University of Naples Federico II
Civil, Architectural and Environmental Engineering Dept.

University of Thessaly
Civil Engineering Dept.

5th International Conference EWaS
12-15 July 2022
"Centro Congressi" University of Naples Federico II
Via Partenope 36, 80121 Naples (IT)

Simulation Framework for Pipe Failure Detection and Replacement Scheduling Optimization

Panagiotis Dimas [1]; Dionysios Nikolopoulos [1]; Christos Makropoulos [1]

[1]National Technical University of Athens (Greece)

Urban Water Management and Hydroinformatics Group (UWMH)

REGIONE CAMPANIA
COMUNE DI NAPOLI
CONFINDUSTRIA Campania
ANCE CAMPANIA
ASSOCIAZIONE IDROTECNICA ITALIANA
E.N.T.D. - ENTE NAZIONALE PER LA TRASFORMAZIONE DIGITALE
cea

Structure of the presentation

The presentation is structured in distinct units as follows:

1. Posing the question: How to predict which pipes to replace?
2. Answering the question
3. Case studies: Mourati zone & C-Town
4. C-Town model: expanding its pipe characteristics database
5. Training the pipe failure probability prediction models
6. Pipe replacement methodology
7. Performance assessment
8. Results
9. Conclusions

Posing the question: How to predict which pipes to replace?

Optimal replacement strategies must ensure:

1. Technical & Socioeconomic specs.
2. Overall resilience of the Urban Water System.
3. Low associated costs.

The usual practices include:

- Pipeline forecasting.
- Identification of water leaks.
- Identification of pipeline ruptures.

Advantages of the traditional methods:

1. Easy to operate.
2. Zero knowledge of the pipe fracture mechanism.

Disadvantage: Depend on specifics of the studied area.

The internal components are hard to identify and remain obscure to the stakeholders.

Traditional methods

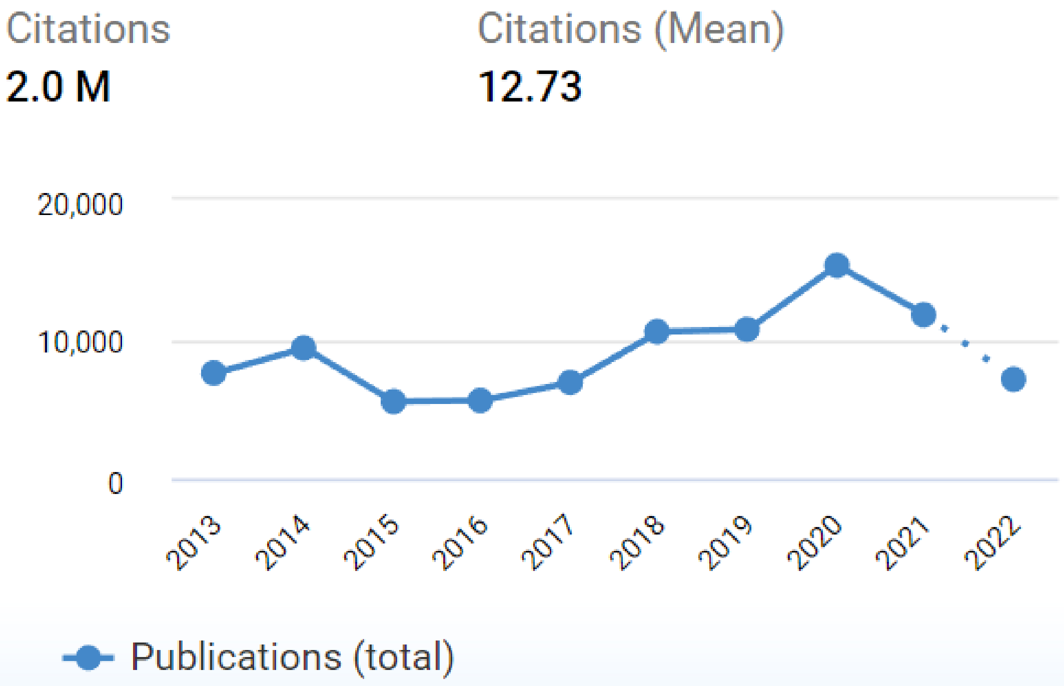
- Statistical models
- Genetic programming
- Data-driven techniques
- Data mining prediction systems

