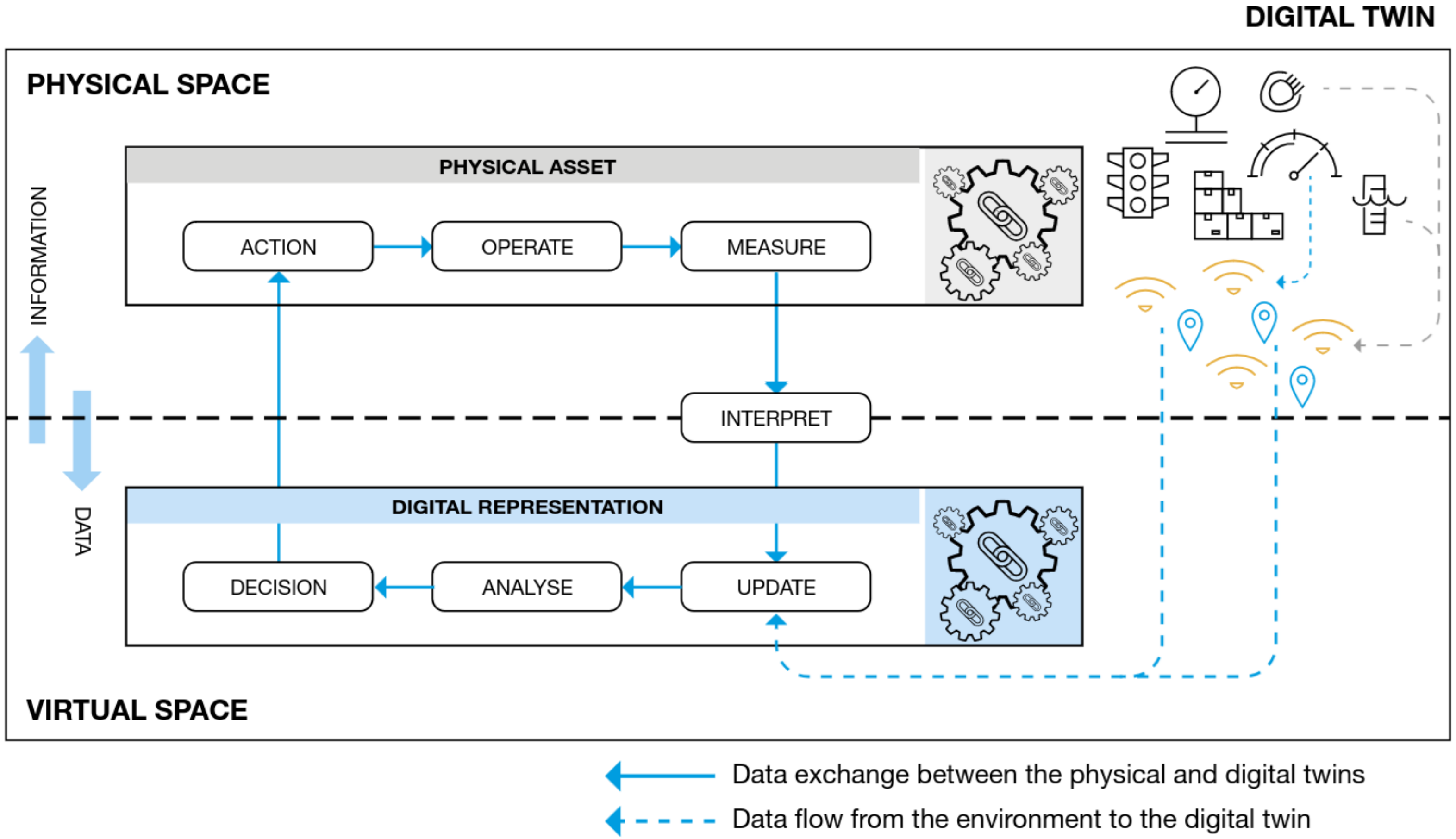
SummerSOC, 21 June 2025

Engineering Digital Twins of Critical Infrastructures

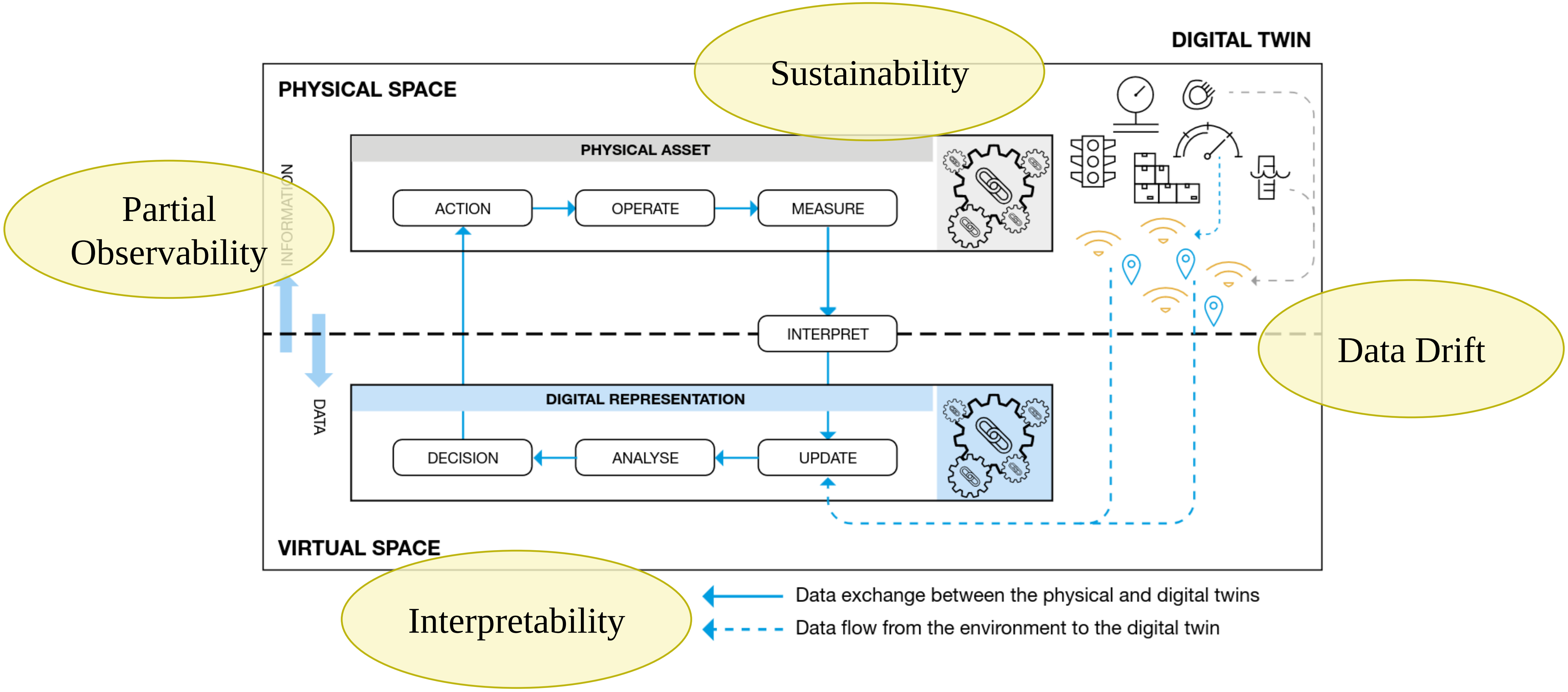
Victoria Degeler

Digital Twins in Industries & Critical Infrastructures



Tello, A., & Degeler, V. (2022). Digital Twins: An enabler for digital transformation. In Digital Transformation: A Guide for Managers (pp. 176–203). Groningen Digital Business Centre (GDBC), University of Groningen. doi: 10.5281/zenodo.7647493

Challenges

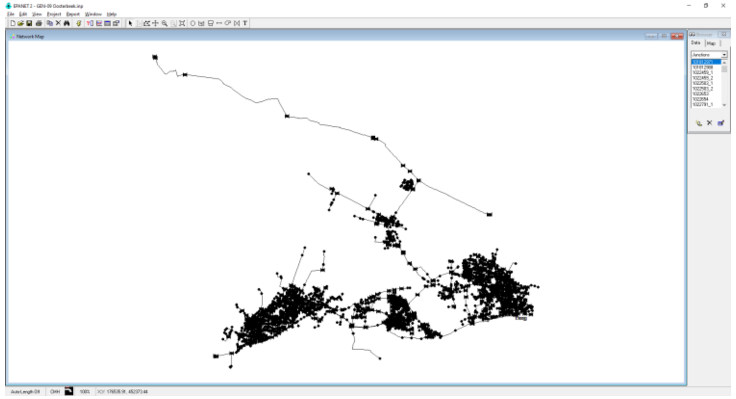


Sustainability
DIGITAL TWIN
PHYSICAL SPACE
PHYSICAL ASSET
ACTION
OPERATE
MEASURE
Partial Observability
INFORMATION
INTERPRET
Data Drift
DATA
DIGITAL REPRESENTATION
DECISION
ANALYSE
UPDATE
VIRTUAL SPACE
Interpretability
Data exchange between the physical and digital twins
Data flow from the environment to the digital twin

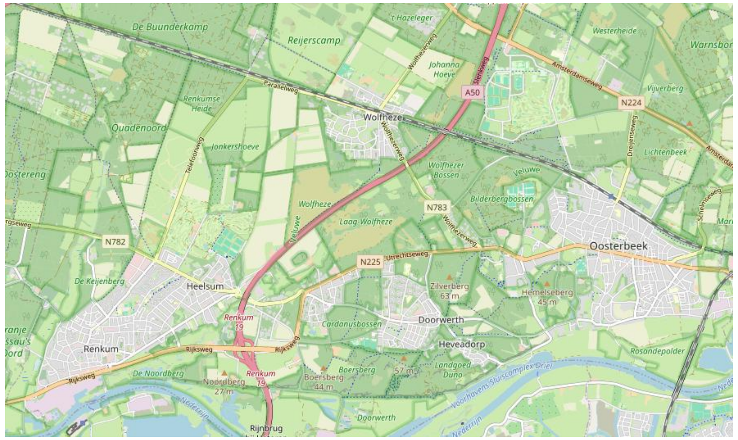
University of Amsterdam

DiTEC in a nutshell

Digital Twin for
Evolutionary Changes in
Water Networks



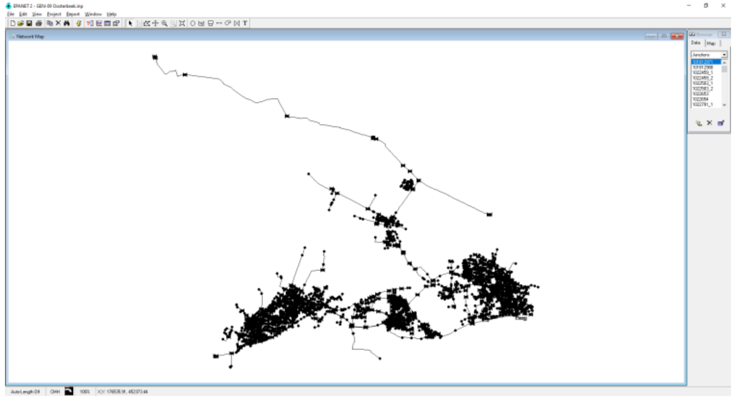
EPANET + WNTR
Water Network Tool for Resilience



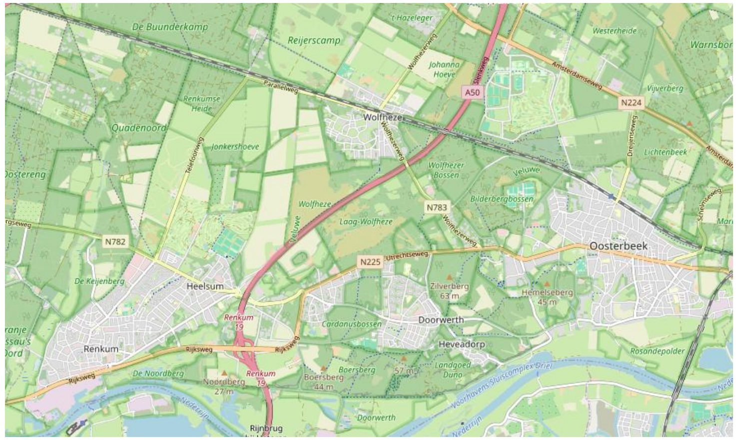
Oosterbeek, Gelderland

DiTEC in a nutshell

Digital Twin for
Evolutionary Changes in
Water Networks



EPANET + WNTR
Water Network Tool for Resilience



Oosterbeek, Gelderland

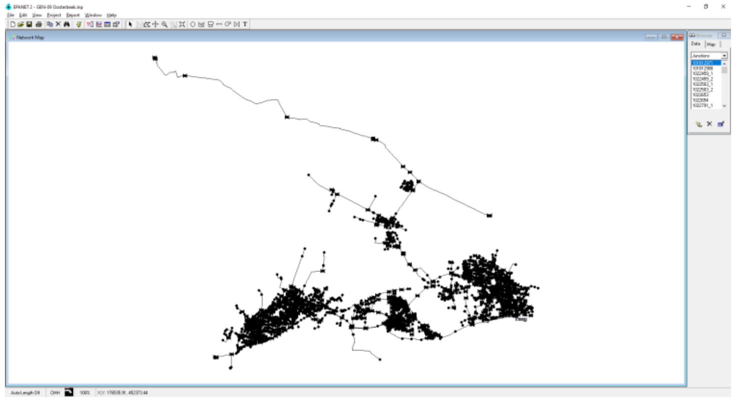
DiTEC in a nutshell

Digital Twin for

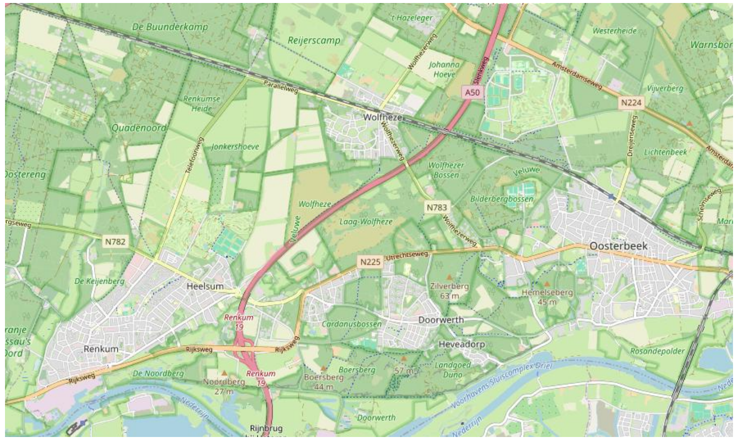
- Network conditions estimated based on minimal external sensor information
- Changes in the network are reflected in a DT in real-time
- Deep Learning for state estimation

Evolutionary Changes in

Water Networks



EPANET + WNTR Water Network Tool for Resilience



Oosterbeek, Gelderland

DiTEC in a nutshell

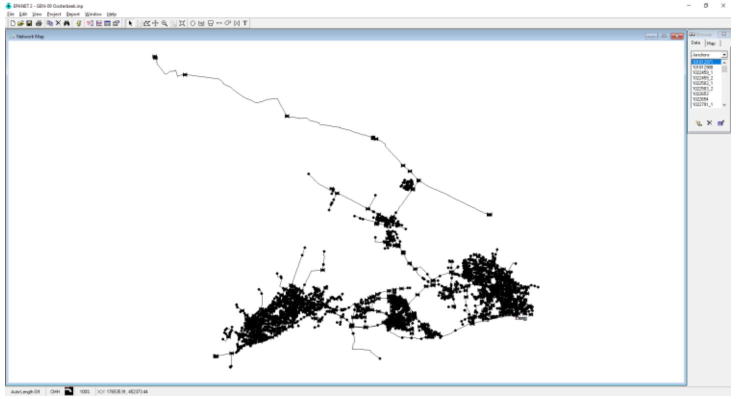
Digital Twin for

- Network conditions estimated based on minimal external sensor information
- Changes in the network are reflected in a DT in real-time
- Deep Learning for state estimation

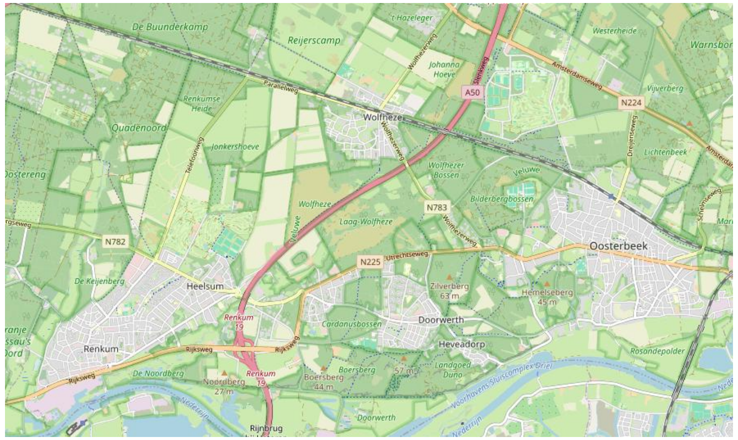
Evolutionary Changes in

- Discrepancy detection: leaks, sensor issues, valves
- Alerts and monitoring system
- Reasoning on sensor data
- Logical explanations of the causes of changes

Water Networks



EPANET + WNTR Water Network Tool for Resilience



Oosterbeek, Gelderland

UNIVERSITY OF AMSTERDAM

DiTEC in a nutshell

Digital Twin for

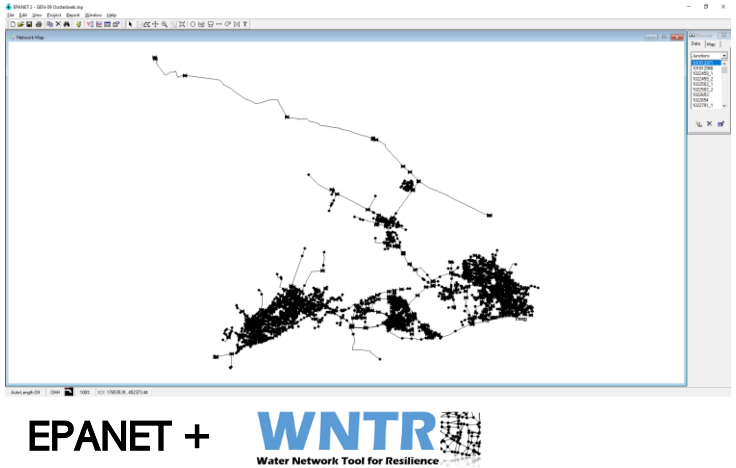
- Network conditions estimated based on minimal external sensor information
- Changes in the network are reflected in a DT in real-time
- Deep Learning for state estimation

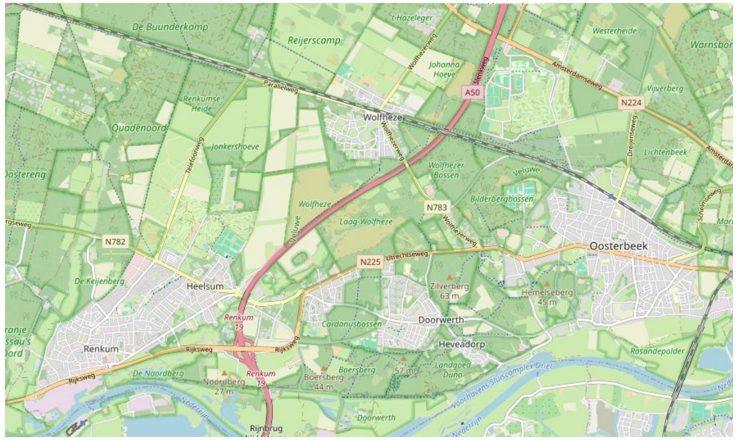
Evolutionary Changes in

- Discrepancy detection: leaks, sensor issues, valves
- Alerts and monitoring system
- Reasoning on sensor data
- Logical explanations of the causes of changes

