

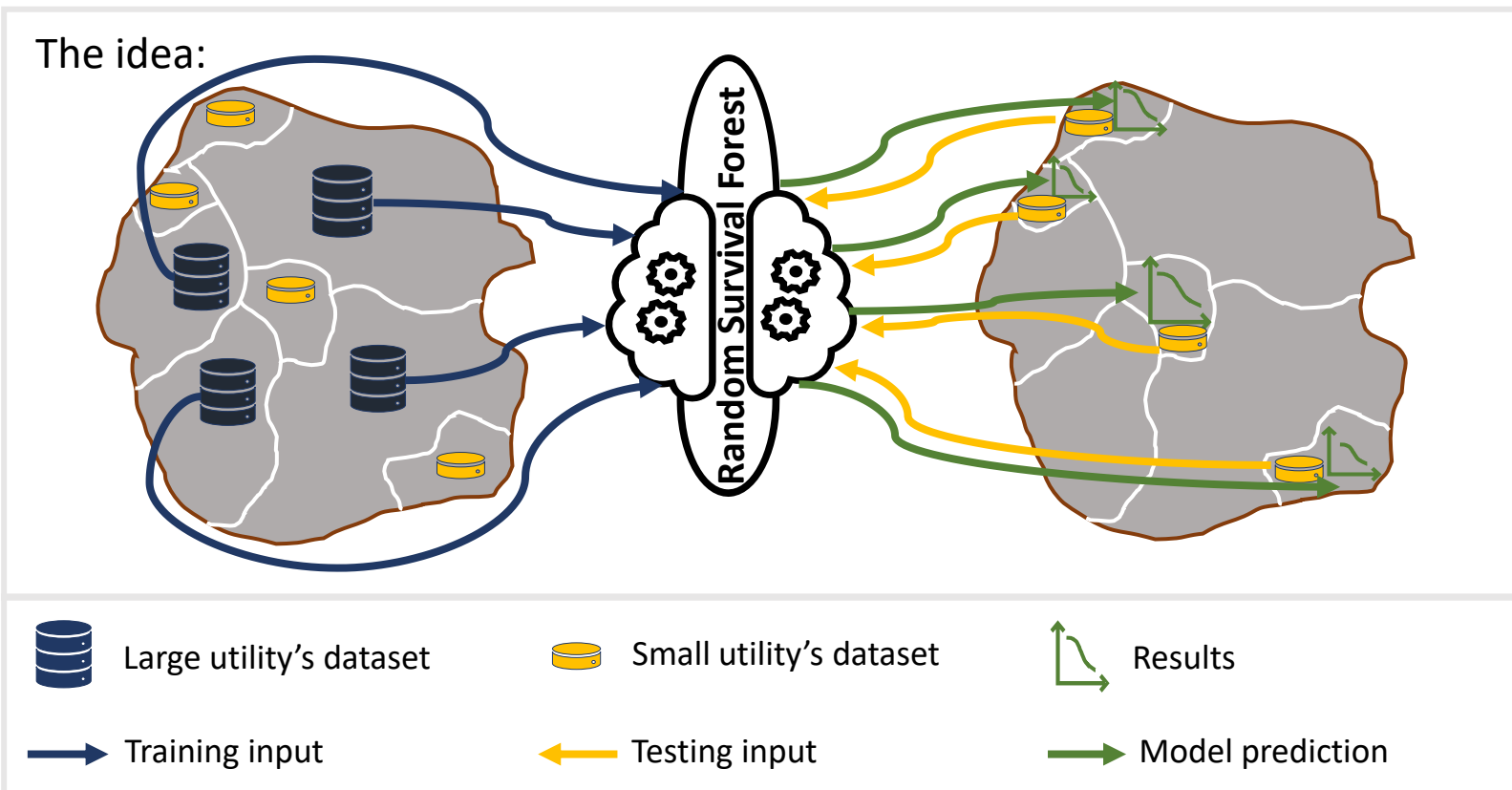
Building a machine learning-based national Norwegian model for predicting water pipe breaks & its application in Klepp Kommune

Shamsuddin Daulat (NTNU)
Marius Møller Rokstad (NTNU)
Stian Bruaset (SINTEF)
Jeroen Langeveld (TU Delft)
Franz Tscheikner-Gratl (NTNU)