Machine Learning & AI

- ML imputation methods combined with survival analysis
- Foretelling water mains with increased breaking likelihood
- Fast and cost effective models

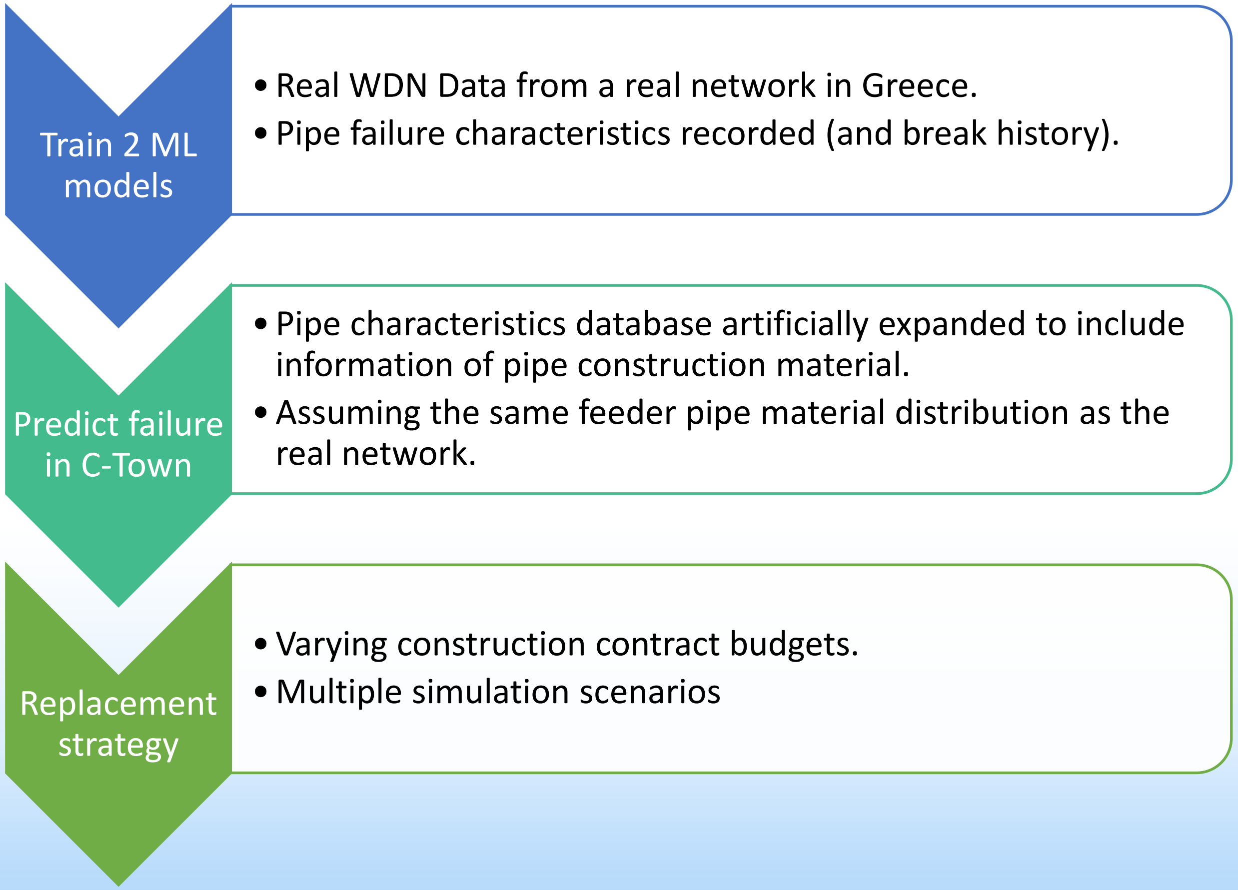
Answering the question

- The number of ML techniques in pipe failure detection has increased during the last years.
- Only **few of them** are utilizing data generated from **Water Distribution Networks (WDN) simulations**.
- **Even fewer are operating under complex water usage conditions.**

