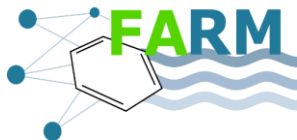


Assessment of groundwater vulnerability to pesticides in Germany – findings from harmonised monitoring data and machine learning

Dr. Anne-Karin Cooke, Dr. Kristof Dorau, Ephraim Erkens, Felix Joger, Dr. Sandra Willkommen,
Dr. Stefan Broda – BGR

Dr. Wolfram König, Mark Rauch, Stephan Marahrens – UBA



Cremona, 03.09.2026

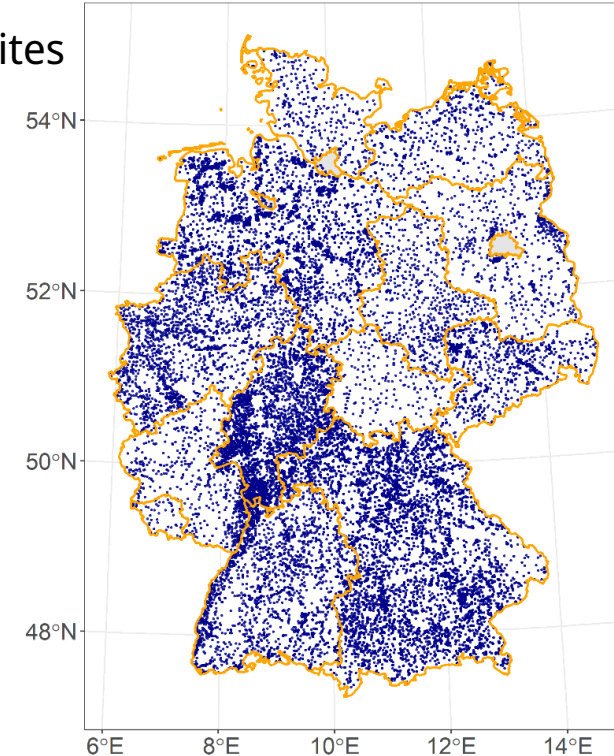
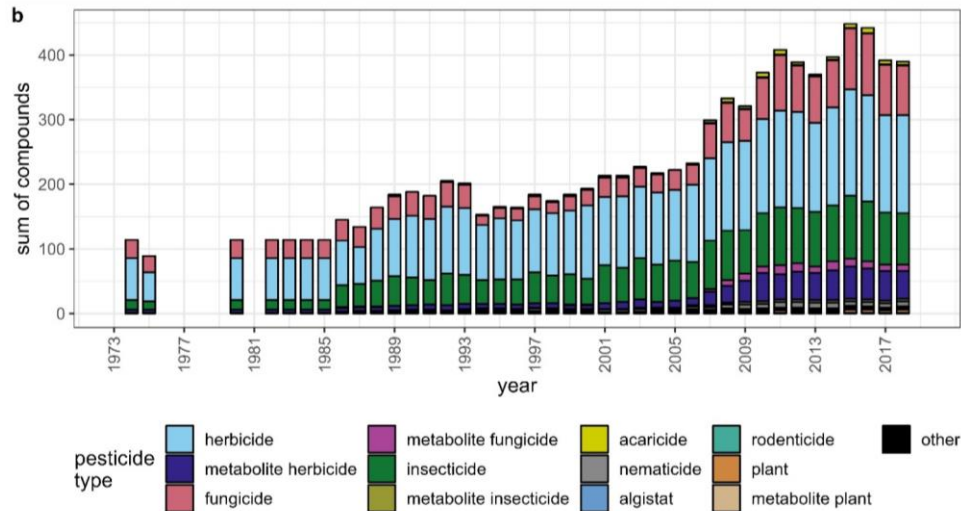
1. Rationale & Objectives



- Assembling and harmonising nationwide dataset based on groundwater monitoring data and additional variables (geospatial data etc.)
- Data-driven approach to assessing groundwater vulnerability to active substances and metabolites based on public monitoring data
- Development of vulnerability index for pesticides
- Comparison of monitoring data with existing vulnerability methods

2. Data – Overview & Challenges

- Monitoring data 1973 - 2021, 2010 onwards largest spatial coverage
- In total 26.192 sites with 530 active substances and metabolites
- Temporal and spatial coverage very heterogeneous
- 2007 onwards increased monitoring of metabolites