20.10.2023

Purpose and Objectives

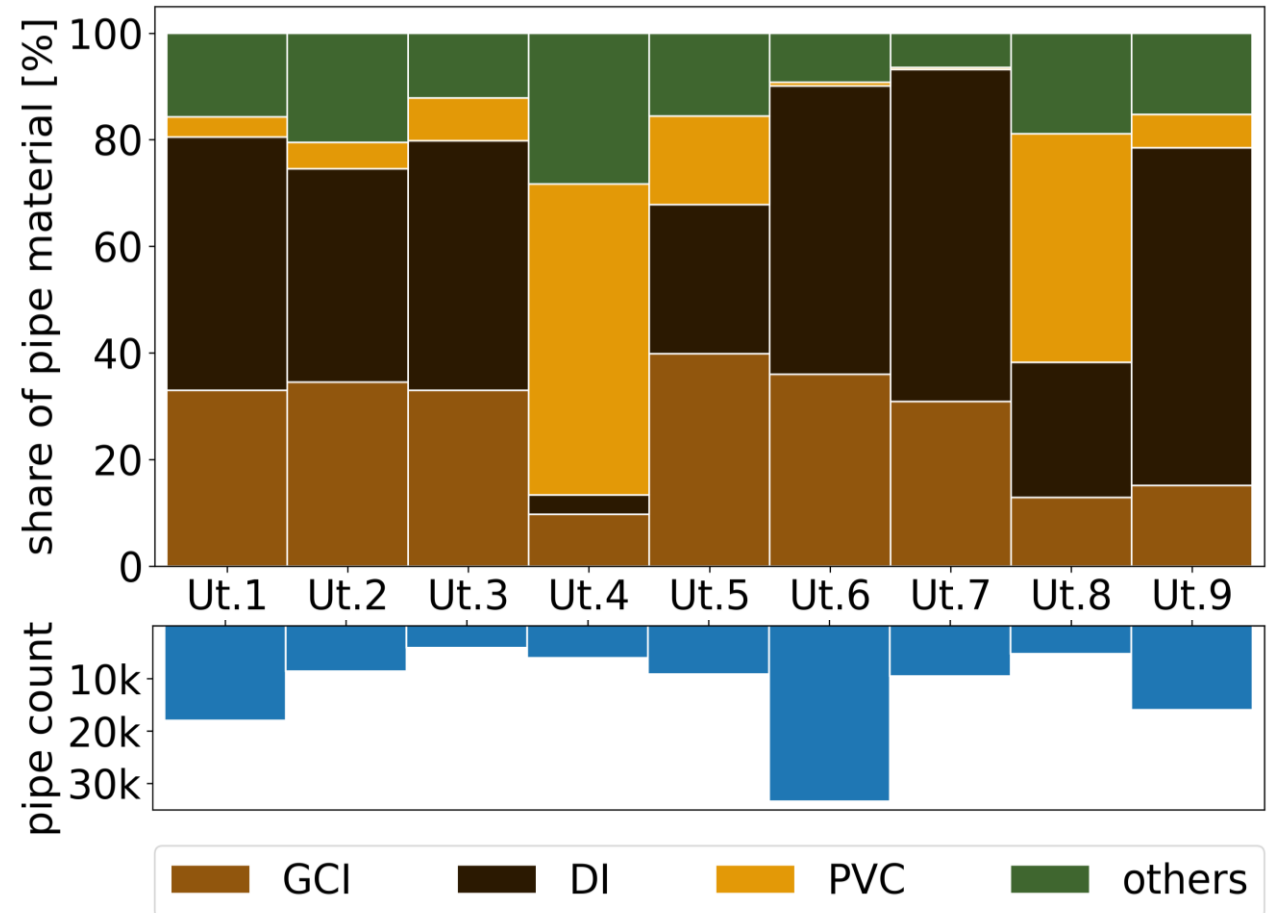
The problem: **Small utilities** are left behind in utilizing the predicting power of machine learning models due to their insufficient data quantity.



But will the national model have high enough accuracy in predicting any utility's pipe break?

Data

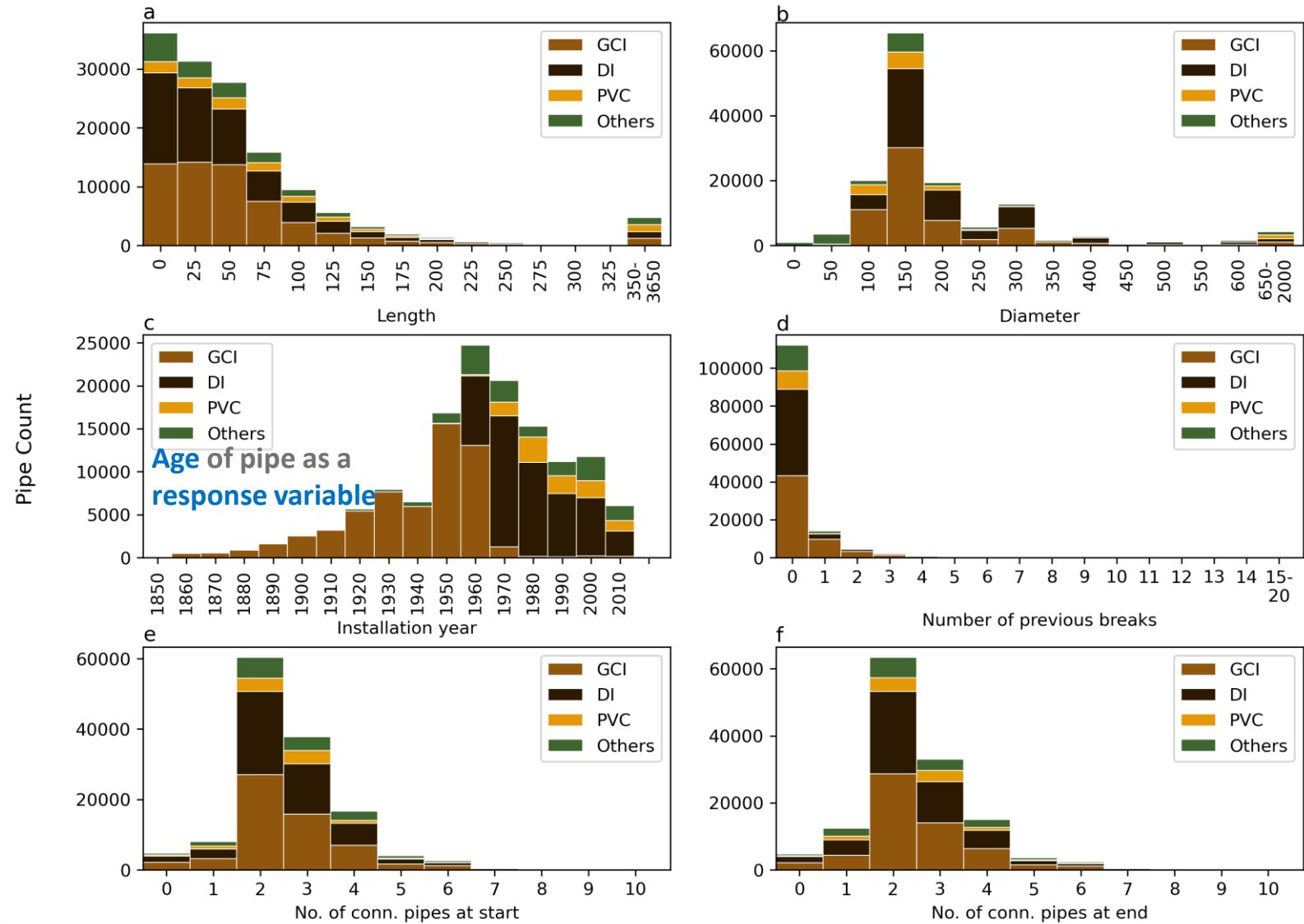
- Predictions based on historical break data
- Water distribution network of 9 municipalities of Norway
- Total length of pipes = ~ 7000 km