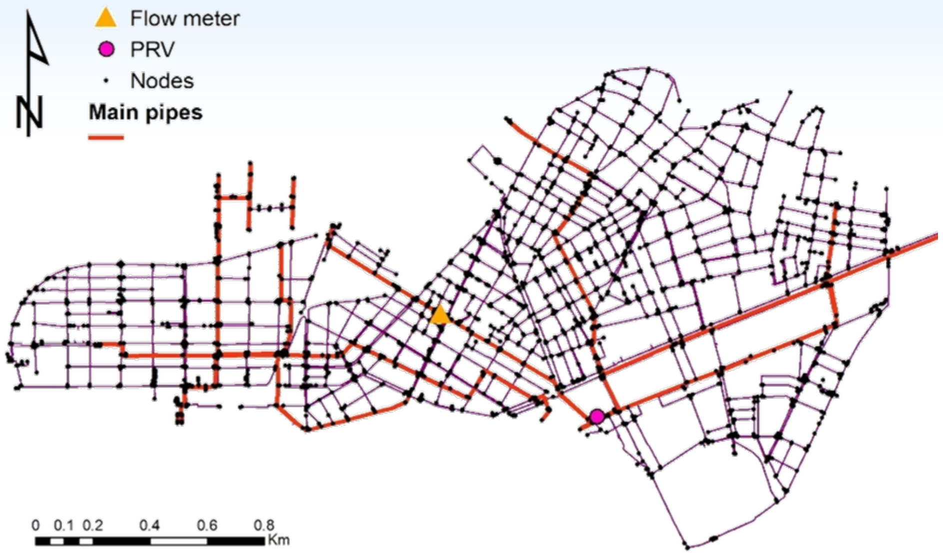


Results returned from the Dimensions AI tool by entering the keywords ML & pipe failure
https://app.dimensions.ai/discover/publication

How did the current work address the issue?



Case studies: Mourati zone & C-Town



- A **DMA** serving ~ 3km^2 and multiple uses.
- **1640 pipes** of external diameters ranging from Ø50 to Ø900.
- **Materials:** *(i)* gray cast iron, *(ii)* asbestos cement, *(iii)* galvanized steel, *(iv)* straight seam steel, *(v)* PVC and *(vi)* polyethylene MRS100 (encoded in the range [1, 6]).
- Main pipes are **mainly manufactured from *i*, *iv* and *ii*.**

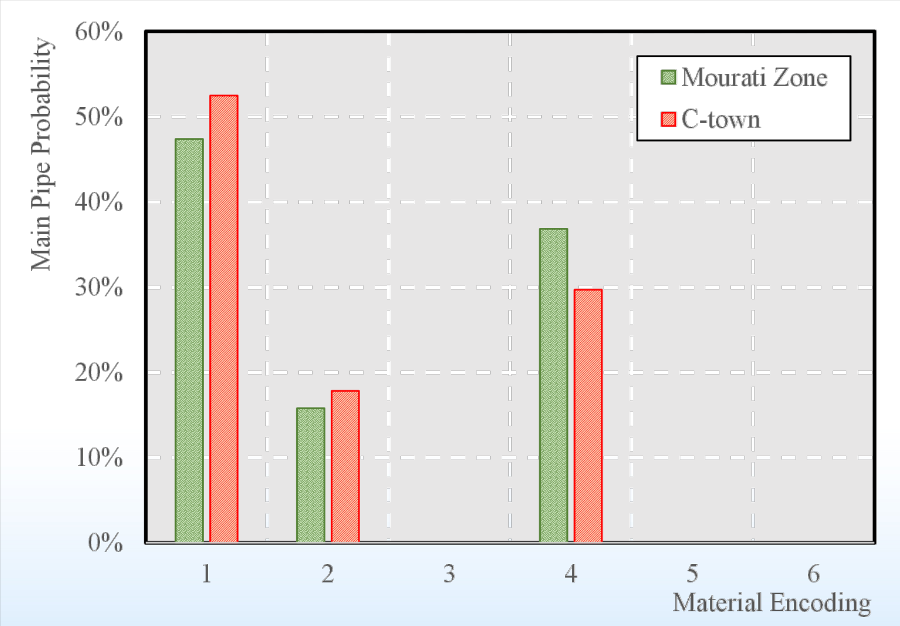
- Based on a real-world medium sized network.
- **1 reservoir, 7 tanks, 388 demand junctions, 429 pipes, 11 head pumps and 4 valves** [3 pressure relief valves (PRV), 1 flow control valve (FCV)].
- The C-town model **does not include any pipe material information**.

