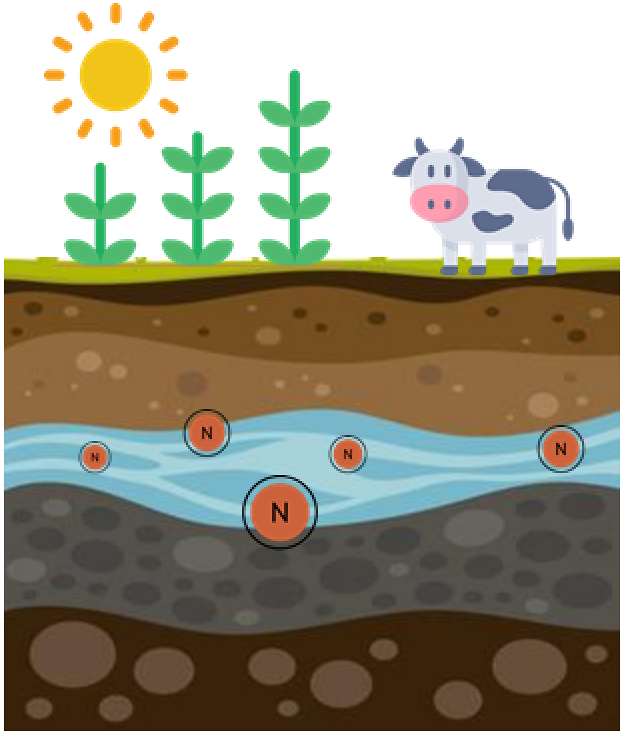
Spatial Modeling of Nitrate Leaching in the Netherlands

AI-driven Data Engineering and Reusability for Earth and Space Sciences (DARES) @ ECAI 2025

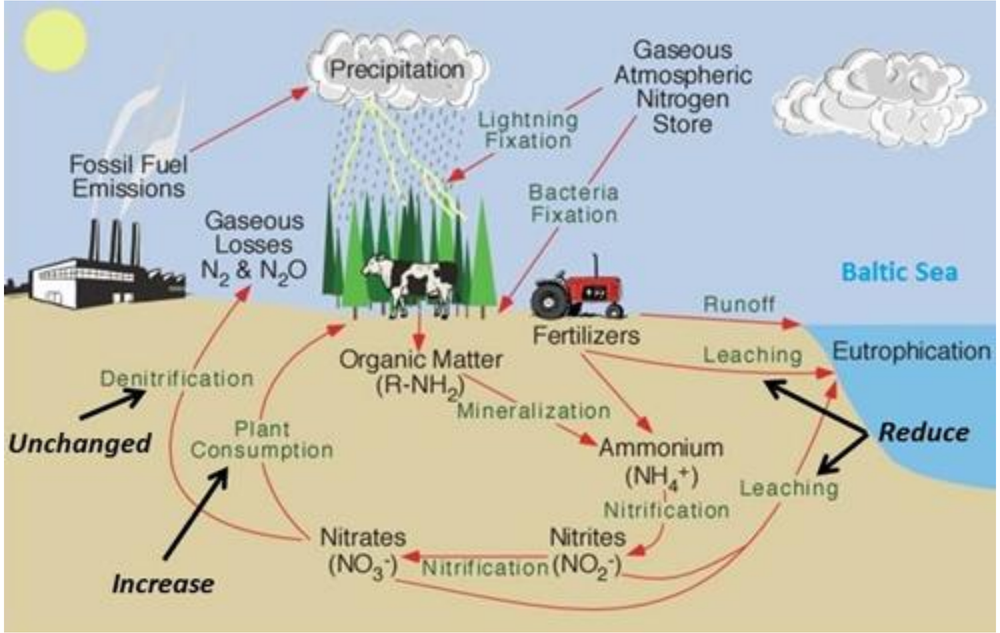


Authors: Iulia Capralova, Juan Cardenas-Cartagena.

1. Motivation

Nitrate leaching → The process of **washing nitrogen out** through the soil with water.

→ Happens when the soil contains **more nitrate than plants can absorb**.



Source: https://ragnabodata.se/drainage_irrigation/nitrogencycle.html

1. Motivation

Nitrate leaching → The process of **washing nitrogen out** through the soil with water.

→ Happens when the soil contains **more nitrate than plants can absorb**.

Why is it a problem?

- Leads to **groundwater pollution**
- Causes algal blooms and fish die-off
- Risks **public health** (e.g., blue baby syndrome)

Algal bloom

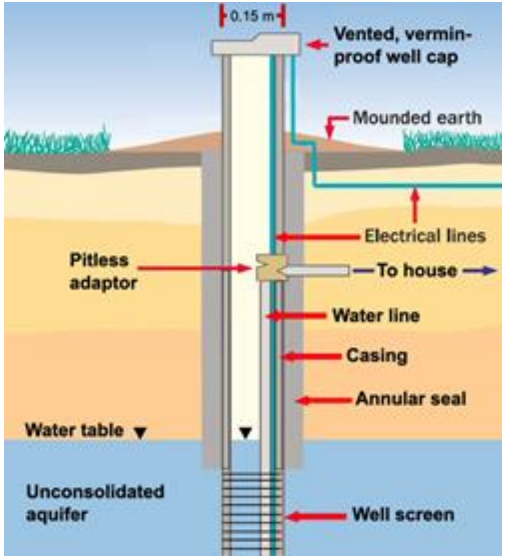
Source:
https://nypost.com/2018/06/22/toxic-algae-blooms-becoming-more-common-across-us/

1. Motivation

Nitrate leaching → The process of **washing nitrogen out** through the soil with water.

