
    <turboCompressor drive="P_CS3_M3" id="T_CS3_M3">
  </compressors>
  <drives>
    <gasTurbine id="P_CS3_M1">
    <gasTurbine id="P_CS3_M3">
  </drives>
  <configurations>

    <configuration confId="1" nrOfSerialStages="2">
      <stage stageNr="1" nrOfParallelUnits="1">
      <stage stageNr="2" nrOfParallelUnits="1">
    </configuration>

    <configuration confId="3" nrOfSerialStages="1">
      <stage stageNr="1" nrOfParallelUnits="1">
    </configuration>

    <configuration confId="2" nrOfSerialStages="1">
      <stage stageNr="1" nrOfParallelUnits="2">
    </configuration>

    <configuration confId="4" nrOfSerialStages="1">
      <stage stageNr="1" nrOfParallelUnits="1">
    </configuration>

  </configurations>
</compressorStation>
```

Compressor station characteristic diagrams



Alternative compressor machine configurations



Decision-Making with Open Data -2

Data Sets	Topology	Flow
ENTSOG IP [4]	High-Level (cumulative cap.)	Aggregated@IP
ENTSOG SOS [5]		Aggregated@BZ
Germany LKD-EU [8]	Pipeline Network (approximate cap.)	Subregion-Based
Germany gas network dataset [10, 11]: **Germany LKD-EU [8]** + *TSO data* *Compressor station data*	Pipeline Network + Active components (computed cap. + gas flow directions)	Node-Based

Complementary Data
ENTSOG IP/node matching
GIE Storage data [9]

[12] AGSI+ Transparency Platform, https://agsi.gie.eu/ Last accessed 06.07.2023

[11] Yueksel-Erguen et al. Modeling the transition of the multimodal pan-European energy system including an integrated analysis of electricity and gas transport. Technical Report 22-17. ZIB, Takustr. 7, 14195 Berlin, 2022.

Operational:
A routing of gas in the network, resulting gas pressure at the nodes, and operational state of the complex facilities of the network

Gas Network – Scenario Generator

*We require scenarios for a selected smaller region in Europe that are***:**

- ***reproducible*** using open data and open models,
- ***consistent*** with the limitations of the *pan-European network infrastructure* given *the pan-European gas supply*,
- ***robust*** against data uncertainties,
- ***effectively span*** the feasible scenario space, and
- able to ***reflect the future uncertainties*** from what-if scenarios.

We developed
a mathematical modeling-based scenario generator
to generate meaningful realistic scenarios.

[13] Inci Yueksel-Erguen, Thorsten Koch, Janina Zittel. Consistent flow scenario generation based on open data for operational analysis of European gas transport networks. Accepted for publication by OR 2023 Proceedings.
[14] 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. Technical Report 24-03. Zuse Institute Berlin, Takustr. 7, 14195 Berlin, 2024.

Scenario Generator

Scenario Generation Tool

- Find the gas in-/out-flow of a region:
  - use ENTSOG IP and SOS data to model the high-level pan-European gas network
- Disaggregate the gas in-/out-flow of the specified region to its entry-exit nodes by ensuring:
  - correct system/node associations
  - meeting sub-region demand distribution of the specified region
  - feasibility given approximate maximum pipeline gas capacity

M1: High-level Pan-European Gas Network Model

Cumulative gas in-/out- flow of the selected region based on IPs

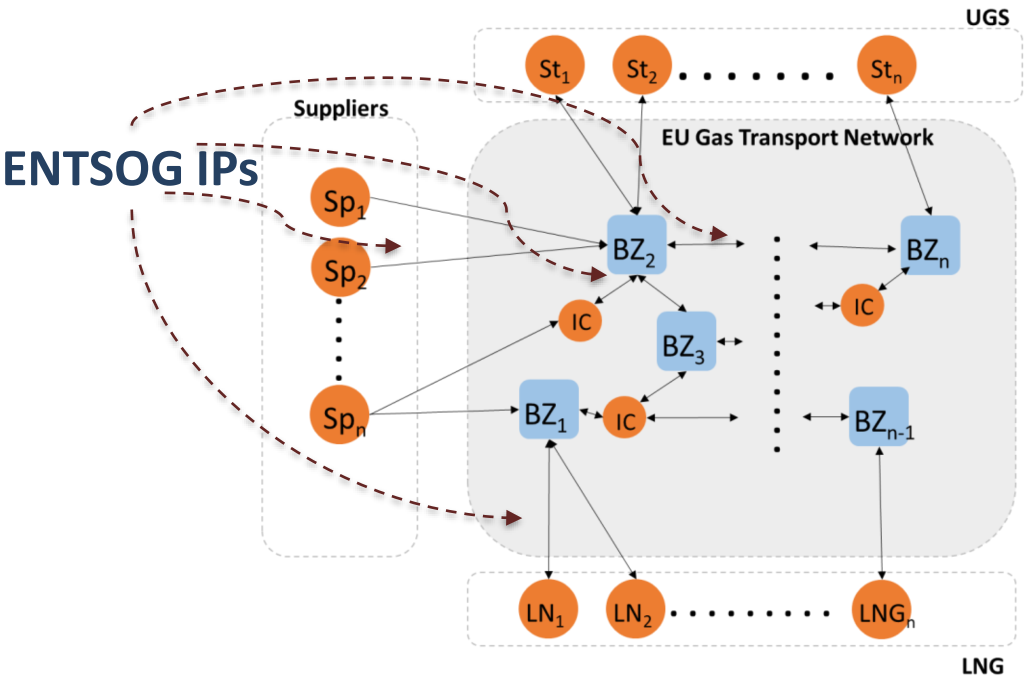
M2: Regional Gas Transport Network Model (Linear)

