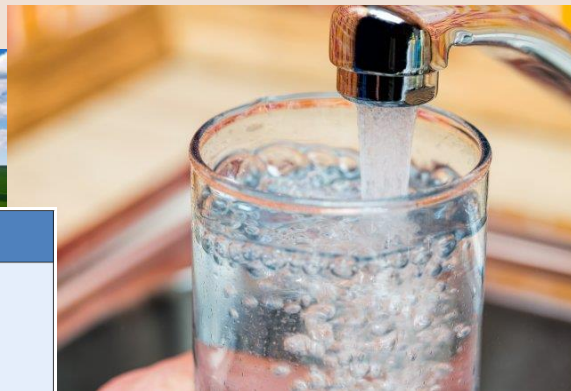
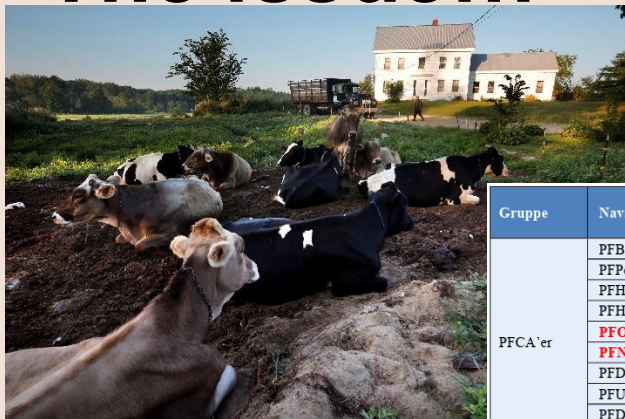


Development of a Machine Learning Model for Assessing PFAS Vulnerability in Danish Groundwater

Nordrocs, Helsinki – September 2026
Bastian Germundsson, Senior Project Manager, COWI

The issue...