Water Networks

- Water Distribution Network in Oosterbeek, Gelderland



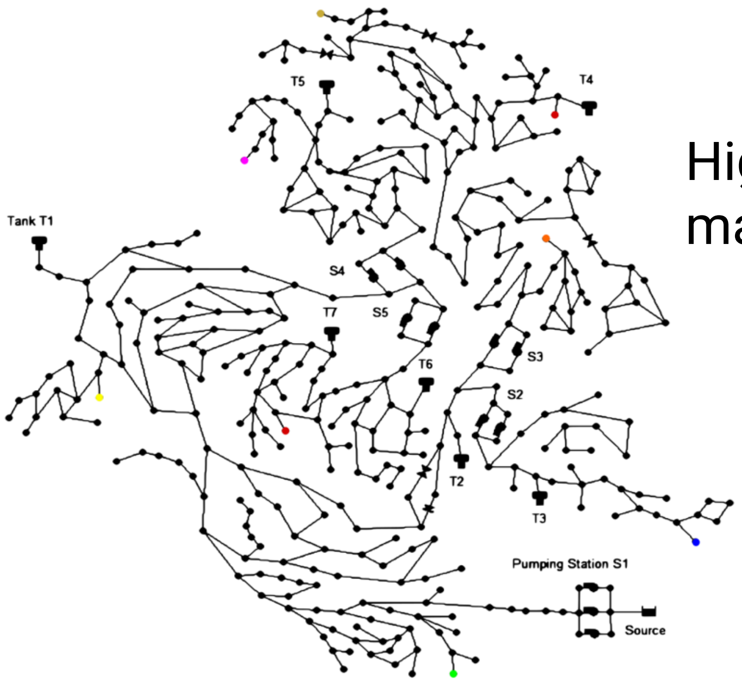


Oosterbeek, Gelderland

UNIVERSITY OF AMSTERDAM

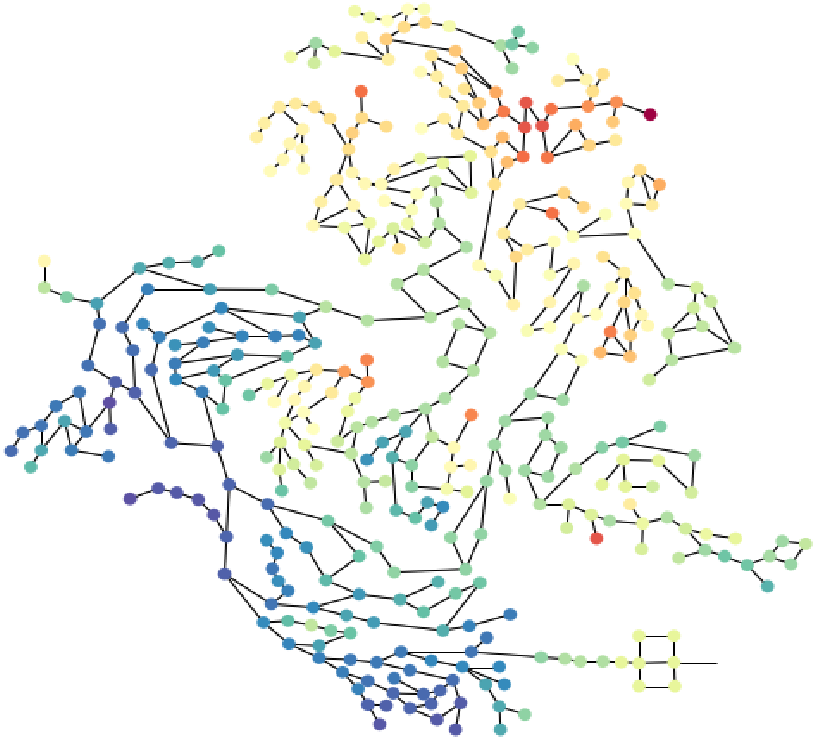
Problems: Partially Observable Data

Real Sensor Coverage



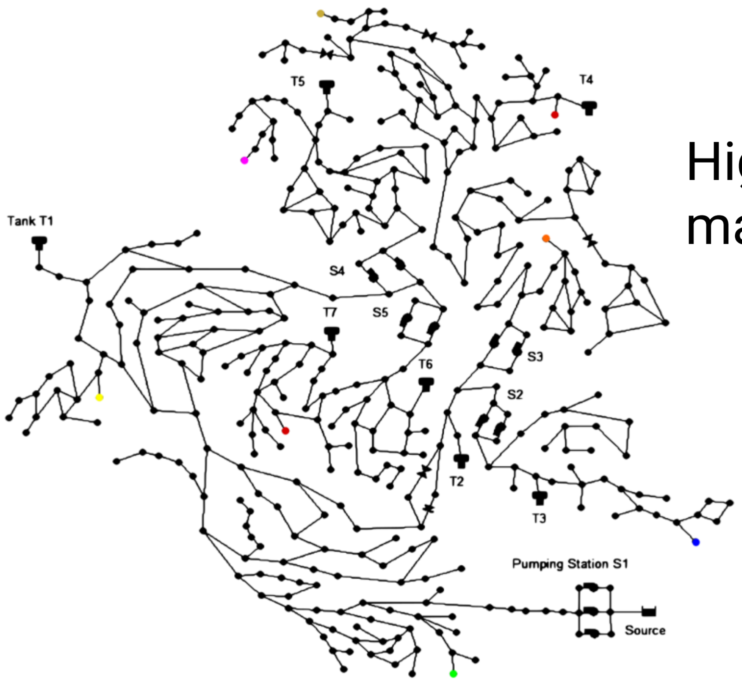
High installation & maintenance costs

Ideal Sensor Coverage



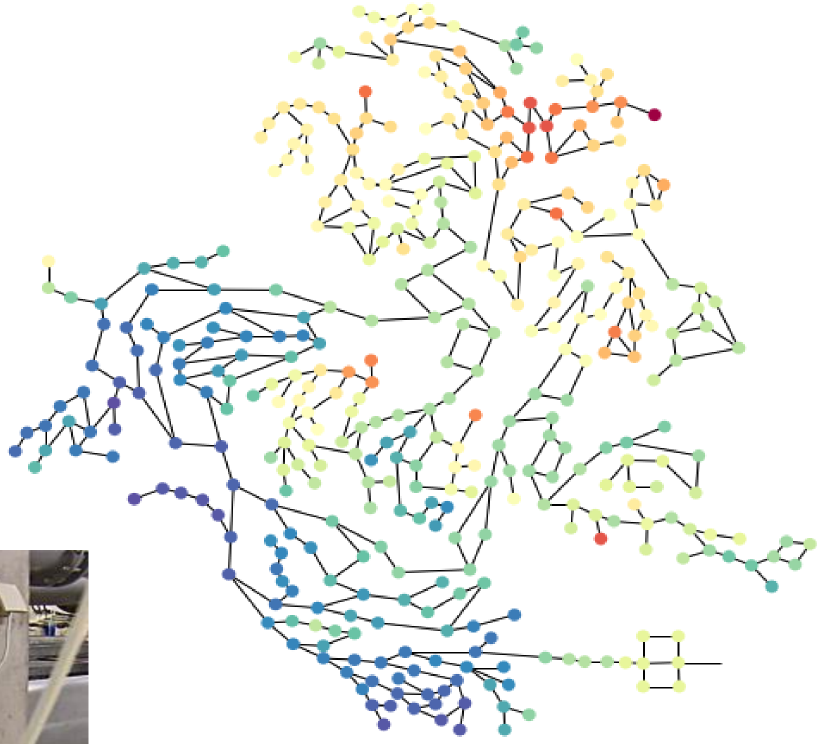
Problems: Partially Observable Data

Real Sensor Coverage

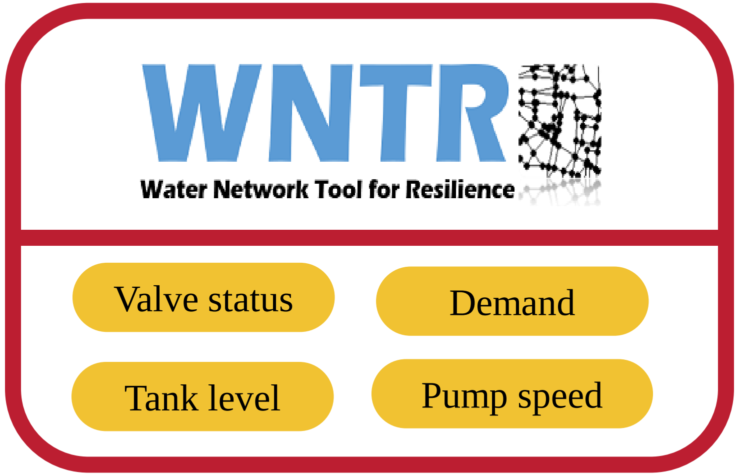


High installation & maintenance costs

Ideal Sensor Coverage

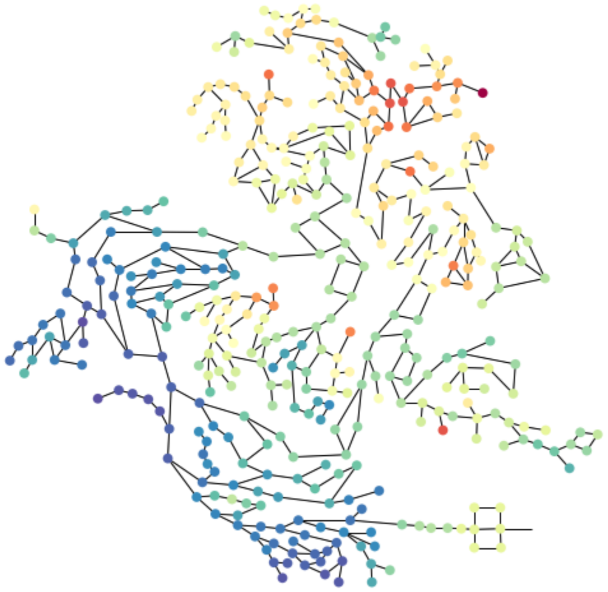


Network state reconstruction: Hydraulic Models

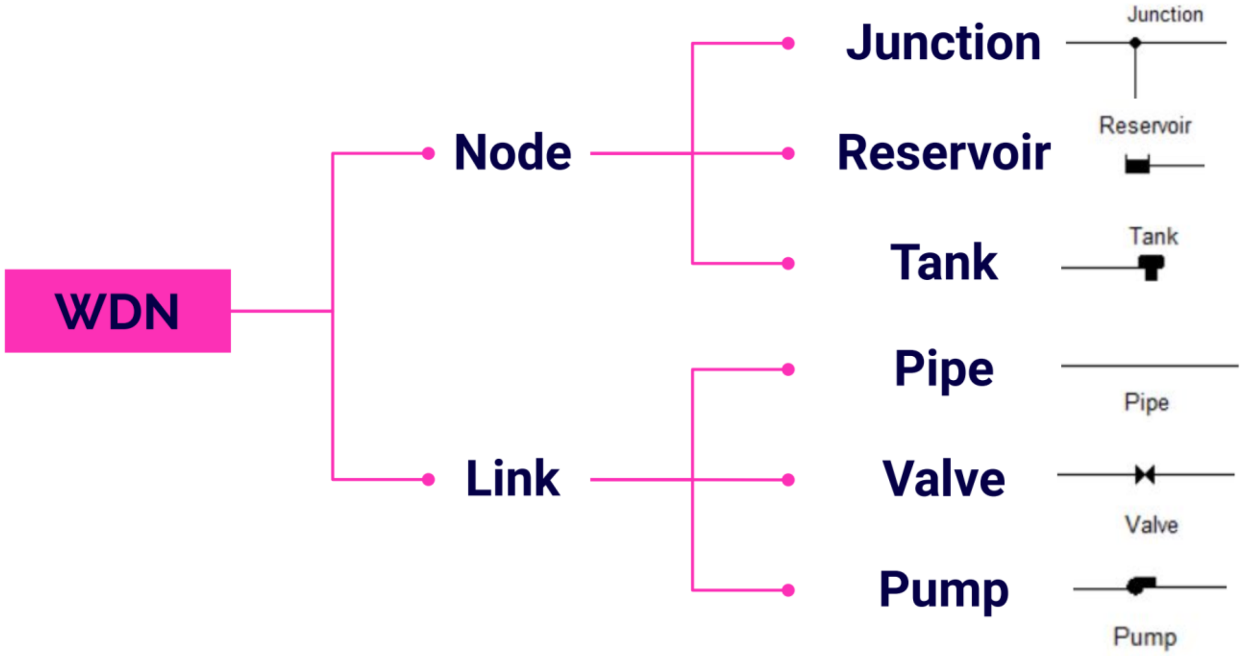


Physical simulation
(needs to know correct parameters)

Real network state
(cannot be observed directly!)



Parameters



WDN
Node
Link
Junction
Reservoir
Tank
Pipe
Valve
Pump
Junction
Reservoir
Tank
Pipe
Valve
Pump

Junction- Elevation
Elevation
The elevation (m) above common reference (sea level)
Junction- Demand
Demand = Flow out
Flow in
Junction- Demand pattern
Reservoir- primary properties
Total Head
The hydraulic head (elevation + pressure_head) of water
Tank- primary properties
Elevation
The elevation above a common reference point
Initial level
Height of water surface at the start of the simulation.
Min level
Minimum height of the water surface above the bottom elevation that will be maintained.
Max level
Maximum height of the water surface above the bottom elevation that will be maintained.
Diameter
The diameter of the tank
Tank- primary properties
diameter
max level
init level
min level
Elevation
Pipe- primary properties
Roughness
The roughness coefficient of the pipe
Diameter
The pipe diameter
Length
The actual length of the pipe
length
diameter
Roughness
Valve- primary properties
Initial status
Open | Close | Active
Diameter
The pipe diameter
Type
Setting
PRV
PSV
PBV
FCV
TCV
GPV
Pump- Pump curve
Head
Flow
WNTR
Water Network Tool for Resilience

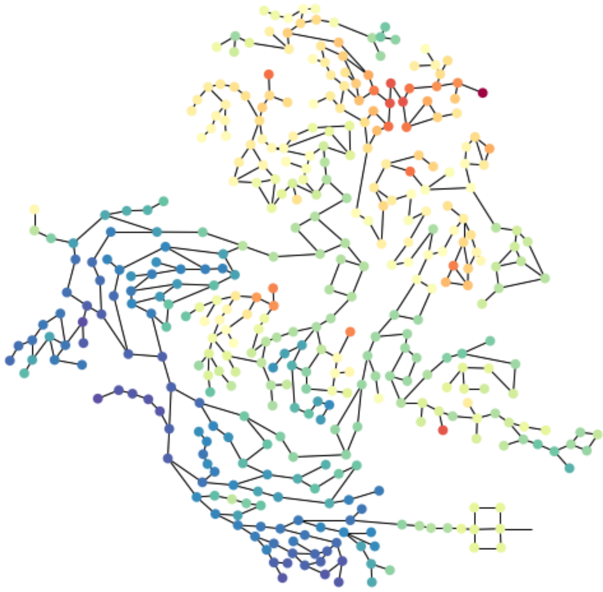
Network state reconstruction

WNTR
Water Network Tool for Resilience

Valve status | Demand
Tank level | Pump speed

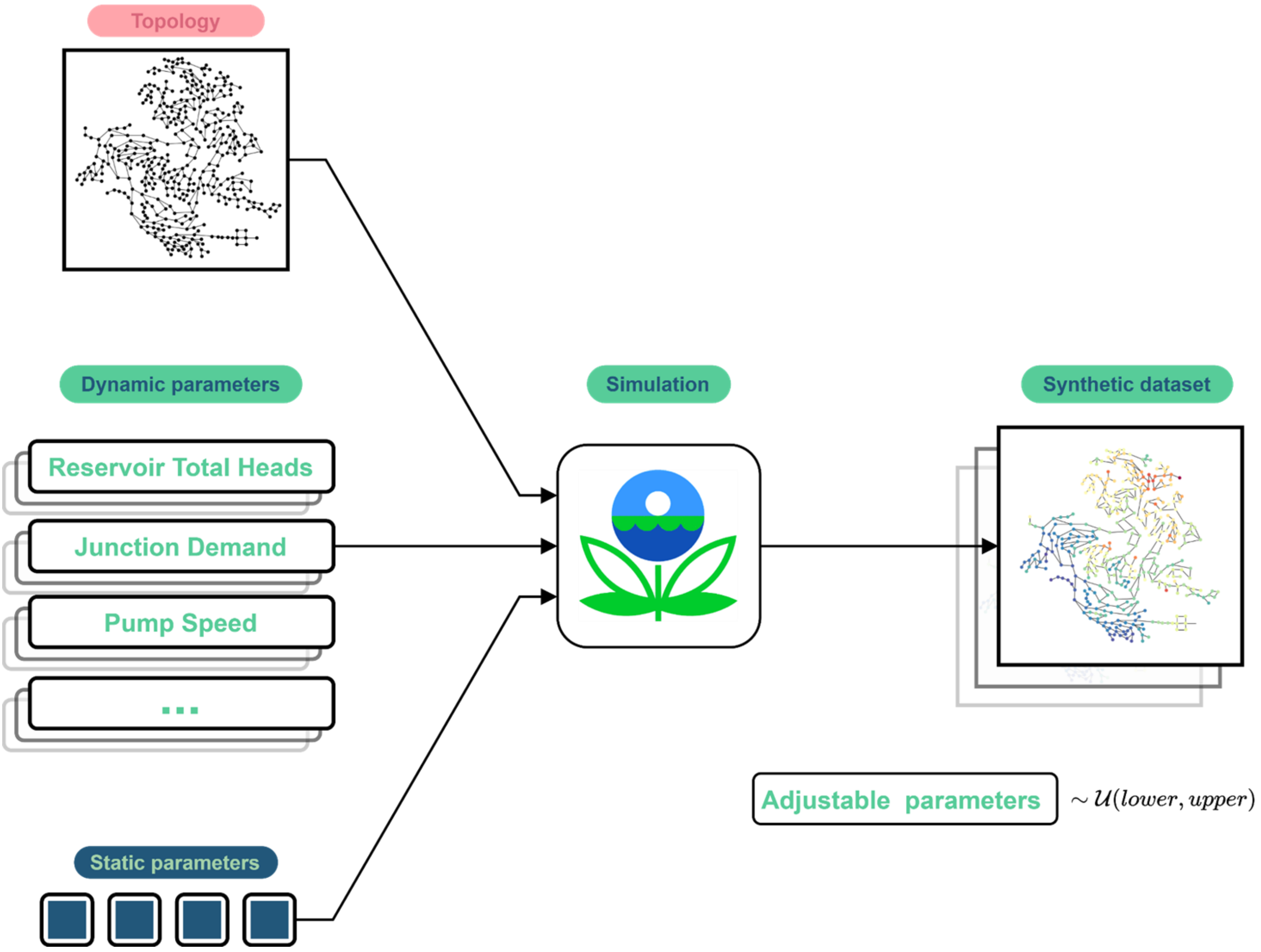
Physical simulation
(needs to know correct parameters)

Real network state
(cannot be observed directly!)