→ Happens when the soil contains **more nitrate than plants can absorb**.

→ Can be monitored by **sampling groundwater** through monitoring wells.

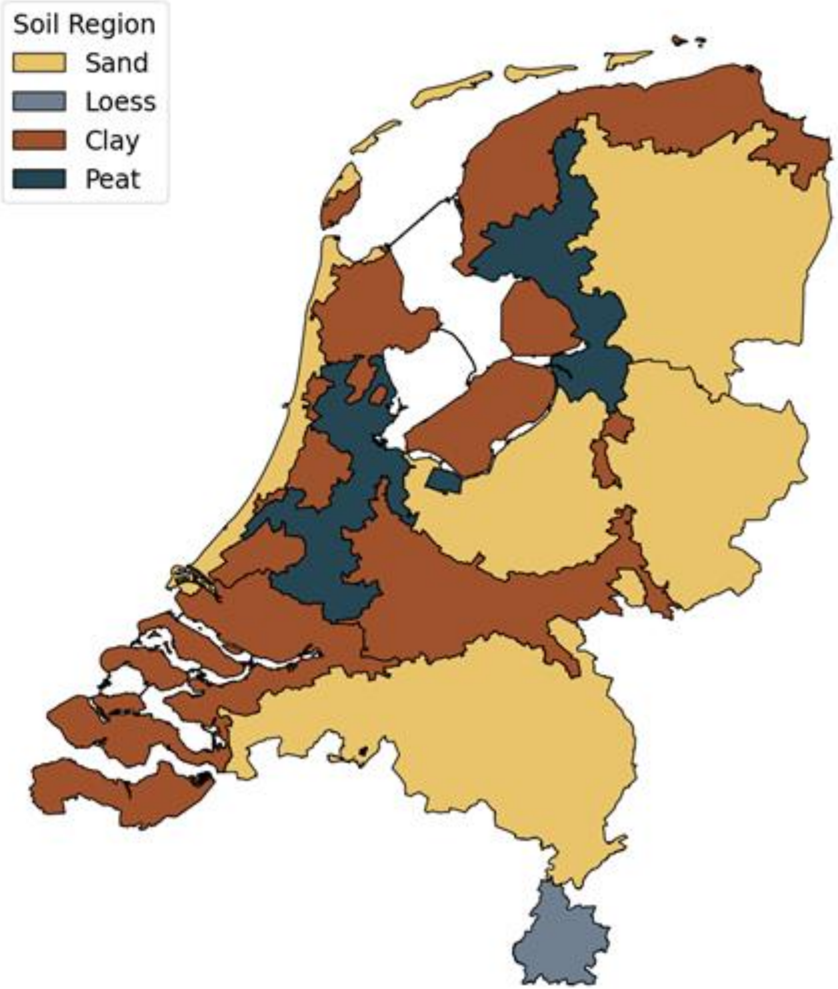


Source: https://books.gw-project.org/domestic-wells-introduction-and-overview/chapter/drilled-wells/

1. Motivation

Relevance for the Netherlands:

- **66%** of the land area is used for **agriculture** [1]
- Highest **livestock density** of EU countries [2]
- High **fertilizer** usage [3]
- **Sandy soils** (50% of area) provoke nitrate leaching [4, 5]



1. Motivation

Objective:

- To develop an **explainable spatial regression model** for predicting nitrate concentrations in groundwater in **the Netherlands**, using spatial and environmental factors.

1. Motivation

Objective:

- To develop an **explainable spatial regression model** for predicting nitrate concentrations in groundwater in **the Netherlands**, using spatial and environmental factors.

Challenges:

- Monitoring wells are **sparse** → many areas with **no direct data**
- Groundwater sampling is **expensive** → **low sampling frequency**

2. Literature review

> *"Everything is related to everything else, but near things are more related than distant things."*
> — *Waldo Tobler's First Law of Geography (1970)*

Traditionally **spatial** and **temporal autocorrelation** is used.

2. Literature review

> *"Everything is related to everything else, but near things are more related than distant things."*
> — *Waldo Tobler's First Law of Geography (1970)*

Relying on geographic proximity and may omit environmental interactions involved in nitrogen cycle:

- ❖ **Spijker et al. (2022)**: → Introduced environmental variables

  → Developed **Random Forest model**

2. Literature review

> *"Everything is related to everything else, but near things are more related than distant things."*
>
> — *Waldo Tobler's First Law of Geography (1970)*

Relying on geographic proximity and may omit environmental interactions involved in nitrogen cycle:

❖ **Spijker et al. (2022)**: → Introduced environmental variables

→ Developed **Random Forest model**

Limitations addressed:

a. **Limited temporal scope** — only year 2017 was analyzed

b. Missing key nitrogen cycle factors – explained **58% of variance**

c. Used only **model-specific** interpretability

3. Methods

a) Study Area & Time Period

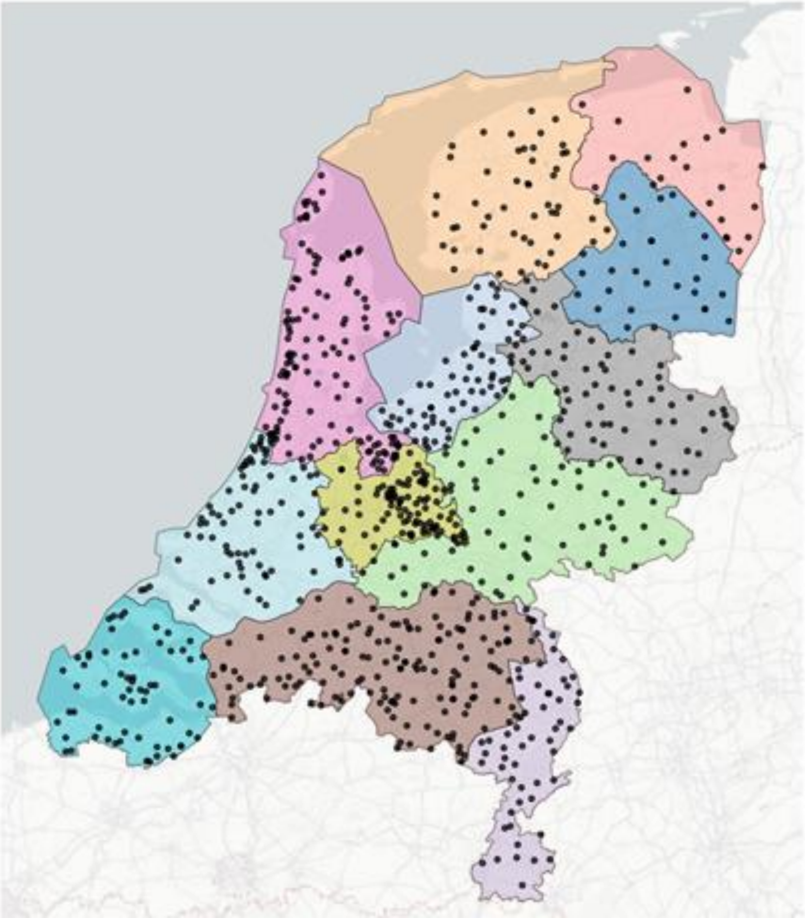
- 12 provinces
- City regions are excluded
- Years: 2008 - 2023

3. Methods

b) Target variable

- **Nitrate** concentration in groundwater
- ~870 locations
- Sampled about twice a year
- Source:

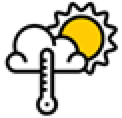
Basic Subsurface Registration (BRO)



3. Methods

c) Explanatory variables

Time Series



Temperature (KNMI)

Precipitation (KNMI)

Groundwater table (BRO)

Spatial

Population density (CBS)

Elevation (AHN)

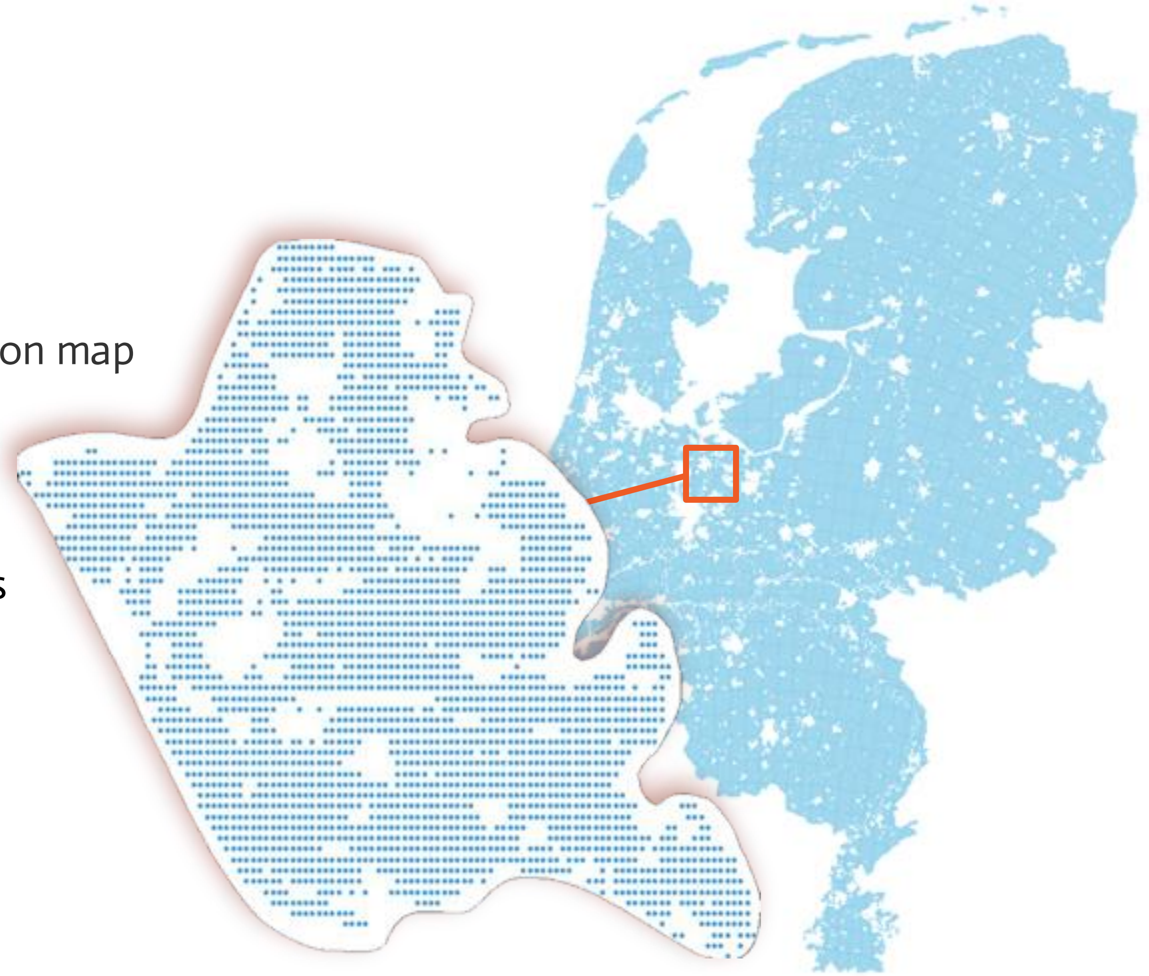
Land use (WER)