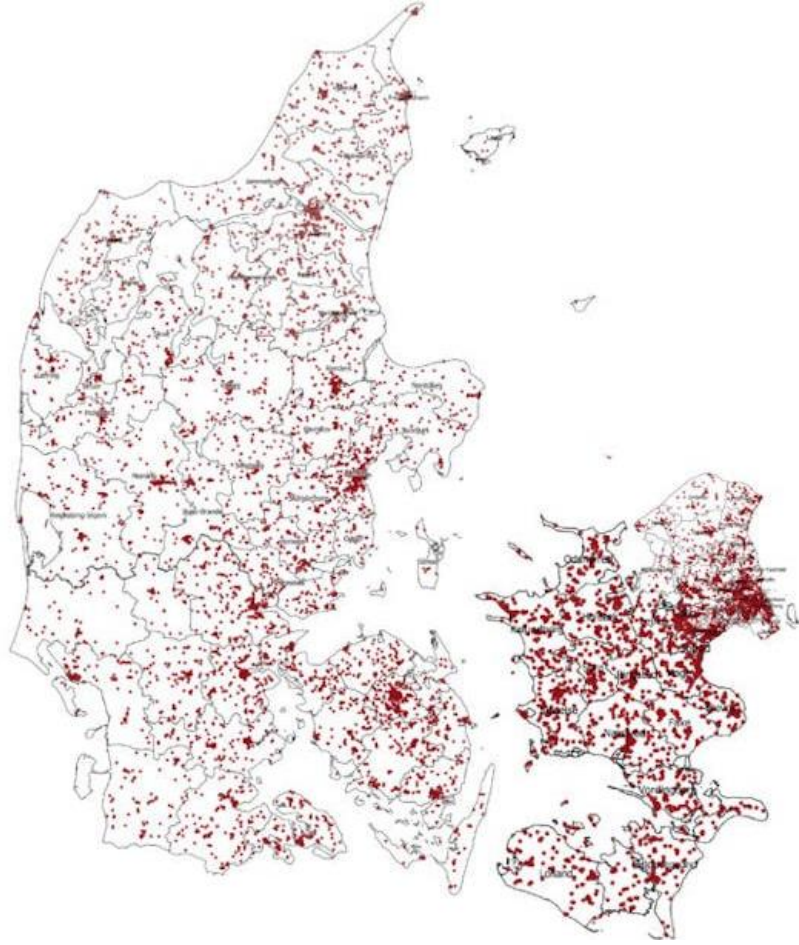


Gruppe	Navn	Formel	CAS	Antal C i C-F- kæden	
PFCA'er	PFBA	C_4F_9COOH	375-22-4	3	
	PFPeA	$C_5F_{11}COOH$	2706-90-3	4	
	PFHxA	$C_6F_{13}COOH$	307-24-4	5	
	PFHpA	$C_7F_{15}COOH$	375-85-9	6	
	PFOA	$C_8F_{17}COOH$	335-67-1	7	
	PFNA	$C_9F_{19}COOH$	375-95-1	8	
	PFDA	$C_{10}F_{21}COOH$	335-76-2	9	
	PFUnDA	$C_{11}F_{23}COOH$	2068-90-8	10	
	PFDoDA	$C_{12}F_{25}COOH$	307-55-1	11	
	PFTriDA	$C_{13}F_{27}COOH$	72629-94-8	12	
PFSA'er	PFBS	$C_4F_9SO_3H$	375-52-5	3	
	PFPeS	$C_5F_{11}SO_3H$	375-91-4	4	
	PFHxS	$C_6F_{13}SO_3H$	355-46-4	5	
	PFHpS	$C_7F_{15}SO_3H$	375-92-8	6	
	PFOS	$C_8F_{17}SO_3H$	1763-23-1	7	
	PFNS	$C_9F_{19}SO_3H$	68259-12-1	8	
	PFDS	$C_{10}F_{21}SO_3H$	335-77-3	9	
	PFUnDS	$C_{11}F_{23}SO_3H$	749786-16-1	10	
	PFDoDS	$C_{12}F_{25}SO_3H$	79780-39-5	11	
	PFTriDS	$C_{13}F_{27}SO_3H$	791563-89-8	12	
Andre PFAS Precursorer	PFOSA	$C_8F_{17}SO_2NH_2$	754-91-6	8	
	6:2 FTS	$C_6F_{13}SO_2H$	27619-97-2	6	

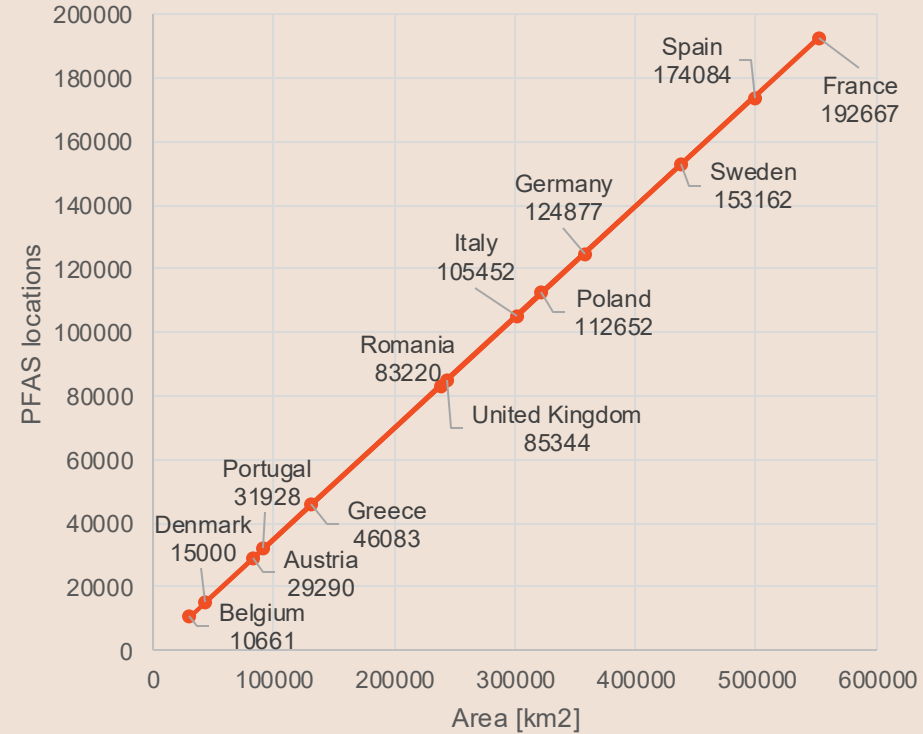
$\Sigma 22 = 0.1 \mu g/L (100 ng/L)$
 $\Sigma 4 = 0.002 \mu g/L (2 ng/L)$

DK-kvalitetskriteriet





15,000 locations



<https://nyheder.tv2.dk/samfund/2022-08-05-her-er-de-15000-steder-der-mistaenkes-for-pfas-forurening>

Expensive

Challenging...