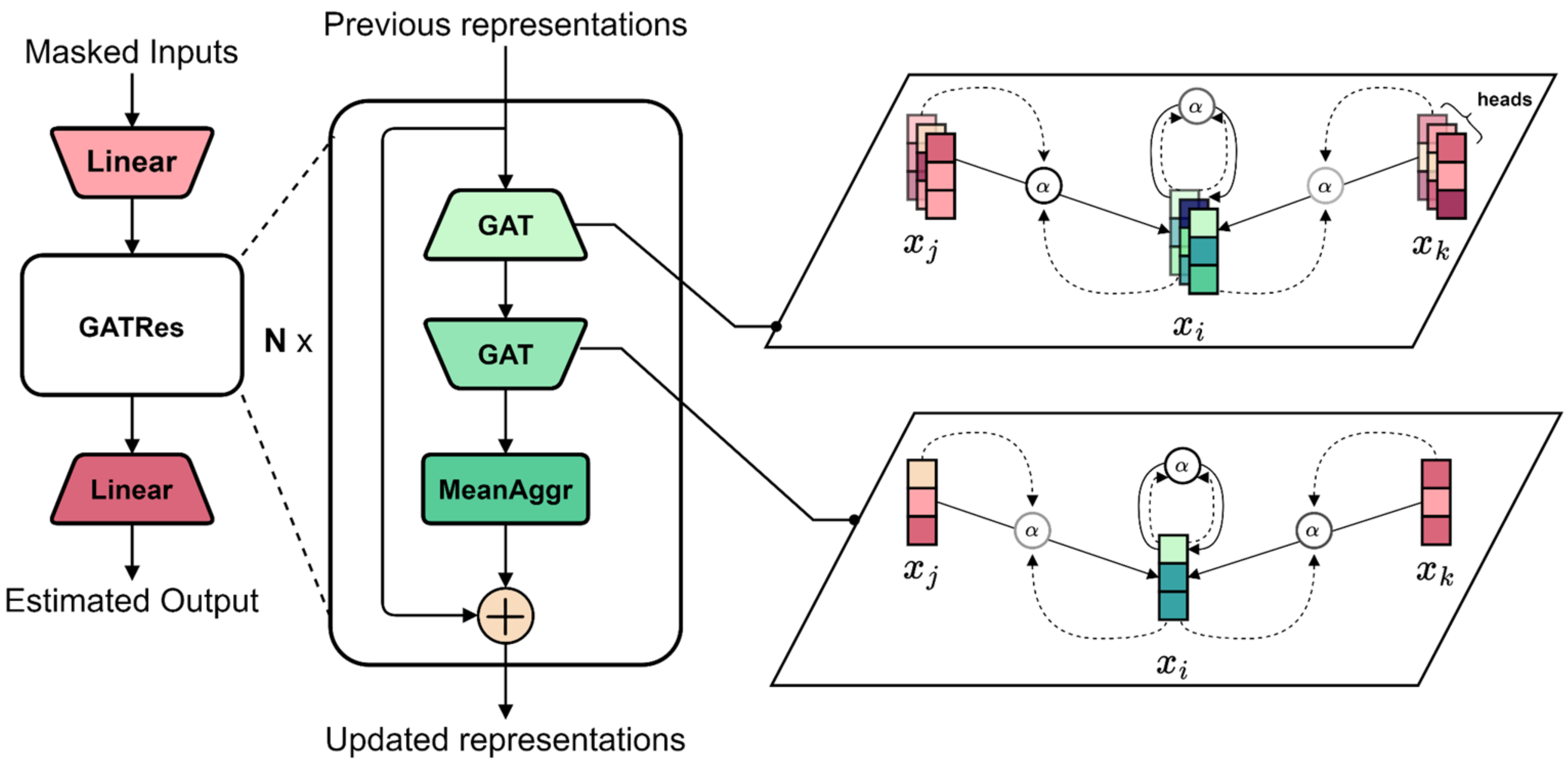


Deep Learning Reconstruction Model
(needs to be trained)

Dataset creation

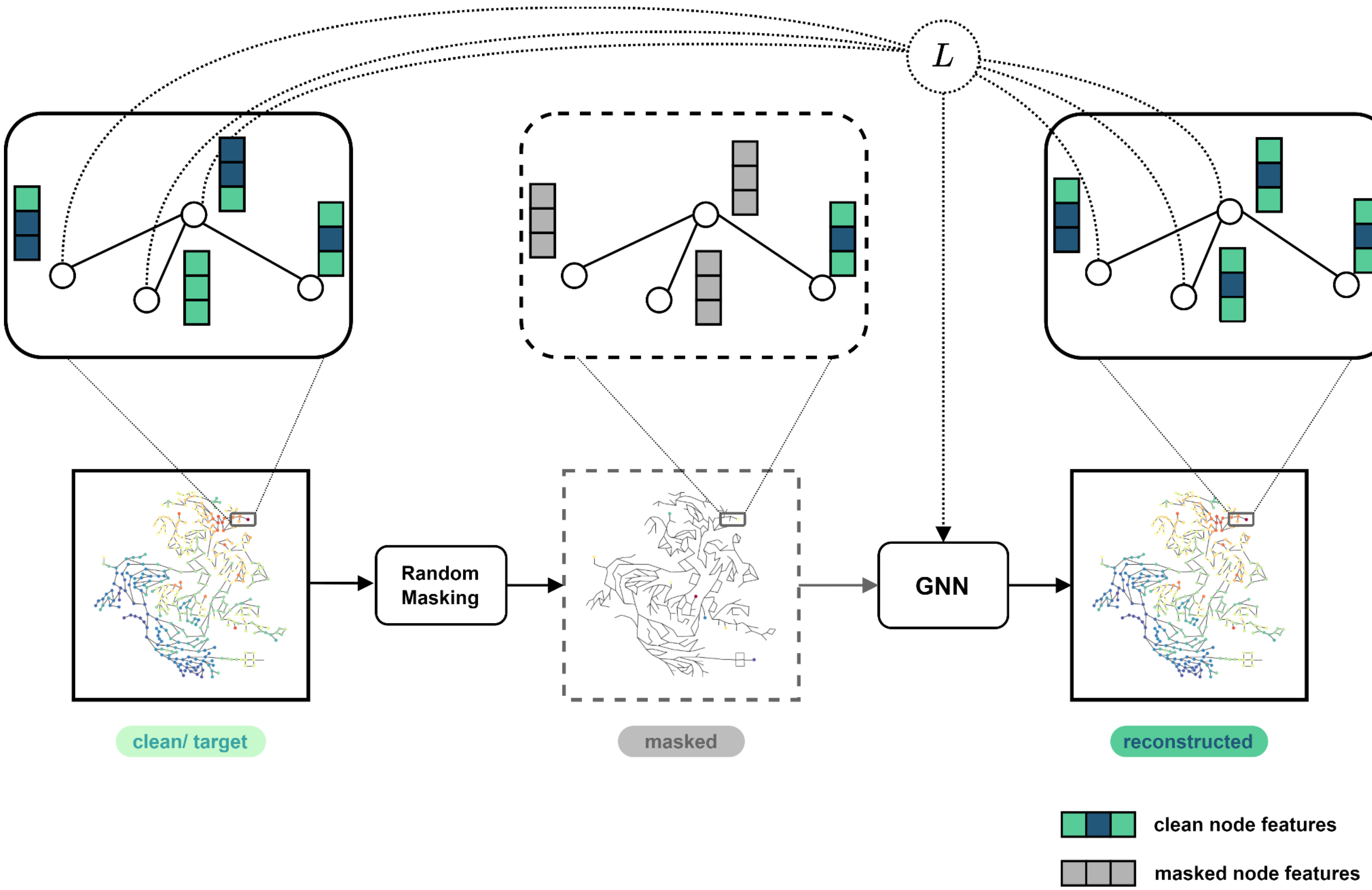


Model architecture - GatRes



H. Truong, A. Tello, A. Lazovik, V. Degeler (2024) Graph Neural Networks for Pressure Estimation in Water Distribution Systems, Water Resources Research, 60, e2023WR036741. doi:10.1029/2023WR036741

Model Training



Results: Oosterbeek

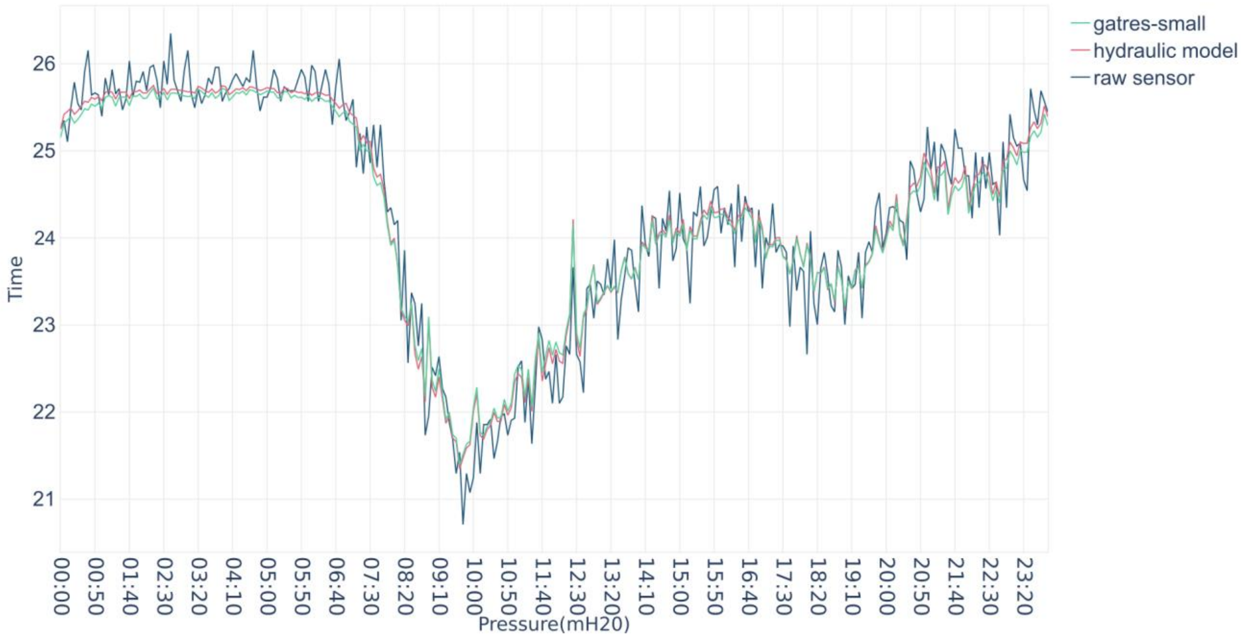
Table 5. Model comparison in the noisy test performed on 24-hour Oosterbeek WDN at 95% masking rate.

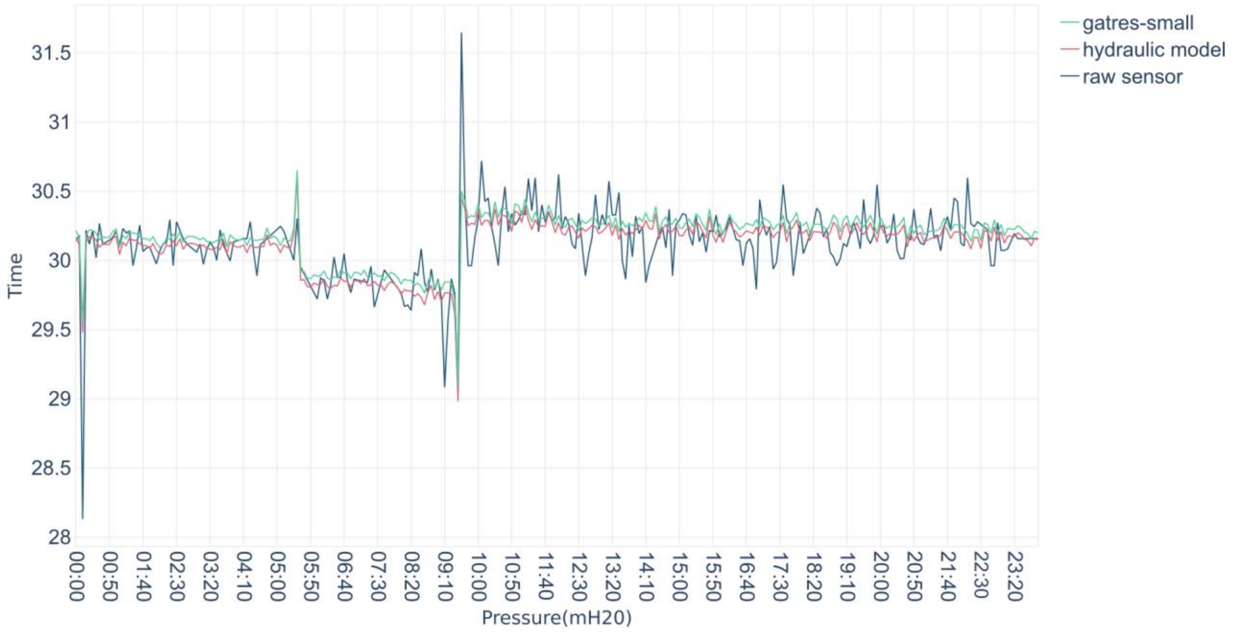
Model	#Milion Params (↓)	MAE(↓)	MAPE(↓)	NSE(↑)	Acc(@0.1)(↑)
GCNii (M. Chen et al., 2020)	0.65	6.696±0.0838	0.2484±0.0552	-0.1064±0.0266	36.02±0.4684
GAT (Veličković et al., 2018)	0.35	4.397±0.3052	0.2112±0.0767	0.1490±0.1153	66.98±1.6290
GraphConvWat (Hajgató et al., 2021)	0.92	3.611±0.1234	0.1551±0.0376	0.5877±0.0370	62.99±1.1600
GraphConvWat-tuned	**0.23**	2.347±0.0252	0.0963±0.0363	0.749±0.0086	81.09±0.3877
mGCN (Ashraf et al., 2023)	2.48	2.188±0.0558	0.0948±0.0155	0.6993±0.0213	82.83±0.4199
GATRes-small (ours)	0.66	**1.964**±0.0301	0.0802±0.0458	**0.778**±0.0113	**86.56**±0.2826
GATRes-large (ours)	1.67	2.115±0.0503	**0.0799**±0.0207	0.7417±0.0140	83.43±0.5044

H. Truong, A. Tello, A. Lazovik, V. Degeler (2024) Graph Neural Networks for Pressure Estimation in Water Distribution Systems, Water Resources Research, 60, e2023WR036741. doi:10.1029/2023WR036741

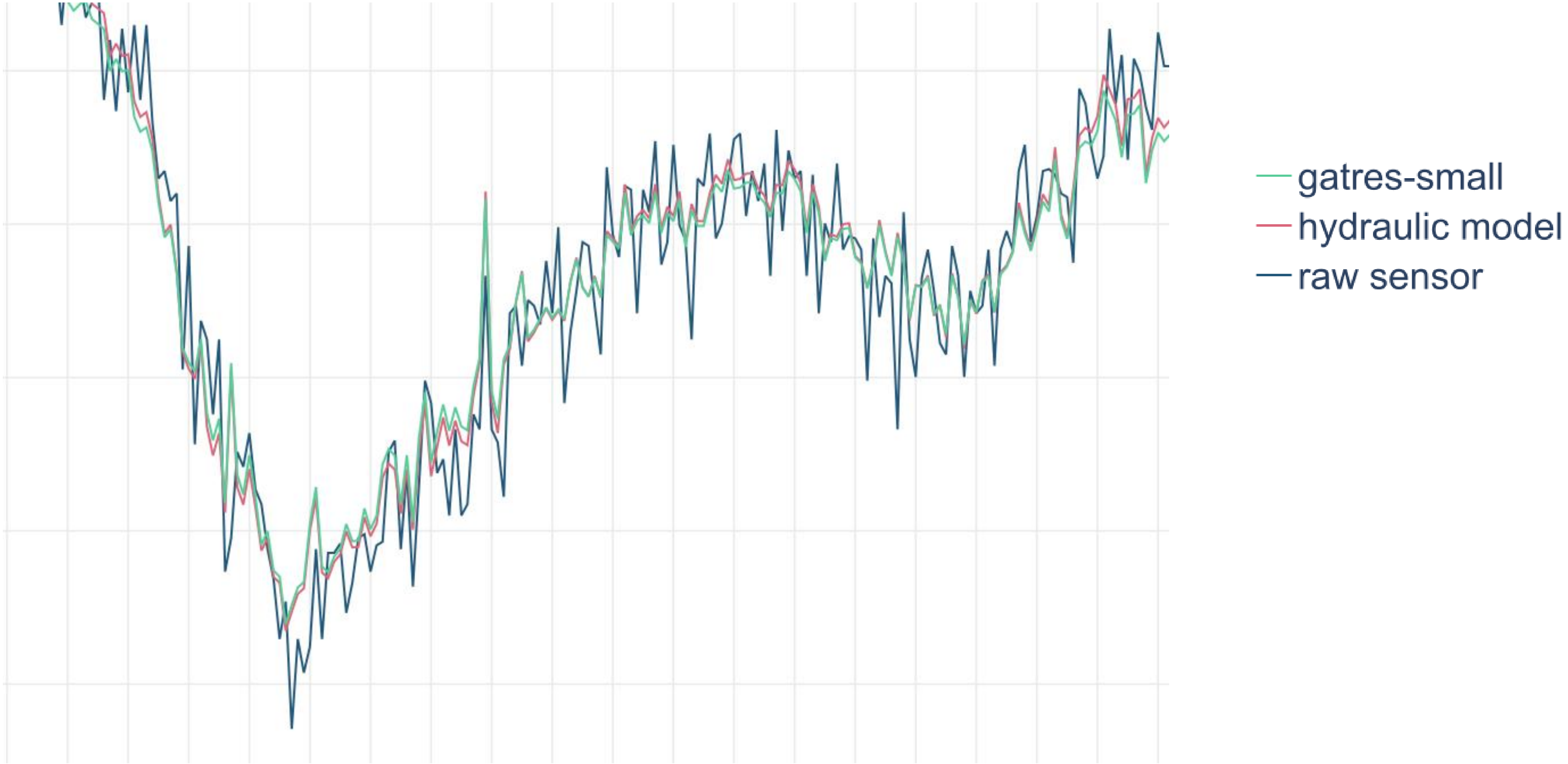
UNIVERSITY OF AMSTERDAM

Results: Oosterbeek





Results: Oosterbeek



DiTEC dataset

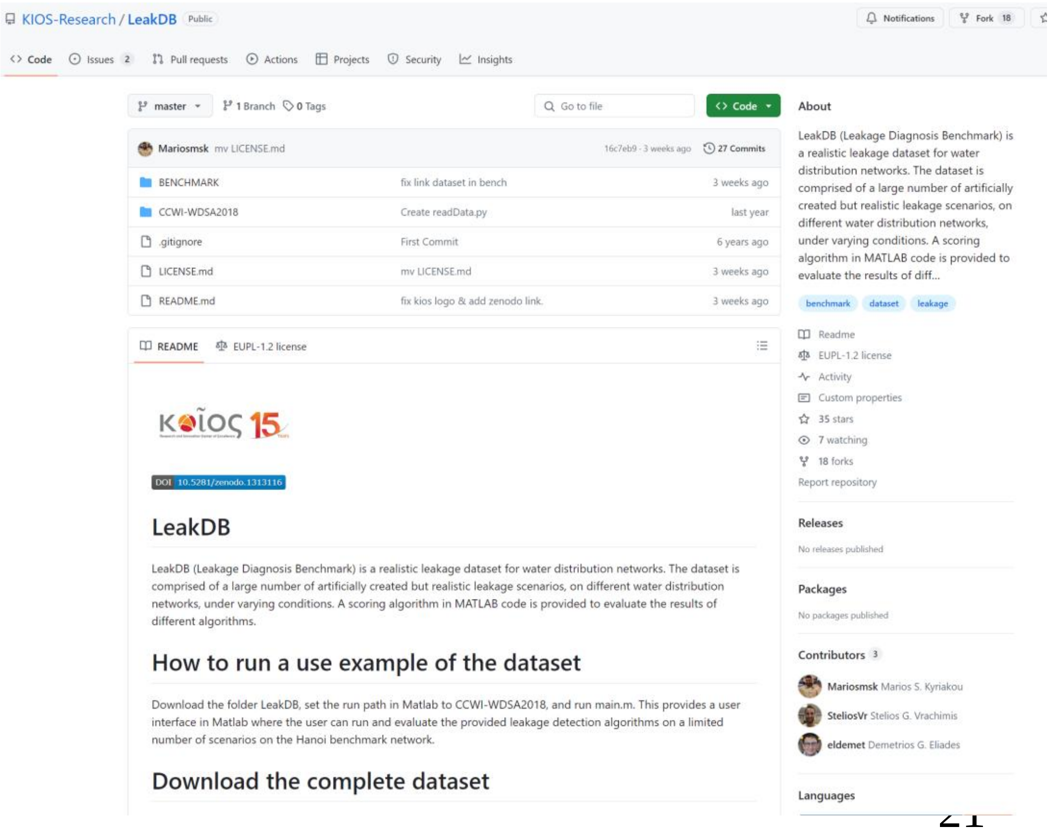
LeakDB 2018

Leakage Diagnosis Benchmark

the only "ready-to-use" published dataset contains a small Hanoi network of 32 nodes and 4 adjustable parameters

