

MODELLING ANNUAL TOTAL ORGANIC NITROGEN CONCENTRATIONS IN STREAMS USING MACHINE LEARNING AT NATIONAL SCALE

Rasmus R. Frederiksen, Søren E. Larsen, Henrik Tornbjerg, Hans Thodsen, Brian Kronvang

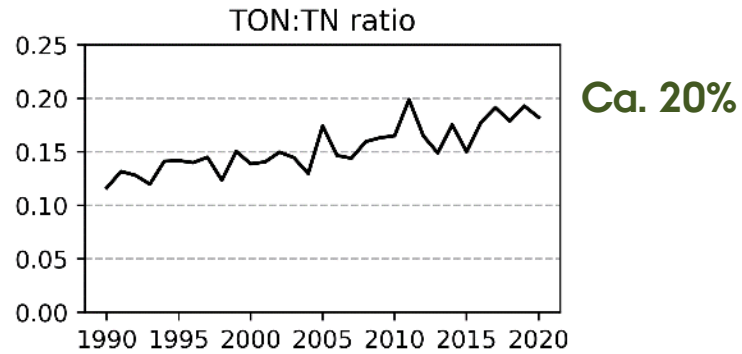
Department of Ecoscience, Aarhus University

Contact: rrf@ecos.au.dk



MOTIVATION

Total Organic Nitrogen (**TON**) constitutes an increasingly larger proportion of TN in streams:



Therefore, for future improvements in water quality, it is **critical** to improve the assessment of TON





BUT it is difficult to measure and model TON in streams because:

1. TON is measured indirectly: $\text{TON} = \text{TN} - \text{NO}_2\text{3N} - \text{NH}_4\text{N}$
2. TON in streams = TON from landscape + TON in-stream production
3. TON in streams = Particulate ON + Dissolved ON

High uncertainty + multiple sources + multiple drivers 😞

CAN WE DERIVE A REASONABLE MODEL?

Objectives:

1. To predict TON in ungauged catchments (ID15 resolution)
2. To improve understanding of key drivers of TON in streams

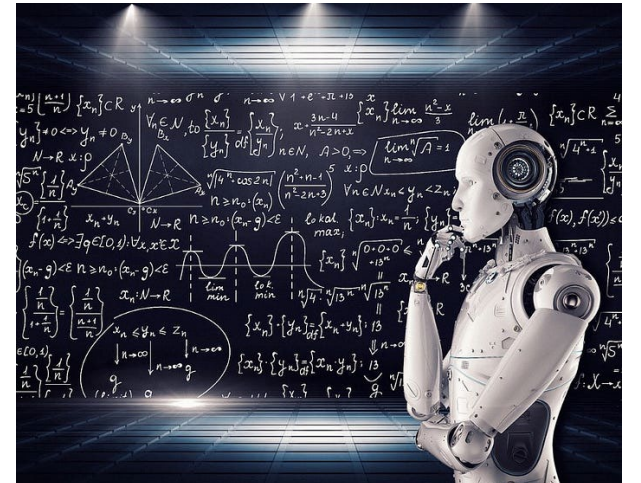


CAN WE DERIVE A REASONABLE MODEL?