Is PFAS the most important
contaminant?

Can we use artificial intelligence?

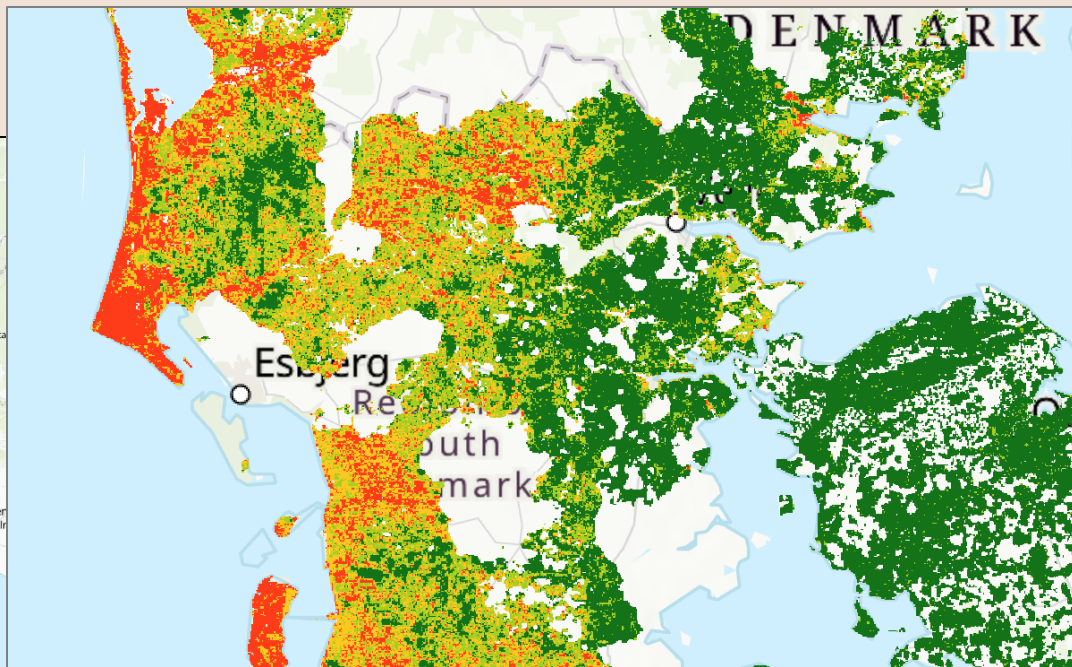
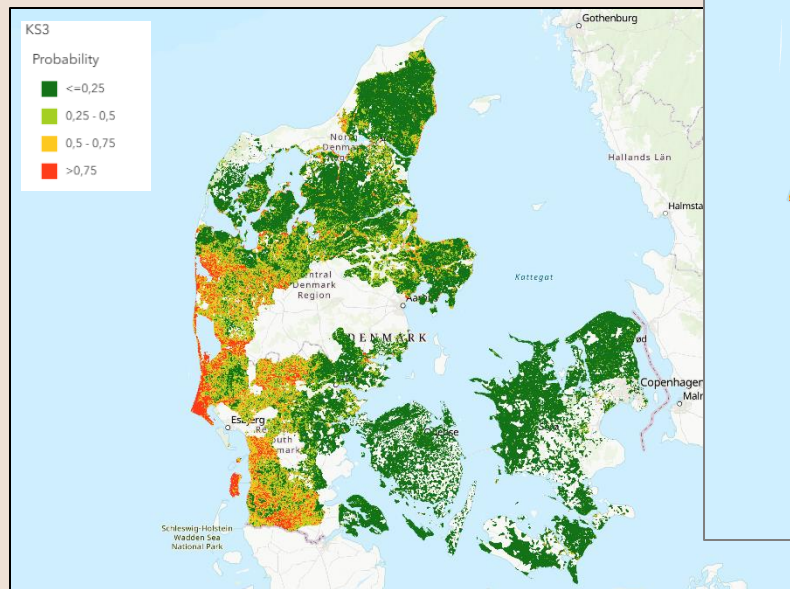
Prioritization?