Soil and geochemical properties (BRO)

3. Methods

d) Prediction grid

- Nationwide 500 × 500 m prediction map
- Excluding city regions
- Each point → location + date + environmental factors
- **Years**: 2010 | 2017 | 2021 | 2023



3. Methods

e) Data Pre-processing

Sources:

KNMI

WAGENINGEN
UNIVERSITY & RESEARCH

Actueel Hoogtebestand
Nederland

Centraal Bureau
voor de Statistiek

Cleaning → Aligning → Feature engineering

3. Methods

f) Model selection

Ridge Regression

- Linear model with **L^2 regularization** to prevent overfitting
- Interpretable by design

Random Forest

- Ensemble of decision trees – non-linearity
- **Model-specific interpretability** (Gini impurity reduction)

XGBoost (Extreme Gradient Boosting)

- Sequentially adds trees to correct previous errors
- **Model-specific interpretability** (Loss-reduction gain)

3. Methods

f) Model selection

Ridge Regression

- Linear model with **L^2 regularization** to prevent overfitting
- Interpretable by design

Random Forest

- Ensemble of decision trees – non-linearity
- **Model-specific interpretability** (Gini impurity reduction)

XGBoost (Extreme Gradient Boosting)

- Sequentially adds trees to correct previous errors
- **Model-specific interpretability** (Loss-reduction gain)

❖ **Model-agnostic** interpretability applied across all models.

❖ **Best-performing** models combined in **Ensemble prediction.**

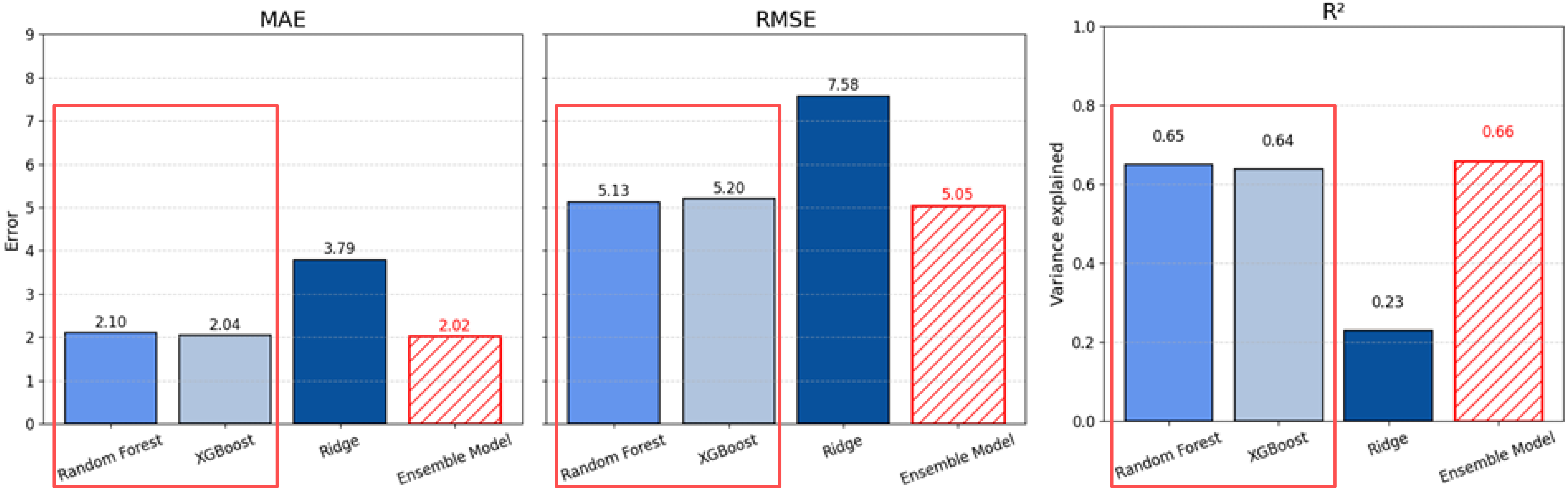
3. Methods

g) Metrics

- Coefficient of Determination **(R^2)** → Goodness of fit
- Mean Absolute Error **(MAE)** → Average error
- Root Mean Squared Error **(RMSE)** → Impact of large mistakes