Machine learning modelling approach: **XGBoost** algorithm

Response variable: Total Organic Nitrogen (TON) in streams

Predictor variables: Dynamic and static catchment characteristics

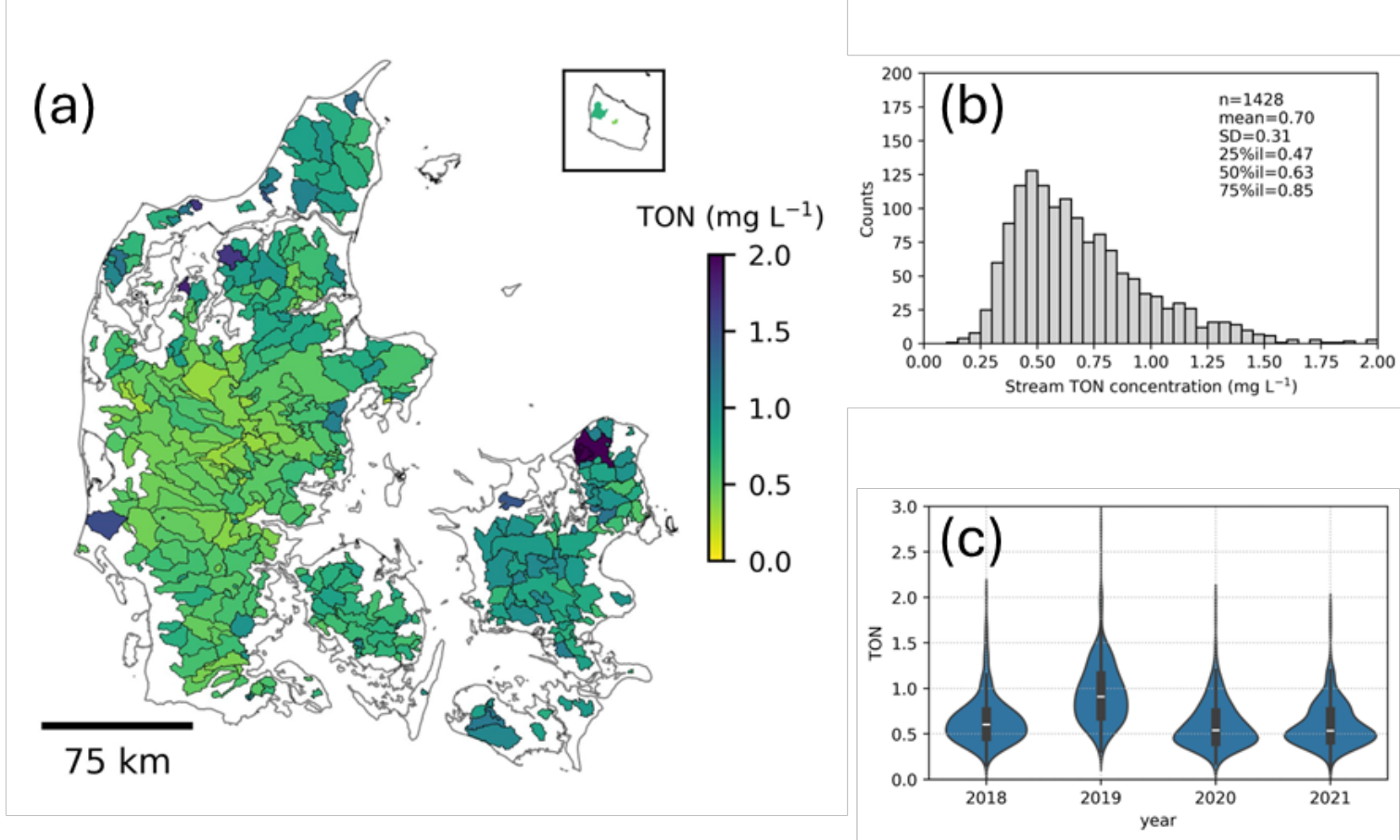


Response variable: Total Organic Nitrogen (TON) in streams

- Annual average TON
- 390 stations/catchments
- 4 year (2018-2021)
- 1428 samples



SUMMARY OF TON SAMPLES



Predictor variables:

- Dynamic variables
- Static variables
- 29 in total

Weather

Topography

Soil type

Land use/cover

Groundwater

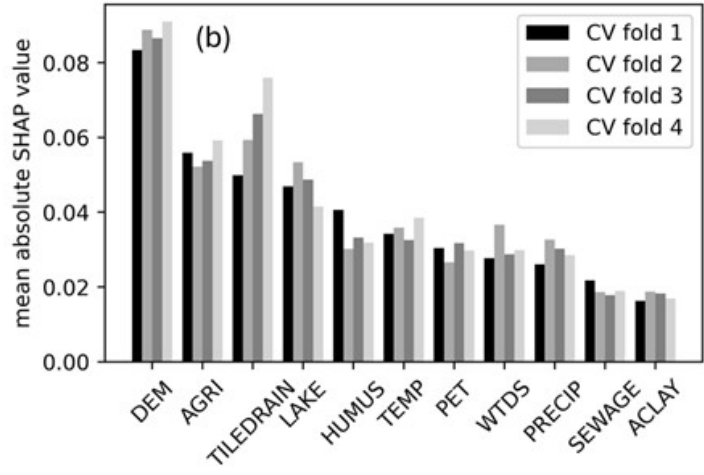
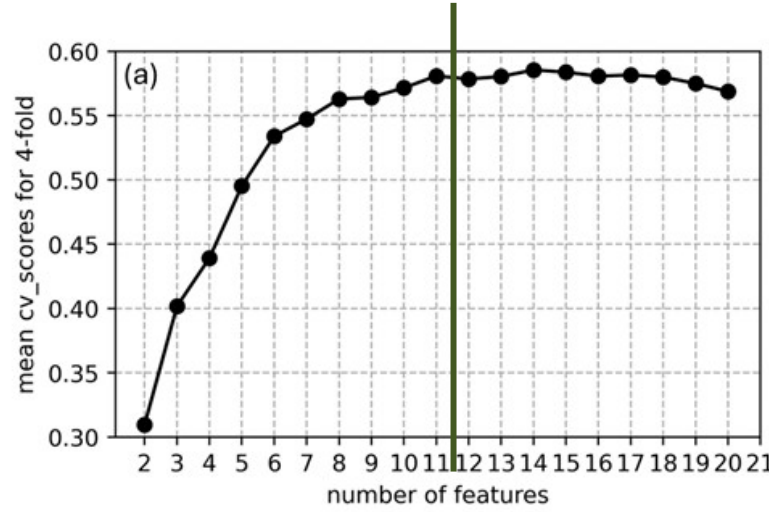
Point sources

Short name	Long name	Units	Source
PRECIP	Annual average precipitation	mm m ⁻¹	Scharling/DMI
PET	Annual average potential evapotranspiration	mm m ⁻¹	Scharling/DMI
TEMP	Annual average temperature	°C m ⁻¹	Scharling/DMI
DEM	Mean elevation	m	SDFI (2021)
SLOPE	Mean surface slope gradient	deg	Derived from dem
ACLAY	Mean clay content (%), 0-30 cm	%	Adhikari et al. (2013)
BCLAY	Mean clay content (%), 30-60 cm	%	Adhikari et al. (2013)
CCLAY	Mean clay content (%), 60-100 cm	%	Adhikari et al. (2013)
DCLAY	Mean clay content (%), 100-200 cm	%	Adhikari et al. (2013)
ASAND	Mean sand content (%), 0-30 cm	%	Adhikari et al. (2013)
BSAND	Mean sand content (%), 30-60 cm	%	Adhikari et al. (2013)
CSAND	Mean sand content (%), 60-100 cm	%	Adhikari et al. (2013)
DSAND	Mean sand content (%), 100-200 cm	%	Adhikari et al. (2013)
HUMUS	Fraction of soil with organic C content > 3%	%	Beucher et al. (2023)
URBAN	Fraction of urban	%	Levin (2022)
AGRI	Fraction of agriculture	%	Levin (2022)
FOREST	Fraction of forest	%	Levin (2022)
LAKE	Fraction of lakes	%	Levin (2022)
STREAM	Fraction of streams	m ha ⁻¹	Levin (2022)
WETLAND	Fraction of wetlands	%	Levin (2022)
TILEDRAIN	Fraction of tile drained land	%	Møller et al. (2018)
WTDS	Mean water table depth in summer	cm	Koch et al. (2019)
WTDW	Mean water table depth in winter	cm	Koch et al. (2019)
REDOX	Mean redox interface depth	m	Koch et al. (2019)
FISHFARM	TN load from freshwater fish farms	kg ha ⁻¹	Thodsen et al. (2024)
INDUSTRY	TN load from industry plants	kg ha ⁻¹	Thodsen et al. (2024)
SEWAGE	TN load from wastewater plants	kg ha ⁻¹	Thodsen et al. (2024)
RBU	TN load from stormwater outlets	kg ha ⁻¹	Thodsen et al. (2024)
BANKEROSION	Sediment load from bank erosion	kg ha ⁻¹	Andersen og Heckrath (2020)