Gas in-/out- flow of the selected region based on physical entry/exit nodes

Regional Gas Transport Network Model (GNO-Nonlinear)

M1 High-level Pan-European Gas Network Model

main challenge: mapping the open data to a network topology which is consistent with the data provider's data semantics



Input:

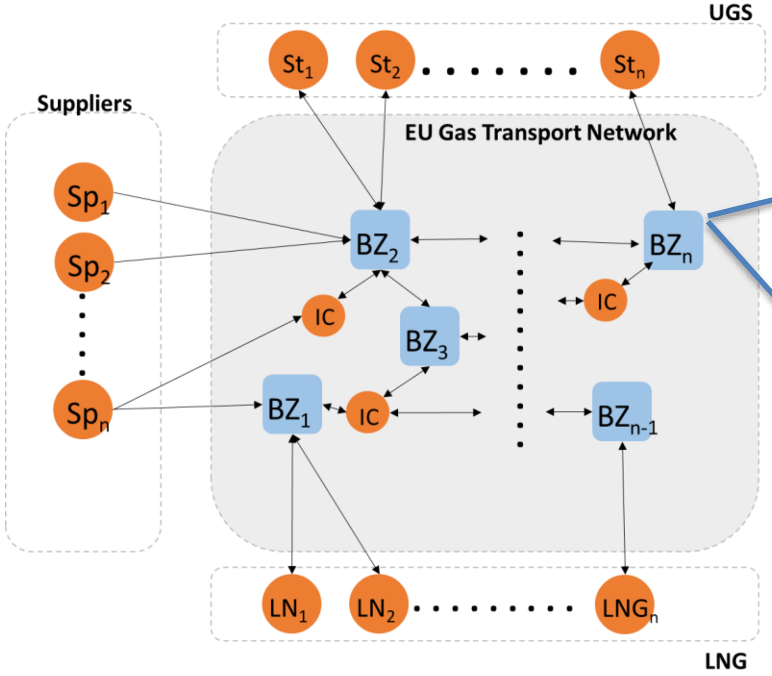
- The high-level network topology data from ENTSOG IP
- Gas supply to Europe *via imports, LNG, underground gas storage facilities, and production*
- Gas demand *of countries or balancing zones, gas injected to the storages, exports, LNG*
- Storage flow models *incl. working capacity, injection and withdrawal capacities based on gas volume in the storages*

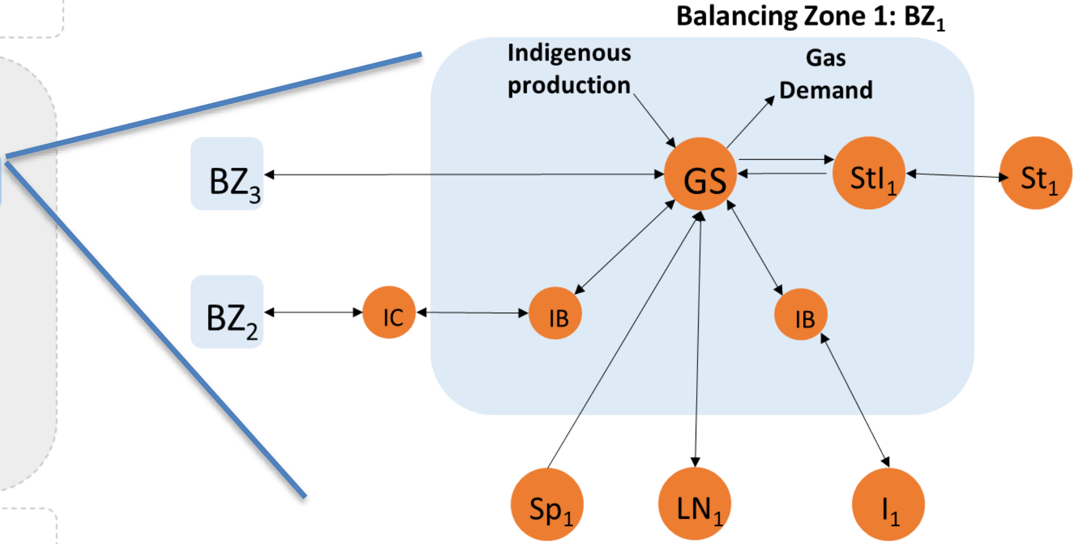
Output:

- Minimum possible demand curtailment
- Gas distribution to the high-level gas network, *i.e., utilization of the pipeline capacities and storage facilities, 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

M1 High-level Pan-European Gas Network Model

main challenge: mapping the open data to a network topology which is consistent with the data provider's data semantics





Input:

- The high-level network topology data from ENTSOG IP
- Gas supply to Europe *via imports, LNG, underground gas storage facilities, and production*
- Gas demand *of countries or balancing zones, gas injected to the storages, exports, LNG*
- Storage flow models *incl. working capacity, injection and withdrawal capacities based on gas volume in the storages*

Output:

- Minimum possible demand curtailment
- Gas distribution to the high-level gas network, *i.e., utilization of the pipeline capacities and storage facilities, 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

M1 – 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 $i \in V$

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

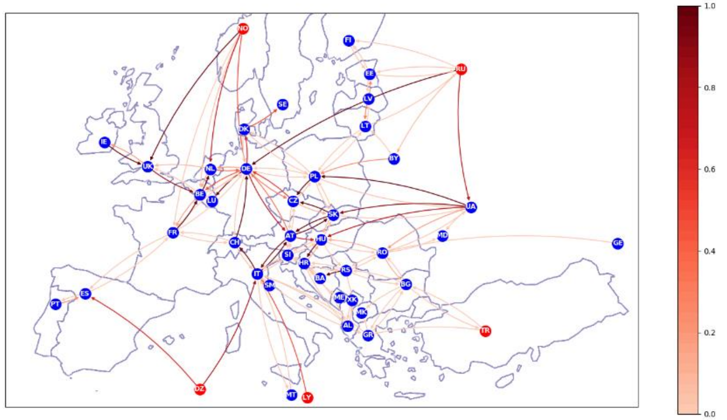
subject to

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

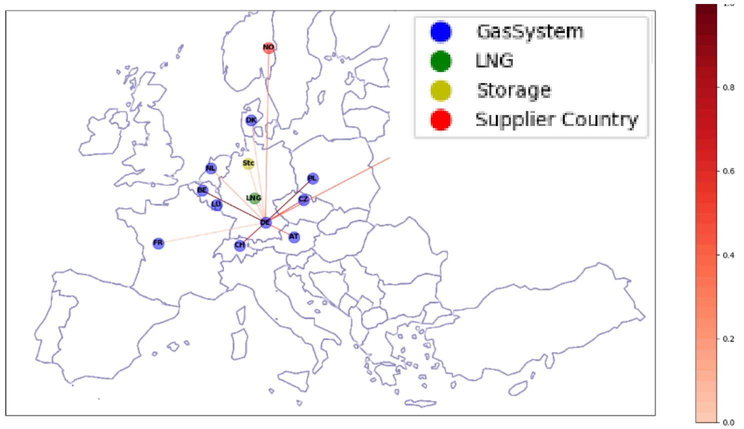
$$\sum_{i \in V_B^b} \left(\sum_{\substack{j \notin V_B^b \\ (j,i) \in A}} x_{ji} - \sum_{\substack{j \notin V_B^b \\ (i,j) \in A}} x_{ij} + y_i - u_i \right) = \sum_{i \in V_B^b} (d_i - s_i), \forall b \in B$$

$$y_i \geq 0, \forall i \in V_{GS} \cap V_{IB};\ y_i = 0, \forall i \in V \setminus V_{GS} \cap V_{IB};\ u_i = 0, \forall i \in V \setminus V_{St}$$

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



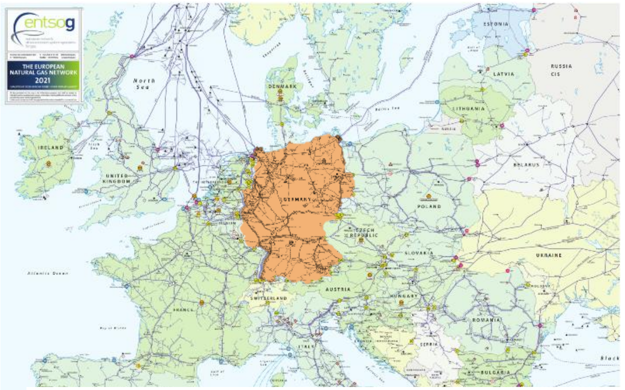
Capacity utilization of connection between the EU and the non-EU countries



Capacity utilization of connections that carries gas to Germany

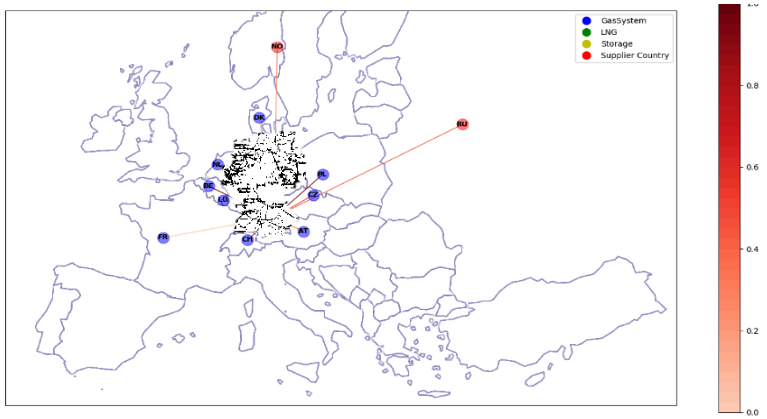
M2: Regional Gas Transport Network Model (Linear)

Dispatch to ≈70 entry & ≈900 exit points

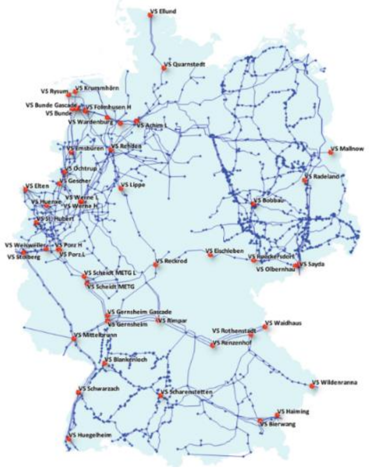


[1] ENTSOG. Transmission Capacity Map. Retrieved from https://www.entsog.eu/maps#transmission-capacity-map-2021. Accessed on 31.10.2022

M1



M2



[11] Yueksel-Erguen et al. Modeling the transition of the multimodal pan-European energy system including an integrated analysis of electricity and gas transport. Technical Report 22-17. ZIB, Takustr. 7, 14195 Berlin, 2022.
[10] Kunz et al. Reference Data Set: Electricity, Heat, and Gas Sector Data for Modeling the German System (V1.0.0), 2017. https://doi.org/10.5281/zenodo.1044463.