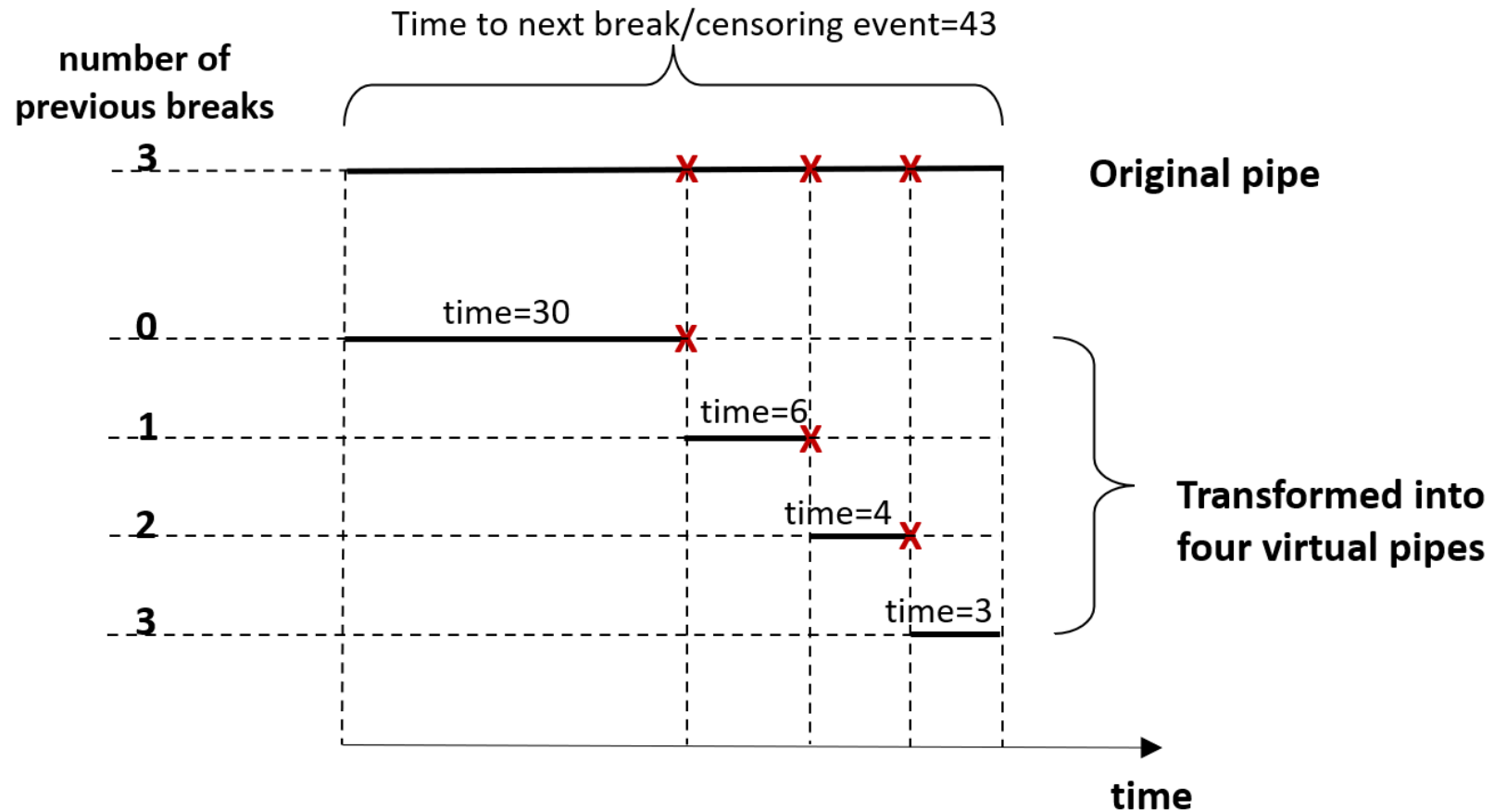
Pipe material distribution and count of pipes in in each utility