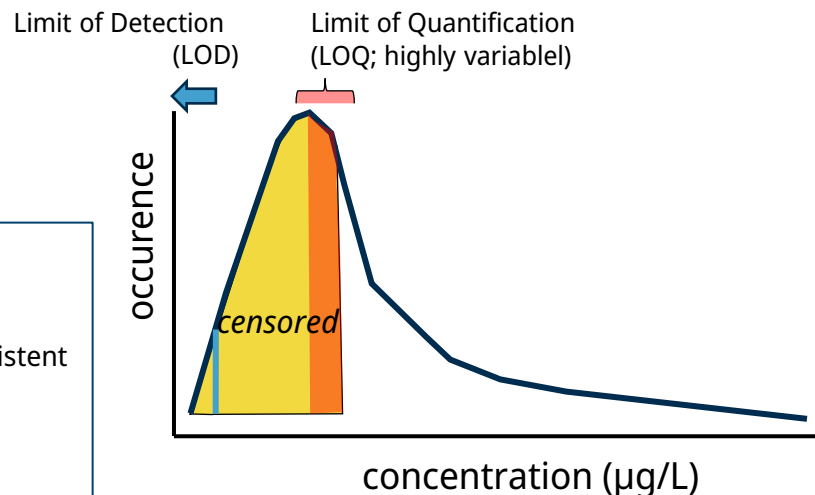
2. Data – Compound Selection

Metabolites

- Metolachlor Esa (persistent, very mobile)
- Metolachlor Oxa (persistent, very mobile)
- Metolachlor Metabolite NOA 413173 (persistent, very mobile)
- Metazachlor Esa (persistent, very mobile)
- Metazachlor Oxa (moderately persistent, very mobile)
- Desphenyl Chloridazone (persistent, mobile)
- Methyldesphenyl Chloridazone (persistent, mobile)
- Trifluoracetat (TFA) (very persistent, very mobile)

- ~97 % der of measurements below limit of quantification (LOQ) (left-censored data)

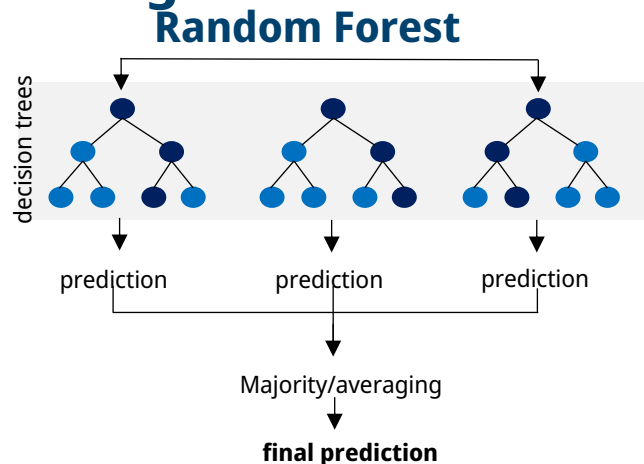
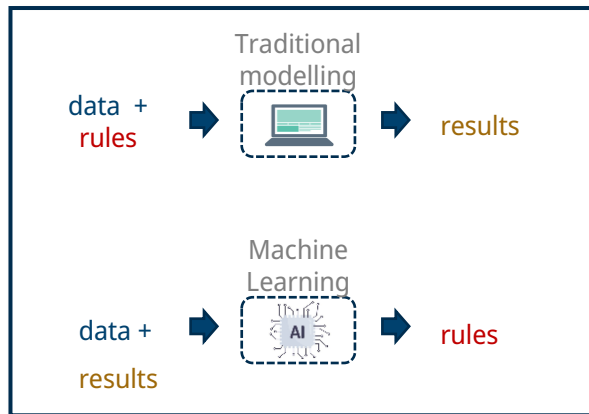
Mobility : kf_{oc} [mL g⁻¹]	Persistence: dt_{50} [Tage]
• < 30 ~ 'very mobile',	• $dt_{50} > 0$, < 30 ~ not persistent
• >30 < 100 ~ 'mobile',	• $dt_{50} > 30$, < 100 ~ moderately persistent
• > 100 < 500 ~ 'moderately mobile',	• $dt_{50} > 100$, < 356 ~ persistent
• > 500 ~ 'not mobile'	• $dt_{50} > 365$ ~ very persistent



2. Data – Target Variable & LOQ

- **Metabolites:** Average concentrations are assigned to **concentration classes**
 - Binary model
 -  1 ($x \leq 0.05 \mu\text{g/L}$)
 -  2 ($0.05 \mu\text{g/L} < x$)
 - 3 class model
 -  1 ($x \leq 0.05 \mu\text{g/L}$)
 -  2 ($0.05 \mu\text{g/L} < x \leq 0.5 \mu\text{g/L}$)
 -  3 ($0.5 \mu\text{g/L} < x$)
- **Active substance groups:** exceedance of **0.01 $\mu\text{g/L}$** by at least one active substance throughout timeseries
- **Active substances grouped by 4 mobility groups:** 1) very mobile, 2) mobile, 3) moderately mobile, 4) little mobile