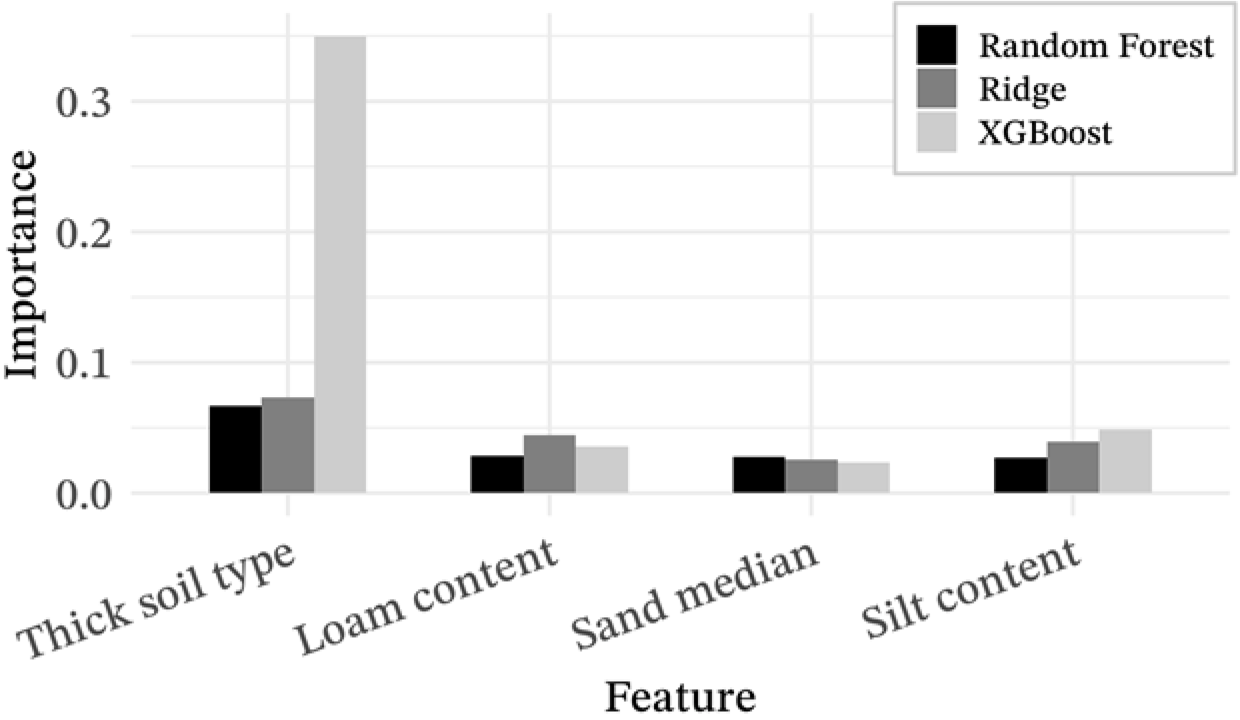
4. Results

a) Performance evaluation



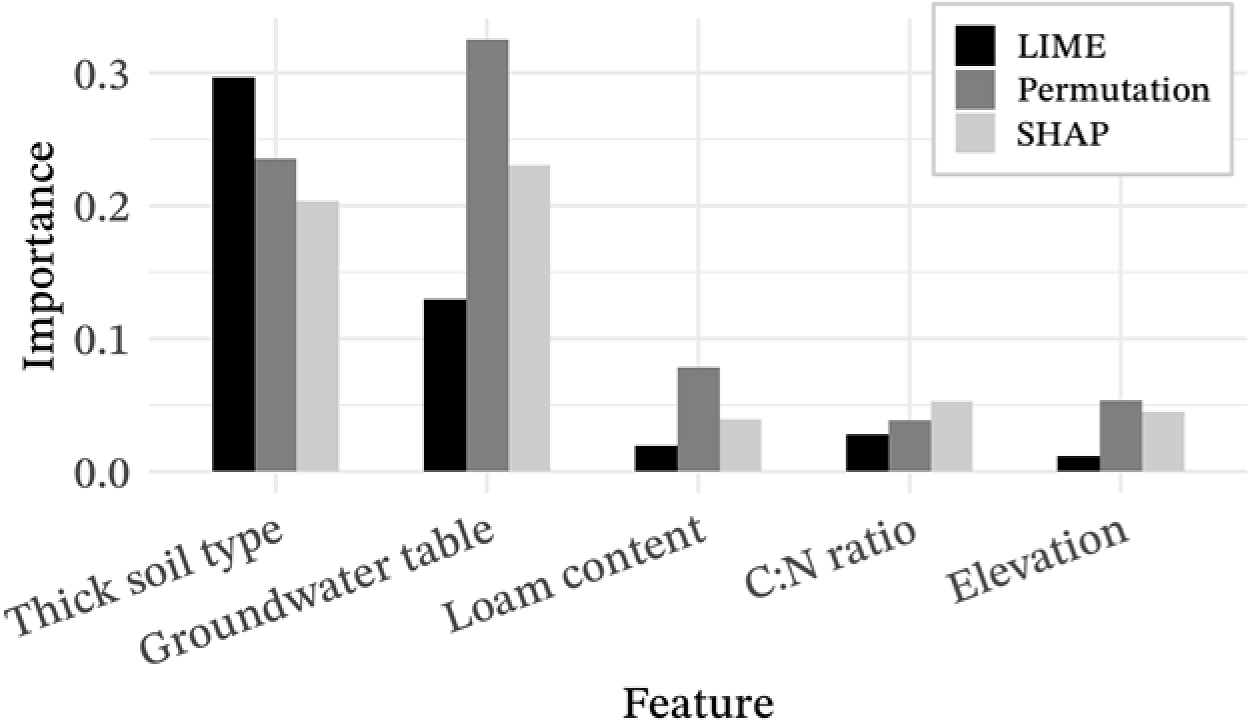
4. Results

b) Feature Importances - Model Specific