Data

Shared available
variables among
the utilities

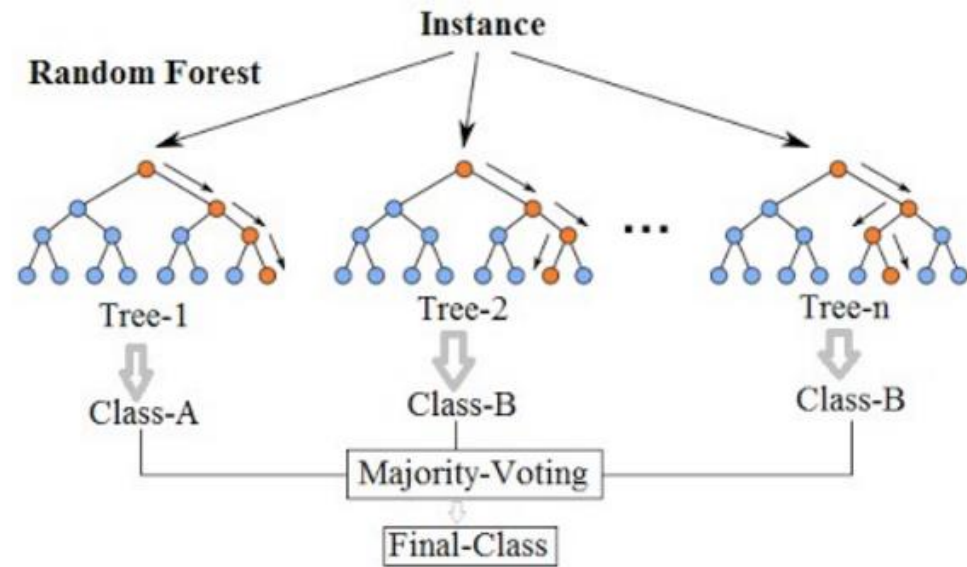


Utilizing previous break data

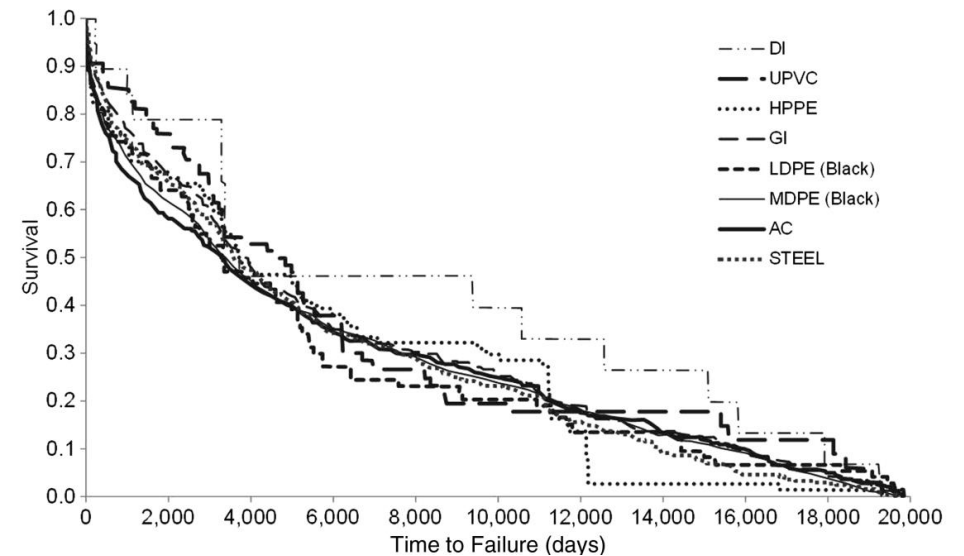


Method: Random Survival Forest (RSF)

Random Forest



Survival modeling



Random Survival Forest (RSF): A combination of Random Forest and survival models

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The annals of applied statistics*, 2(3), 841-860

Random Survival Forest - Performance evaluation

Concordance index (C-index) is the metric to evaluate the performance of RSF

ID	Observed 1 = failed	Predicted Risk score
P1	1 (at 40)	46.1
P2	1 (at 40)	12.5
P3	1 (at 50)	10.5

ID	Observed 0 = not failed	Predicted Risk score
P4	0 (at 50)	37.9
P5	0 (at 50)	10.5
P6	0 (at 30)	8.6

C-index and values:
1.0 – perfect model
0.5 – random model
0.0 – perfectly wrong model

	Concordants (1 point)	Discordants (0 point)	Tied (0.5 point)	Incomparable pair
1	46.1 > 37.9	12.5 < 37.9	10.5 = 10.5	P1 <?> P6
2	46.1 > 10.5	10.5 < 37.9		P2 <?> P6
3	12.5 > 10.5			P3 <?> P6
4	46.1 > 10.5			
5	12.5 > 10.5			

$$\text{C-index} = (5+0.5)/8 = 0.69$$

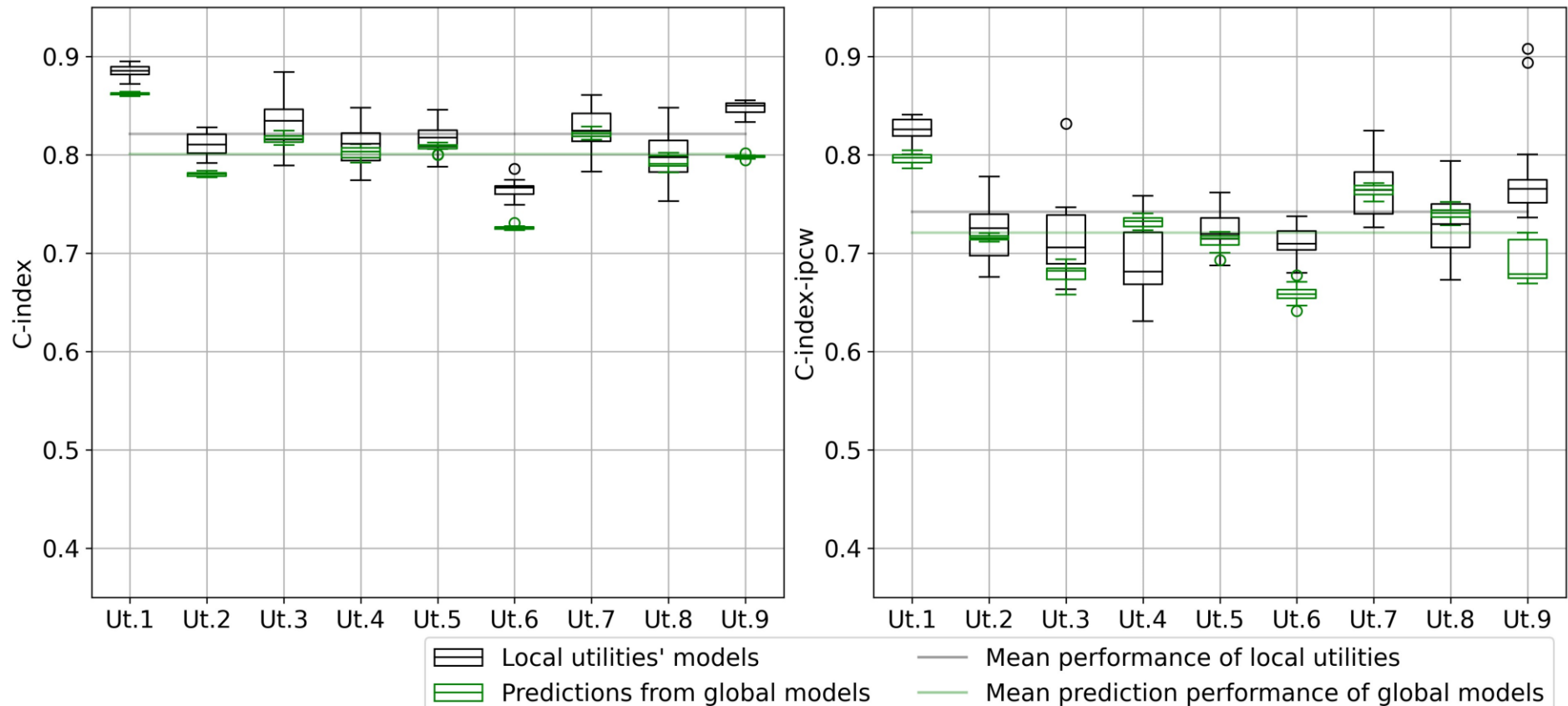
C-index-ipcw
For high amount of censored data:
C-index overestimates performance, use instead **C-index-ipcw**

Results

C-index and C-index-ipcw values:

- 1.0 – perfect model
- 0.5 – random model
- 0.0 – perfectly wrong model

National (global) vs local model's performance

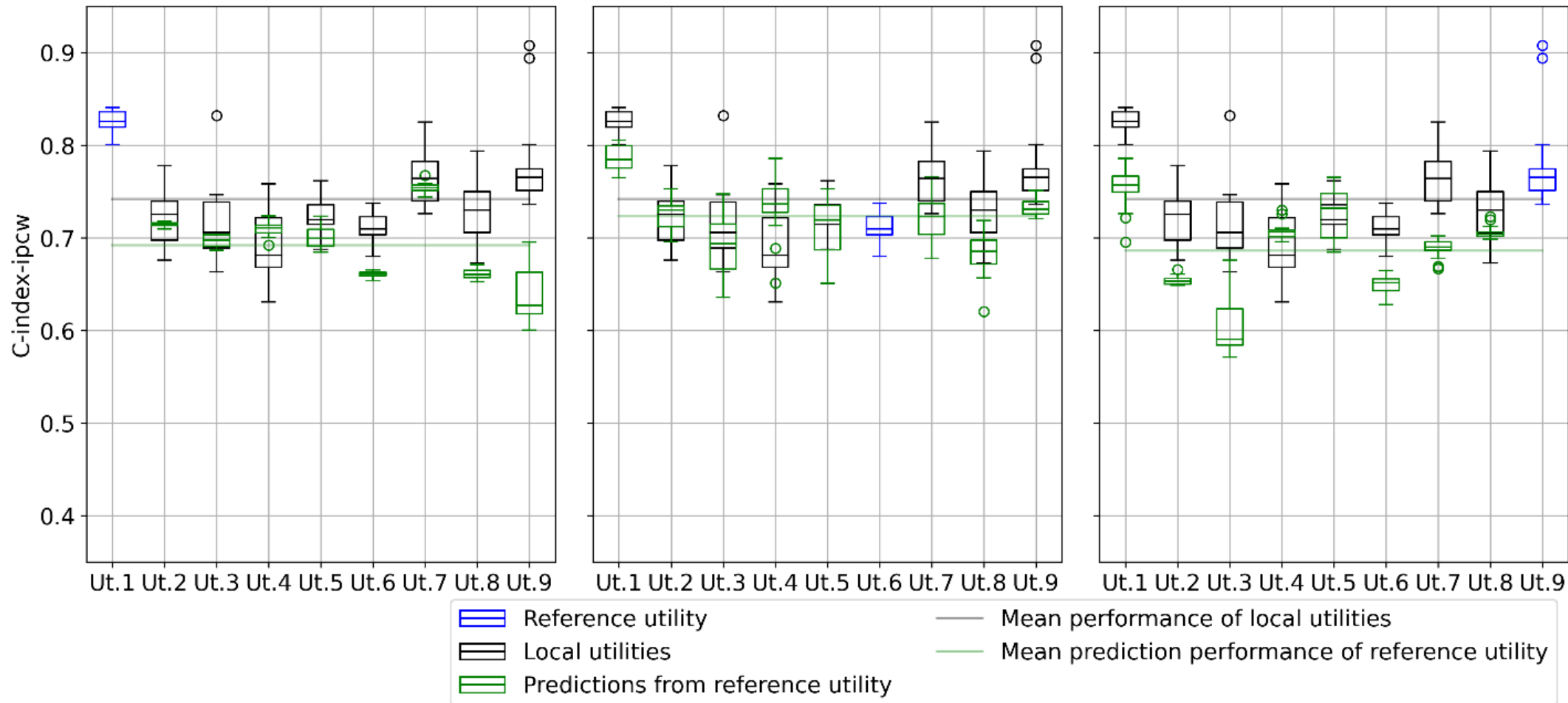


Results

C-index and C-index-ipcw values:

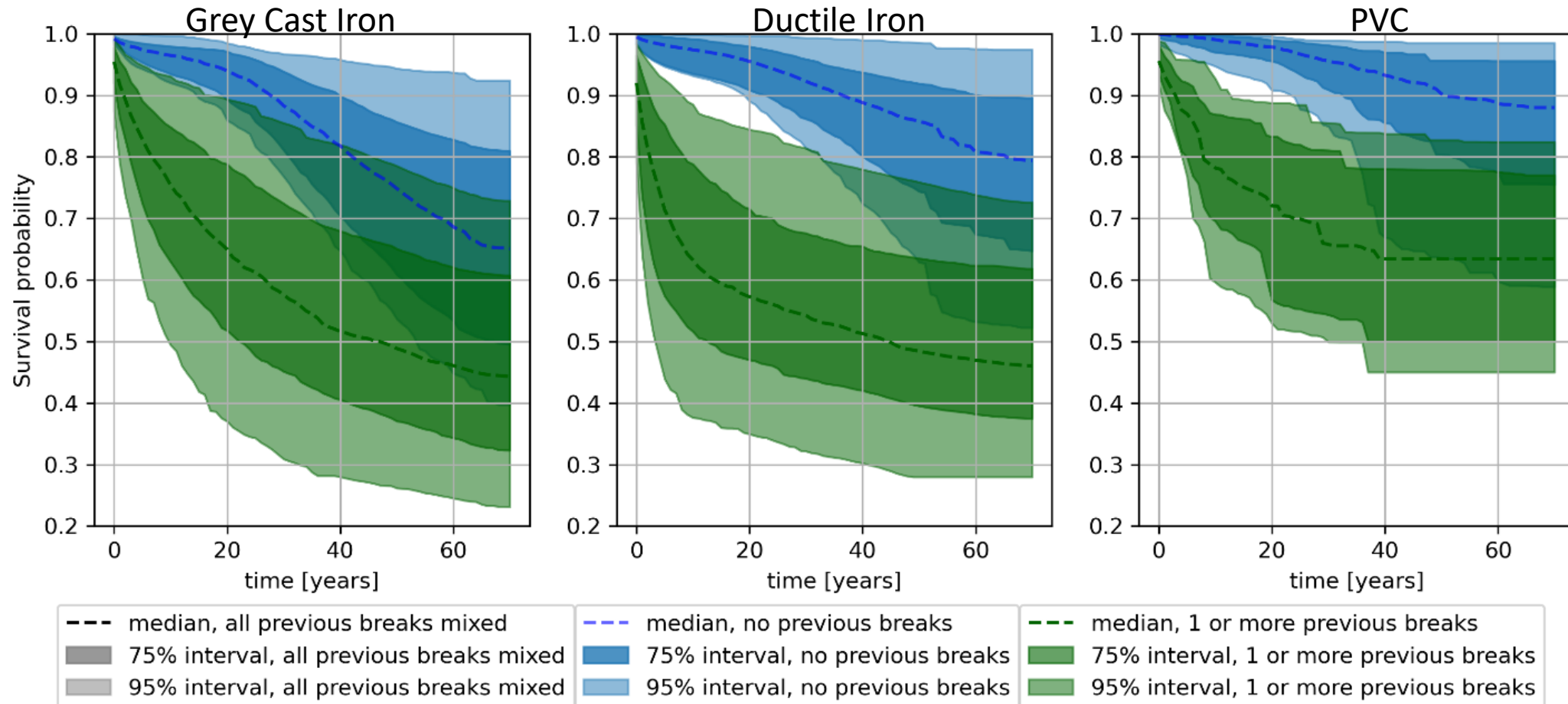
- 1.0 – perfect model
- 0.5 – random model
- 0.0 – perfectly wrong model