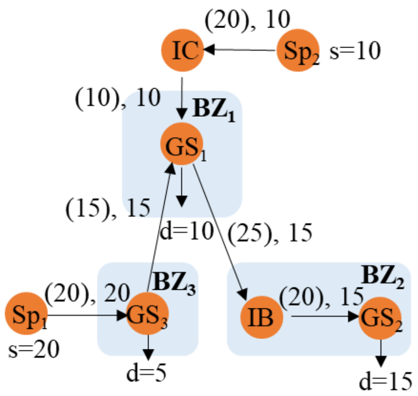
Input:

- Amount of gas demand and supply of the region as a result of M1
- Network topology data of the region including physical entry, exit and inner nodes, and pipelines with predefined capacity upper bounds
- Physical entry-exit node association to the high-level ENTSOG data set given in M1
- Gas demand distribution to NUTS-3/postal code regions

Output

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

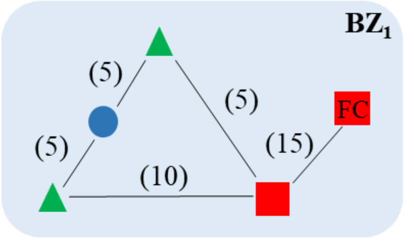
C2F Network Transformations (NT)



A feasible solution on European IP network: <(capacity), flow>

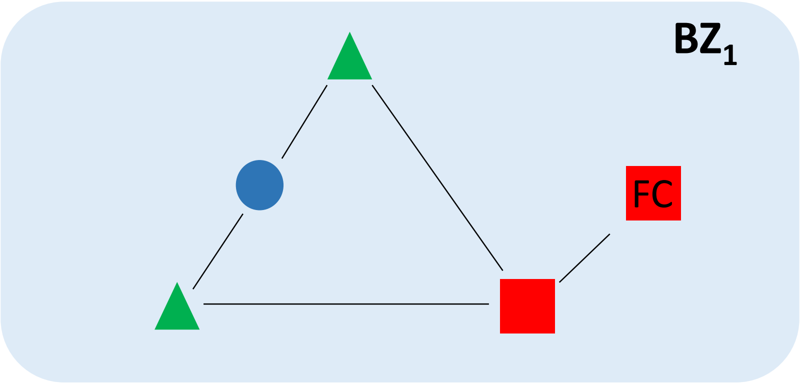
$$G_E = (V, A)$$

- Entry node
- Transshipment node
- Exit node
- FC Final consumer exit node

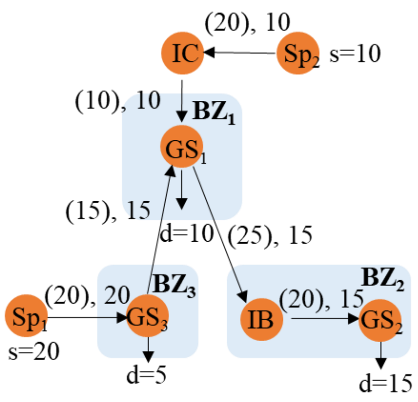


Graph representing BZ_1's physical network

$$G_P = (V_P, E_P)$$

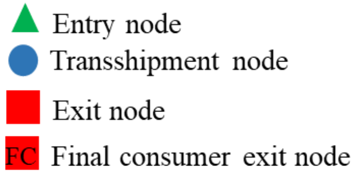


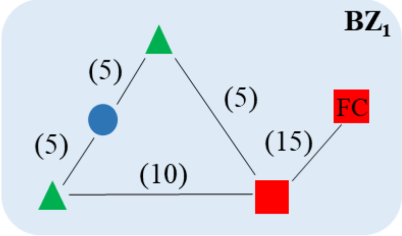
C2F Network Transformations (NT)



A feasible solution on European IP network: <*(capacity), flow*>

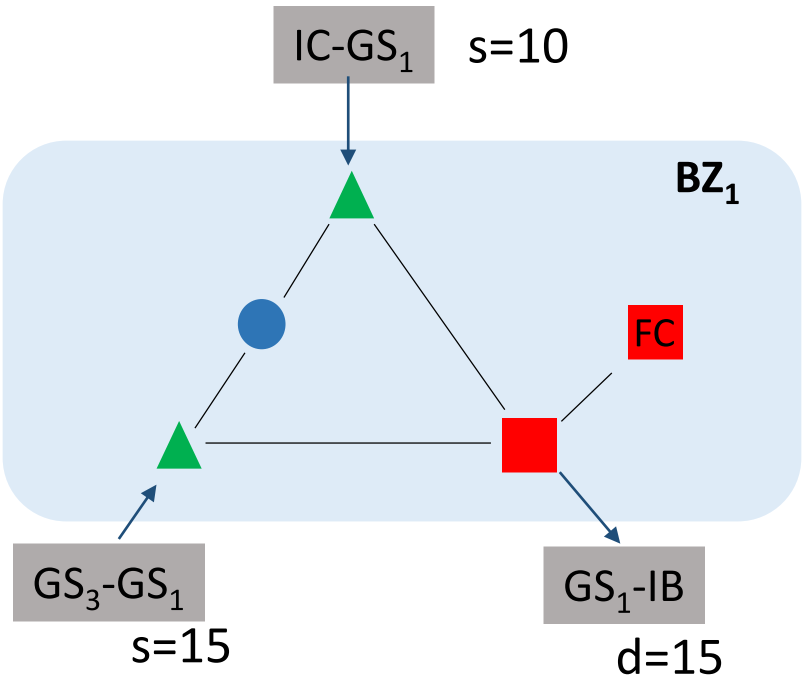
$$G_E = (V, A)$$





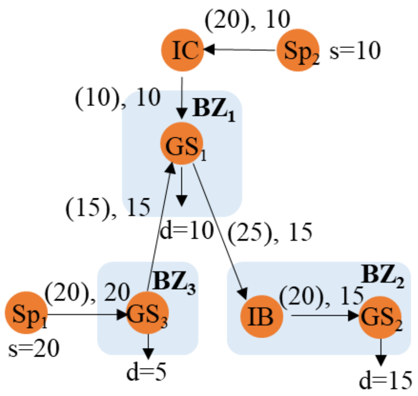
Graph representing BZ_1's physical network

$$G_P = (V_P, E_P)$$



NT1: Matching the boundary nodes

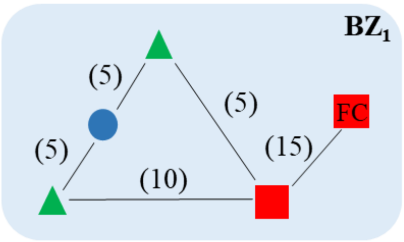
C2F Network Transformations (NT)



A feasible solution on European IP network: <(capacity), flow>

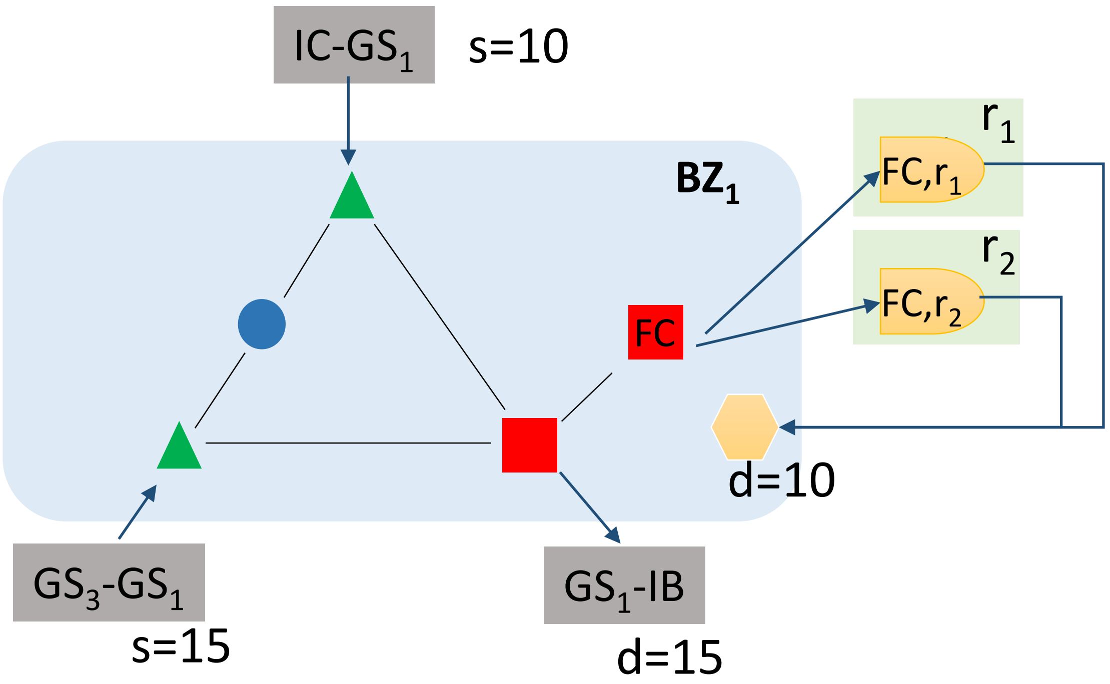
$$G_E = (V, A)$$

- Entry node
- Transshipment node
- Exit node
- FC Final consumer exit node



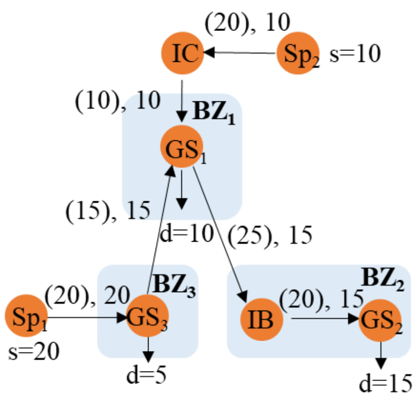
Graph representing BZ_1's physical network

$$G_P = (V_P, E_P)$$



NT1: Matching the boundary nodes
NT2: Geographical dispatch

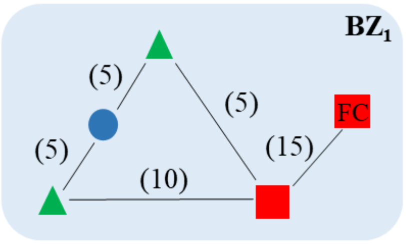
C2F Network Transformations (NT)



A feasible solution on European IP network: *<(capacity), flow>*

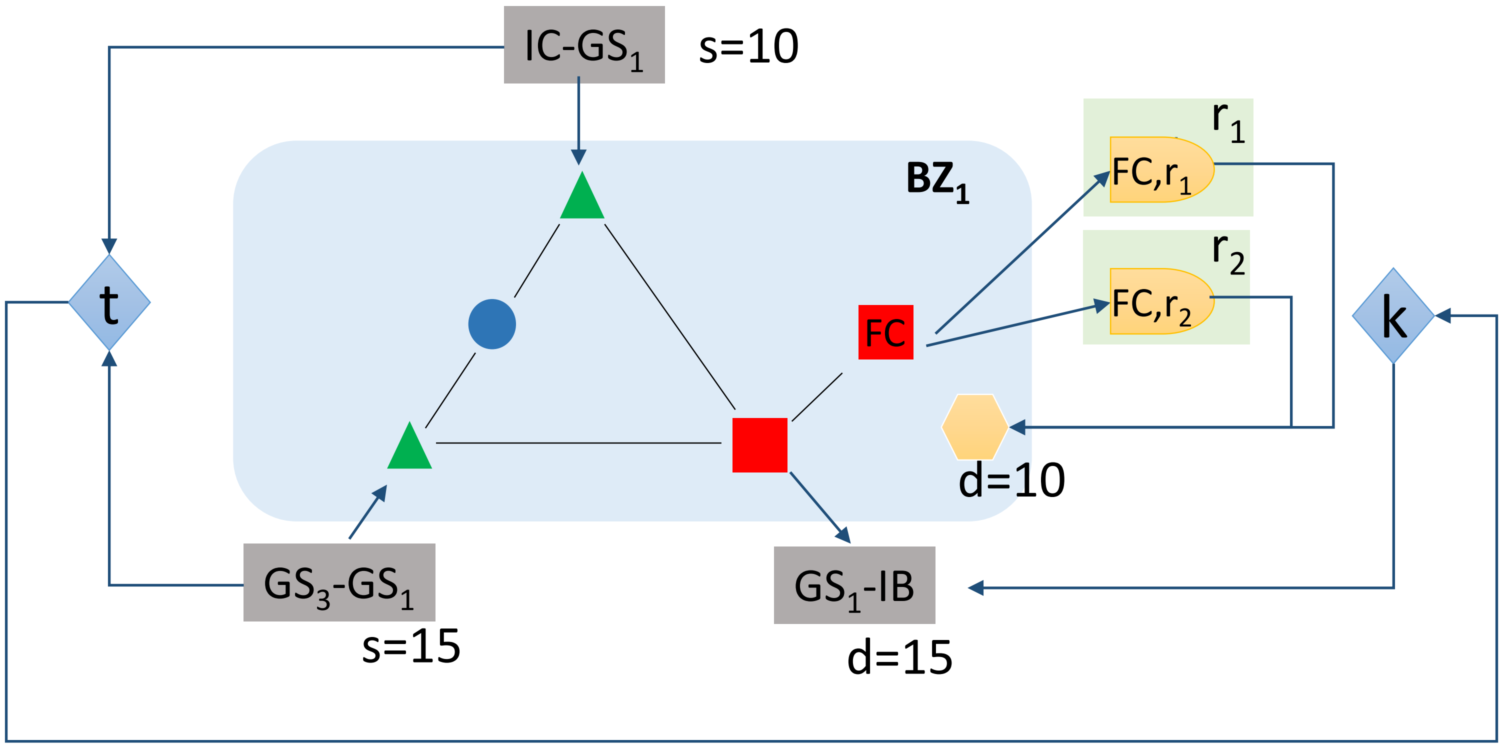
$$G_E = (V, A)$$

- Entry node
- Transshipment node
- Exit node
- FC Final consumer exit node



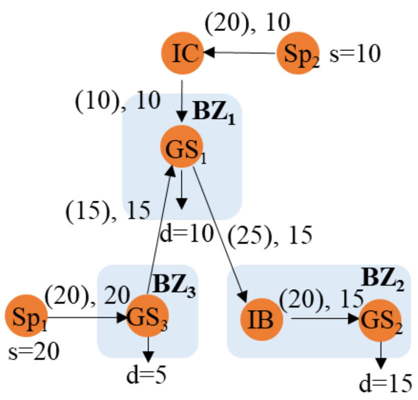
Graph representing BZ_1's physical network

$$G_P = (V_P, E_P)$$



NT1: Matching the boundary nodes
NT2: Geographical dispatch
NT3: Exploring artificial feasible solutions

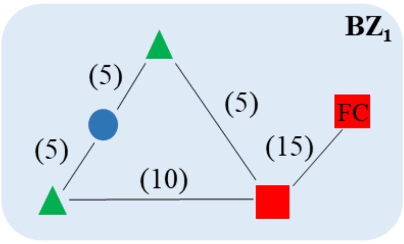
C2F Network Transformations (NT)



A feasible solution on European IP network: <(capacity), flow>

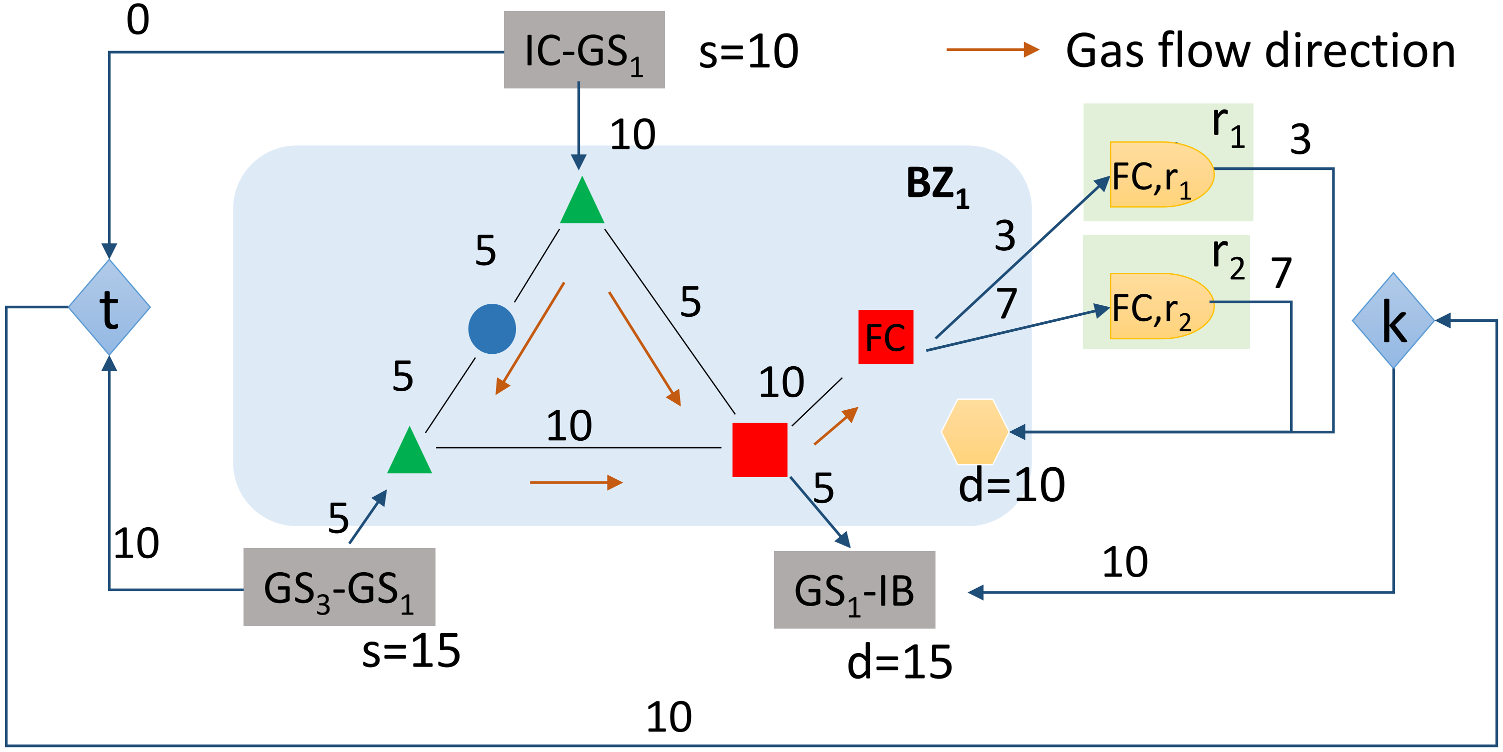
$G_E = (V, A)$

- ▲ Entry node
- ● Transshipment node
- ■ Exit node
- FC Final consumer exit node



Graph representing BZ_1's physical network

$G_P = (V_P, E_P)$



$G_R = (V_R, E_R, A_{Ar})$

NT1: Matching the boundary nodes

NT2: Geographical dispatch

NT3: Exploring artificial feasible solutions

C2F - Results

- NT1 + NT2
  - filters the infeasible scenarios before operational analysis
  - the resulting feasible scenarios from M2 are still feasible in M1
- NT1 + NT2 + NT3
  - the scenarios are feasible for both M1 and M2 if the $x_{tk}=0$
  - otherwise, the reason of infeasibility should be investigated for individual cases by checking the non-negative flows on the arcs artificial arcs added by NT3

M2 – Mathematical Model

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 $r, r \in S$

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

subject to

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