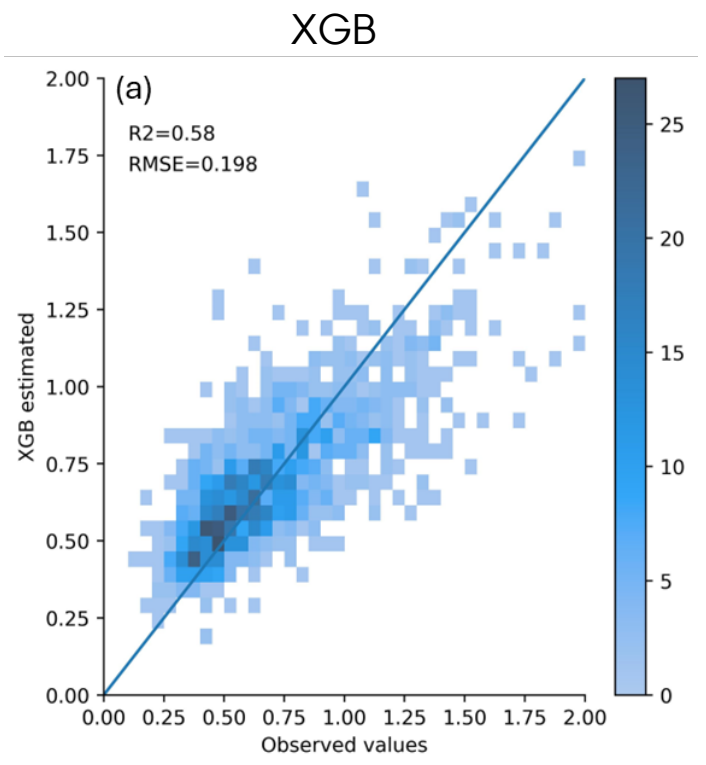


ONLY 11 PREDICTOR VARIABLES WERE NEEDED

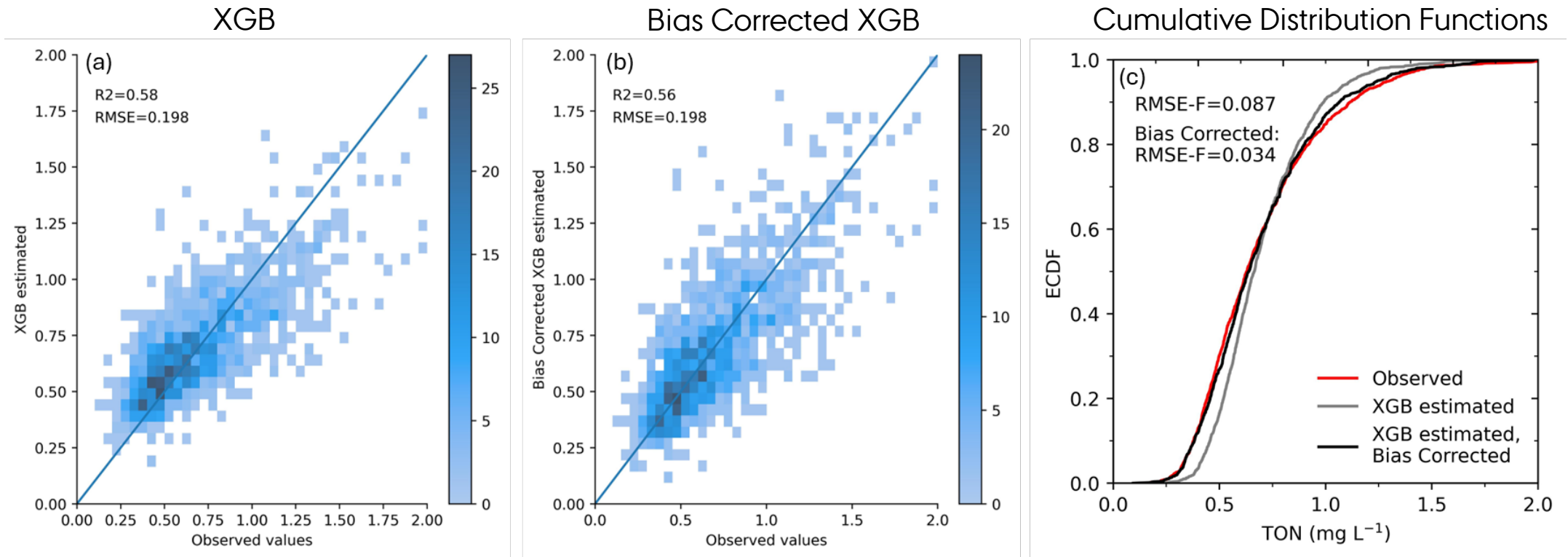
Forward Feature Selection