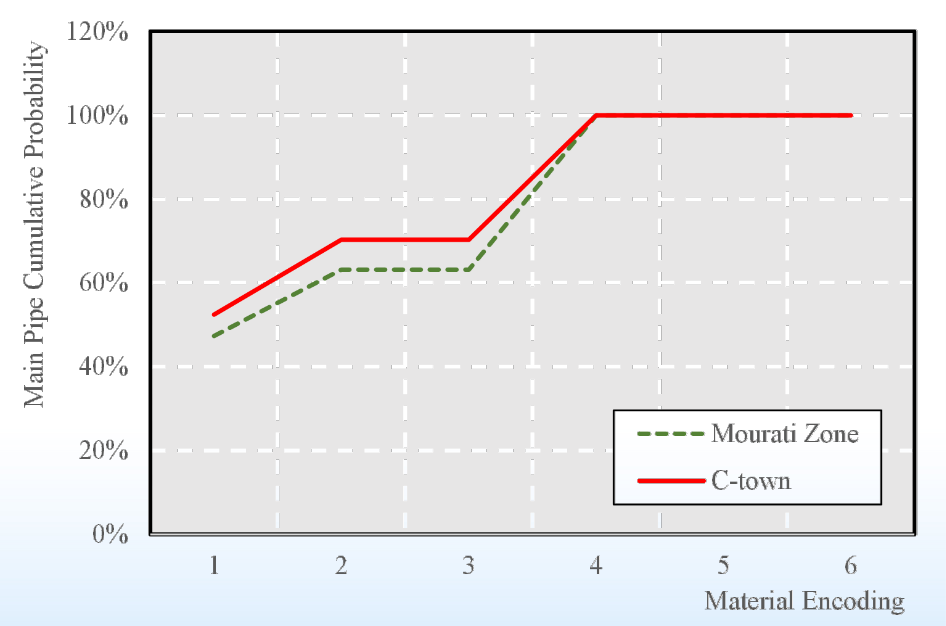
C-Town model: expanding its pipe characteristics database

- The C-town model **does not include any pipe material information**.

It is assumed that:

1. **The main pipes** are following the **same materials distribution** as the ones of Mourati zone**.**
2. **The secondary & tertiary pipes** are assigned to a material by a **random choice** of uniform selection.





Training the pipe failure probability prediction models (1/3)

Two models are developed and trained using Scikit-learn in Python:

- Regression model based on k-nearest neighbors (**kNN-R**).
- The Decision Tree regression model (**DTR**).

The dataset is including:

1. **Numerical features:**

   (i) the pipe material encoding;

   (ii) the main flagging (assigned a value of 1 if the pipe is a main one, or 0 if the pipe is secondary or tertiary);

   (iii) the pipe external diameter;

   (iv) the pipe length.

2. **The label:**

   Indicates if the corresponding pipe is faulty (broken) or not (with values of 1 and 0).

Training the pipe failure probability prediction models (2/3)

The percentage of pipes that failed is only a small subset of the dataset **(~ 6%).**

Unbalanced dataset: most of the ML models could ignore such results or **exhibit poor performance** in the minority class

What can we do?

Oversample the minority class to create a balanced dataset.

Synthetic Minority Oversampling Technique (SMOTE)

Advantages

The new synthetic examples from the minority class are plausible, i.e., relatively close to real examples!