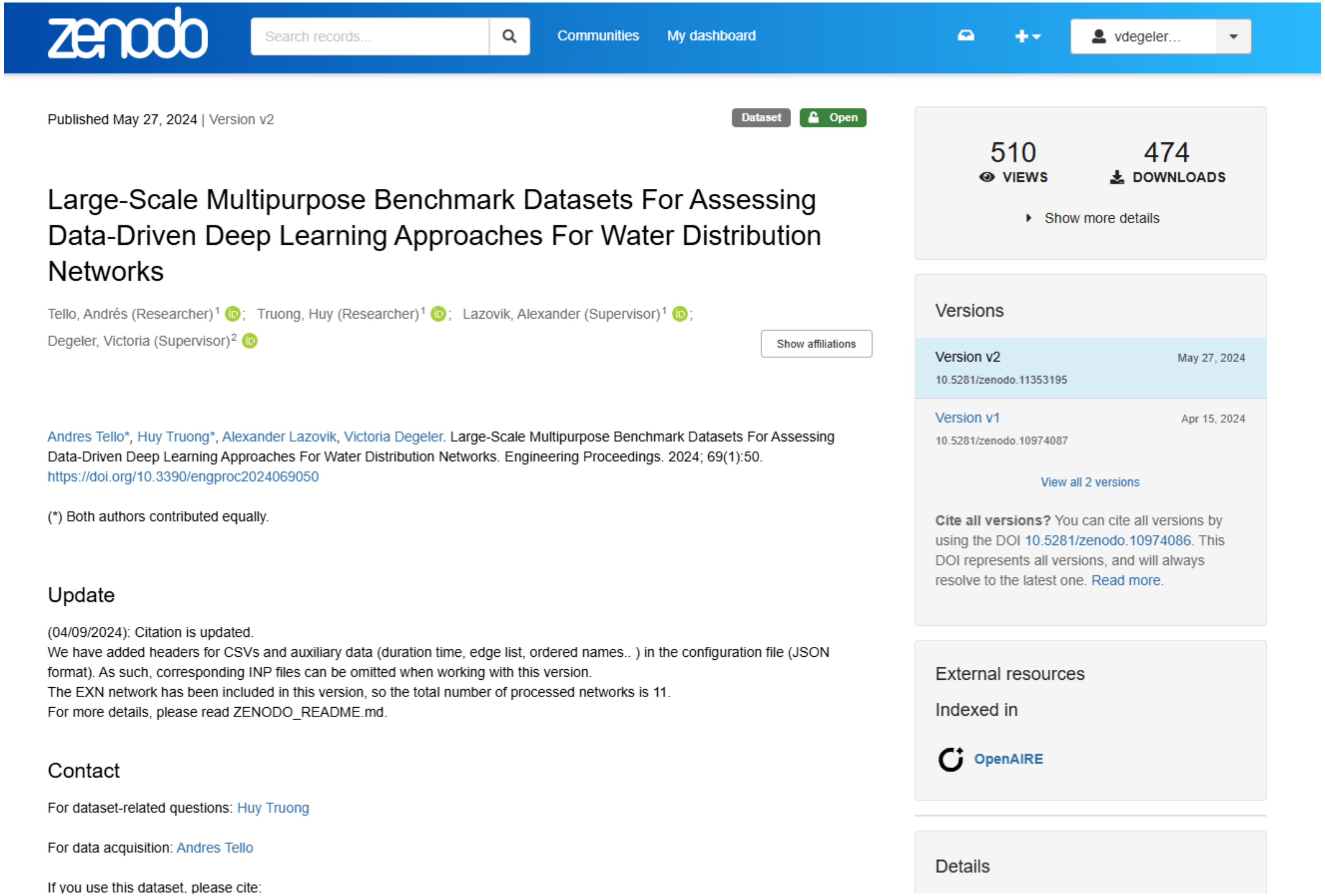
Why do we need new data?

- Publicly available WDN networks are provided as a collection of asset configuration files
- Need to run simulations for full data generation
- Existing data rarely include time-dependent patterns
    - Privacy and safety concerns
    - Few patterns reused many times on several nodes in the network
    - Limits the variety of the data and robustness of models to uncertainties



UNIVERSITY OF AMSTERDAM

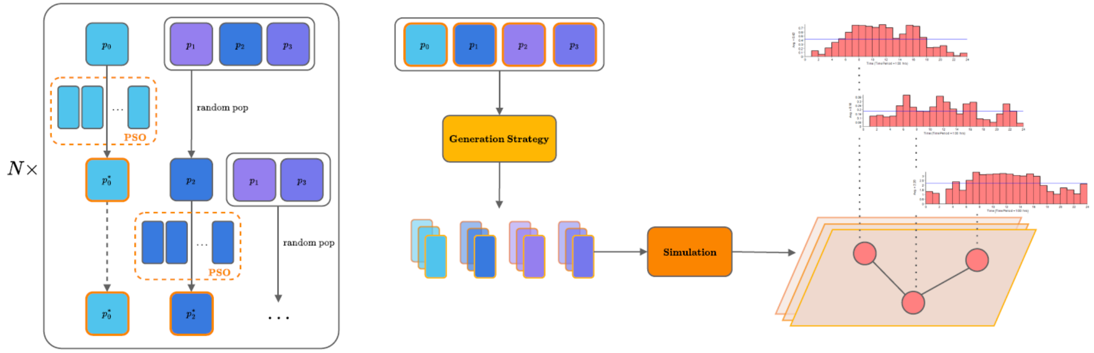
DiTEC initial dataset



A. Tello, H. Truong, A. Lazovik, V. Degeler (2024) Large-Scale Multipurpose Benchmark Datasets For Assessing Data-Driven Deep Learning Approaches For Water Distribution Networks. Engineering Proceedings. 2024; 69(1):50. doi: 10.3390/engproc2024069050; Dataset: 10.5281/zenodo.10974086

DiTEC-WDN Large-scale Dataset

- To facilitate research into data driven WDN models on a wide range of tasks, such as:
    - foundation models
    - surrogate modelling
    - state estimation
    - demand forecasting



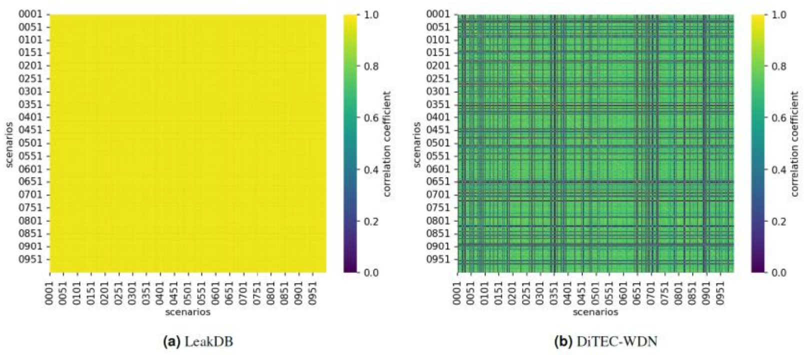
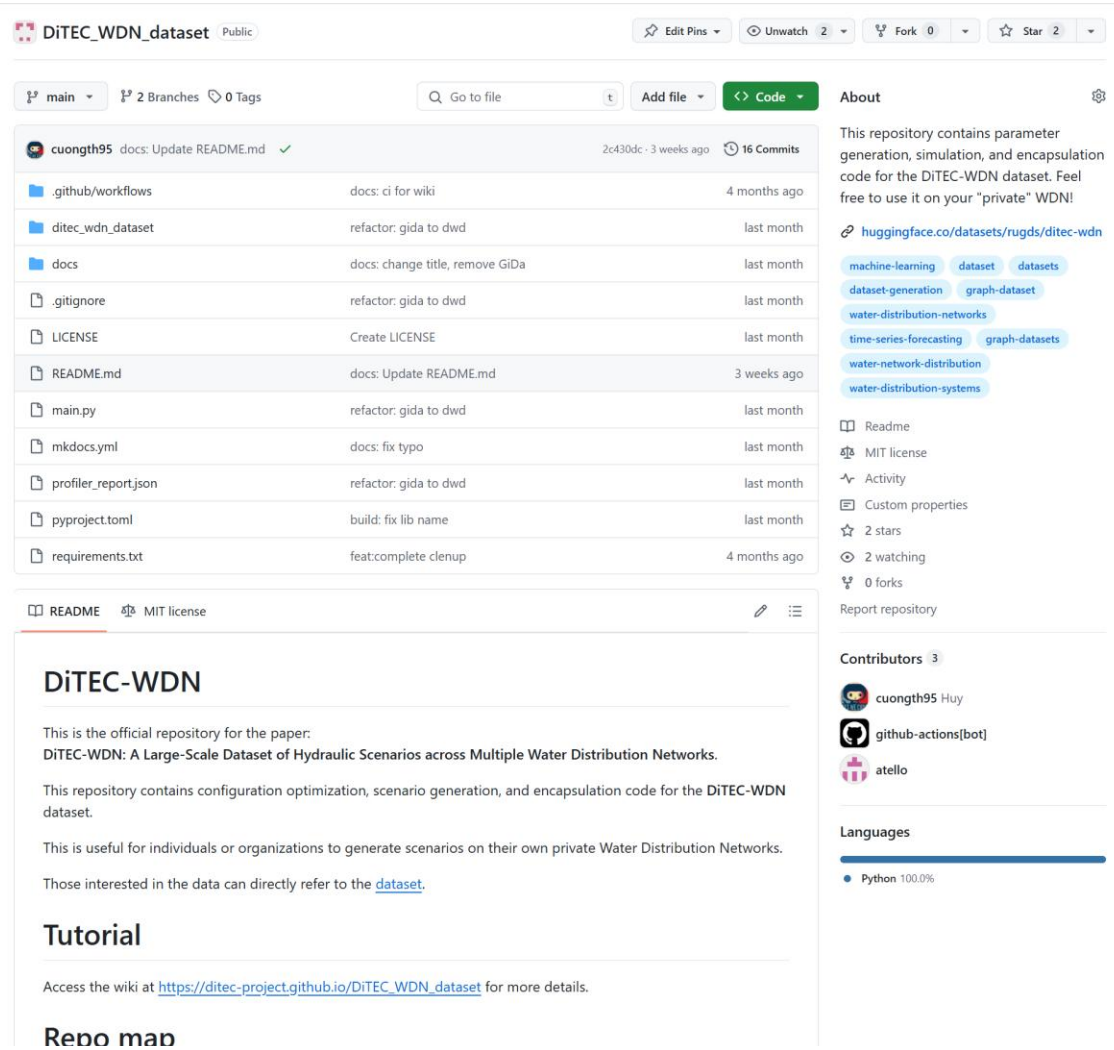


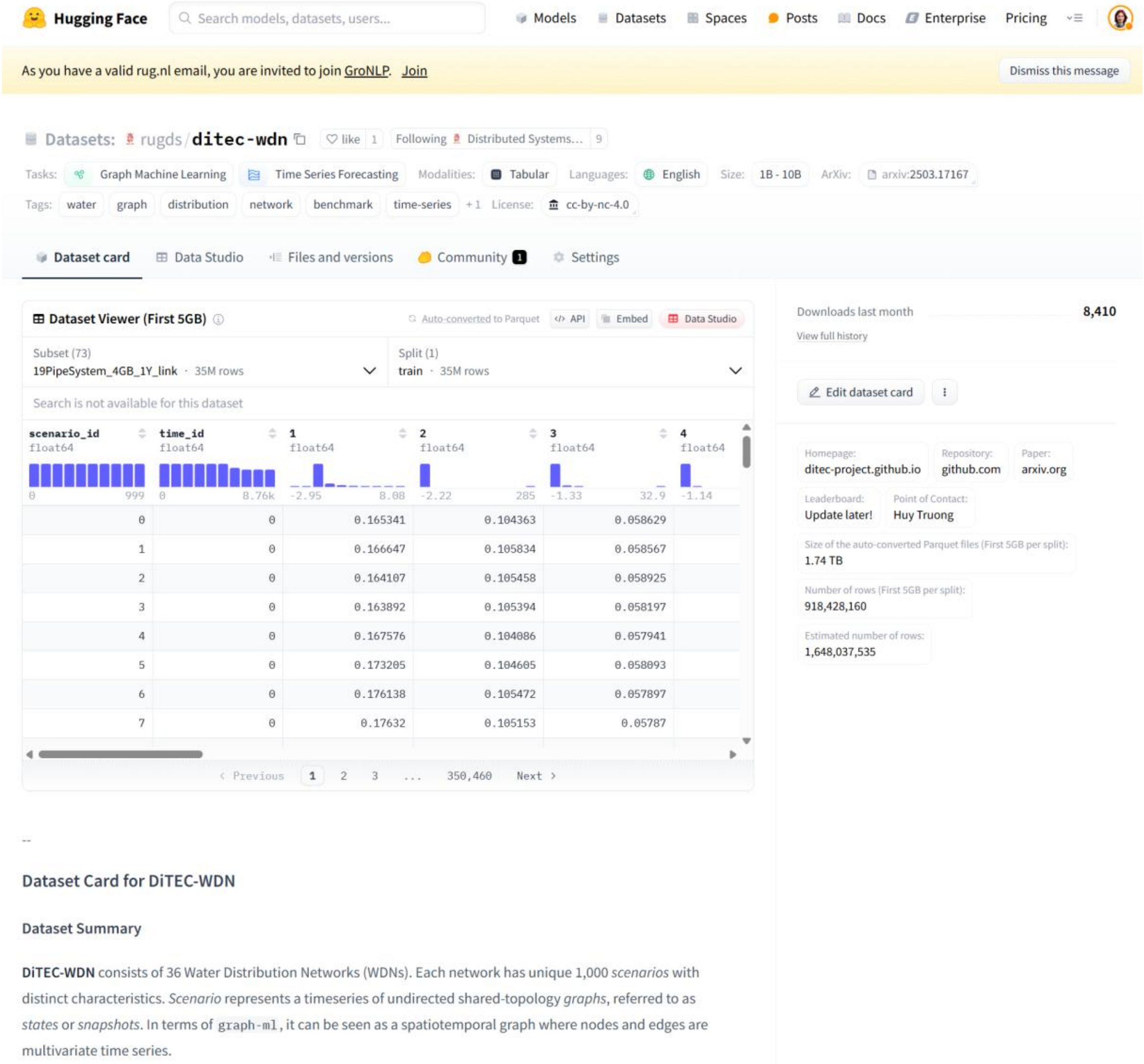
Figure 4. Correlation matrices of generated demands between all scenarios in Hanoi WDN. The left figure (a) shows the correlation between scenarios in the data generated in LeakDB[8]. The right figure (b) shows the correlation between the scenarios in our dataset. Both matrices include all 1,000 scenarios, each containing 1-year of demand data. The low correlation between scenarios in our dataset shows the diversity of the data, contrary to the similarity observed across LeakDB scenarios.