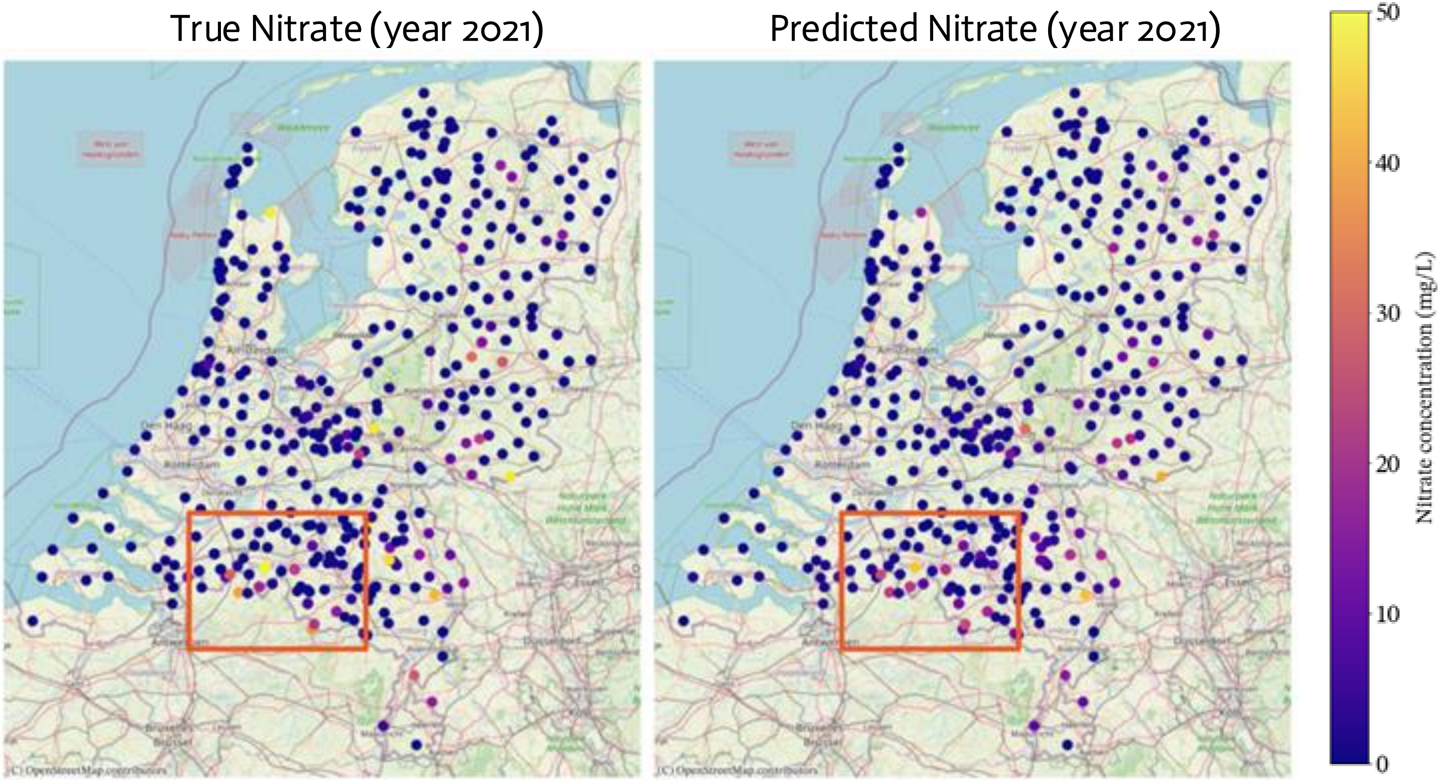
4. Results

c) Feature Importances - Model Agnostic



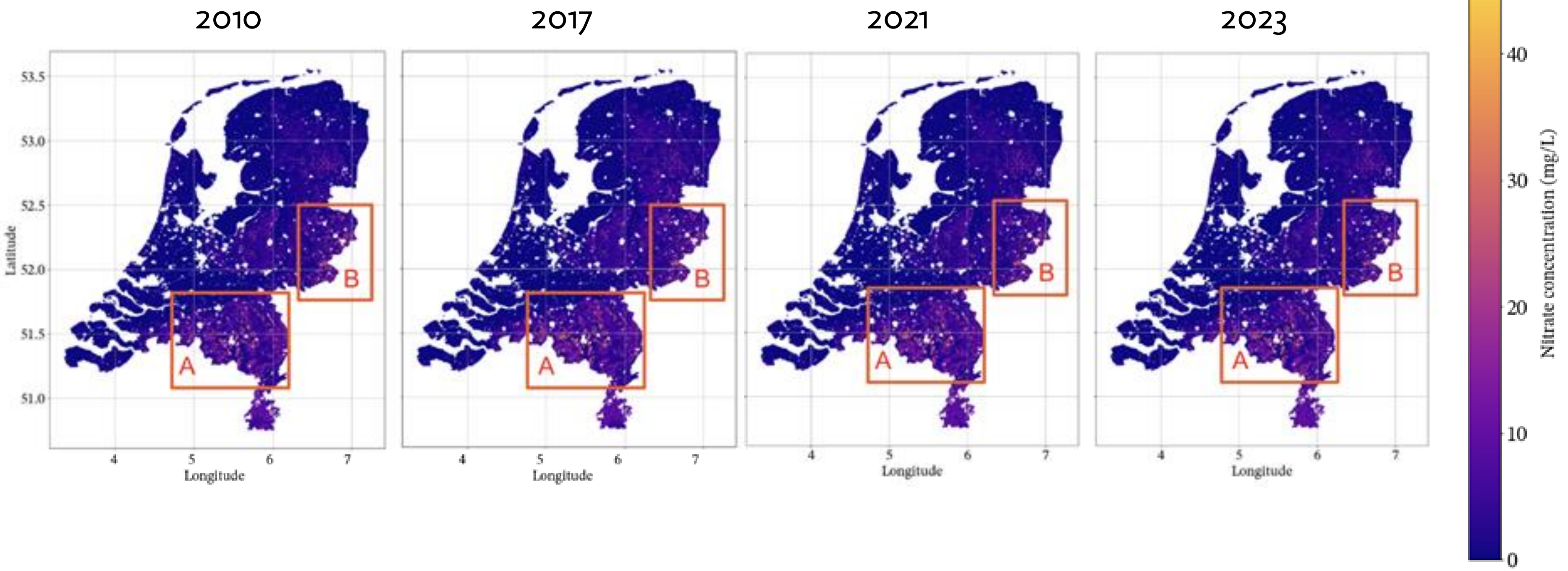
4. Results

d) Test set predictions (year 2021)