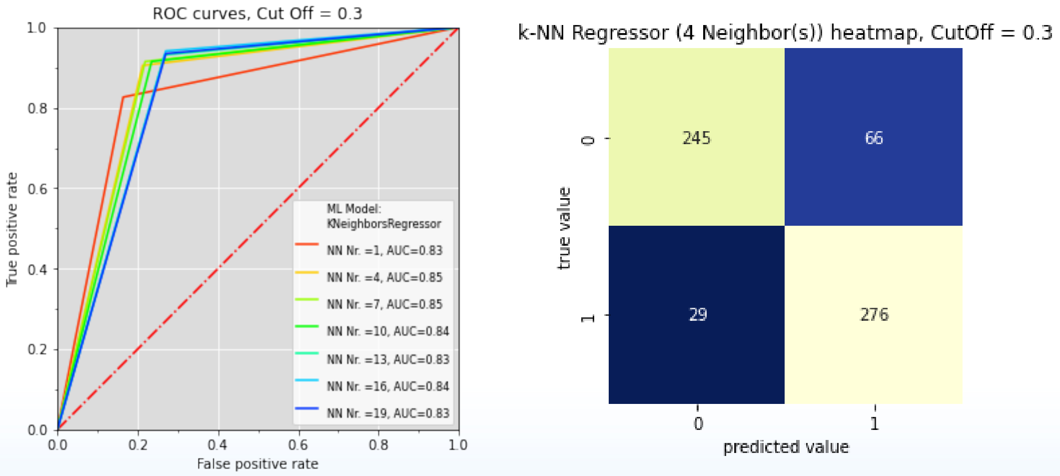
Disadvantages

Synthetic examples are created without considering the majority class →

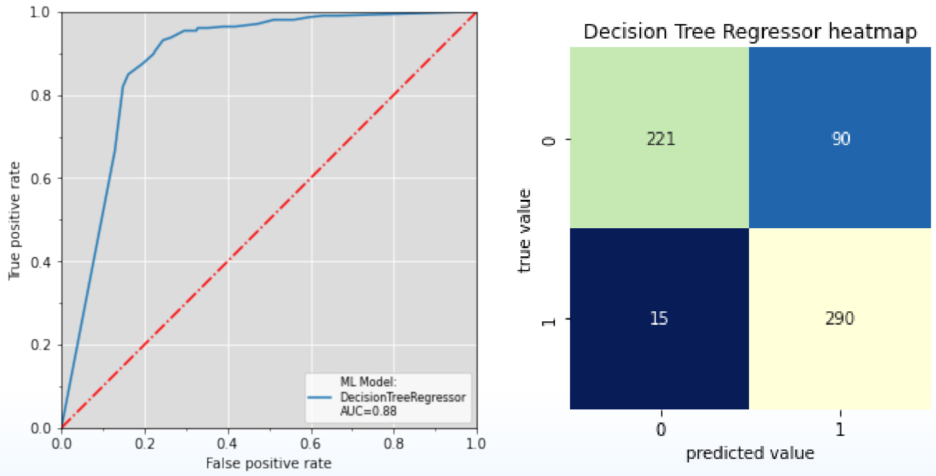
Ambiguous examples if there is a strong overlap between the classes

Training the pipe failure probability prediction models (3/3)

- The two ML models (**kNN-R, DTR**) are trained in the oversampled dataset.
- **The two models are regressive ones: they produce real numbers (i.e., failure probabilities).**
- To make them **binary (e.g. translate the probability to failure or non-failure)** we choose a **threshold (cut-off value).**
- **It is treated as a hyperparameter and tuned accordingly!**



kNN-R



DTR

Pipe replacement methodology (1/2)

Strategy conceptualization:

- The **available budget** for pipe replacements is distributed **amongst a 5-year construction contract**.
- The contract is allocated to **annual sub-contracts**.
- A **proxy metric** of **pipe replacement cost** is utilized (where *L* the length and *D* the diameter).

$$C_i = L_i {D_i}^2$$

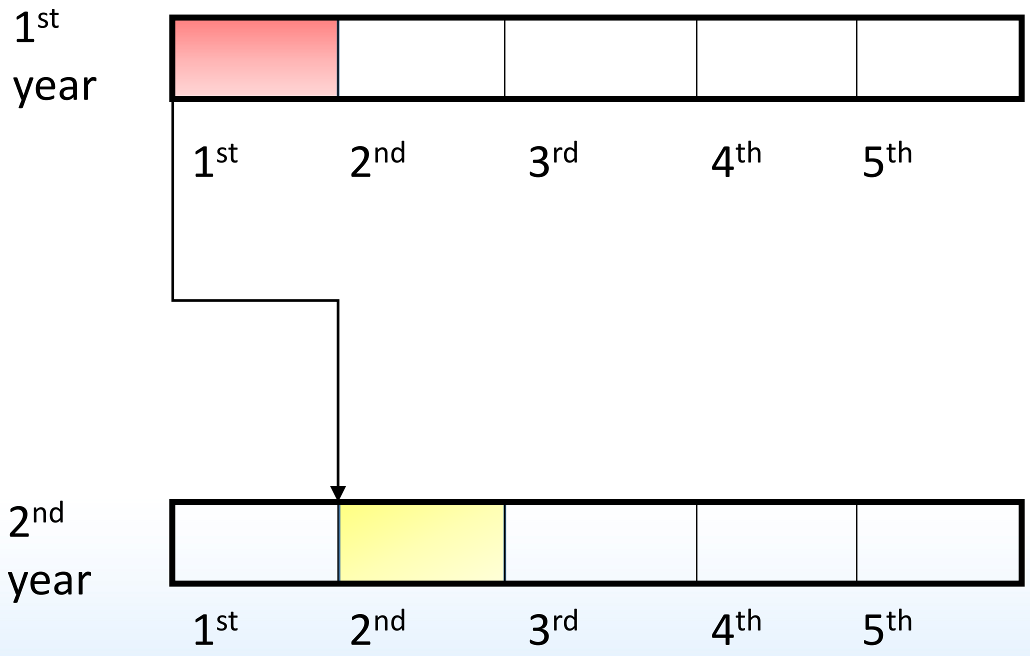
- The **available budget for the contract** is assumed to be a proportion (***p***), i.e., **10%, 15% or 20%,** of the total replacement cost ($\sum C$) for all WDN's pipes → **three contract cost levels are produced.**
- Pipes of the WDN are **sorted** with their **failure probability (as predicted by the 2 ML models)**.
- In total we get **6 sets** of annual schedules (**2 models x 3 contract cost levels**).