Reference utilities' model performance



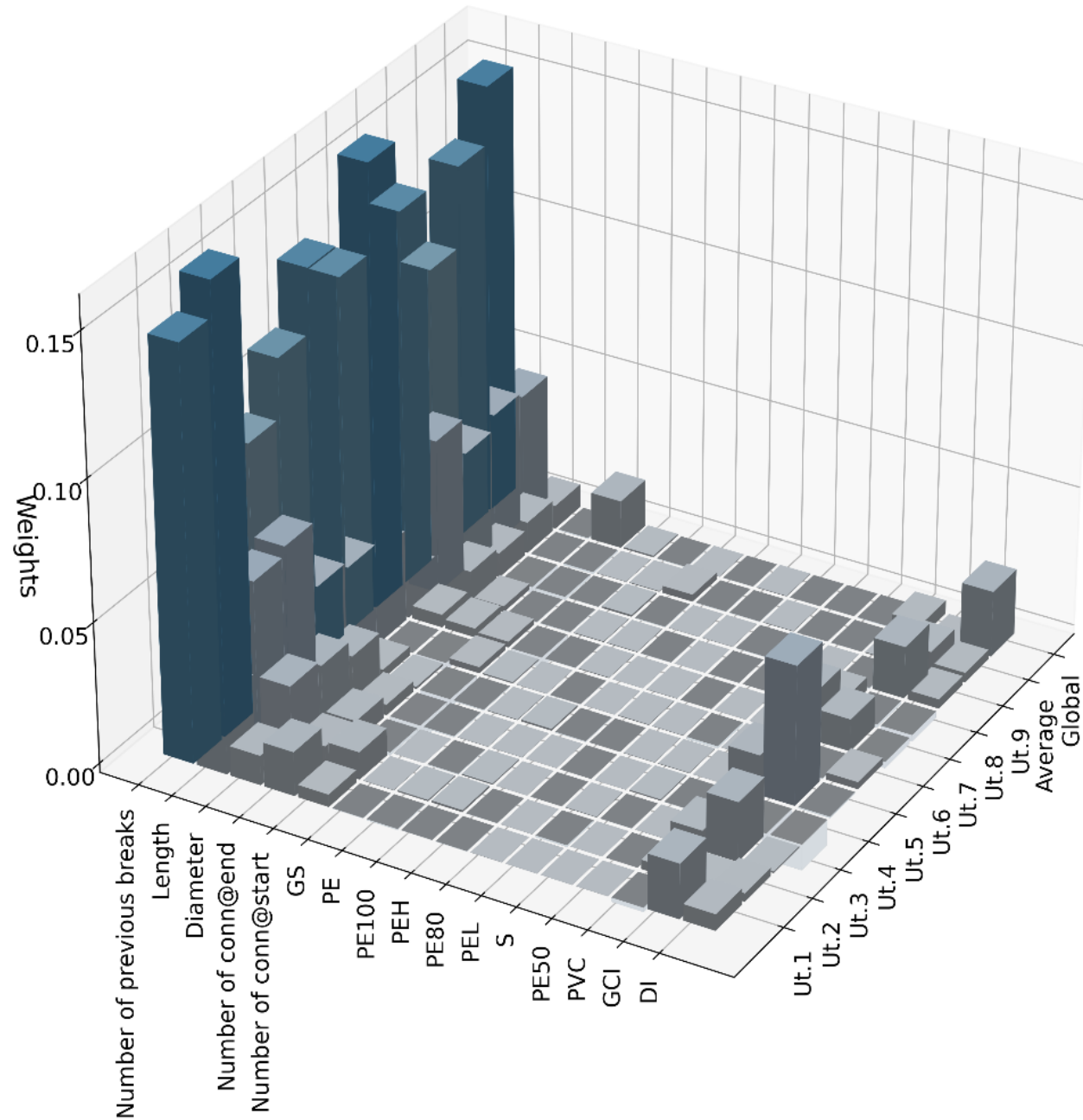
Results

Group survival curves



Analysis of variable importance

Number of previous breaks is the most important variable!



Further work on the model

- The transferability of the model **to other countries is not tested.**
- Also, the accuracy of **regional models** with similar climate and environment, instead of one national model, needs to be tested.
- **Hyperparameters will be tuned** to increase the accuracy even more.