- 36 publicly available networks
- 228 million of network state snapshots
- 38 hydraulic parameters are considered

H. Truong, A. Tello, A. Lazovik, V. Degeler (2025) DiTEC-WDN: A Large-Scale Dataset of Hydraulic Scenarios across Multiple Water Distribution Networks, arxiv preprint: 10.48550/arXiv.2503.17167

DiTEC-WDN dataset

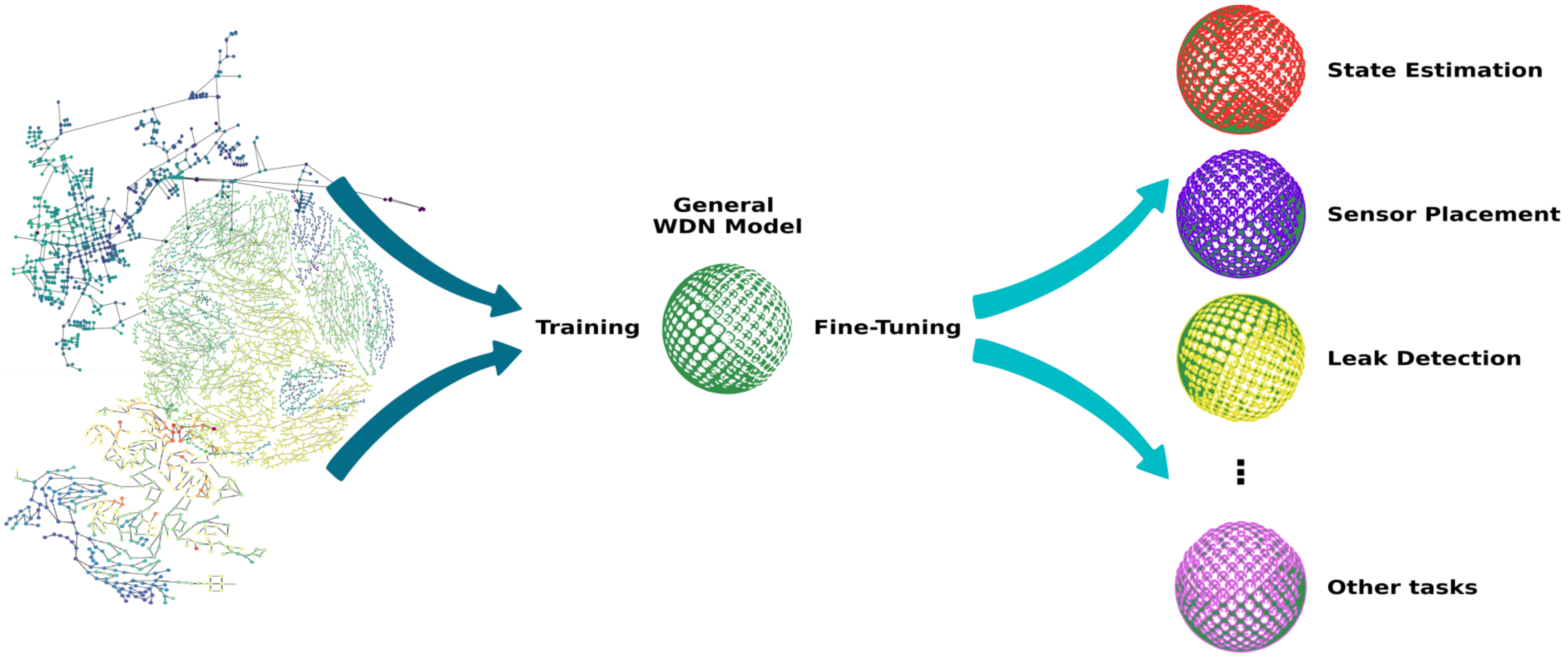




H. Truong, A. Tello, A. Lazovik, V. Degeler (2025) DiTEC-WDN: A Large-Scale Dataset of Hydraulic Scenarios across Multiple Water Distribution Networks, arxiv preprint: [10.48550/arXiv.2503.17167](https://doi.org/10.48550/arXiv.2503.17167)

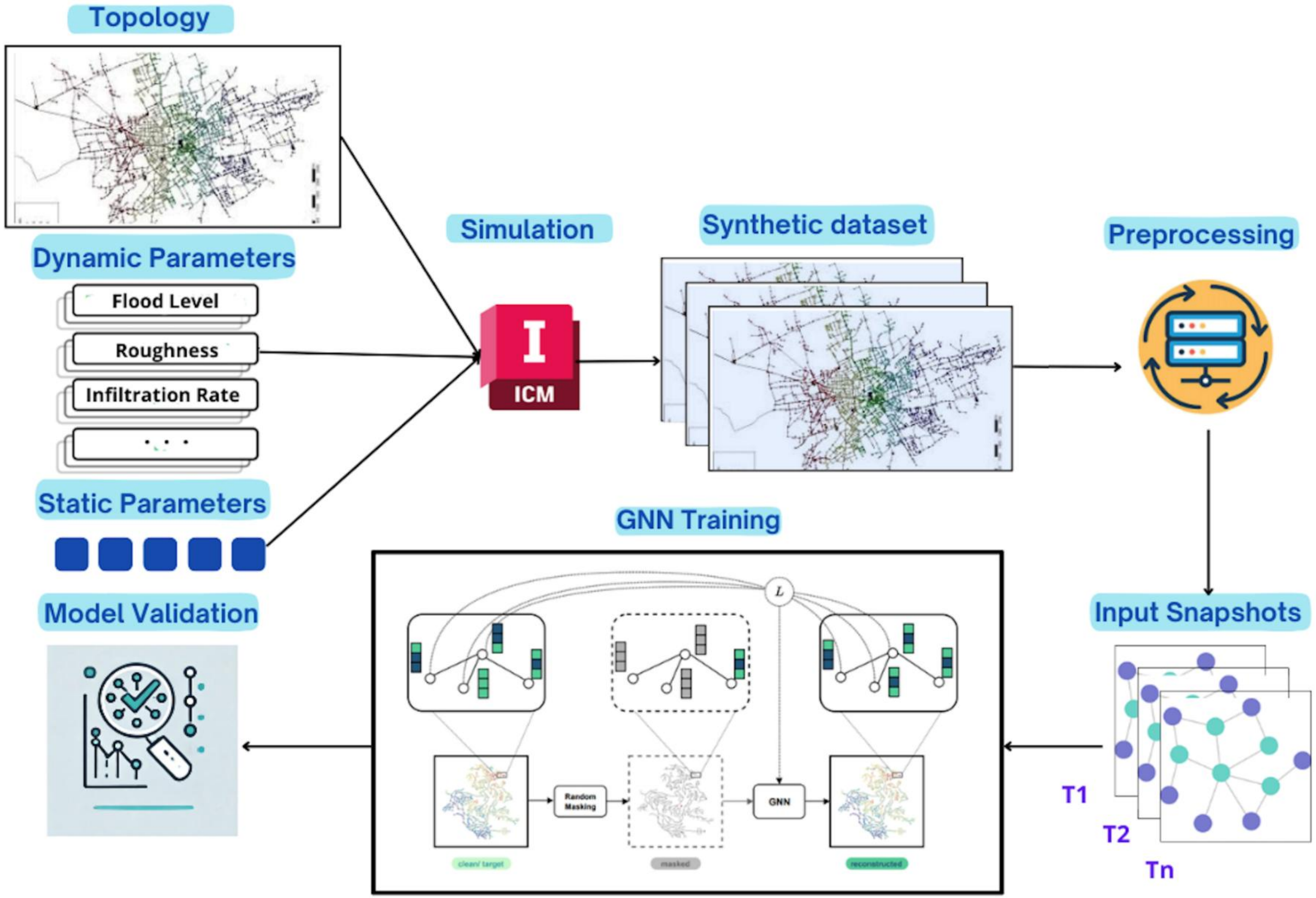
Training foundation models

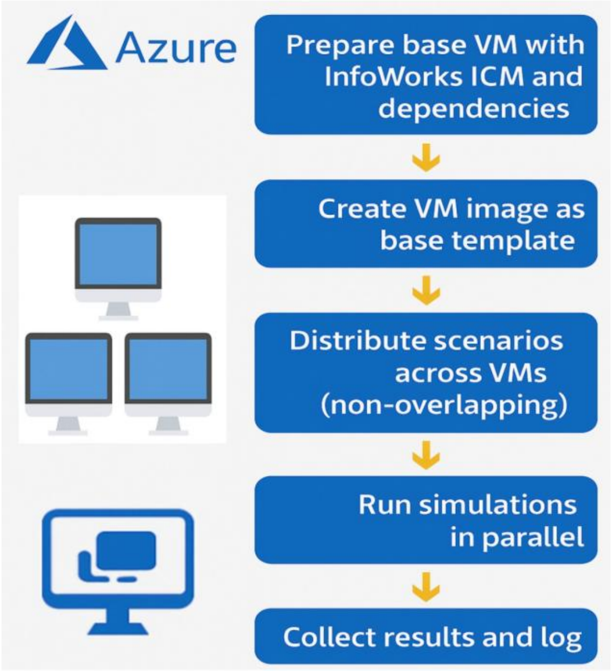
Foundation model's principles



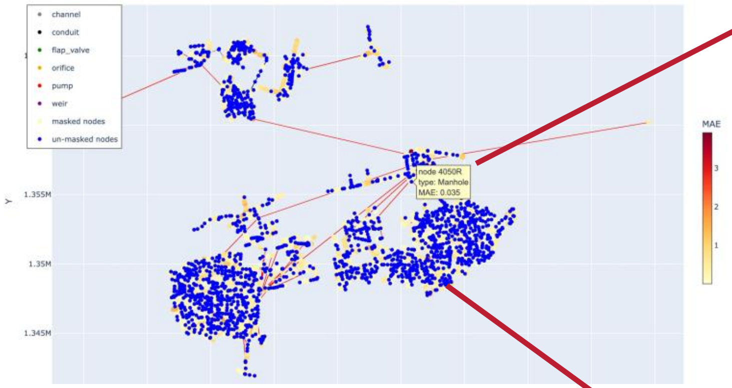
University of Amsterdam

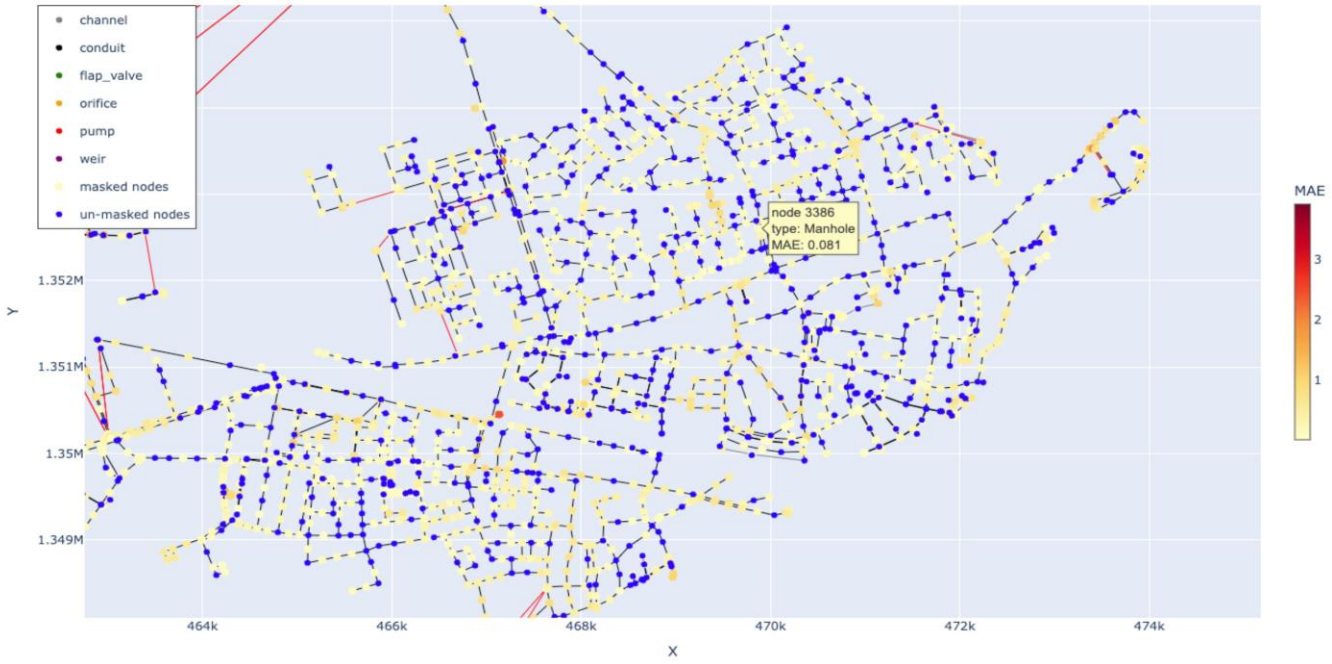
Recreating the pipeline in a wastewater domain

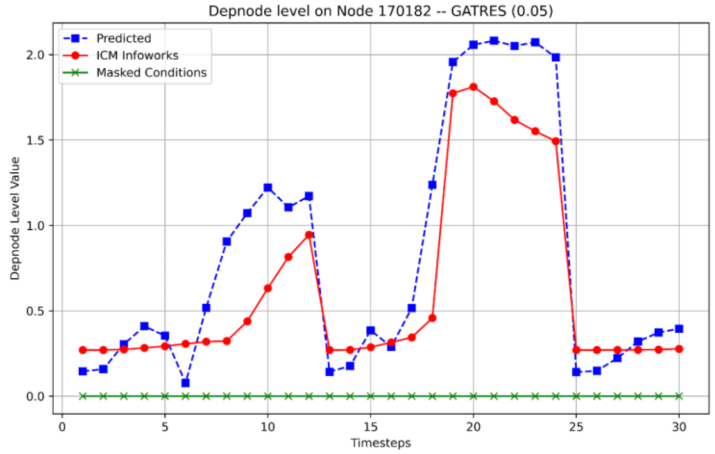




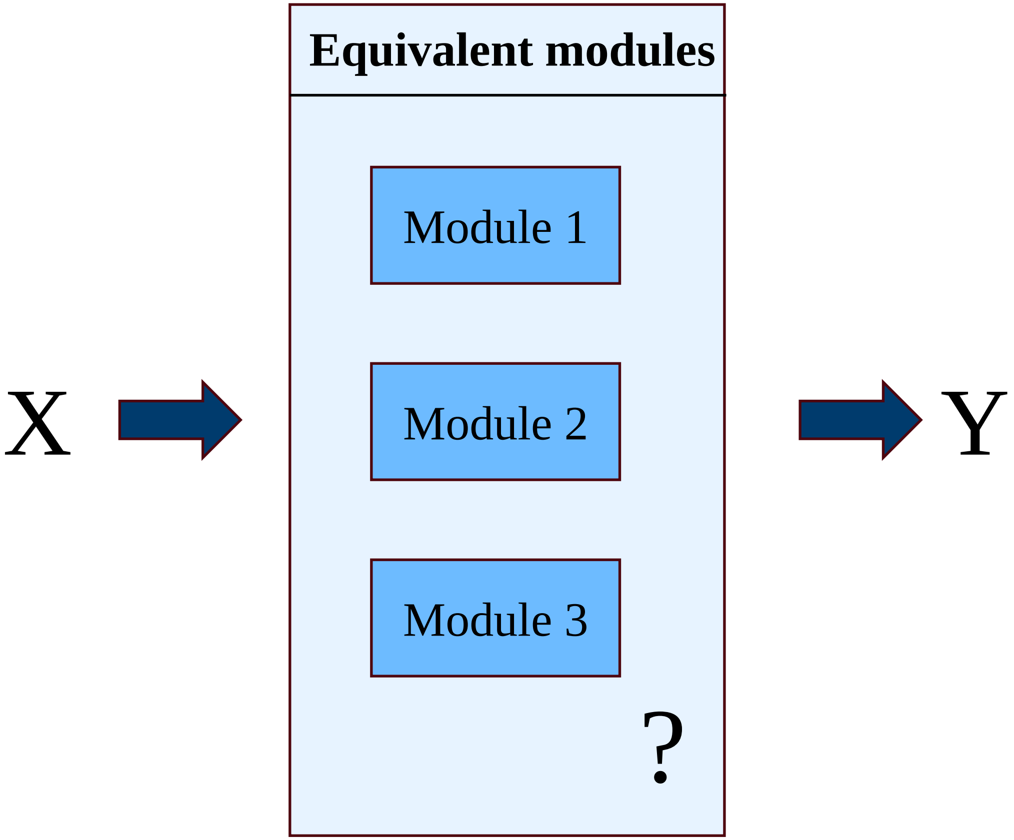
GatRes results in Heusden (NL)

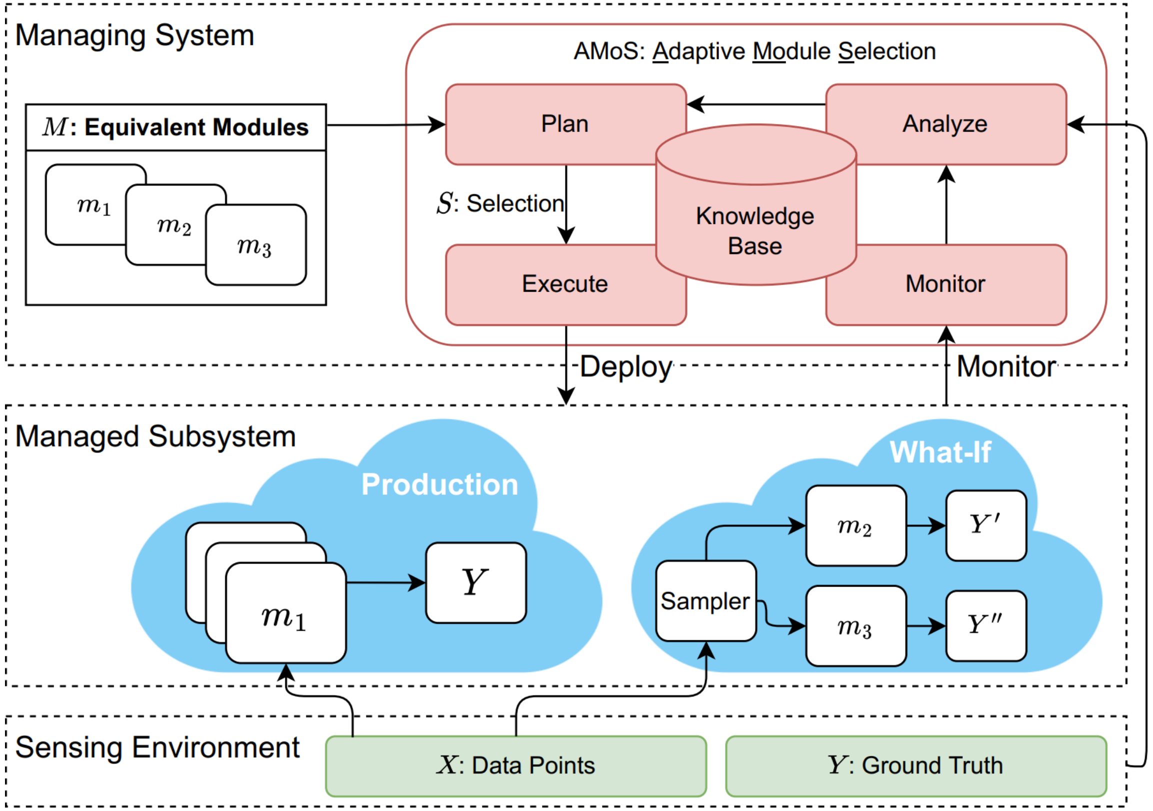






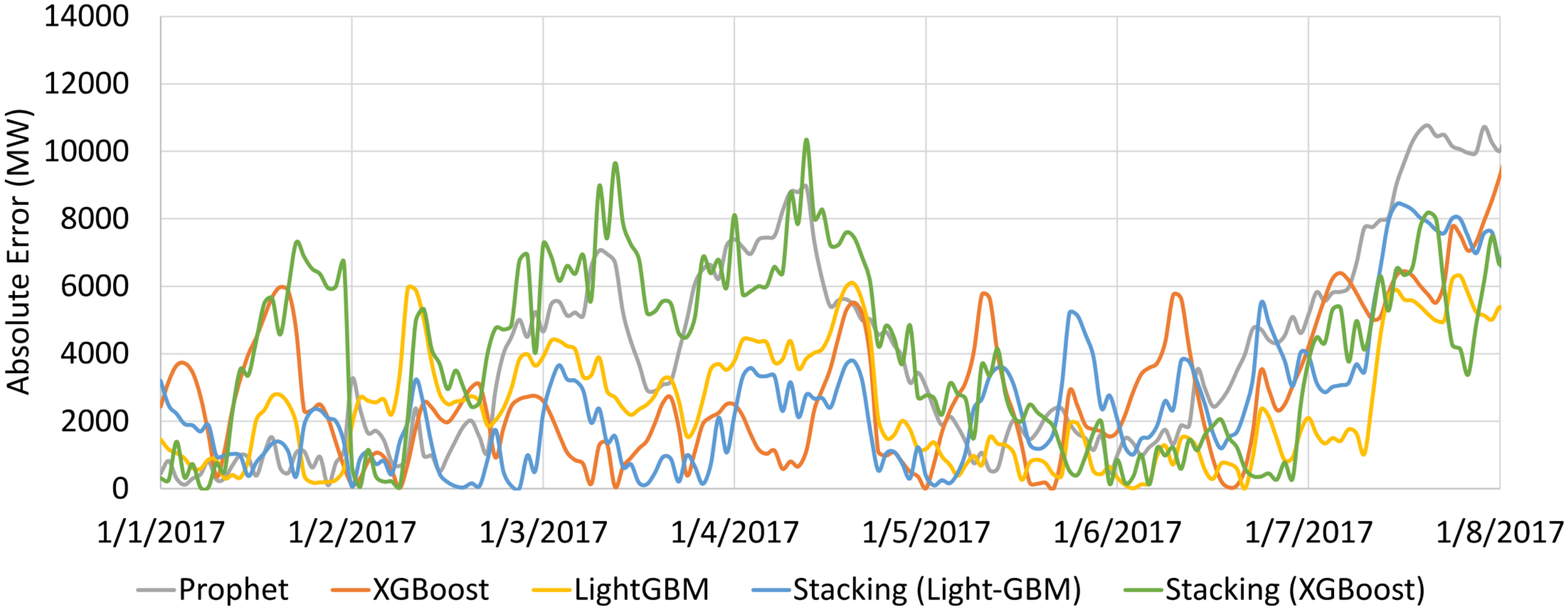
Adaptive Model Selection





M. H. Nejad Yousefi, V. Degeler and A. Lazovik, "Self-Adaptive Service Selection for Machine Learning Continuous Delivery," 2024 IEEE International Conference on Web Services (ICWS), Shenzhen, China, 2024, pp. 1048-1056, doi: 10.1109/ICWS62655.2024.00123.