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

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

$$0 \leq f_{ij} \leq c_{ij}, \forall (i,j) \in A_{Ar}$$

$$-c_{ji} \leq f_{ij} \leq c_{ij}, \forall \{i,j\} \in E_P$$

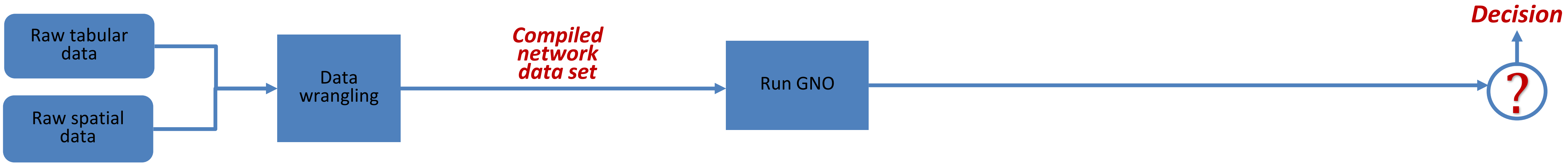
$$0 \leq \pi_r^+, 0 \leq \pi_r^-, \forall r \in S$$


```xml
<?xml version="1.0" encoding="UTF-8"?>
<boundaryValue xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
               xmlns="http://gaslib.zib.de/Gas"
               xsi:schemaLocation="http://gaslib.zib.de/Gas
                                   http://gaslib.zib.de/schema/Scenario.xsd"
               xmlns:framework="http://gaslib.zib.de/Framework">

  <scenario id="ExampleScenario">

    <node type="entry" id="entry01">
      <flow bound="lower" value="160.00" unit="1000m_cube_per_hour"/>
      <flow bound="upper" value="160.00" unit="1000m_cube_per_hour"/>
    </node>

    <node type="entry" id="entry02">
      <flow bound="lower" value="140.00" unit="1000m_cube_per_hour"/>
      <flow bound="upper" value="140.00" unit="1000m_cube_per_hour"/>
    </node>

    <node type="exit" id="exit01">
      <flow bound="lower" value="200.00" unit="1000m_cube_per_hour"/>
      <flow bound="upper" value="200.00" unit="1000m_cube_per_hour"/>
    </node>

  </scenario>

</boundaryValue>
```

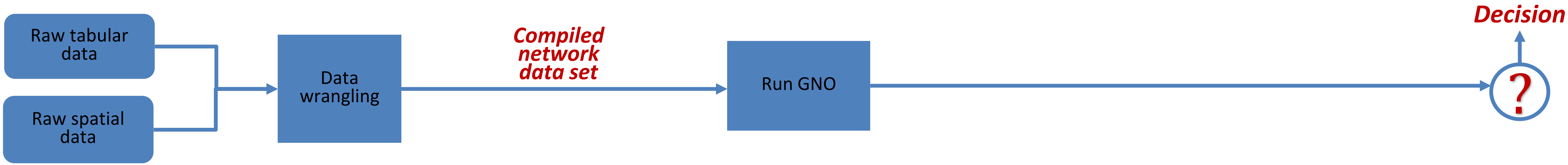
Scenario Generation with LP

- Enhancing objective function by introducing auxiliary variables and parameters
  - Storage levels, sub-region demand flexibility
- Generating artificial feasible solutions by NT3
  - To explore feasible scenarios in the neighborhood of the base scenario
  - Commenting on the infeasibility of an upper-stream scenarios
- Generating alternative solutions using information from NT3 and adding constraints
  - To generate solutions sharing the flow amount between pipelines before reaching their capacities
- Incorporating side constraints
  - To explore effects of potential future regulations

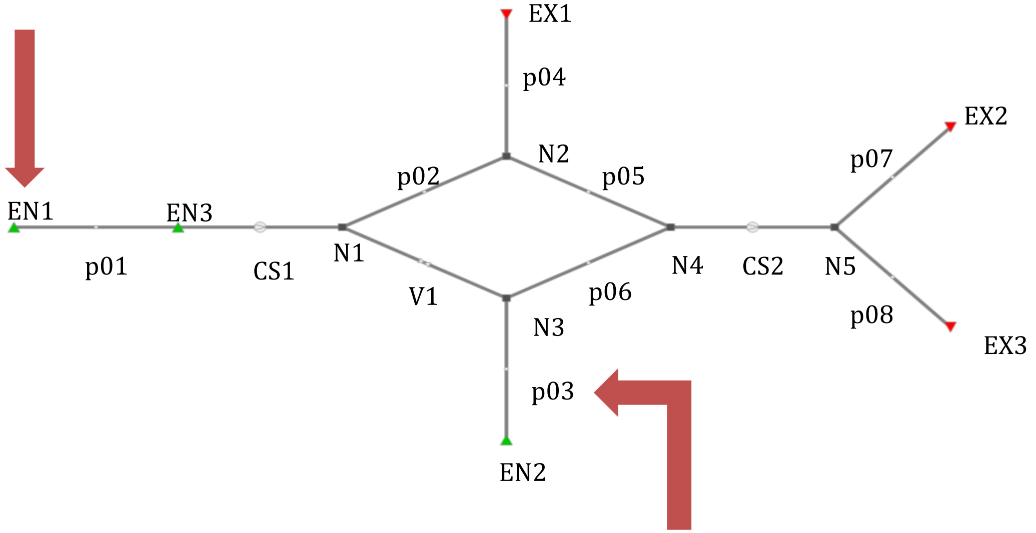
Analysis Results with the Compiled Data Set



Analysis Results with the Compiled Data Set



Inject a "data error" to GasLib11 instance - .net file by $\overline{p_{EN1}}$=70 → 50



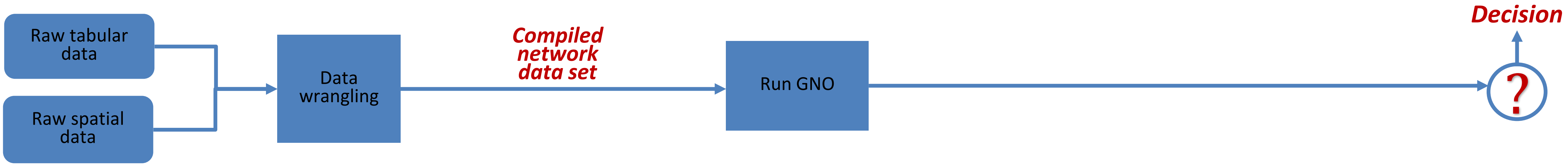
ERROR: INFEASIBILITY: Preprocessing detected that problem is infeasible. Reason:

Lower Bound of flow on arc pipe03_entry02_N03 = 186.1060148 [1000 Nm^3 / h]
Upper Bound of flow on arc pipe03_entry02_N03 = 140 [1000 Nm^3 / h]

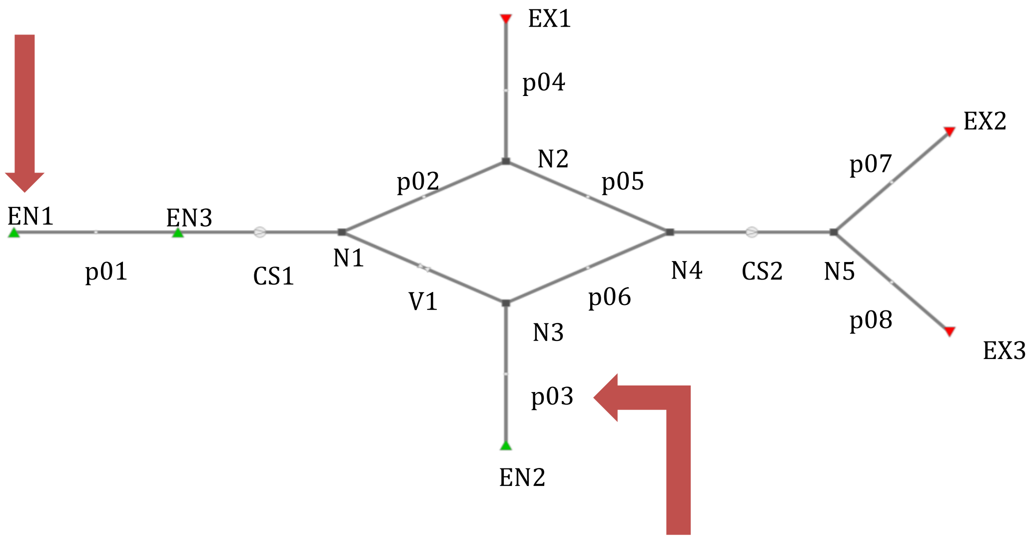
Infeasible what-if scenarios:

- inconsistency between upper/lower bounds

Analysis Results with the Compiled Data Set



Inject a “data error” to GasLib11 instance - .net file by $\overline{p_{EN1}}$=70 → 50

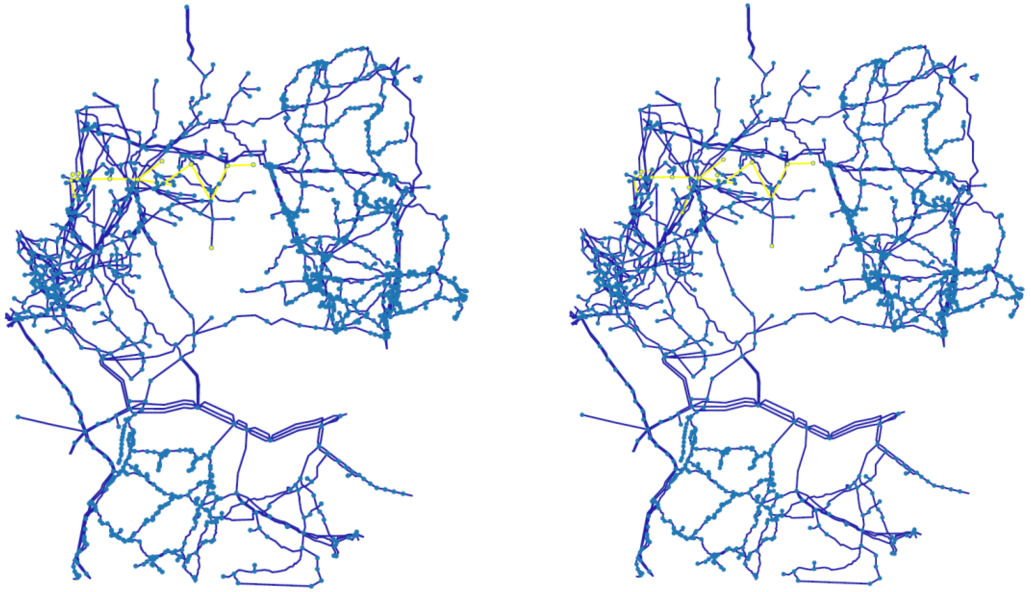


ERROR: INFEASIBILITY: Preprocessing detected that problem is infeasible. Reason:

Lower Bound of flow on arc pipe03_entry02_N03 = 186.1060148 [1000 Nm^3 / h]
Upper Bound of flow on arc pipe03_entry02_N03 = 140 [1000 Nm^3 / h]

Infeasible what-if scenarios:

- inconsistency between upper/lower bounds
- consistent bottlenecks from various scenarios



[11] Yueksel-Erguen et al. Modeling the transition of the multimodal pan-European energy system including an integrated analysis of electricity and gas transport. Energy Systems, 2023.