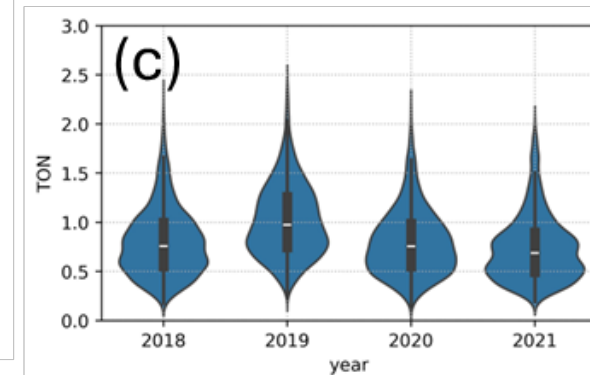
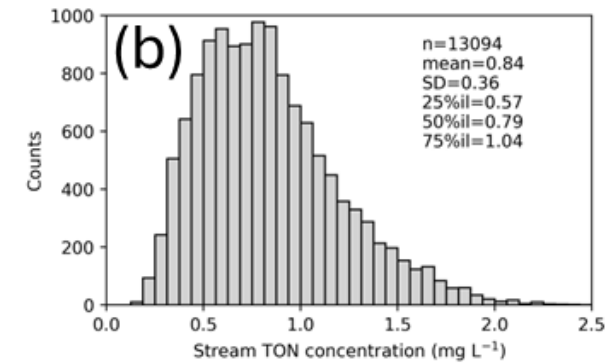
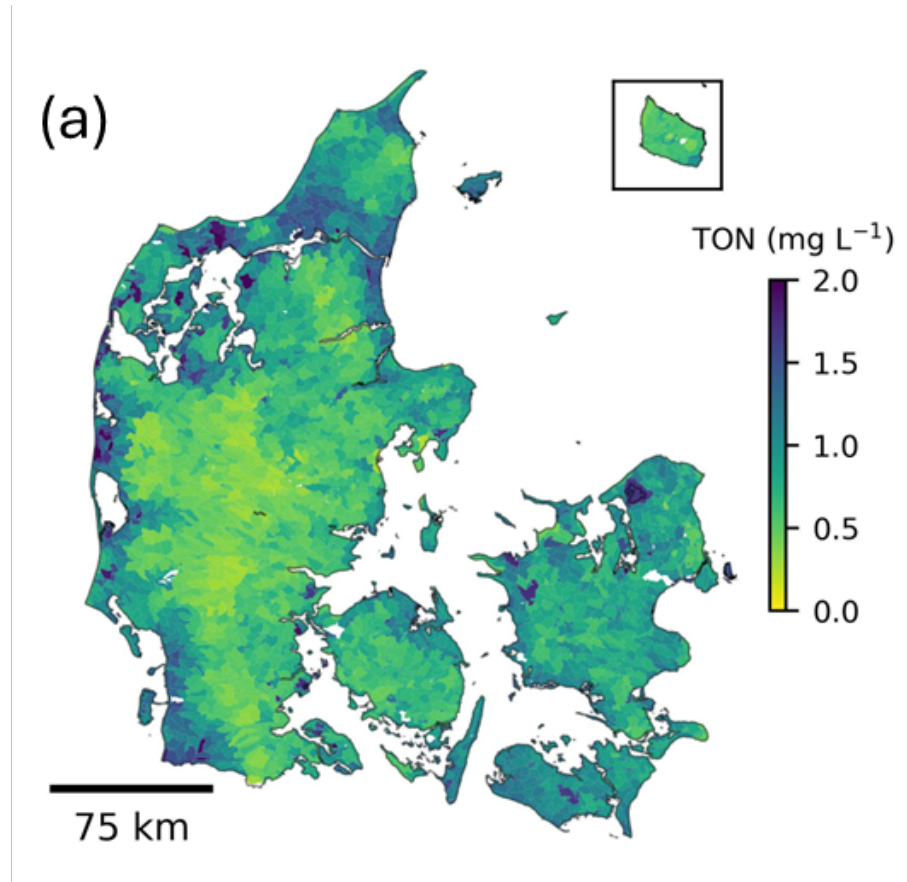
MODEL PERFORMANCE