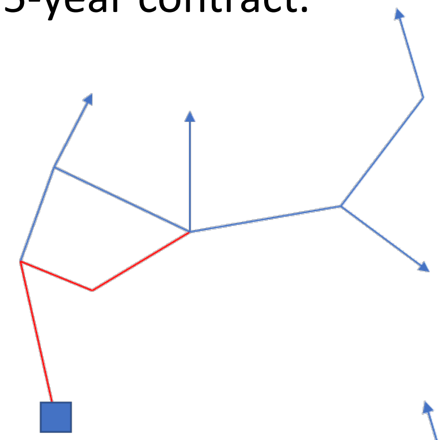
kNN-R	DTR
10%	10%
15%	15%
20%	20%

Pipe replacement methodology (2/2)

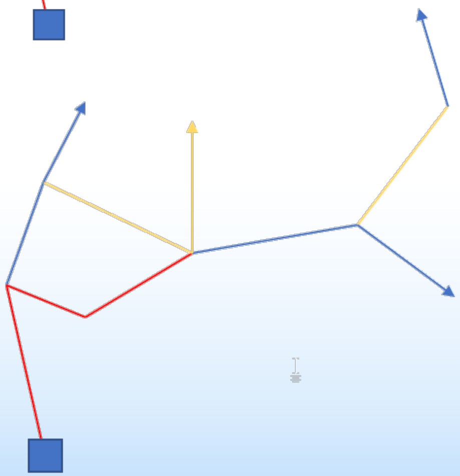
Annual schedule:

- Formed at the **start of the year** from the **set of pipes** that **accumulate the annual construction budget**.
- **They are replaced** with new ones.
- The new pipes are **assumed failure-proof** till the end of the 5-year contract.





$$\sum C_i = \frac{p(\%)\ \sum C_{total}}{5}$$



$$\sum C_i = \frac{p(\%)\ \sum C_{total}}{5}$$

Performance assessment

A Monte Carlo scheme regarding pipe failures is employed (simulations in EPANET):

- A **daily pipe failure probability** is assigned to the WDN $F = 0.0005 \frac{L}{1000} \sqrt{\frac{D}{D_{min}}}$
- An ensemble of **100 realizations of the WDN hydraulic simulation**, with a duration of 1825 days, is formed.
- For **each day and each pipe**, a random probability (p_r) of non-exceedance from the uniform distribution is generated.
- If $p_r > F$ the **pipe brakes → An emitter is introduced at the breaking pipe (with WNTR Python package).**
  - The **emitter flowrate (q)** is calculated from **node's pressure (p)** and a **burst coefficient (b).**
  - The emitter is introduced by **splitting the pipe in two parts of equal length**.