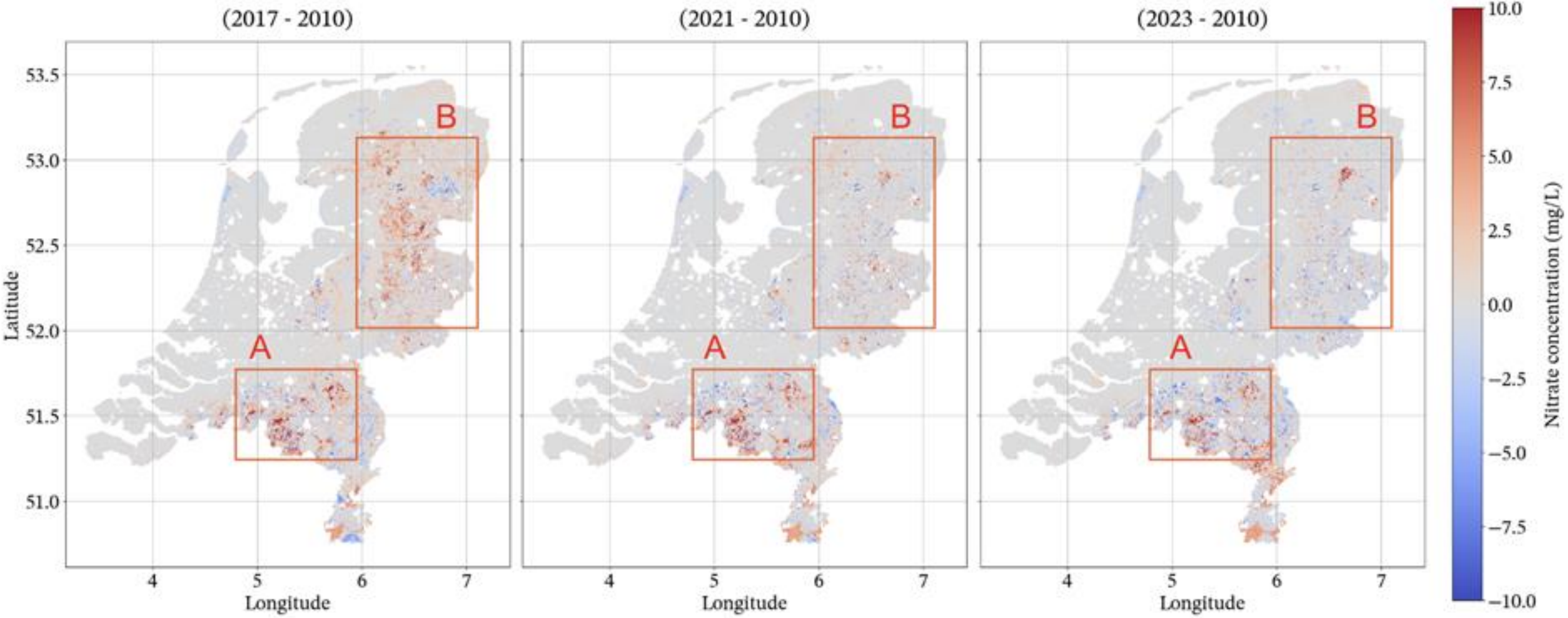
4. Results

e) Grid predictions



4. Results

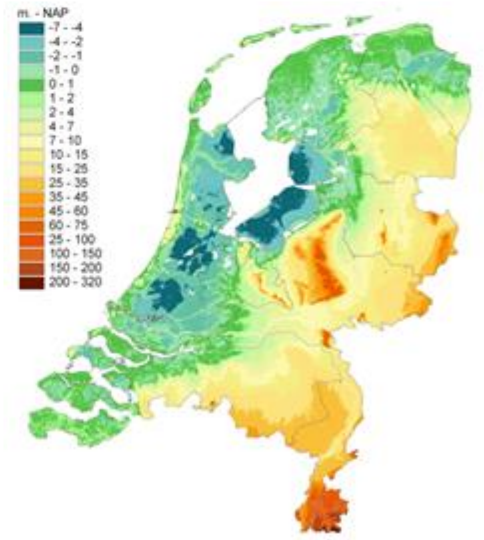
f) Changes over years



5. Discussion

a) Feature importance insights

- **Thick earth soils**
  - Sandy, highly permeable
- **Low loam content**
  - Low values → unbalanced soil
- **Deeper groundwater tables**
  - Proxy for elevation



Elevation

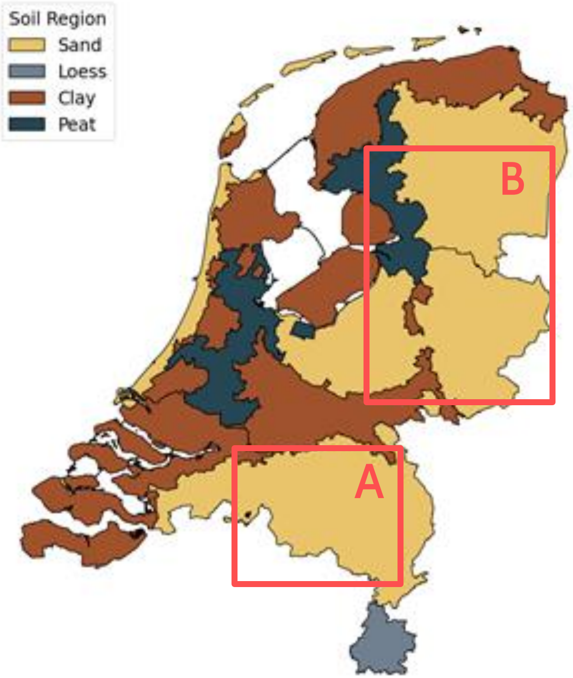
Soil type

5. Discussion

b) Spatial trends

- **High nitrate areas** → nitrogen surplus after **year 2015** [10]
  - **Region A:** dry, sandy soils, and low share of grasslands.
  - **Region B**: more **grasslands**, and higher proportion of **peat.**

- **Anomaly:** northern Drenthe (2023) → study limitation.