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress



Evaluating the generalizability and transferability of water distribution deterioration models

Shamsuddin Daulat^{a,*}, Marius Møller Rokstad^a, Stian Bruaset^b, Jeroen Langeveld^c, Franz Tscheikner-Gratl^a

^a Department of Civil and Environmental Engineering, Norwegian University of Science and Technology (NTNU)

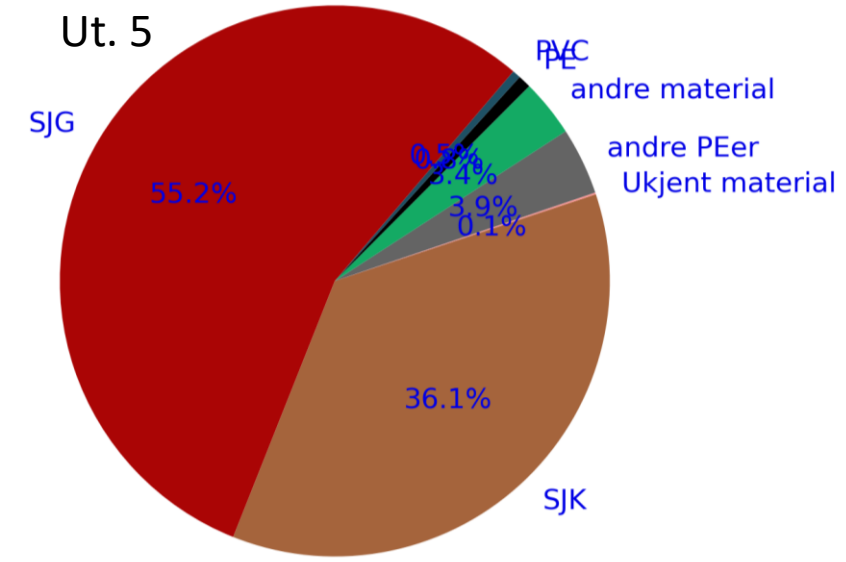
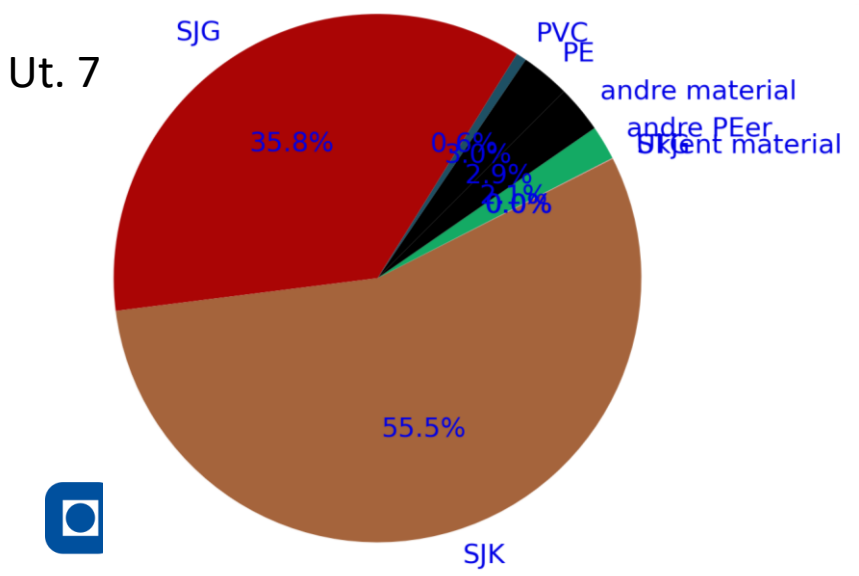
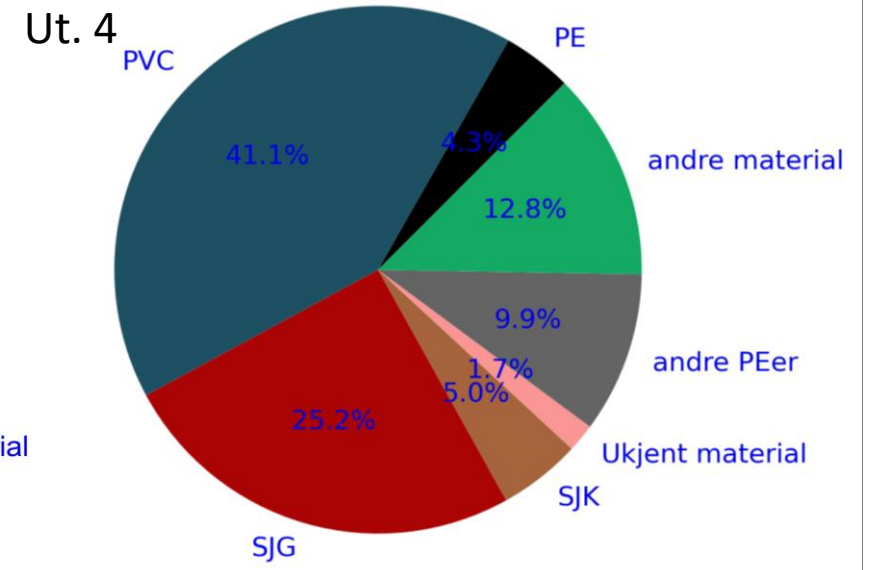
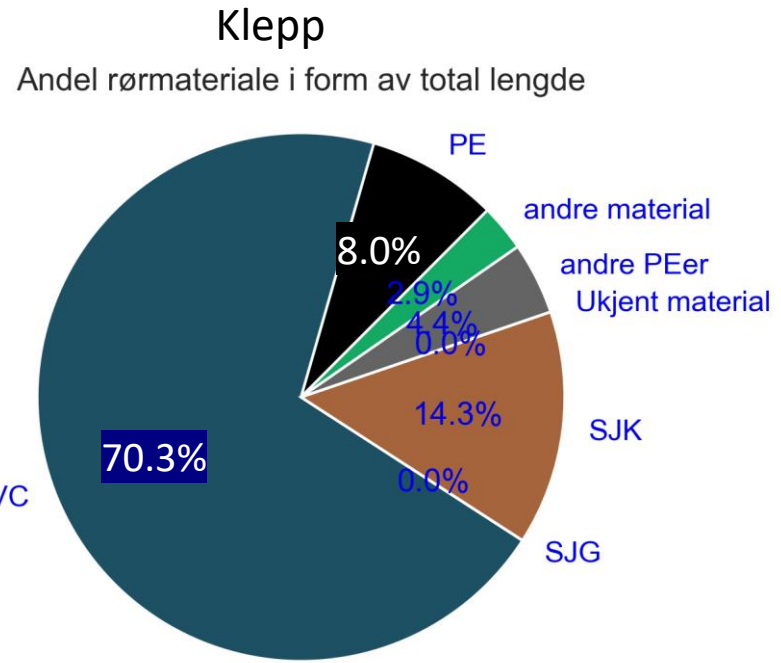
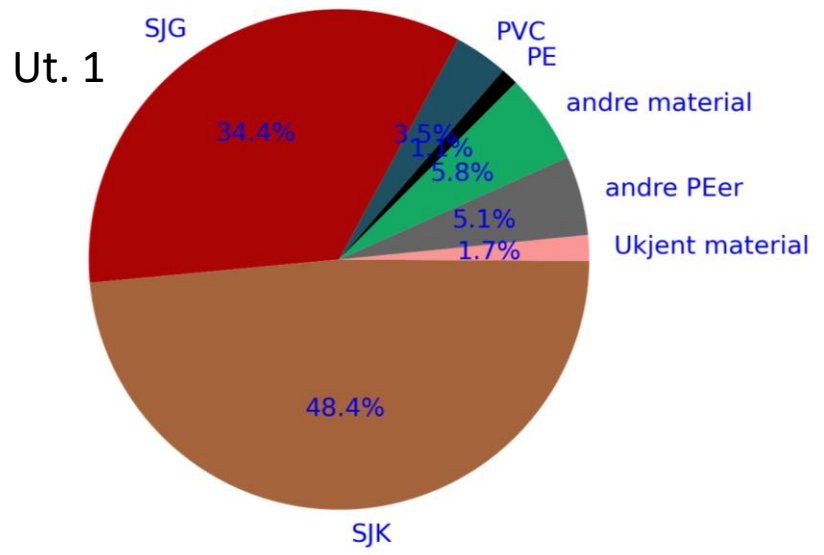
^b SINTEF, Trondheim, Norway