Dealing with Complex Errors in Infrastructure Datasets

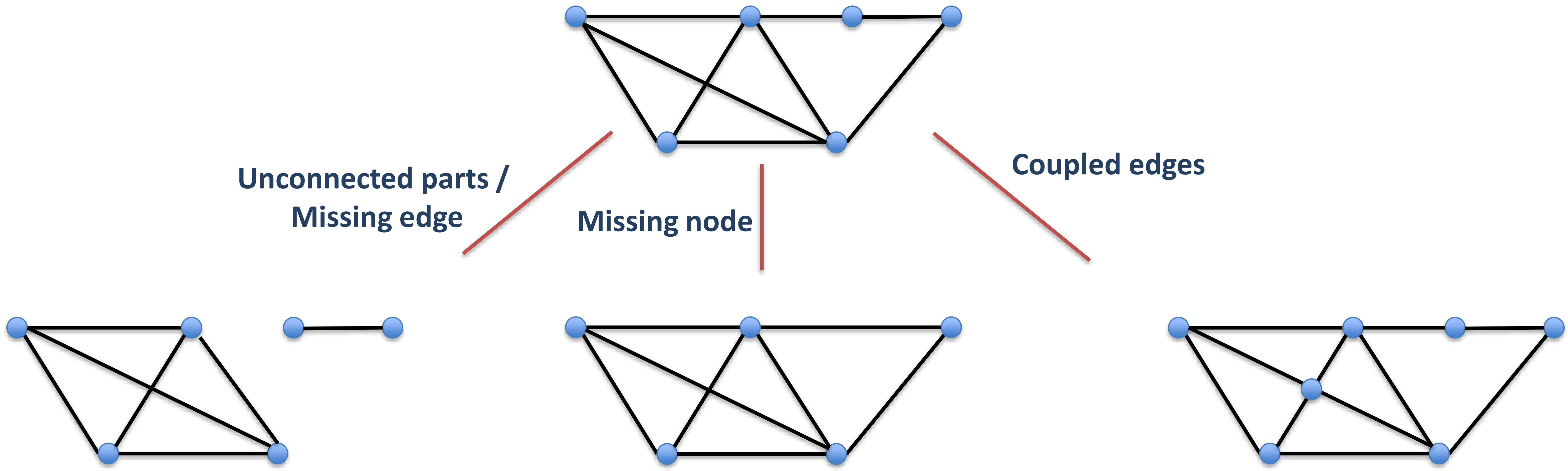
- **Highly-connected** data in industrial applications
  - provided by industrial partners
  - consolidated from public data sources
  - generated using mathematical models

- Data errors **too complex for humans** to understand detected by analysis tools
  - Detection and correction of such errors are extremely difficult

- **Engineered** to fulfill a quality of service (QoS) level
  - Assumptions like iid cannot be made

- A **holistic** quality assessment is required
  - Error in critical components introduce bottlenecks

Network Topology Data Errors



Correction:

- add the missing edge

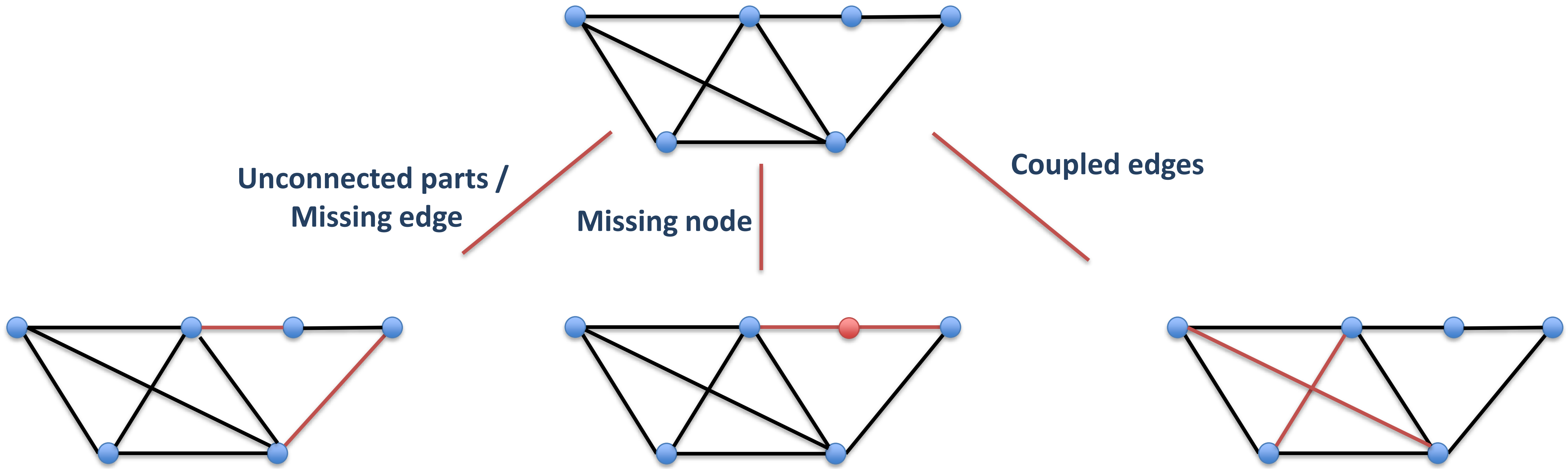
Correction:

- add the missing node to node set
- updating end node of existing edges
- adding two pipes between the end nodes of the existing edges and the added node

Correction:

- deleting the node
- deleting two of the existing edges
- updating end nodes of the two edges

Network Topology Data Errors



Correction:

- add the missing edge

Correction:

- add the missing node to node set
- updating end node of existing edges
- adding two pipes between the end nodes of the existing edges and the added node

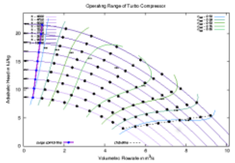
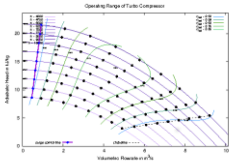
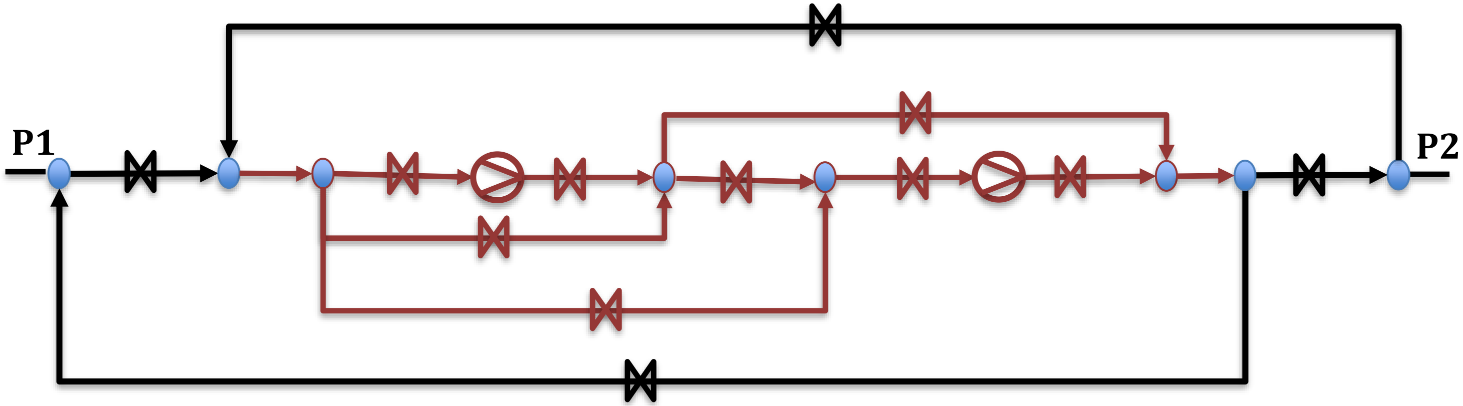
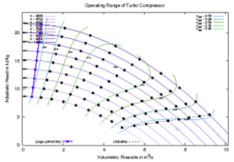
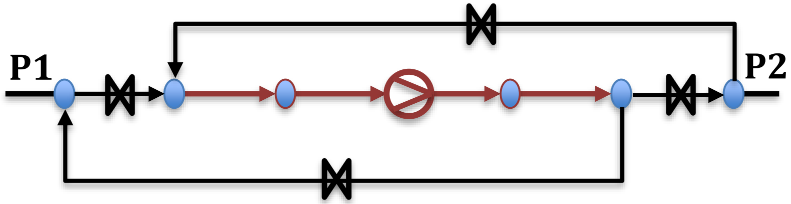
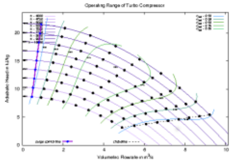
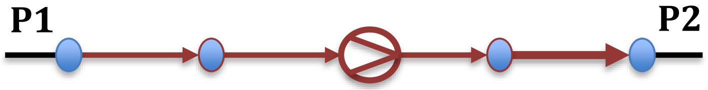
Correction:

- deleting the node
- deleting two of the existing edges
- updating end nodes of the two edges

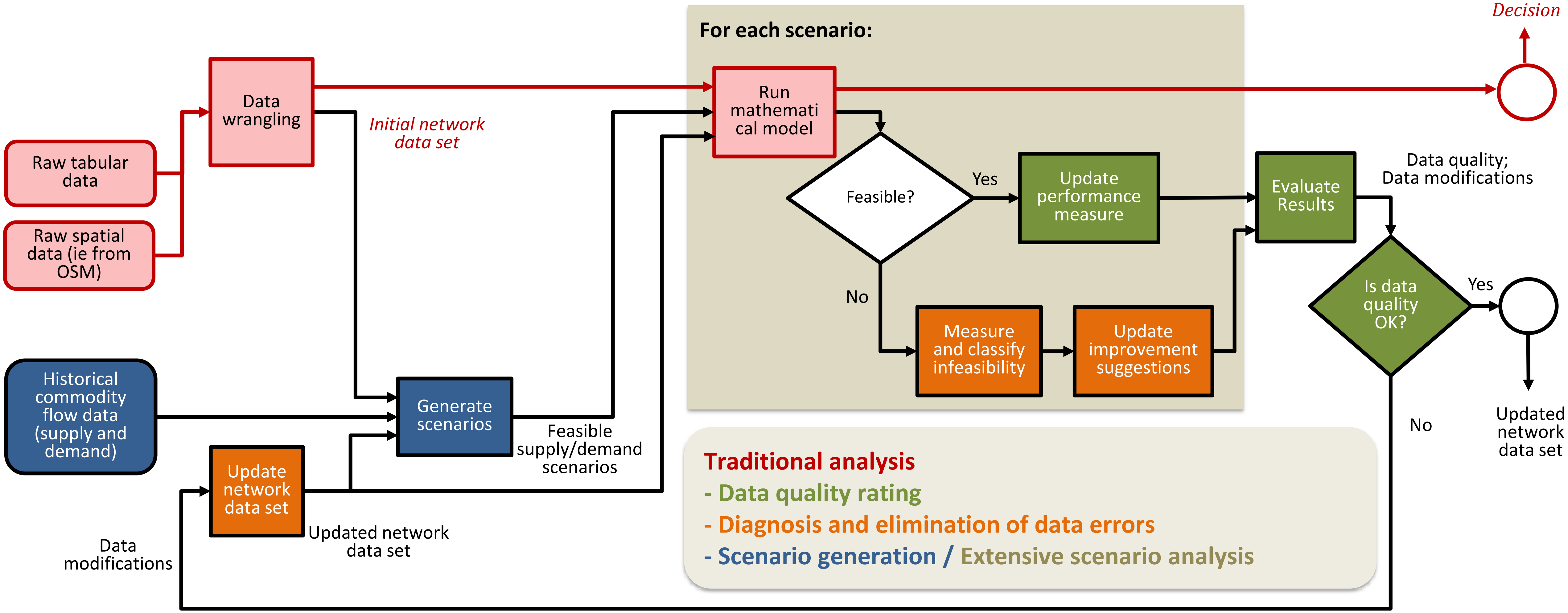
Compressor Station Data Errors

Compressor machine (CM)

Valve

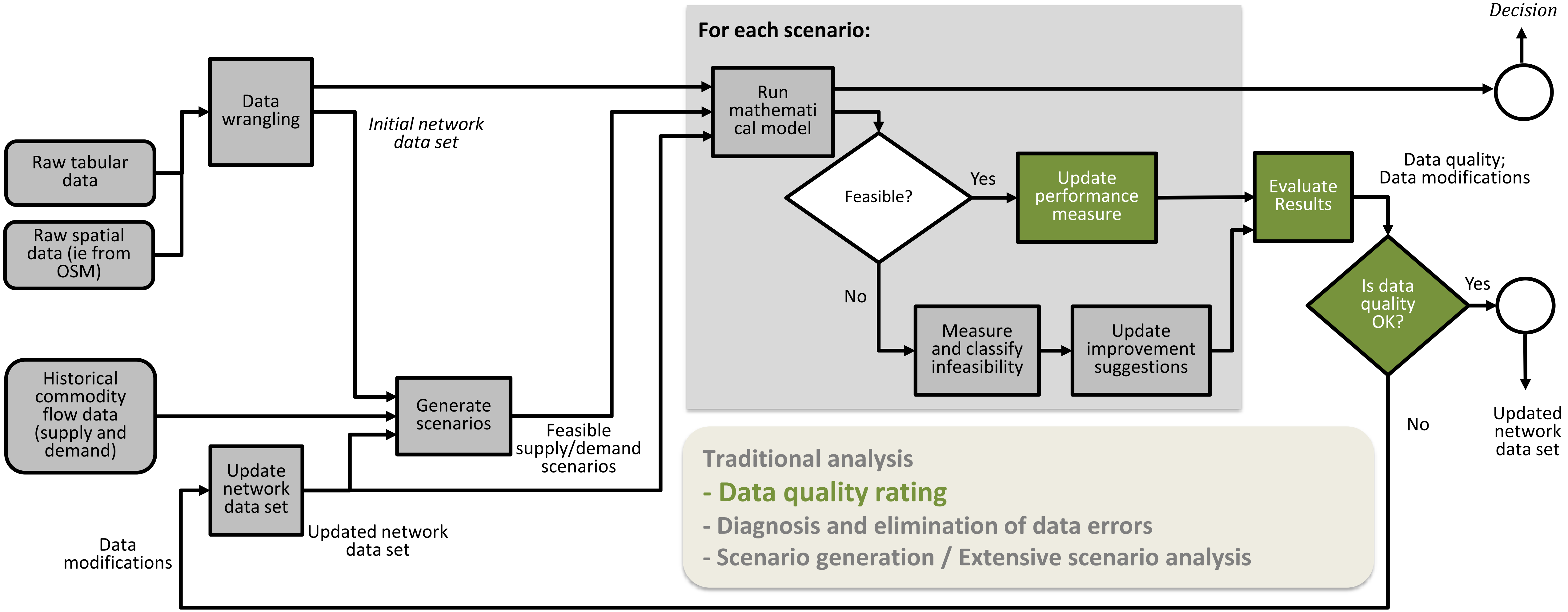
Admissible gas direction	#CM	Characteristic Diagrams	Possible CM Configurations	Resulting Subnetwork for the CS
Bi-directional	2	 	All	
Bi-directional	**1**		**N/A**	
P1 to P2	**1**		**N/A**	