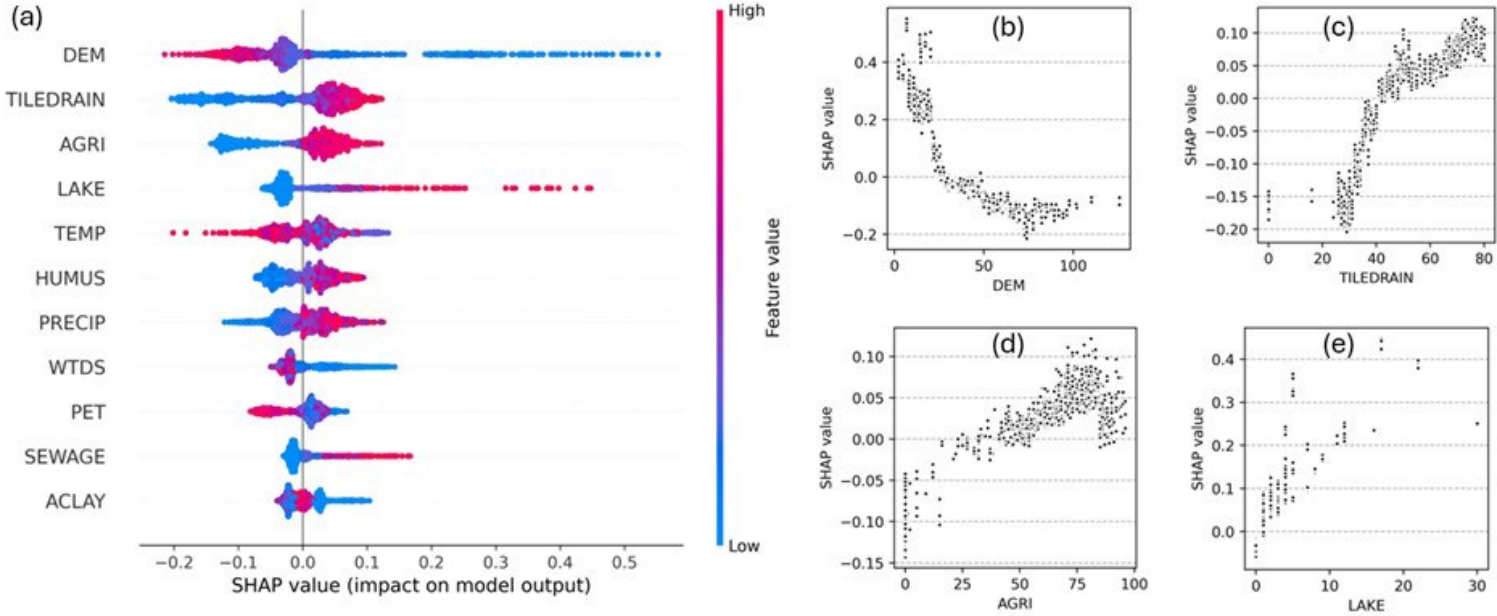
MODEL PERFORMANCE



MODEL PREDICTION: NATIONAL MAP

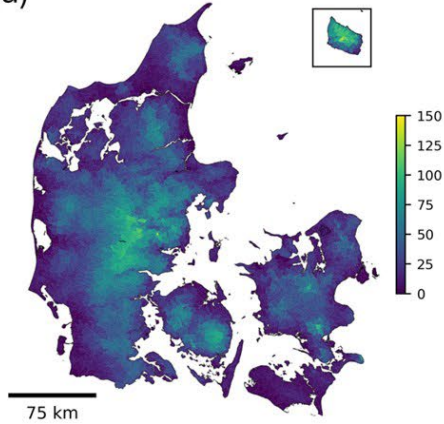


KEY DRIVERS OF TON IN STREAMS IN DENMARK

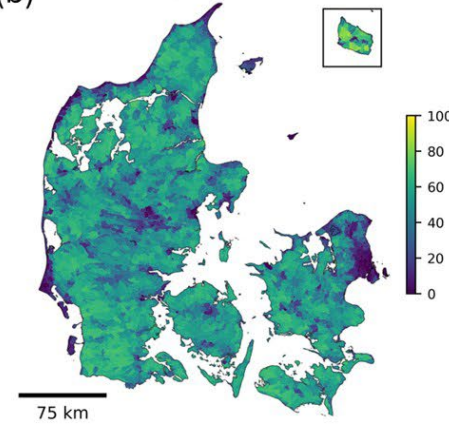