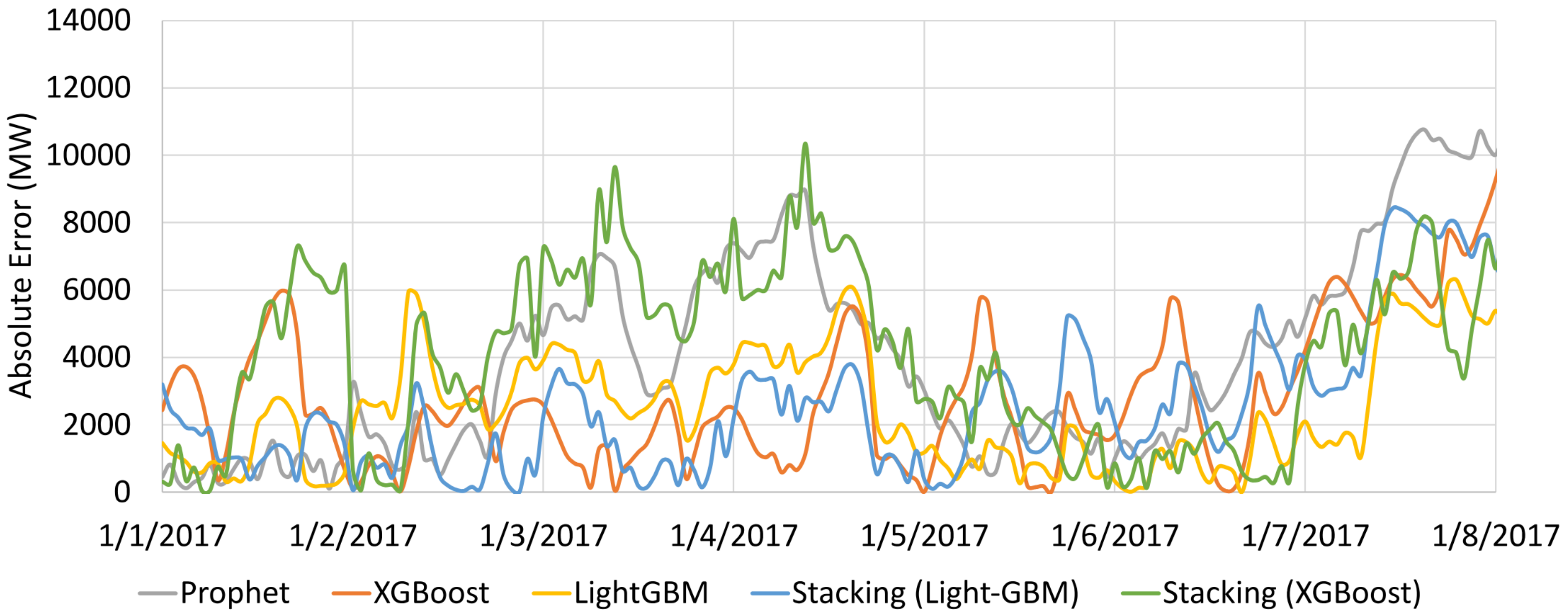
Adaptive Model Selection

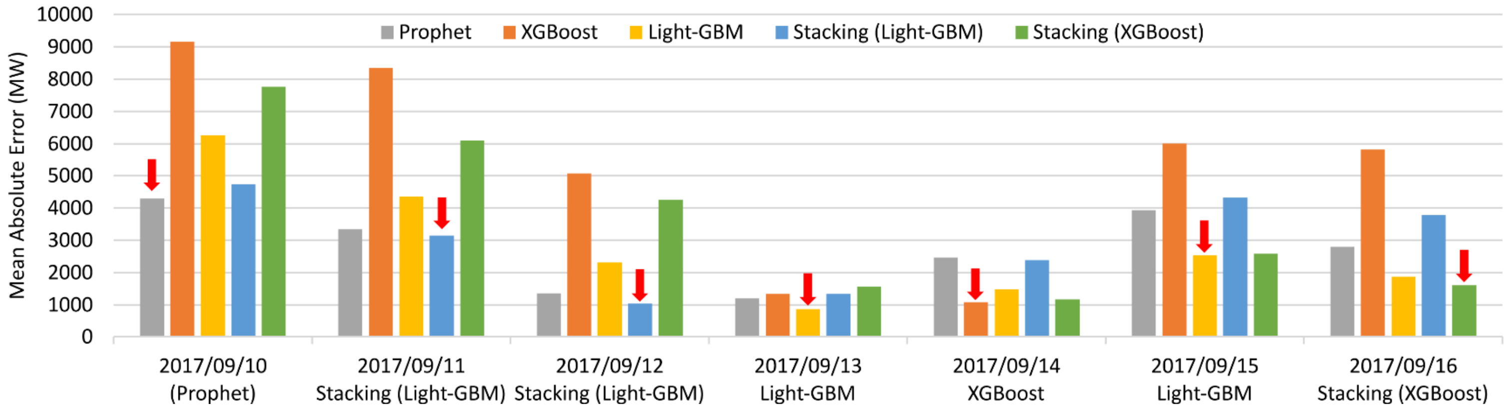


PJM Hourly Energy Consumption dataset, Kaggle

UNIVERSITY OF AMSTERDAM

Adaptive Model Selection

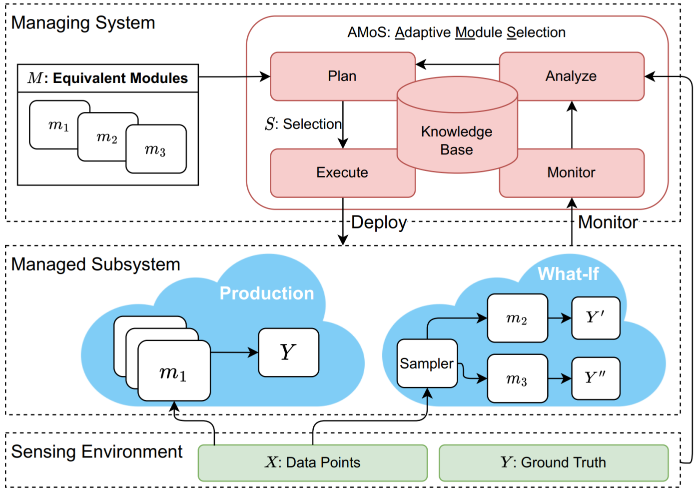


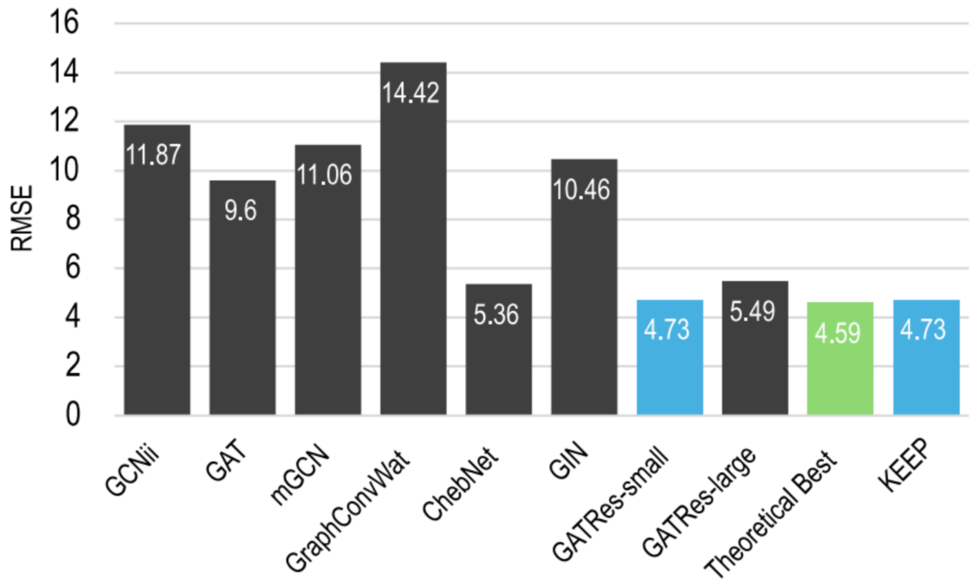


PJM Hourly Energy Consumption dataset, Kaggle

Adaptive Model Selection

- Production environment
  - Runs the best available module
  - Given the highest priority in resources
- What-If environment
  - Operates on ephemeral computing environments with fewer resources
  - A sampling sensor data is given
  - Tests a promising sub-set of modules against new conditions.
    - Learning-Based Selection (LBS) is employed to decide which modules are promising.
- Two strategies are employed:
  - Rolling Average
  - KEEP (Keeping Errors down with Enhanced Persistence)





Explainable Sensor Data

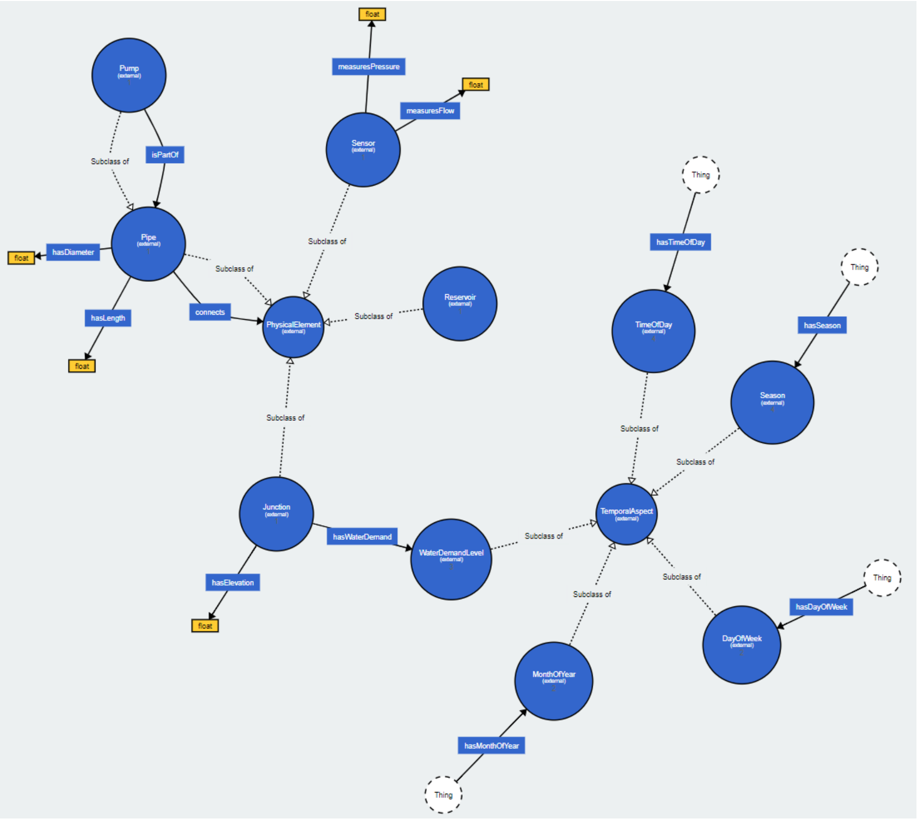
Interpret raw sensor data

- Perform profiling of system’s behavior
- Explain patterns in a clear manner (e.g. with logical notation)
- Create meaningful rules that describe system’s behavior

With a straightforward approach,
logical rules would contain:

- Raw sensors and their measurements
- Supporting entities of the system
- Features from the original ML approach

This is not very explainable!

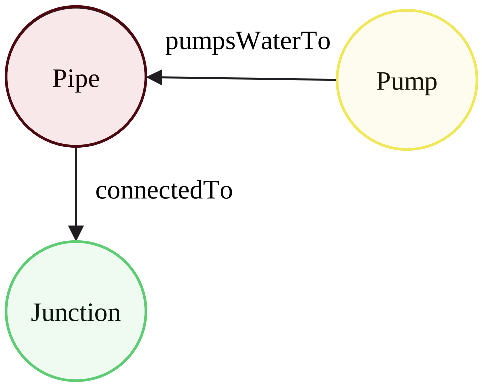


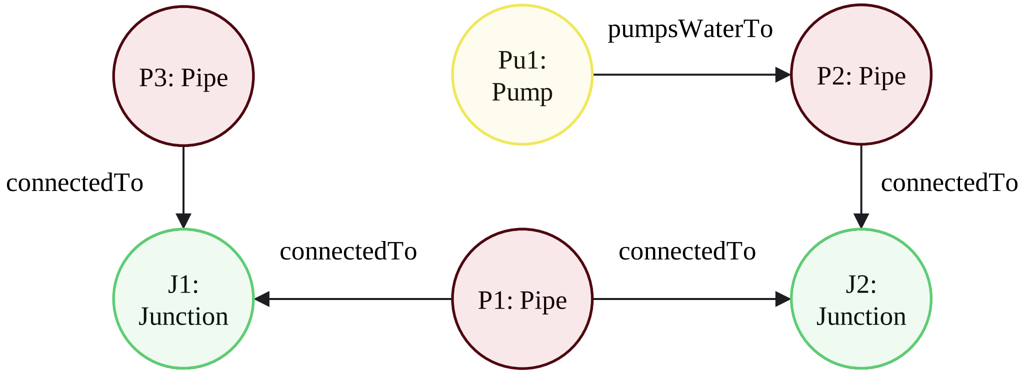
Instead: We can use a system ontology.

Ontologies and Knowledge Graphs

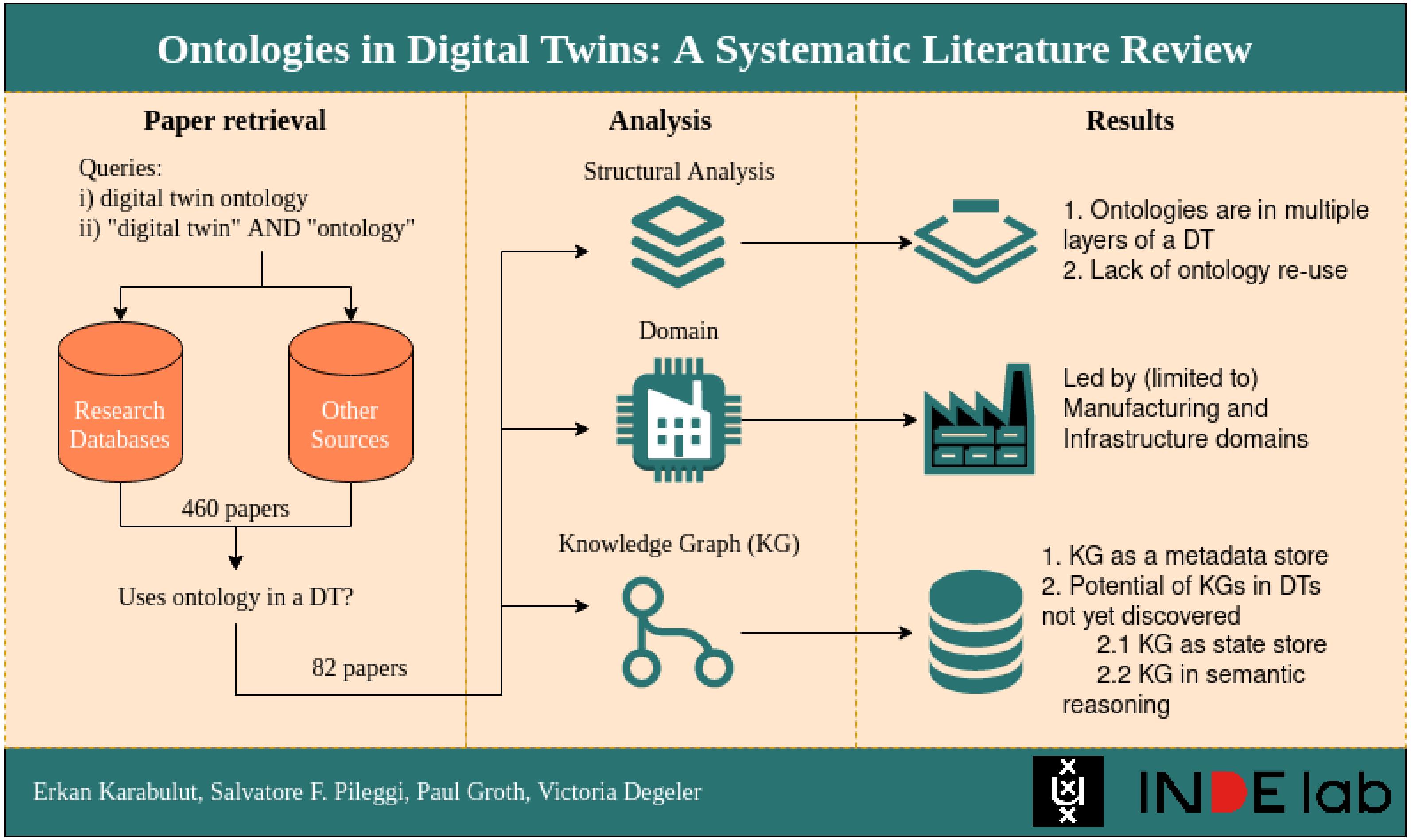
Ontology: An ontology provides a formal machine-processable and human-understandable conceptualization of a given domain.

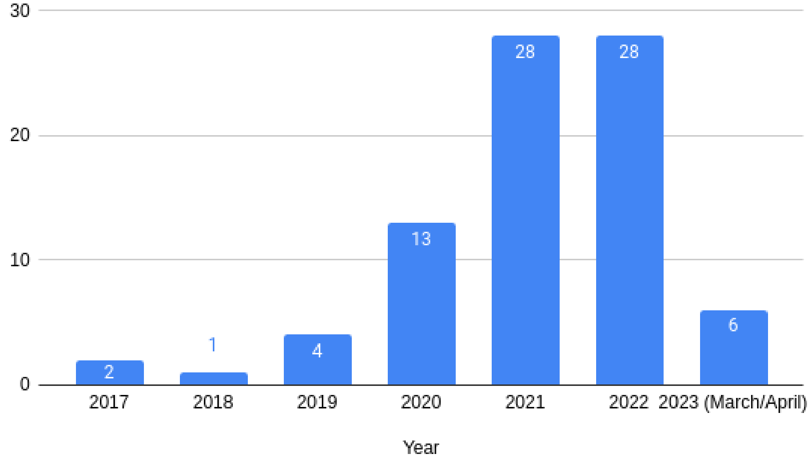
Knowledge Graph: is the data with the ontology as underlying structure.