A Data Quality Assessment and Improvement Framework



[15] Inci Yueksel-Erguen, Thorsten Koch, Janina Zittel. A Quality Assessment and Improvement Framework for Supply Infrastructure and Utility Network Datasets. (Under preparation)

Proposed Framework



Traditional analysis

- **Data quality rating**
- Diagnosis and elimination of data errors
- Scenario generation / Extensive scenario analysis

Data Quality Rating - 1

A data rating measure to **facilitate the automated improvement** by enabling us to

- **quantify** our distance to the **quality level for the aimed analysis**
- **compare** the performance of alternative improvements

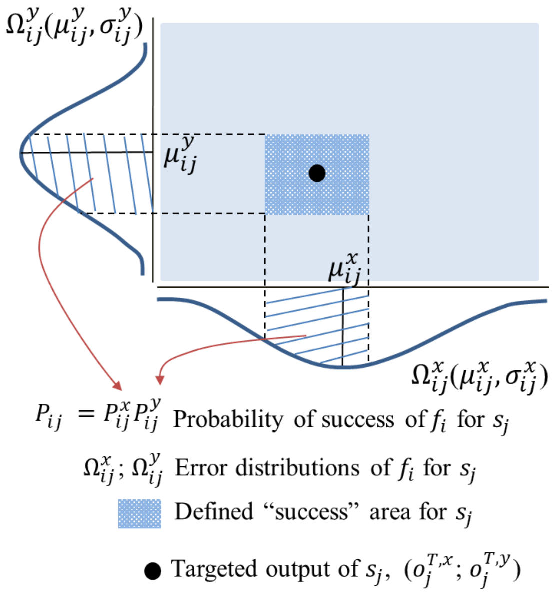
.. and also lead us/the search to **the potential errors and error sources…**

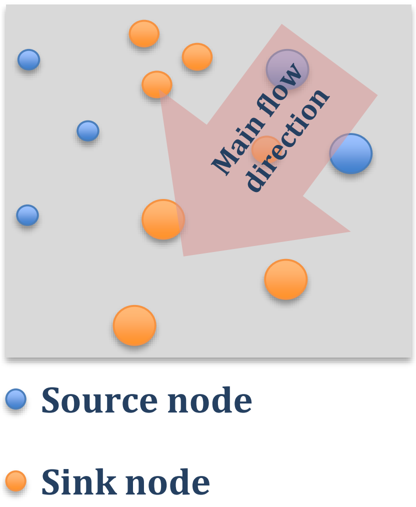
Example: Number of scenarios generated from the historical data that the data set finds a feasible routing

- Not informative enough
- Especially in the very beginning of the improvement process – we may get 0 for all scenario sets

Data Quality Rating - 2

Definition 1: A data set, $\mathrm{DS^i}$, correctly and accurately represents an infrastructure design $\mathrm{f_i}$ if there exists a feasible solution for the mathematical model using $\mathrm{DS^i}$ for all scenarios that can be routed by the infrastructure.





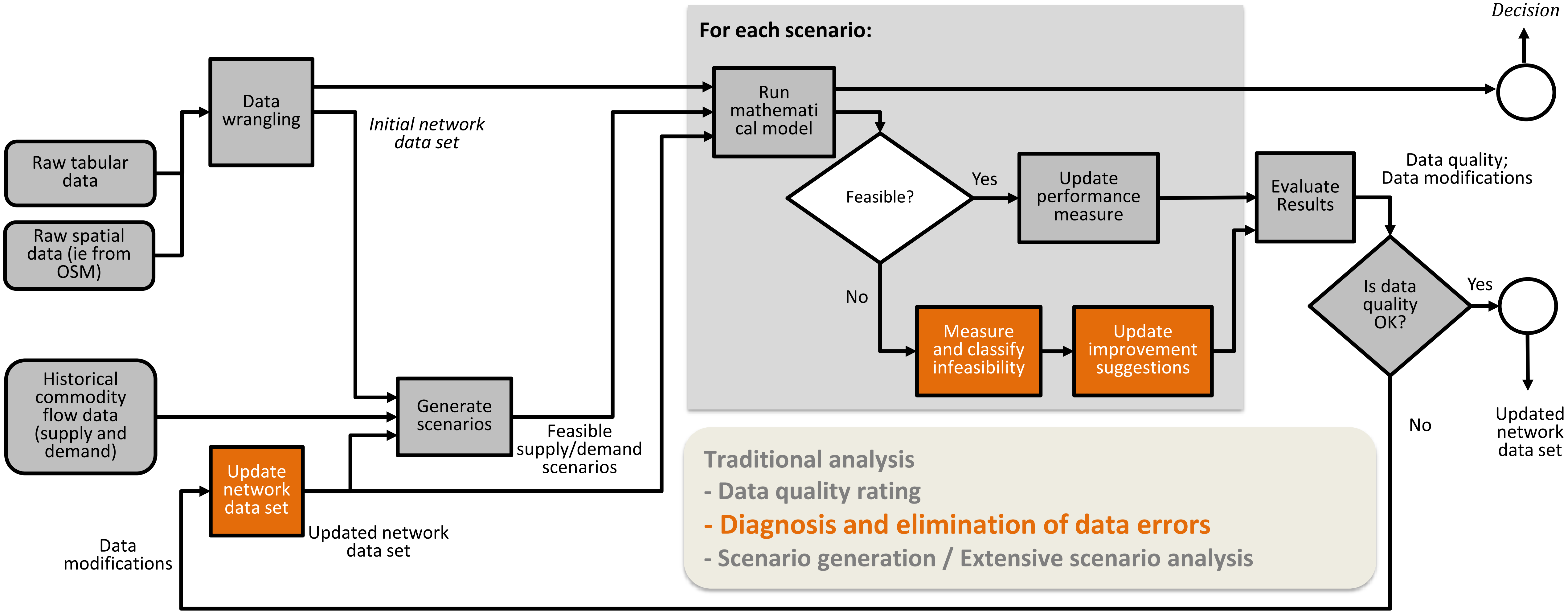
Data Quality Rating

- For S_H, the infrastructure is represented by a data set DS^r correctly and accurately if $\exists[\epsilon^r_R, \epsilon^r_\Phi]$ neighborhood of s_j, $\xi^r_j, \forall s_j \in S_H$ such that $|\ \xi^r_j\ | > 1$; ϵ^r_R and ϵ^r_Φ converges to 0, and the mathematical model finds a feasible solution to $\forall s_k \in \xi^r_j$ using DS^r.
- The corollaries enable exploring the scenario space by ξ^l_j having feasible solutions for all scenarios using an extensive scenario analysis, and compare them with the historical flow scenarios on the basis of $[\epsilon^l_R, \epsilon^l_\Phi]$ to assess the quality of DS^i.