Clean drinking water

15.000 locations

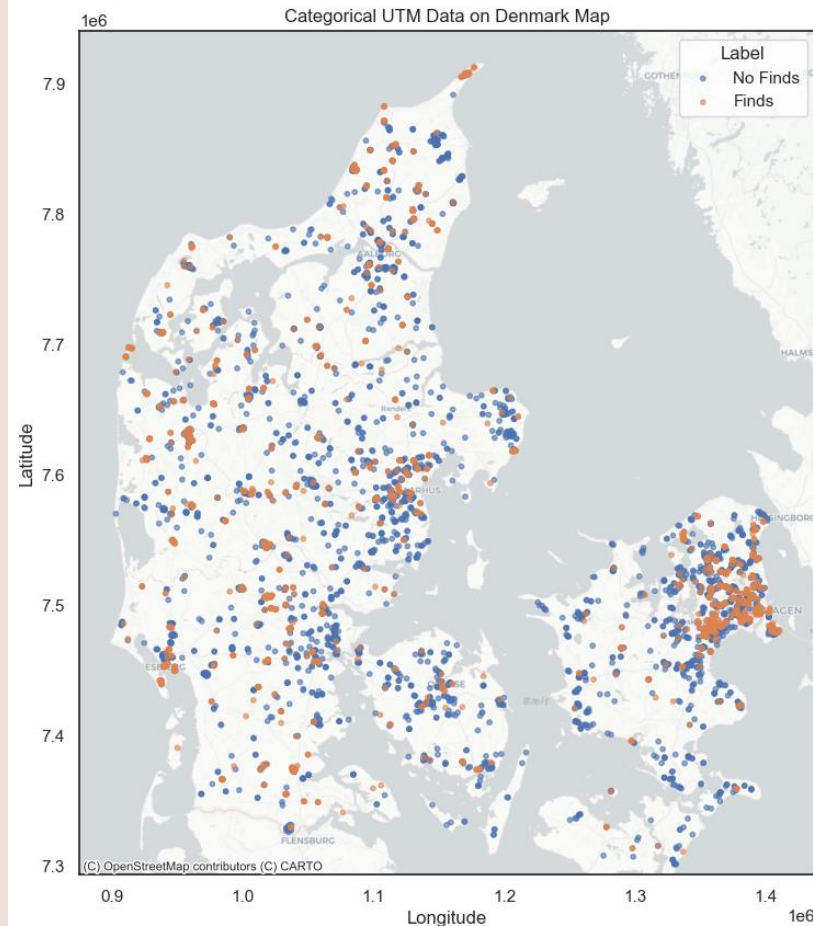
Results - PFAS Map

13 aquifers = 13 maps:



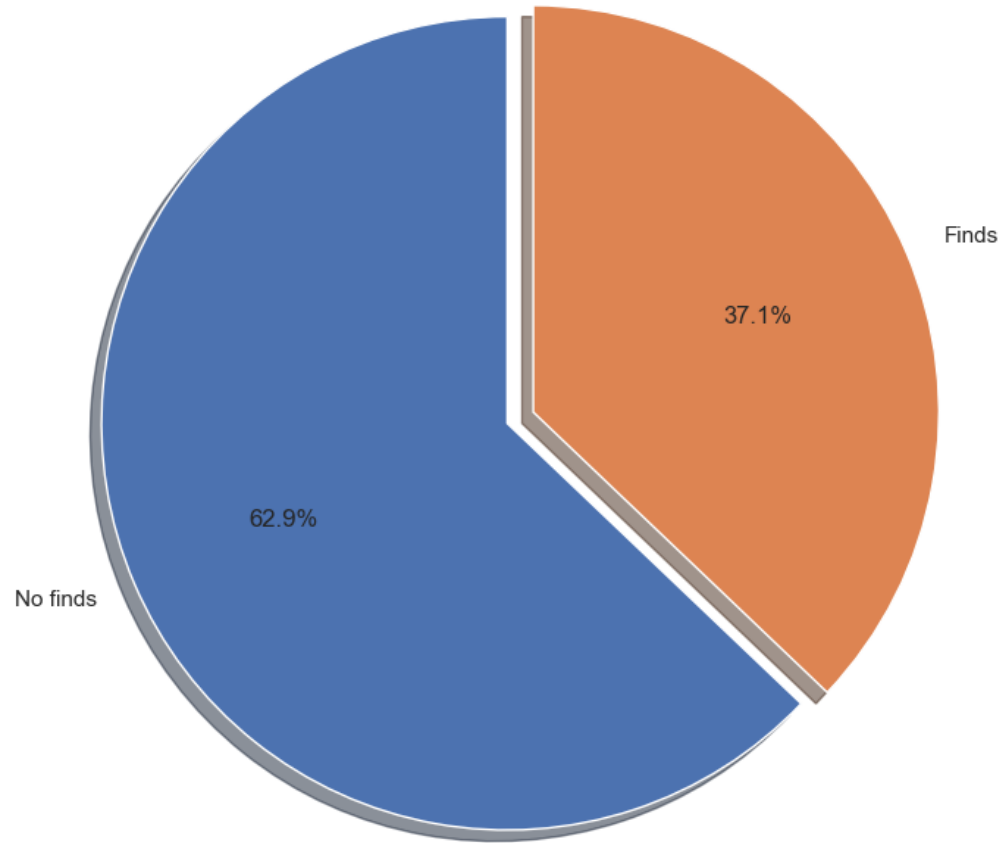
Target value

- The value we want to predict:
Whether there is PFAS present in a given intake/groundwater aquifer or not.
- Rule:
An intake is considered affected by PFAS if PFAS has been detected in the latest analysis, or if PFAS has been detected more than once historically.
- Data source:
GEUS' Jupiter database, 5018 observations → 5008



GEUS: The Geological Survey of Denmark and Greenland

Pie Chart of 5008 PFAS Intakes



Defining parameters / features

Parametre	Data source
Topography	DK-modellen
Aquifer type	DK-modellen
Water level	DK-modellen
Net infiltration	DK-modellen
Thickness of the unsaturated zone	DK-modellen
Recharge	DK-modellen
Depth to aquifer	DK-modellen
Accumulated thickness of clay above intake	DK-modellen
Accumulated thickness of clay above intake, below redox	DK-modellen
Macropore flow	Iversen et al. 2011
Humus content	Adhikari et al. 2013
Land use	Levin et al. 2016

Modelling

- The model is implemented in Python 3.11 using the XGBoost module.
- XGBoost fundamentally aims to maximize accuracy without weighting either 'presence' or 'absence' of detections.

Training: 75% of the data

Validation: 25% of the data

- Evaluation metrics:

Accuracy – overall fit of the model

Precision – correctness of positive predictions

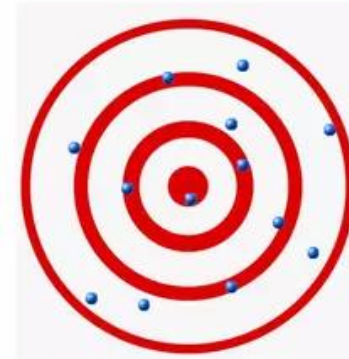
Recall (Sensitivity) – proportion of actual positives correctly identified