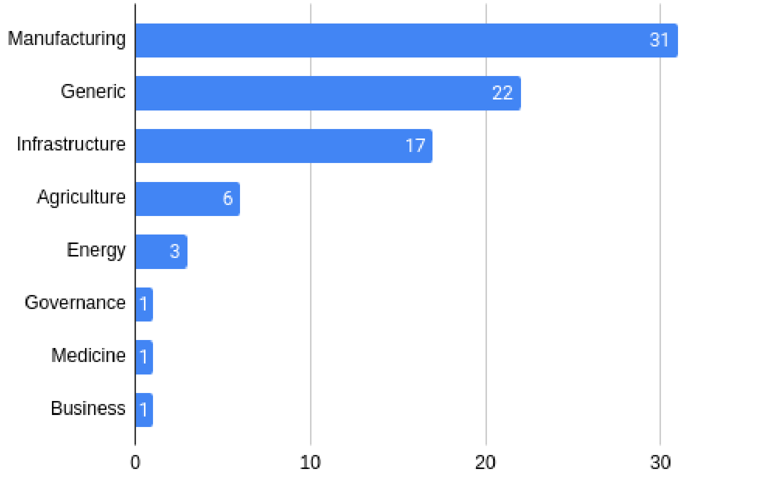




Ontologies in Digital Twins



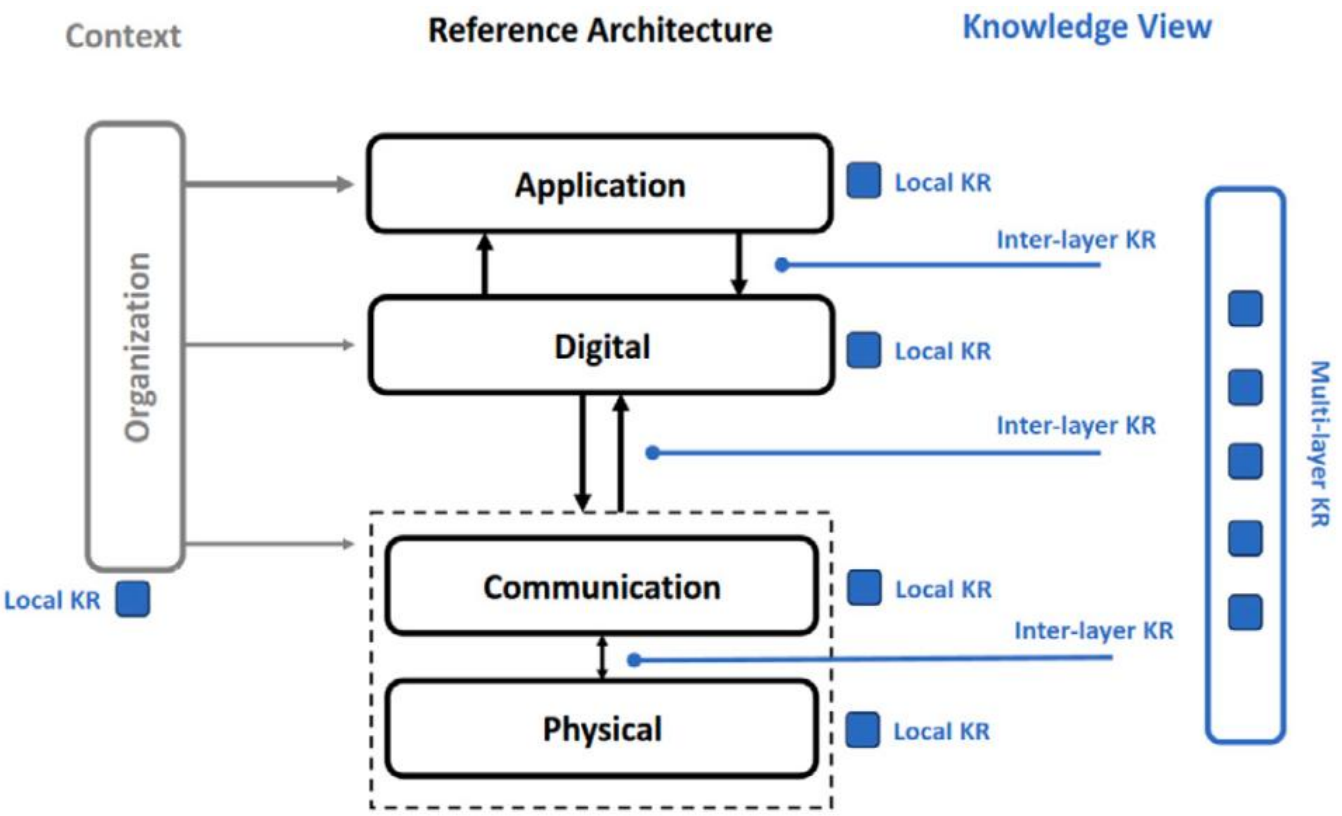




Karabulut, E., Pileggi, S. F., Groth, P., & Degeler, V. (2024). Ontologies in Digital Twins: A Systematic Literature Review. Future Generation Computer Systems, vol. 153, pp. 442-456. doi:10.1016/j.future.2023.12.013

Reference DT Architecture

Production lines,
facilities,
client and order info,
project management,
bridging DT and
non-DT parts



Ranges between task-specific terms to domain-independent application terms

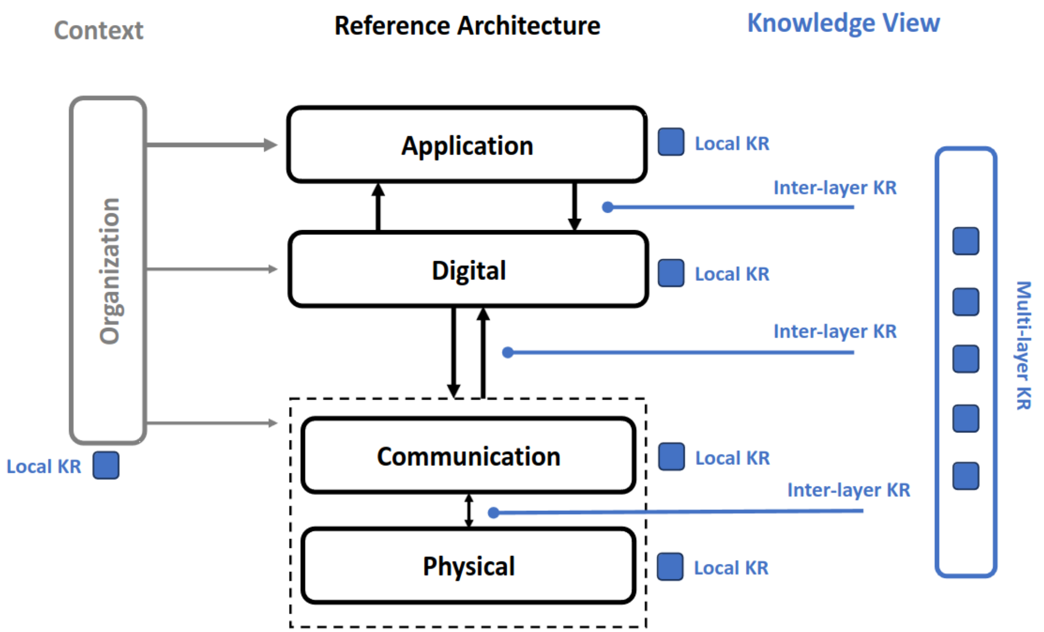
Generic DT concepts, real or abstract/derived digital terms, assets and operations

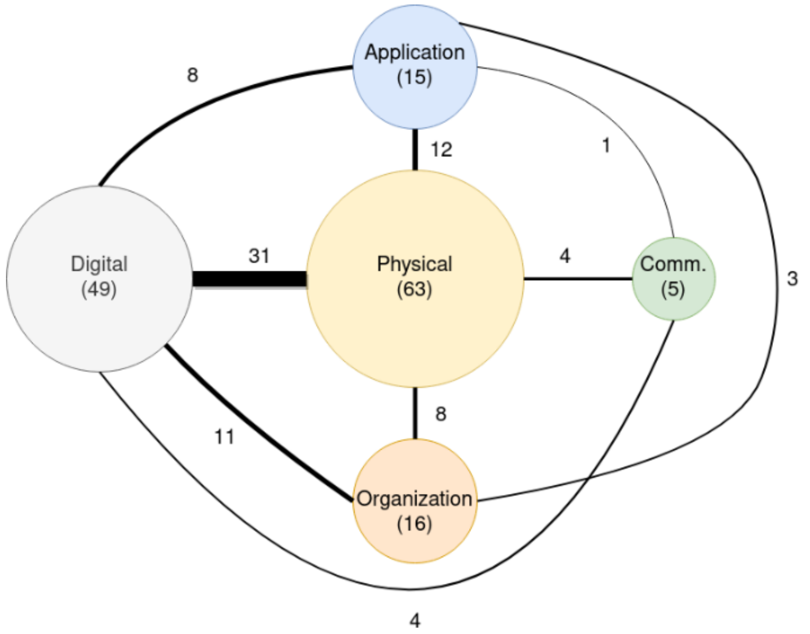
Protocols, access parameters

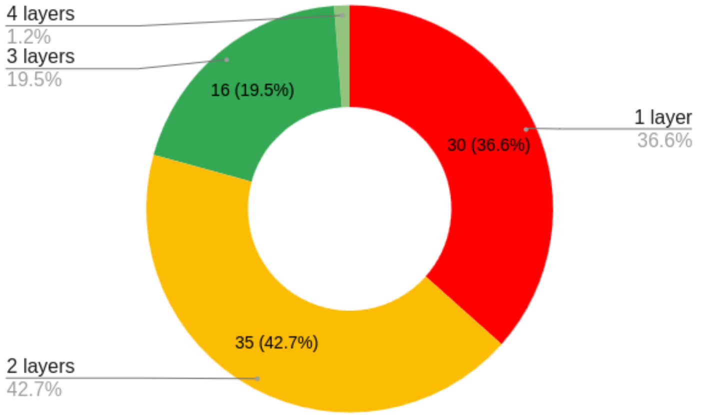
Physical entities, actions and processes

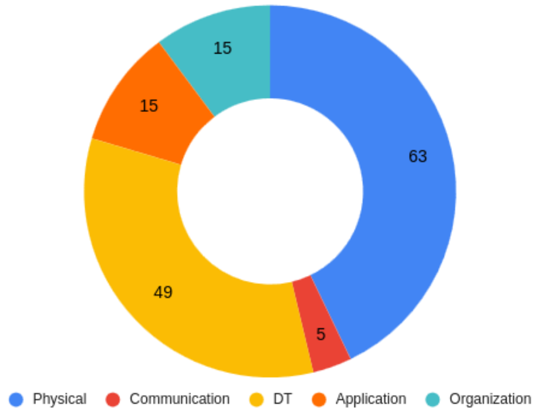
Karabulut, E., Pileggi, S. F., Groth, P., & Degeler, V. (2024). Ontologies in Digital Twins: A Systematic Literature Review. Future Generation Computer Systems, vol. 153, pp. 442-456. doi:10.1016/j.future.2023.12.013

Ontologies in Digital Twins









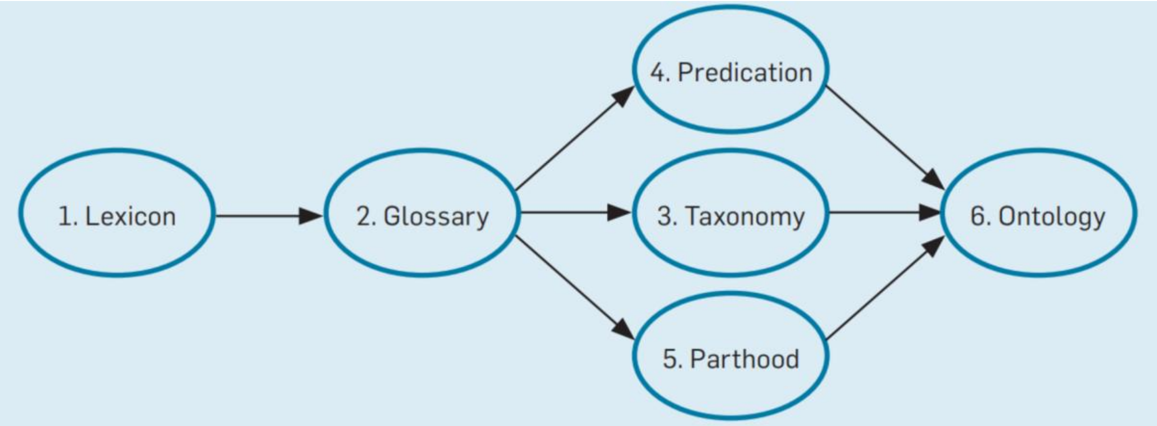
Karabulut, E., Pileggi, S. F., Groth, P., & Degeler, V. (2024). Ontologies in Digital Twins: A Systematic Literature Review. Future Generation Computer Systems, vol. 153, pp. 442-456. doi:10.1016/j.future.2023.12.013

Ontology Learning with LLMs

Traditional manual methodologies for Ontology Construction are investigated

Ontology Methodologies	Time consuming	Consistency	Structure	Descriptions of Steps	Reusability	Error Avoidance Techniques	Strategies for Identifying Concepts	Interoperability
Development 101	High	Yes	Yes	Some details	Yes	No	Developer's consent	No
METHODOLOGY	High	Yes	Yes	Sufficient details	Yes	No	Middle out strategy	No
KBSI IDEF5	High	Yes	Yes	Sufficient details	Yes	Yes	Not clear	No
KACTUS	Low	Yes	No	Insufficient details	Yes	No	Top-down strategy	No
Cyc	Low	Yes	No	Insufficient details	Yes	No	Not clear	No
ONIONS	High	Yes	Yes	Insufficient details	No	No	Not clear	Yes
SENSUS	Moderate	Yes	Yes	Some details	Yes	No	Bottom-up	Yes
UPON	Moderate	Yes	Yes	Sufficient details	Yes	Yes	Middle-out strategy	No
TOVE	Moderate	Yes	Yes	Some details	Yes	No	User-driven	No
Uschold & King	High	Yes	Yes	Some details	Yes	No	Bottom-up	No
On-To-Knowledge	Moderate	Yes	Yes	Some details	No	Yes	Middle out strategy	No
Gruninger & Fox	Moderate	Yes	Yes	Insufficient details	Yes	No	User-driven	No
CommonKADS	Low	Yes	No	Insufficient details	Yes	No	Not clear	No

UPON Life
Ontology Construction



Six Steps to Light Ontology Building

UPON Lite is organized as a sequence of

Step 1. Domain terminology. The domain

Step 2. Domain glossary. The terms of the

Step 3. Taxonomy.[11] Domain terms organi

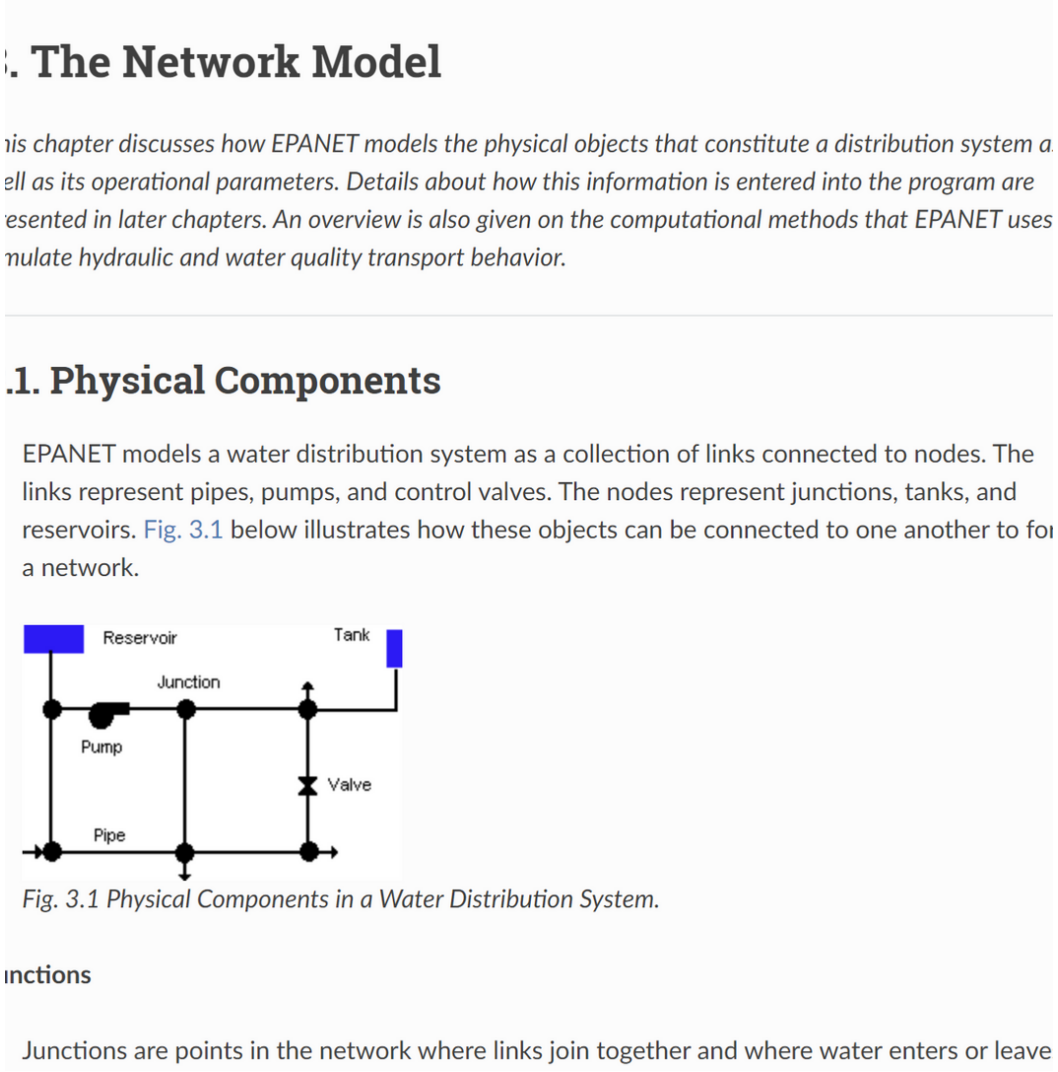
Step 4. Predication.[17] Terms representing

Step 5. Parthood (meronymy).[9] Complex

Step 6. Ontology. This last step produces t

Antonio De Nicola and Michele Missiko. 2016. A lightweight methodology for rapid ontology engineering. Communications of ACM 59, 3 (2016), 79–86.

Retrieval-Augmented Generation (RAG)



:. The Network Model

ıis chapter discusses how EPANET models the physical objects that constitute a distribution system a ell as its operational parameters. Details about how this information is entered into the program are esented in later chapters. An overview is also given on the computational methods that EPANET uses nulate hydraulic and water quality transport behavior.

.1. Physical Components

EPANET models a water distribution system as a collection of links connected to nodes. The links represent pipes, pumps, and control valves. The nodes represent junctions, tanks, and reservoirs. Fig. 3.1 below illustrates how these objects can be connected to one another to foı a network.

Fig. 3.1 Physical Components in a Water Distribution System.