A $[\epsilon^l_R, \epsilon^l_\Phi]$ be a scenario neighborhood of a historical flow scenario s_j, $s_j \in S_H$

$$\xi^l_j = \{s_k : |\ R^l_k - R^T_j\ | \le \epsilon^l_R ; |\ \Phi^l_k - \Phi^T_j\ | \le \epsilon^l_\Phi\}$$

Proposed Framework



Traditional analysis
- Data quality rating
- **Diagnosis and elimination of data errors**
- Scenario generation / Extensive scenario analysis

Mathematical Modeling Methods for Infeasibility Analysis

Slack Formulations

- Different aspects of the formulations can be relaxed with slack variables
- The objective is to minimize the deviation from the original model → zero objective function
- The smallest distance from the feasibility

(Minimum) Irreducible Infeasible Subsystems (IIS)

An IIS of a MIP is a subset of the constraints and bounds that is infeasible and becomes feasible if a single constraint or bound is removed

- Isolates the infeasibility by variables and constraints
- Not an explicit reason why a IIS is infeasible
- Long computation times for large-scale MIPs
- A trivial IIS is not informative

Mathematical Modeling Methods for Infeasibility Analysis

For each infeasible scenario,

Explore a scenario where SMM is also infeasible

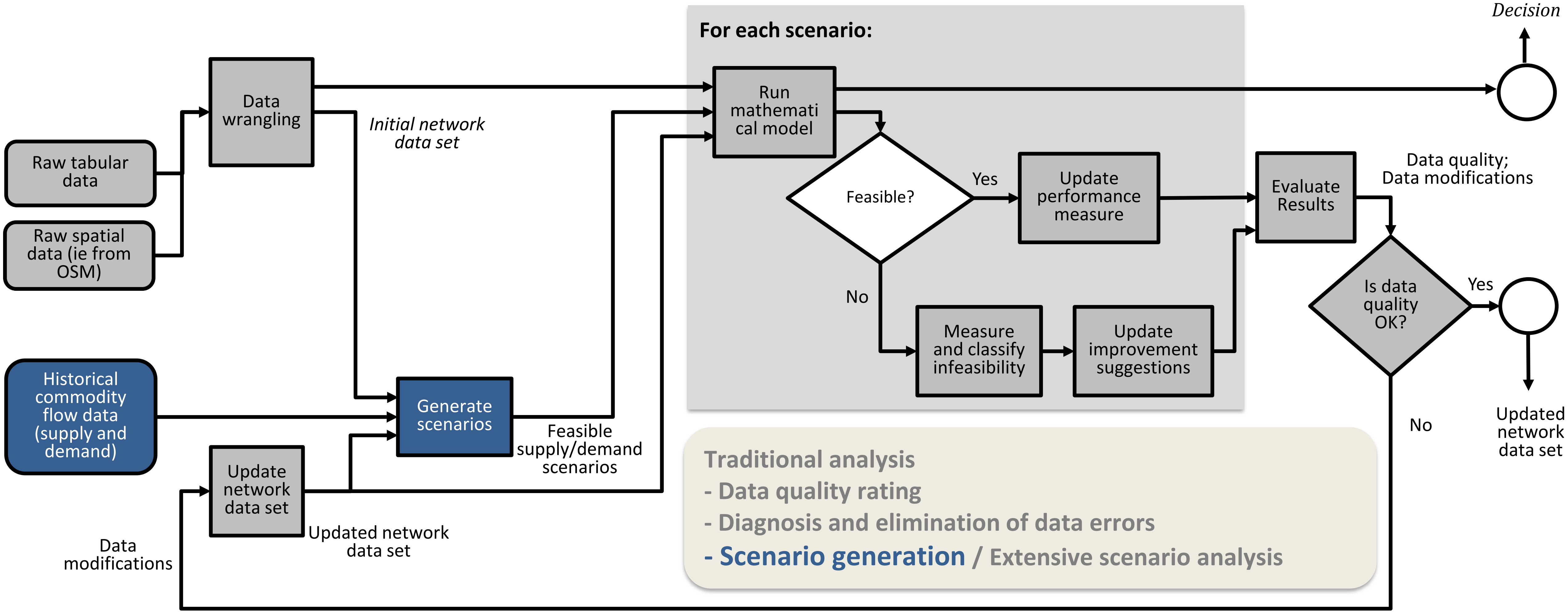
Search for a “non-trivial” minIIS pointing an inconsistency

Procedure 1: Data error detection by infeasibility analysis

Input: D_0: input dataset; S_0: Scenario set; P: Potential error sources; T':Evaluated data errors
Output: T: Set of data error corrections to D given P

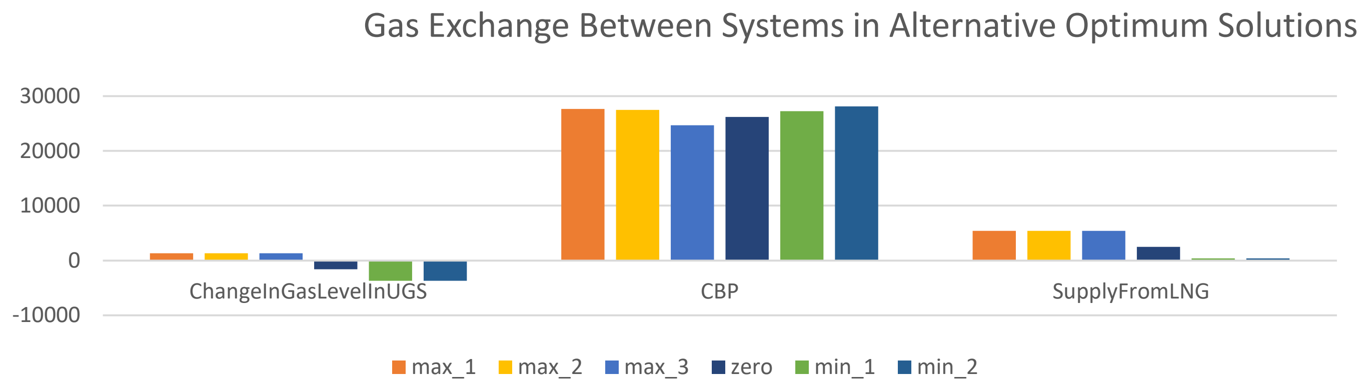
begin
- $D \leftarrow D_0; S \leftarrow S_0; T \leftarrow \emptyset;$
- **while** $|S| > 0$ **do**
  - Select a scenario s_j from S;
  - $S \leftarrow S \setminus s_j$;
  - Relax all non-linear constraints in the slack formulation SMM;
  - Solve MM: $sol_{MM} \leftarrow MM(D, s_j)$;
  - **If** sol_{MM} is *infeasible*
    - Solve SMM: $sol_{SMM}, obj_{SMM} \leftarrow SMM(D, s_j)$;
    - **while** sol_{SMM} is *feasible* **do**
      - **if** $obj_{SMM} > 0$
        - Update s_j with obj_{SMM};
      - **else**
        - Tighten one set of non-linear constraints in SMM
      - Solve SMM: $sol_{SMM}, obj_{SMM} \leftarrow SMM(D, s_j)$;
    - Solve $minIIS$ of LP relaxation of MM: $sol_{IIS} \leftarrow minIIS_{LP}$;
    - **while** sol_{IIS} is *feasible* **do**
      - Save sol_{IIS};
      - Scale s_j: $s_j \leftarrow 0.95 s_j$;
      - Solve $minIIS$ of the LP relaxation of MM: $sol_{IIS} \leftarrow minIIS_{LP}$;
- Analyze the saved $minIIS$ solutions given P;
- Propose potential errors T: $T \neq T'$;
- **return** T

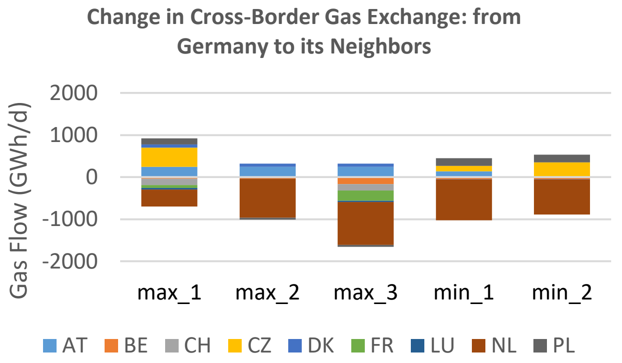
Proposed Framework

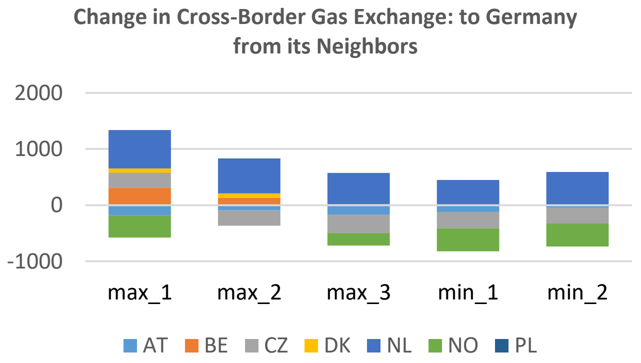


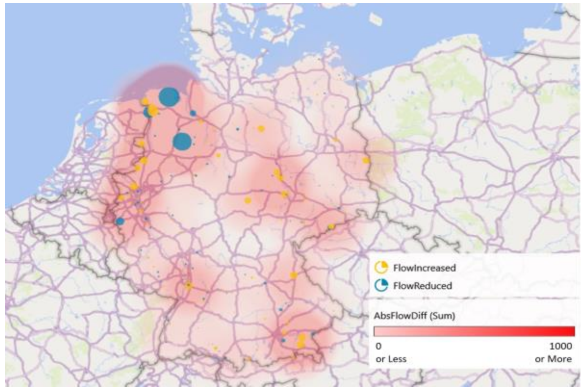
Traditional analysis
- **Data quality rating**
- **Diagnosis and elimination of data errors**
- **Scenario generation** / **Extensive scenario analysis**

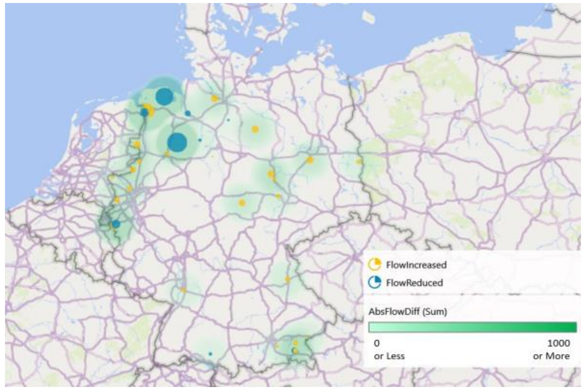
Gas Network Scenario Generator











M1: High-level Pan-European Gas Network Model

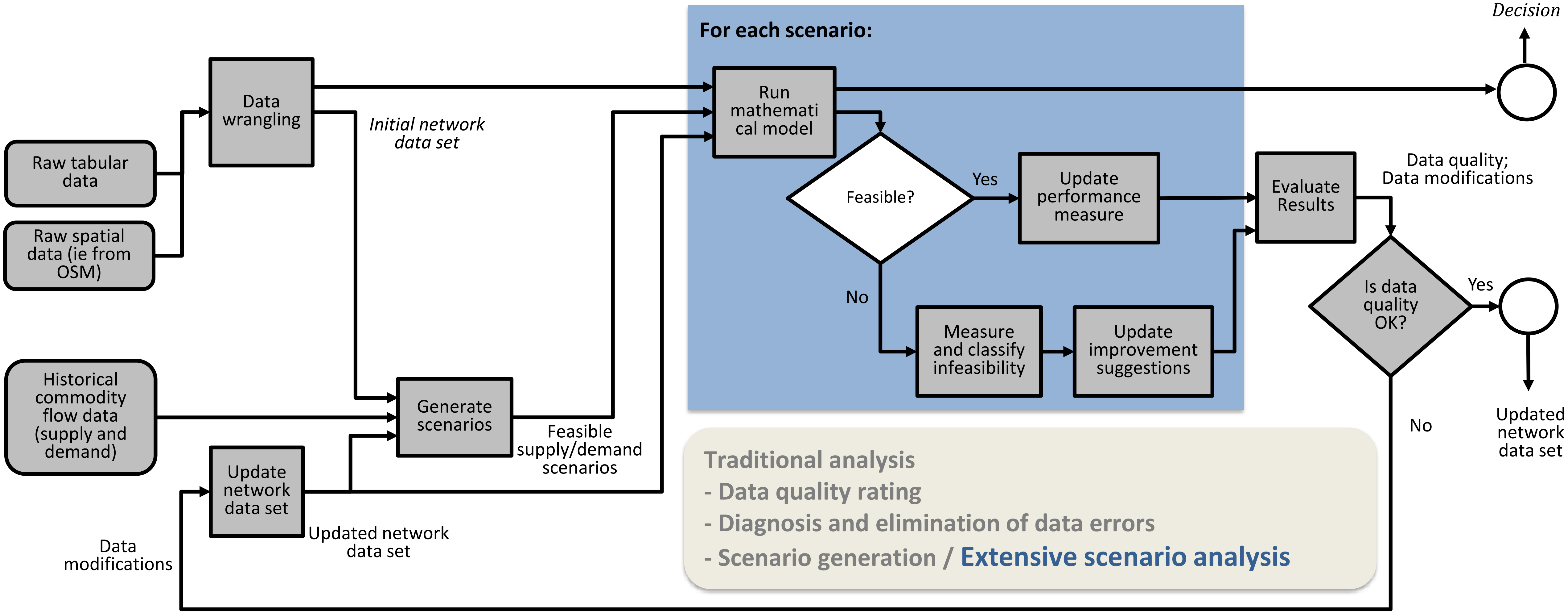
Cumulative gas in-/out- flow of the selected region based on IPs

M2: Regional Gas Transport Network Model (Linear)

Gas in-/out- flow of the selected region based on physical entry/exit nodes

Regional Gas Transport Network Model (GNO-Nonlinear)

Proposed Framework



Traditional analysis
- Data quality rating
- Diagnosis and elimination of data errors
- Scenario generation / **Extensive scenario analysis**

Extensive Scenario Analysis - 1

Set $\epsilon_R = \epsilon_R^0$, where ϵ_R^0 is a sufficiently large initial neighborhood size

while $\epsilon_R > \epsilon_R^T$

Generate $[\epsilon_R , \epsilon_\Phi^l]$ scenario neighborhoods for each $s_j \in S$