A: accurate and precise



B: precise, but not accurate



C: neither accurate nor precise



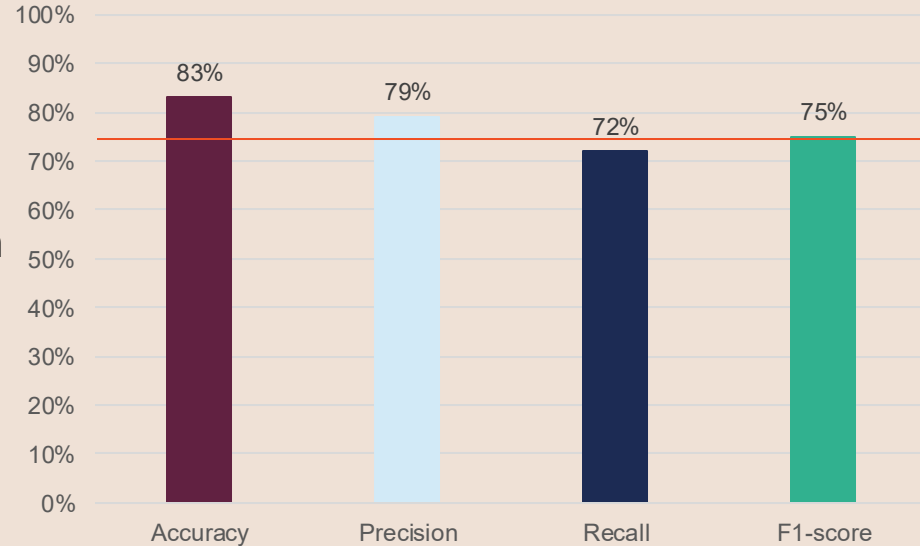
D: accurate, but not precise

The model predicts PFAS presence with high confidence

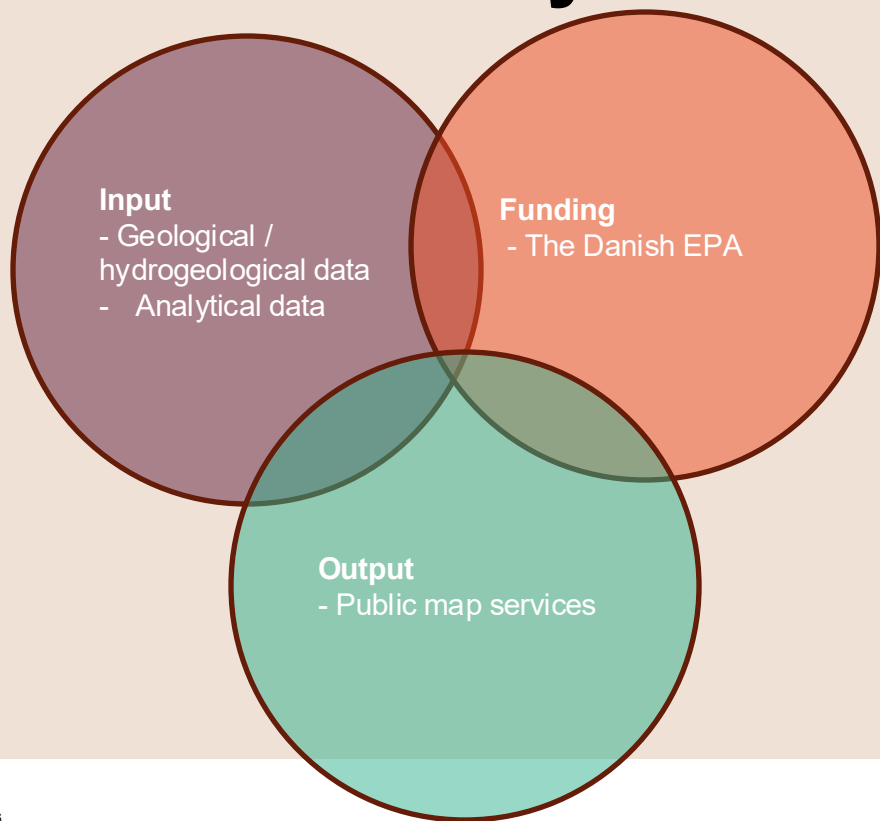
- > 75 % is acceptable
- High accuracy / high precision
- Lower recall: tendency to produce a higher number of false negatives (may improve with more data being available)



- Acceptable model
- High confidence that PFAS is present when the model indicates it



Public availability



Lessons learned

- Machine learning is useful — but **data quality and quantity are everything**
- The difficult part is not necessarily the model, but building the  **between disciplines**
- Environmental data are never truly static — so the solution should not be either
- AI-generated results create interest, questions — and plenty of opinions
- **Start small, learn fast — and just do it!**

Thank you!

Special thanks to:

- Nils Høgsted et al., DMP
- Marianne Jeppesen, HOFOR
- Klaus Mosthaf, DTU
- Kaspar Ruegg, Region Midtjylland
- And the Danish EPA for funding