Emitter flowrate $q = b\sqrt{p}$ *Node's pressure*

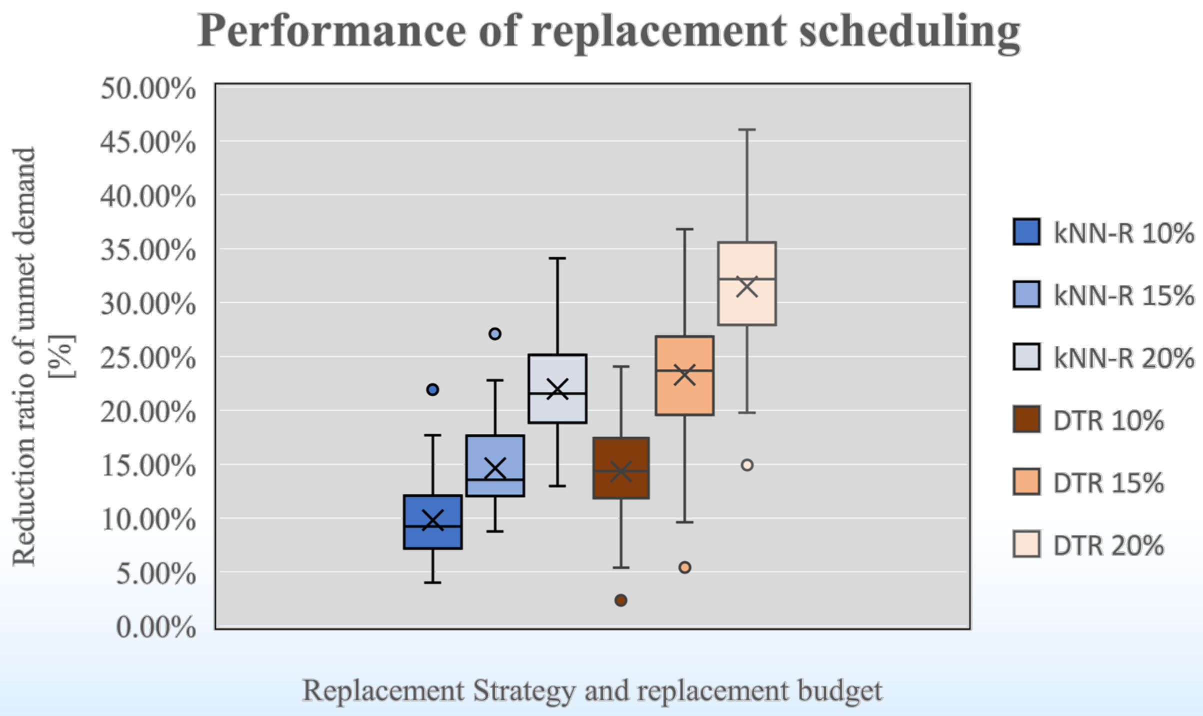
Burst coefficient $b = 0.75A\sqrt{\frac{2}{\rho}}$ *Cross-section area* *Water density*

For the whole simulation period (i.e., all instances of bursts), this **set of pipe failures** tallies to a **total unmet demand.**

- If pipes that burst **are in the replacement schedule** and the replacement **has been already applied at the specific timestep of burst**, the unmet demand that occurs from this burst is tallied to **another variable**, i.e., **unmet demand reduction.**

Results

- The performance of each realization is the **ratio of *unmet demand with the scheduling of pipe replacements* versus *unmet demand***, i.e., the **reduction ratio** of unmet demand.
- After assessing **performance of the whole ensemble of realizations**, for each of the **six discrete sets of annual schedules** we compare the results.



Remarks:

- In the **truly random replacement** of pipes in the network the expected outcome would follow the budget level; **replacing the 10% of pipes** would -more or less- result in **10% unmet demand reduction** ratio compared to no replacements.
- The **kNN-R** strategy (albeit in the limited pool of 100 realizations) seems to offer **no benefit** compared to a random replacement schedule.
- There is an added value of using the **DTR** strategy, as it systematically offers better than randomly expected performance.

Conclusions

1. A novel **coupling of ML prediction models** for pipe bursts in WDNs, with **strategies of replacement,** is presented**.**
2. We assess their performance with **Monte Carlo hydraulic simulations in EPANET**, to address uncertainty.
3. The methodology is demonstrated **by training the ML models in a real-world system** (lacking demand supply and control data).
4. Then the **replacement strategies are assessed** in a synthetic WDN.
5. It is an **example case** far from a real-world system.
6. Acts as an **early schematic prototype** of a promising methodology.

University of Naples Federico II
Civil, Architectural and Environmental Engineering Dept.

University of Thessaly
Civil Engineering Dept.

Thank you!

pn.dimas@gmail.com