^c Faculty of Civil Engineering and Geosciences, TU Delft



Applying the model in Klepp Kommune

Water pipes in Klepp vs other municipalities

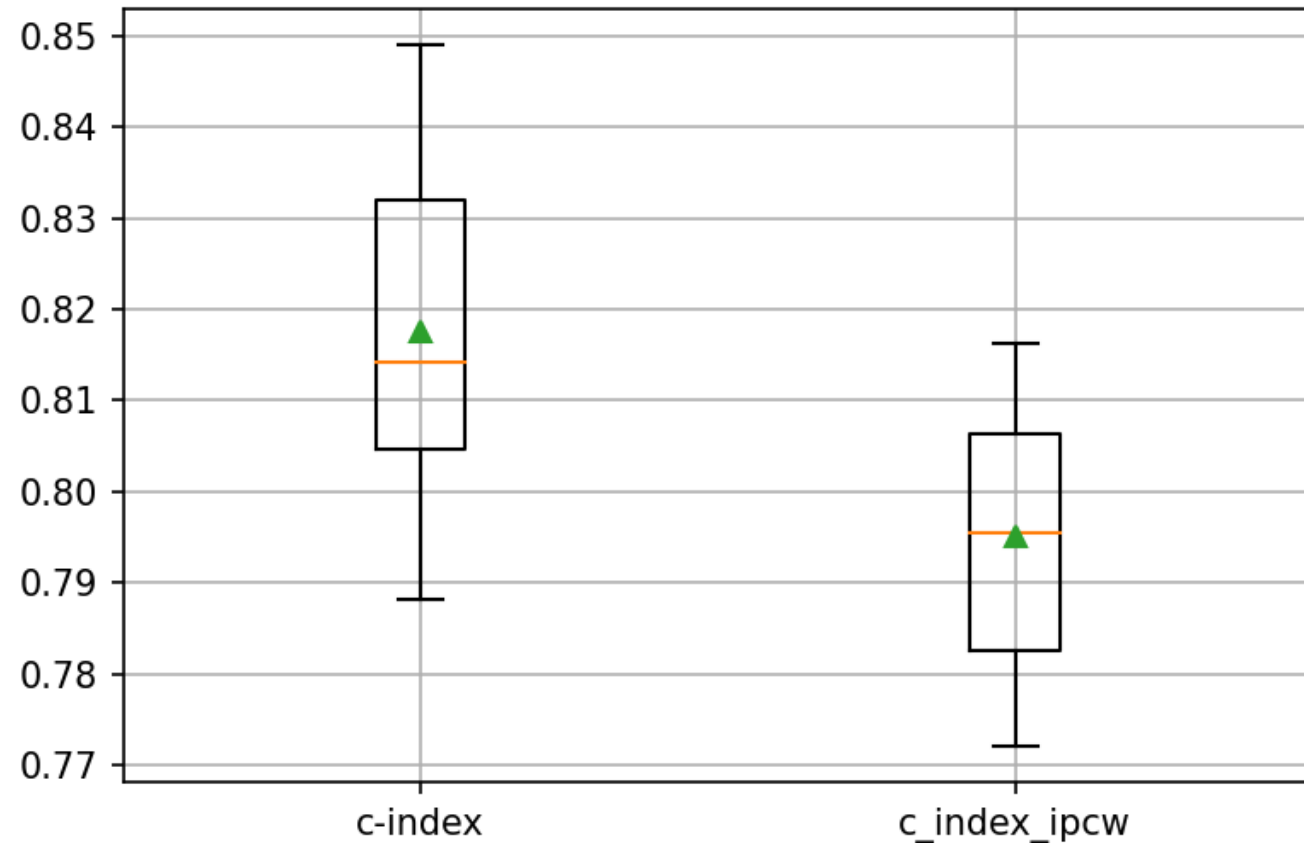


Applying the model in Klepp

C-index and C-index-ipcw values:

- 1.0 – perfect model
- 0.5 – random model
- 0.0 – perfectly wrong model

National model's performance when predicting water pipe breaks in Klepp



Take home messages

- We **can use the national model** in (all) Norwegian municipalities.
- Historic **break data is the dominant data** that can help us predict future breaks.
- **Municipalities and Volue** should focus on proper registration of breaks.
- Hyperparameters are **not yet** tuned but is recommended.

Acknowledgement

This work was supported by the Norwegian Research Council under the project B for VA-net (project number 309604).

Shamsuddin Daulat

PhD Candidate

Department of Civil and Environmental Engineering

shamsuddin.daulat@ntnu.no

+4746902145

Vassbygget, 220, Valgrinda

[ResearchGate](#)

[Google Scholar](#)



Or at:

Shamsuddin.Daulat@Klepp.Kommune.no