5. Discussion

c) Limitations

- Spatial mismatch between groundwater & nitrate datasets ($r \approx 0.6$).
- Models **underpredict nitrate values** above **6 mg/L** due to skewed nitrate dataset
- ~34 % of variance remains **unexplained** → Lack of fertilizer input data

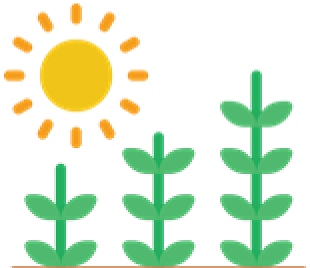
6. Conclusion

- **Random Forest** and **XGBoost** outperformed Linear Regression
- **Ensemble model** performed best → **R^2 = 0.66**
- **Key predictors:** thick soil types, loam content, and groundwater depth.
- **Highest nitrate risk:** South and North-East of the Netherlands.

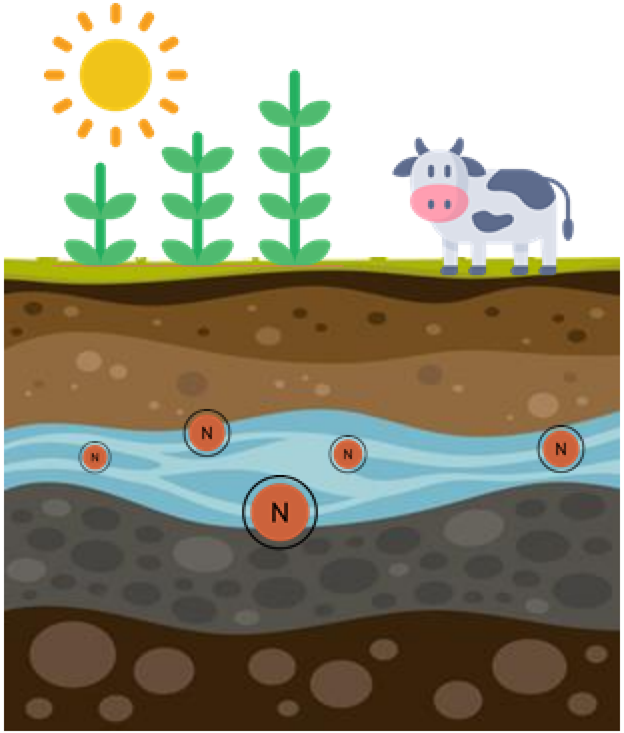
Future work:

- **Add agricultural practice, and fertilizer data**
  → Capture finer-scale variability and improve predictions.

Thank you!



- Iulia Capralova
  - i.capralova@student.rug.nl
- Juan Cardenas-Cartagena
  - j.d.cardenas.cartagena@rug.nl



References

1. European Commission. (2025). *CAP Strategic Plan: Netherlands*. Retrieved from https://agriculture.ec.europa.eu/cap-my-country/cap-strategic-plans/netherlands_en
2. Eurostat. (2022). *Agri-environmental indicator – livestock patterns*. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Agri-environmental_indicator_-_livestock_patterns
3. Eurostat. (2024). *Fertiliser consumption (aei_fm_usefert)*. Retrieved from https://ec.europa.eu/eurostat/databrowser/view/aei_fm_usefert/default/bar?lang=en
4. Delta Programme. (2024). *Elevated Sandy Soils*. Retrieved from https://english.deltaprogramma.nl/areas/elevated-sandy-soils
5. Van Drecht, G. (1993). Modelling of regional scale nitrate leaching from agricultural soils, The Netherlands. *Applied geochemistry*, *8*, 175-178.
6. Verloop, J., van den Brink, C., & Gielen, J. (2025). *Effectiveness of Voluntary Nutrient Management Measures to Reduce Nitrate Leaching on Dairy Farms Using Soil N Surplus as an Indicator*. *Water*, *17* (3), 455.
7. D'Agostino, V., Greene, E. A., Passarella, G., & Vurro, M. (1998). Spatial and temporal study of nitrate concentration in groundwater by means of coregionalization. *Environmental geology*, *36*(3), 285-295.
8. Boumans, L. J., Fraters, D., & Van Drecht, G. (2005). Nitrate leaching in agriculture to upper groundwater in the sandy regions of the Netherlands during the 1992–1995 period. *Environmental Monitoring and Assessment*, *102*, 225-241.
9. Spijker, J., Fraters, D., & Vrijhoef, A. (2021). A machine learning based modelling framework to predict nitrate leaching from agricultural soils across the Netherlands. *Environmental Research Communications*, *3*(4), 045002.
10. B. Fraters, A. Hooijboer, A. Vrijhoef, A. Plette, N. Van Duijnhoven, J. Rozemeijer, M. Gosseling, C. Daatselaar, J. Roskam, H. Begeman, Agricultural practices and water quality in the netherlands; status (2016-2019) and trend (1992-2019) (2021).