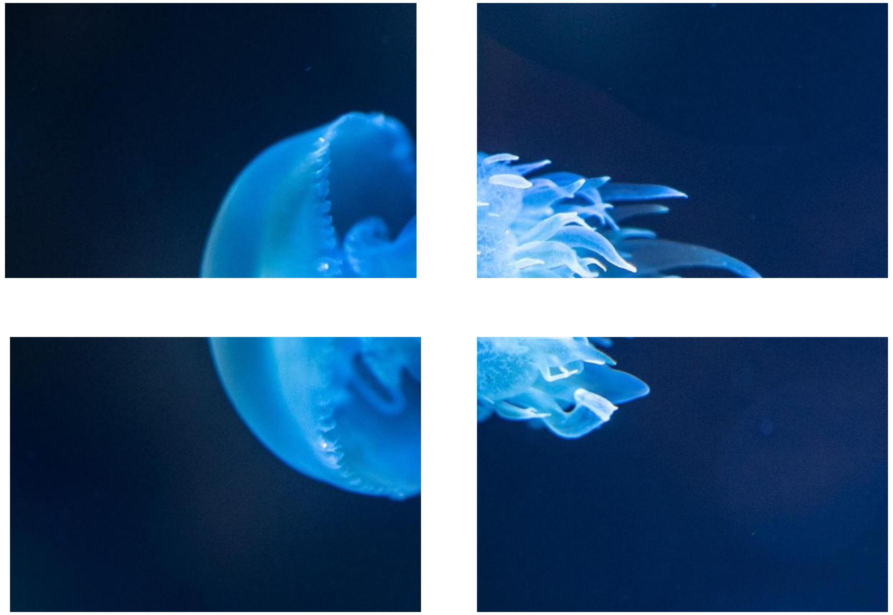
ınctions

Junctions are points in the network where links join together and where water enters or leave

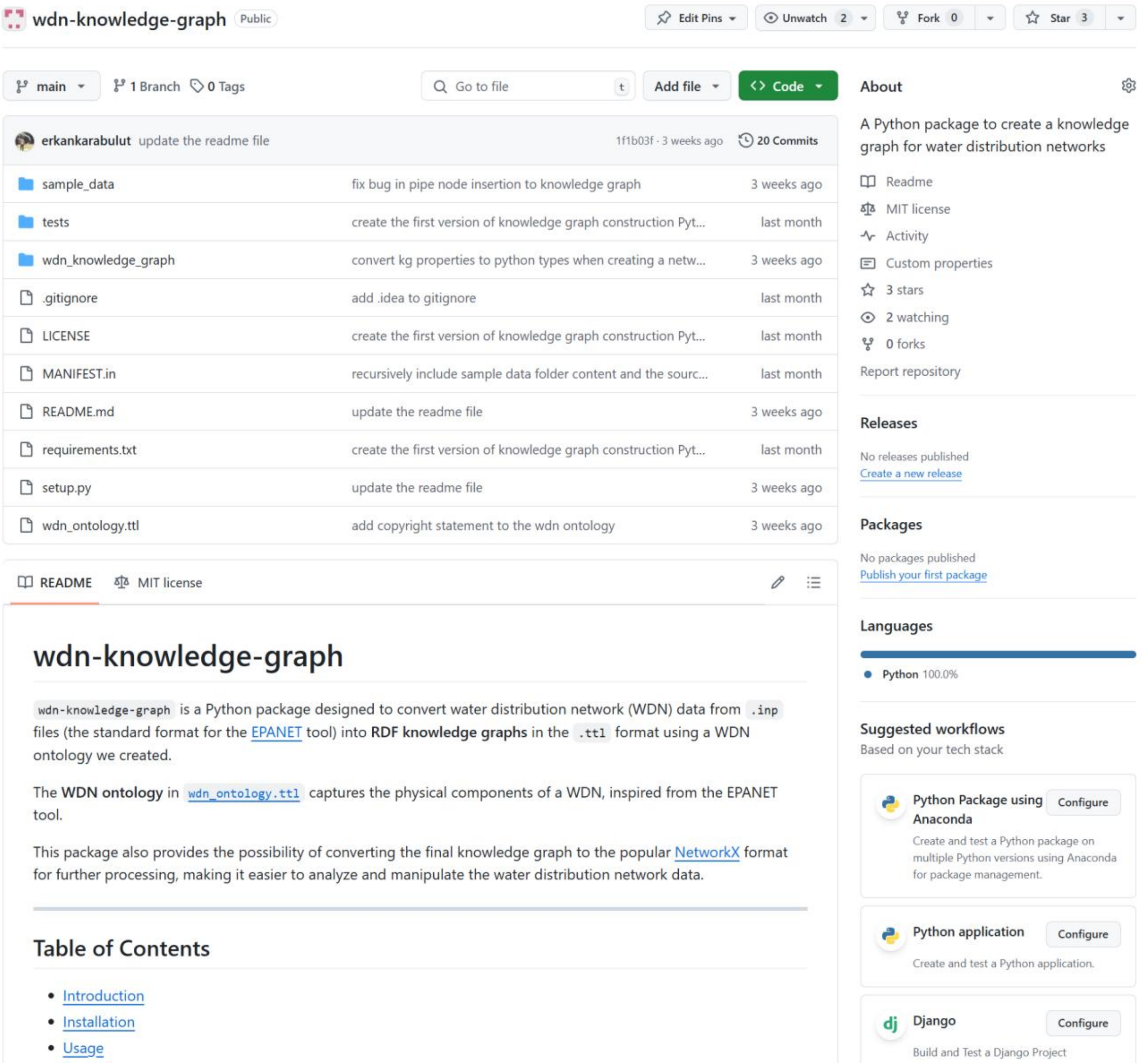
Retrieval-Augmented Generation (RAG) incorporates external documents into the context of a query to mitigate LLM hallucinations

Input document

The Network Model from EPANET

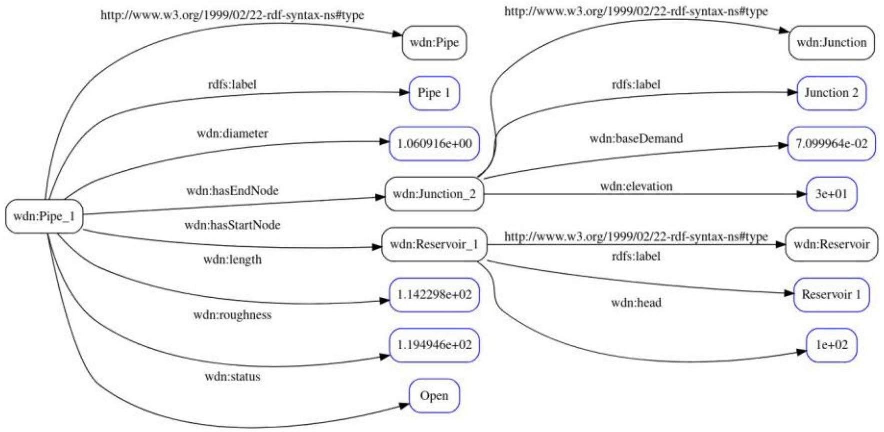


WDN Knowledge Graph



Converts EPANET *.inp files into RDF (Resource Description Framework), to work with semantic graph data

Sample output from Pipe_1 of LeakDB

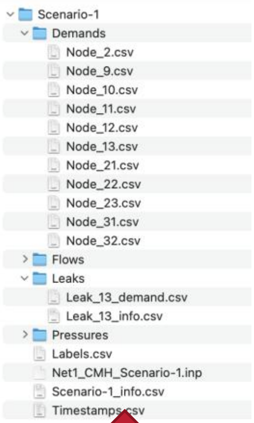


Rule learning from sensor data: finding missing connection pieces with knowledge graphs

IoT Sensor Network

Source: https://www.sensorsexpo.com/sites/sensorsexpo/files/til_image/iotecosystem_icon_final_0.png

Raw sensor data



Ontologies and Knowledge Graphs

ConceptNet
An open, multilingual knowledge graph

BabelNet

yago select knowledge

Microsoft Concept Graph

WIKIDATA

Entities and relationships

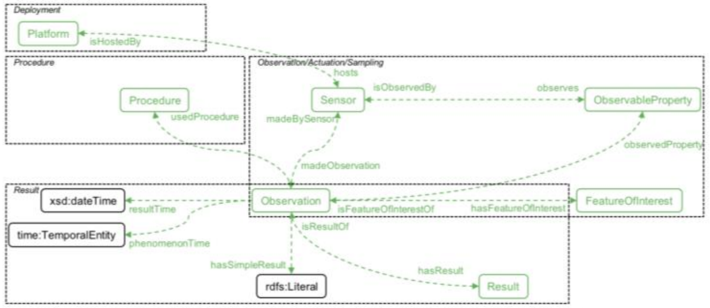


Figure 3 Overview of the SOSA classes and properties (observation perspective)

Automated environmental rules extraction

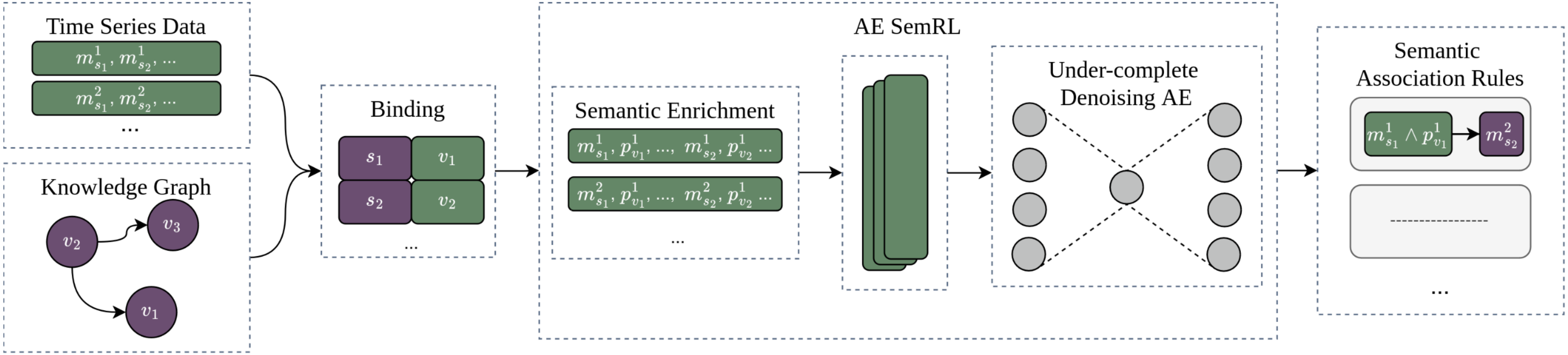
No semantics:

if *sensor1* measures *V1*, then *sensor2* measures *V2*

Semantically-enriched:

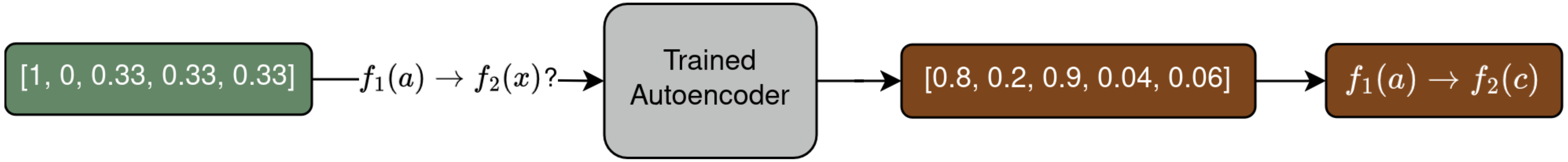
if a *flow sensor* placed in a *pipe P1* with *length > L1* measures *V1*, then a *pressure sensor* that is placed in a *junction J1* connected to *P1* measures *V2*

Semantic Rule Learning



Example:

$f_1 = \{a, b\}, f_2 = \{c, d, e\}, threshold = 0.8$



E. Karabulut, V. Degeler, P. Groth "Neurosymbolic Association Rule Mining from Tabular Data", preprint: arXiv:2504.19354

Downstream classification tasks

Table 5: Rule-based interpretable ML models with Aerial+ achieve higher or comparable accuracy with significantly lower execution time. Bold indicates the highest performance.

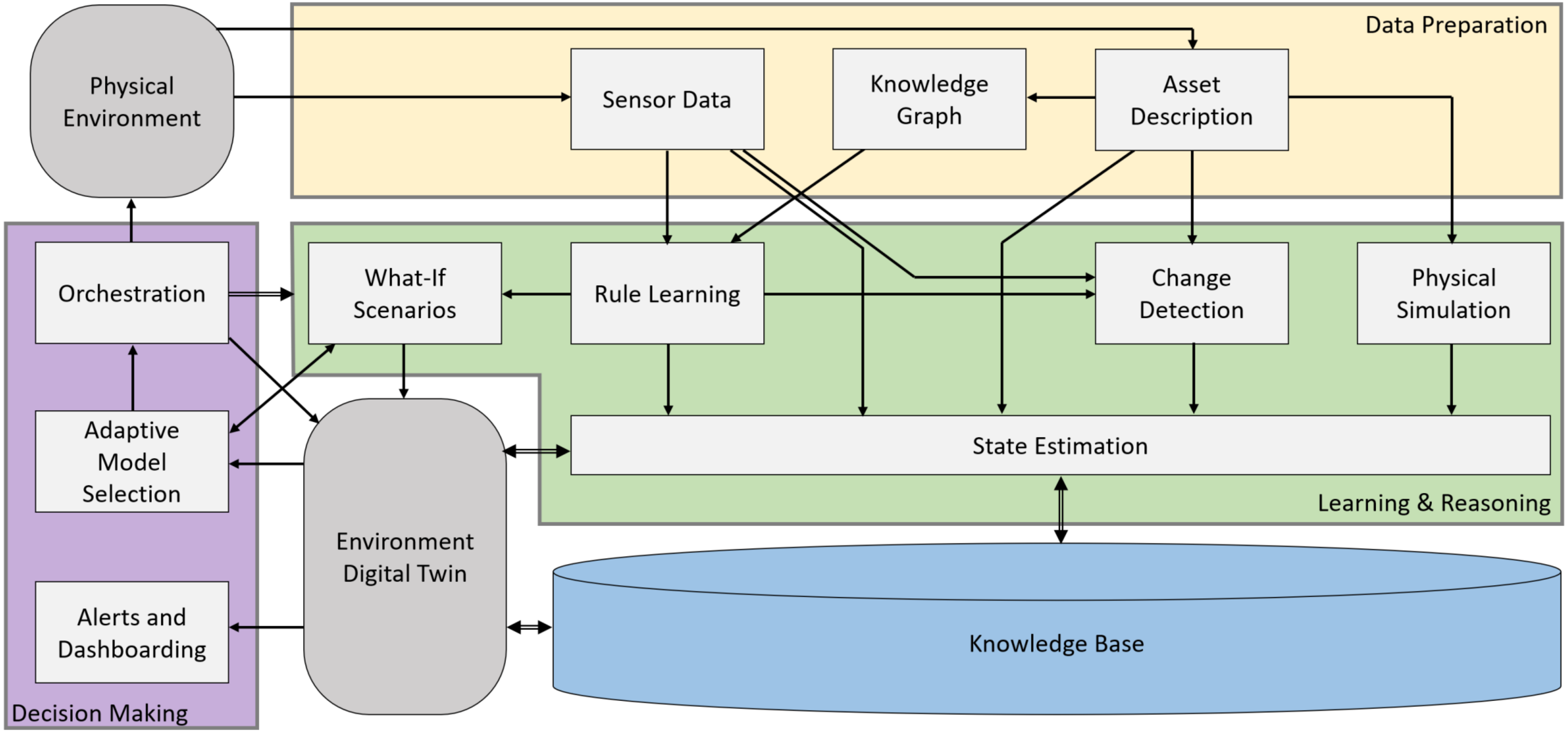
Dataset	Algorithm	# Rules or Items	Accuracy	Exec. Time (s)
		Exhaustive | Aerial+	Exhaustive | Aerial+	Exhaustive | Aerial+
Congressional Voting Records	CBA	3437 | **1495**	91.91 | **92.66**	0.34 | **0.14**
	BRL	2547 | **57**	**96.97** | **96.97**	15.37 | **9.69**
	CORELS	4553 | **61**	**96.97** | **96.97**	3.04 | **0.17**
Mushroom	CBA	27800 | **2785**	**99.82** | **99.82**	1.75 | **1.30**
	BRL	5093 | **493**	**99.87** | 99.82	244 | **167**
	CORELS	23271 | **335**	90.14 | **99.04**	61 | **2**
Breast Cancer	CBA	695 | **601**	66.42 | **71.13**	**0.08** | 0.28
	BRL	2047 | **290**	71.13 | **71.46**	16.82 | **14.5**
	CORELS	2047 | **369**	73.69 | **75.82**	1.42 | **0.40**
Chess	CBA	49775 | **34490**	**94.02** | 93.86	24.31 | **6.24**
	BRL	19312 | **1518**	**96.21** | 95.93	321 | **119**
	CORELS	37104 | **837**	81.1 | **93.71**	106 | **3.87**
Spambase	CBA	125223 | **33418**	84.5 | **85.42**	23.87 | **7.56**
	BRL	37626 | **5190**	72.78 | **84.93**	1169 | **431**
	CORELS	275003 | **1409**	85.37 | **87.28**	1258 | **5.23**

Effect of adding semantic information on rules

Table 5. Comparison of ARM on sensor data with semantics (w-s, our pipeline) and without (wo-s), showing a significant increase in support and rule coverage (cov.) with semantics (FP-G = FP-Growth, Conf = Confidence).

	# Rules	**Support**	**Cov.**	**Conf.**
	w-s|wo-s	w-s|wo-s	w-s|wo-s	w-s|wo-s
		LeakDB		
FP-G	103K|9K	**0.41**|0.19	**0.43**|0.2	0.95|0.97
Aerial	554|2.5K	**0.54**|0.25	**0.59**|0.3	0.91|0.87
		L-Town		
FP-G	25K|5K	**0.86**|0.36	**0.9**|0.38	0.96|0.96
Aerial	1K|2.5K	**0.59**|0.39	**0.65**|0.45	0.91|0.88
		LBNL		
FP-G	7K|2K	**0.84**|0.73	**0.85**|0.75	0.98|0.99
Aerial	73|258	**0.74**|0.65	**0.74**|0.66	1.0|0.99

Digital Twin Architecture



V. Degeler, M. Hadadian, E. Karabulut, A. Lazovik, H. van het Loo, A. Tello, H. Truong (2024) DiTEC: Digital Twin for Evolutionary Changes in Water Distribution Networks. Leveraging Applications of Formal Methods, Verification and Validation. Application Areas. ISoLA 2024. Lecture Notes in Computer Science, vol 15223. Springer. doi: 10.1007/978-3-031-75390-9_5

Q&A

- V. Degeler, M. Hadadian, E. Karabulut, A. Lazovik, H. van het Loo, A. Tello, H. Truong (2024) DiTEC: Digital Twin for Evolutionary Changes in Water Distribution Networks. Leveraging Applications of Formal Methods, Verification and Validation. Application Areas. ISoLA 2024. Lecture Notes in Computer Science, vol 15223. Springer. doi: 10.1007/978-3-031-75390-9_5
- H. Truong, A. Tello, A. Lazovik, V. Degeler (2024) Graph Neural Networks for Pressure Estimation in Water Distribution Systems, Water Resources Research, 60, e2023WR036741. doi: 10.1029/2023WR036741
- H. Truong, A. Tello, A. Lazovik, V. Degeler (2025) DiTEC-WDN: A Large-Scale Dataset of Hydraulic Scenarios across Multiple Water Distribution Networks, arxiv preprint: 10.48550/arXiv.2503.17167
- Karabulut, E., Pileggi, S. F., Groth, P., & Degeler, V. (2024). Ontologies in Digital Twins: A Systematic Literature Review. Future Generation Computer Systems, vol. 153, pp. 442-456. doi: 10.1016/j.future.2023.12.013
- E. Karabulut, P. Groth, V. Degeler (2025) "Neurosymbolic Association Rule Mining from Tabular Data", preprint: 10.48550/arXiv.2504.19354
- M. H. Nejad Yousefi, V. Degeler and A. Lazovik, "Self-Adaptive Service Selection for Machine Learning Continuous Delivery," 2024 IEEE International Conference on Web Services (ICWS), Shenzhen, China, 2024, pp. 1048-1056, doi: 10.1109/ICWS62655.2024.00123