Use Procedure 1 to detect and correct errors until there exists a feasible solution for all scenarios in $[\epsilon_R , \epsilon_\Phi^l]$ for each $s_j \in S$

Update $\epsilon_R = \beta\epsilon_R$, where $\beta < 1$ is a non-negative neighborhood size shrinking factor

Gas Transport Network Quality Rating

Representative flow scenarios: ENTSOG Security of supply simulation (2017)

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

- Text mining, fuzzy queries
- Schema validation
- Data augmentation
- Data generation
- Consistency check heuristics

Less than 40% of representative historical flow scenarios at 90% commodity level

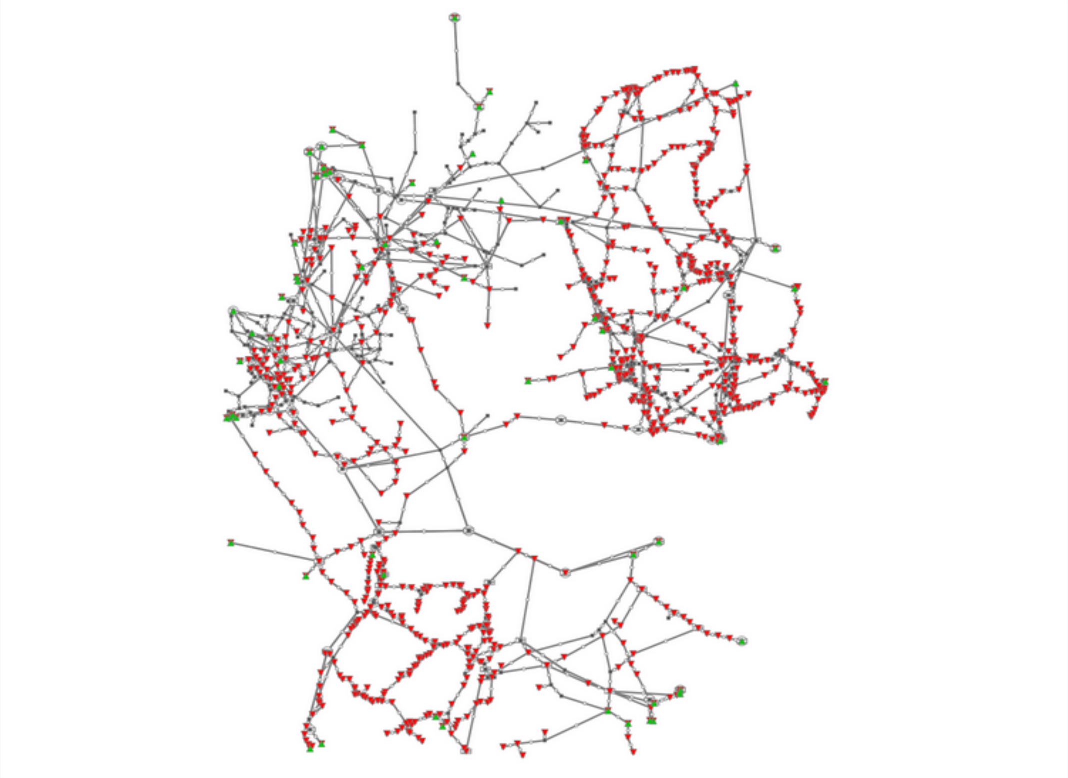
- Extensive scenario analysis

All representative historical flow scenarios at 95% commodity level

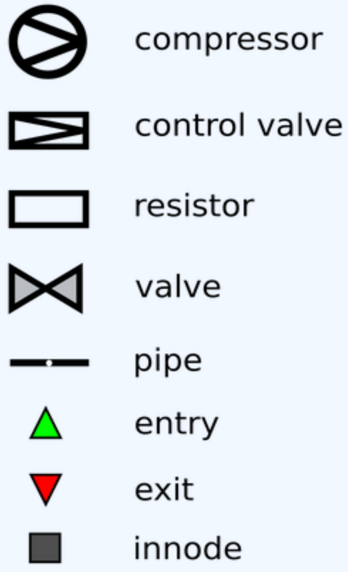
German Gas Transport Network Data Set

GasLib-2607

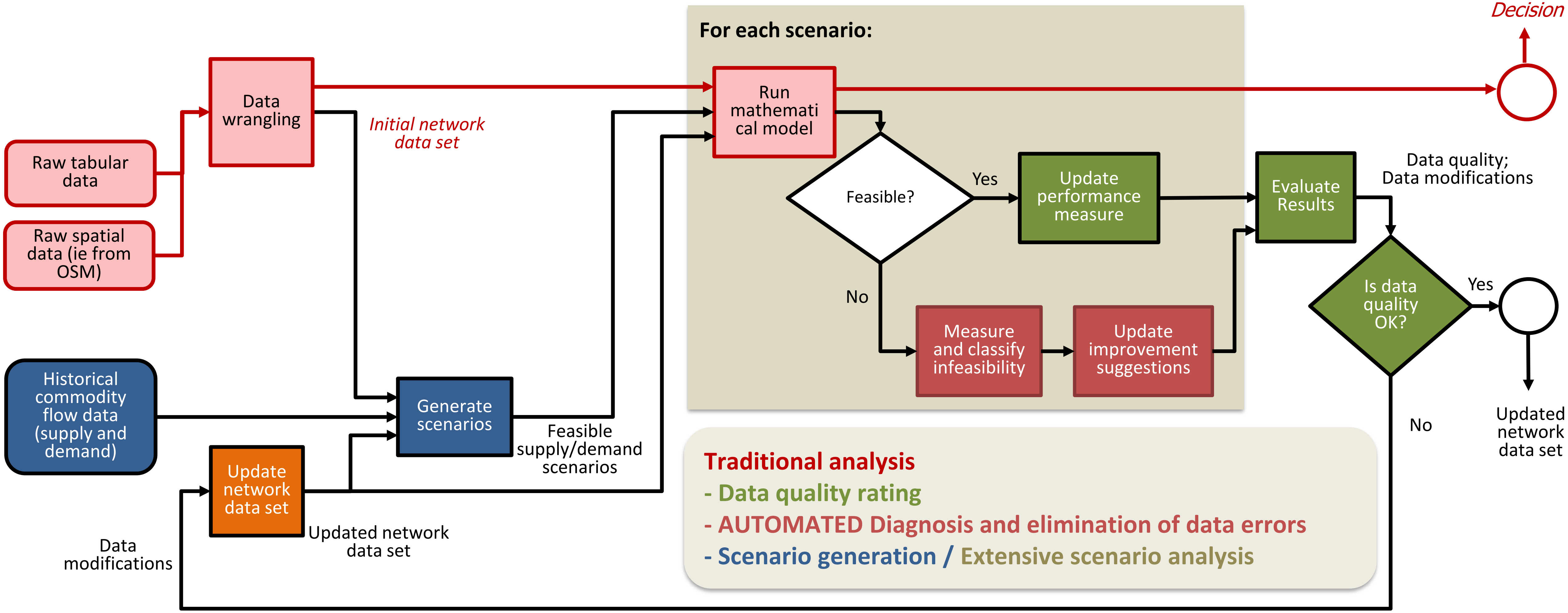
https://gaslib.zib.de/download/data/GasLib-2607-v1-20240708.zip



- network file (.net)
- compressor file (.cs)
- visualization (.svg)
- nominations (.zip archive of .scn)
- complete GasLib-2607 archive (.zip)
- changelog for GasLib-2607 (plaintxt)



Future Outlook



Traditional analysis
- **Data quality rating**
- **AUTOMATED Diagnosis and elimination of data errors**
- **Scenario generation / Extensive scenario analysis**

References

[1] ENTSOG. Transmission Capacity Map. Retrieved from https://www.entsog.eu/maps#transmission-capacity-map-2021. Accessed on 31.10.2022

[2] Thorsten Koch, Benjamin Hiller, Marc E. Pfetsch, and Lars Schewe. 2015. Evaluating Gas Network Capacities. Society for Industrial and Applied Mathematics, USA.

[3] Schmidt, M. et al.: GasLib—A Library of Gas Network Instances. Data, 2(40) (2017)

[4] Nicola Askham, Denise Cook, Martin Doyle, Helen Fereday, Mike Gibson, Ulrich Landbeck, Rob Lee, Chris Maynard, Gary Palmer, Julian Schwarzenbach. The six primary dimensions for data quality assessment. Technical Report, 2013, DAMA United Kingdom.

[5] Inci Yueksel-Erguen, Janina Zittel, Ying Wang, Felix Hennings, Thorsten Koch. Lessons learned from gas network data preprocessing. Technical Report 20-13(v2). Zuse Institute Berlin, Takustr. 7, 14195 Berlin, 2020. Available online https://opus4.kobv.de/opus4-zib/frontdoor/index/index/docId/7826

[6] ENTSOG TP, https://transparency.entsog.eu/, Last accessed 25.08.2023

[7] ENTSOG SOS, https://www.entsog.eu/security-of-supply-simulation#

[8] ENTSOG TYNDP, https://www.entsog.eu/tyndp#, , Last accessed 25.08.2023

[9] SciGRID Gas, https://www.gas.scigrid.de, Last accessed 06.07.2023

[10] Kunz, F. et al.: Reference Data Set (V1.0.0), https://doi.org/10.5281/zenodo.1044463, Last accessed 06.07.2023

[11] Inci Yueksel-Erguen, Dieter Most, Lothar Wyrwoll, Carlo Schmitt, Janina Zittel. Modeling the transition of the multimodal pan-European energy system including an integrated analysis of electricity and gas transport. Energy Systems, 2023.

[12] AGSI+ Transparency Platform, https://agsi.gie.eu/ Last accessed 06.07.2023

[13] Inci Yueksel-Erguen, Thorsten Koch, Janina Zittel. Consistent flow scenario generation based on open data for operational analysis of European gas transport networks. Accepted for publication by OR 2023 Proceedings.

[14] 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. Technical Report 24-03. Zuse Institute Berlin, Takustr. 7, 14195 Berlin, 2024. Available online https://opus4.kobv.de/opus4-zib/frontdoor/index/index/docId/9578

[15] Inci Yueksel-Erguen, Thorsten Koch, Janina Zittel. A Quality Assessment and Improvement Framework for Supply Infrastructure and Utility Network Datasets. (Under preparation)

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

[17] GasLib-2607 instance. Available online https://gaslib.zib.de/download/data/GasLib-2607-v1-20240708.zip