KEY DRIVERS OF TON IN STREAMS IN DENMARK

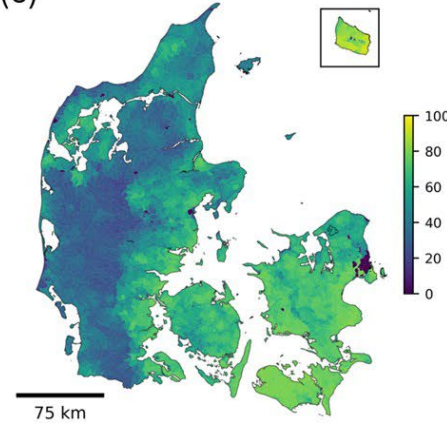
(a) DEM: Mean elevation (m)



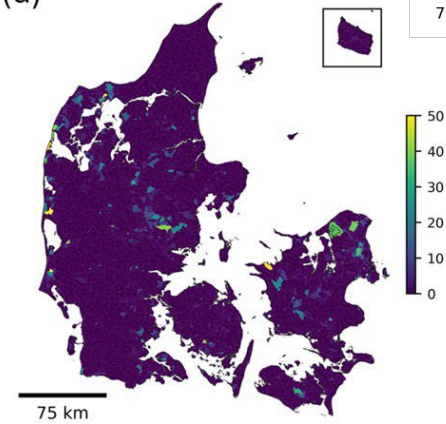
(b) AGRI: Fraction of agriculture (%)



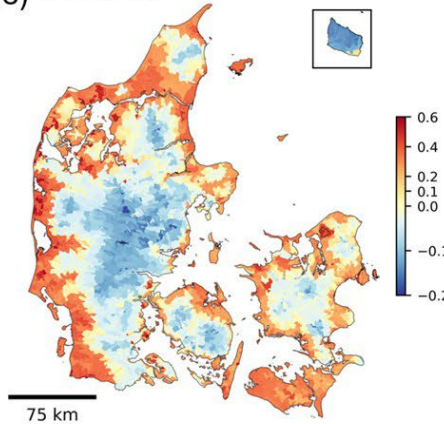
(c) TILEDRAIN: Fraction of tile drained land (%)



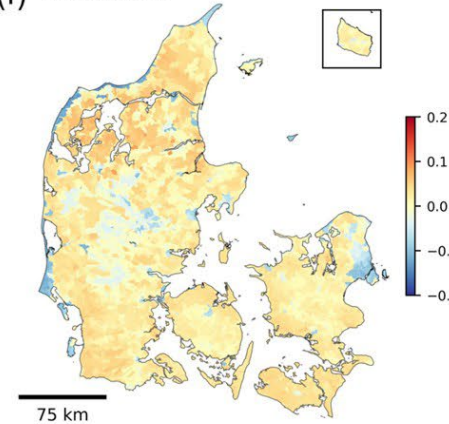
(d) LAKE: Fraction of lakes (%)



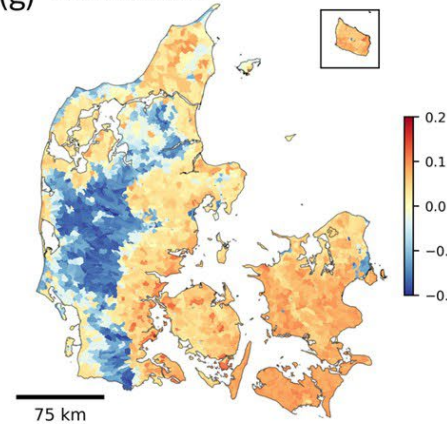
(e) DEM: SHAP value



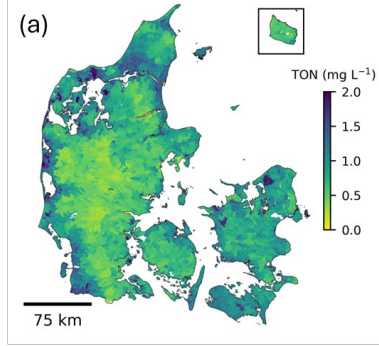
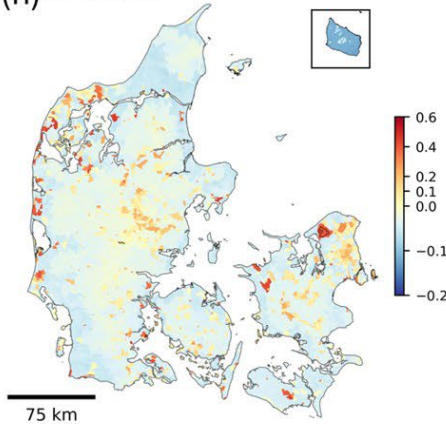
(f) AGRI: SHAP value



(g) TILEDRAIN: SHAP value



(h) LAKE: SHAP value



SUMMARY

Conclusions:

- ❖ We have a model for predicting total organic nitrogen in streams with reasonable accuracy, precision, and transferability
- ❖ The four most significant variables are: terrain elevation, agricultural area, drained area, and lakes

Remember:

- ❖ Garbage in = garbage out! (Sampling and data processing are crucial!)

Perspectives:

- ❖ Direct measurements of TON instead of indirect measurements
- ❖ Measurements of PON and DON and different flow paths
- ❖ Inclusion of more and/or better predictor variables such as crop type distribution, soil organic N content and shallow geology



Thank you for listening

contact: rf@ecos.au.dk



QR CODE FOR PAPER:

—
Feel free to take a picture 😊





AARHUS
UNIVERSITY