2. Method - Machine Learning



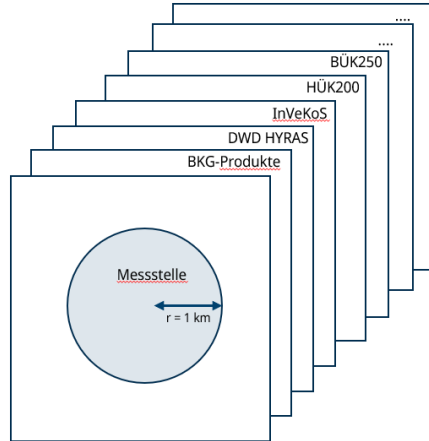
Data split



- Train-Set: Modell learns interactions from training data in an iterative way
- Validation-Set: Optimization of hyperparameter
- Test-Set: data unseen during training is used for performance evaluation

- Which compounds?
- Which target variable?
- What input data? Which explanatory variables?

3. Method – Feature Selection



Number of monitoring sites



Number of features



Test performance / Model performance

Soil

- Sand content
- Soil organic carbon content
- Macropore class

Land use

main crop area within 1 km buffer

Site

- Well screen depth [m below surface]

hydrogeology

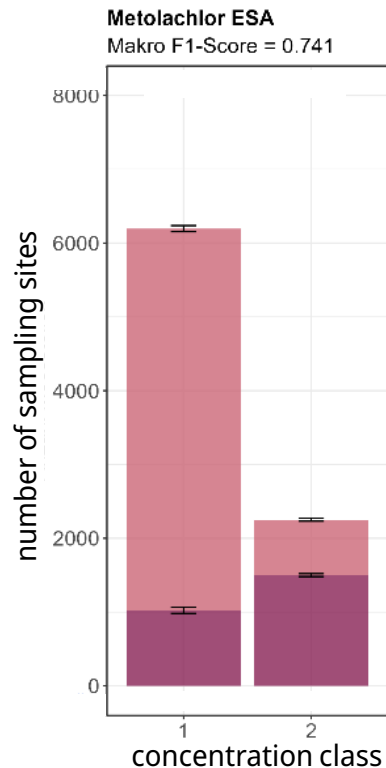
- Cavity type
- Hydraulic conductivity class

Water budget

- Mean groundwater recharge rate

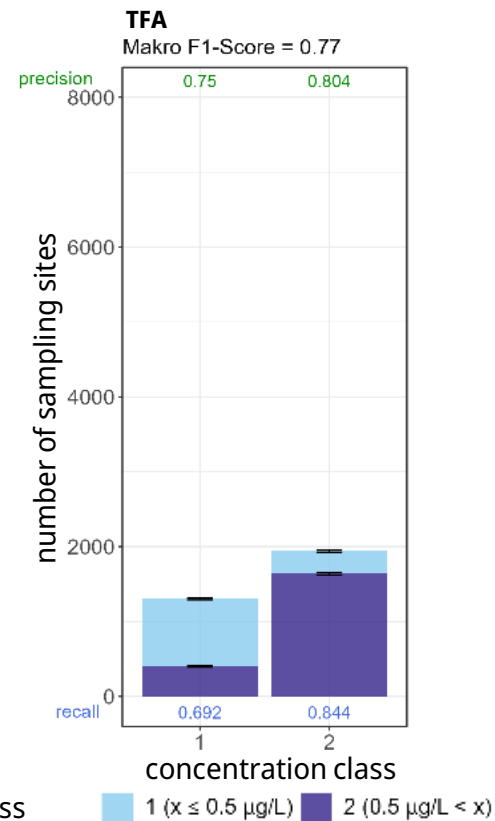
- **Selection of features relevant to the target variable, avoidance of strongly correlated features**
- **Complete data series required (no missing data)**

4. Results – Model Performance (F_1 -score)



- Metolachlor Oxa: **0.728**
- Metolachlor NOA 413173: **0.706**
- Desphenyl-Chloridazone: **0.758**
- Methyldesphenyl-Chloridazone: **0.742**
- Metazachlor Esa: **0.701**
- Metazachlor Oxa: **0.674**
- TFA: **0.77**

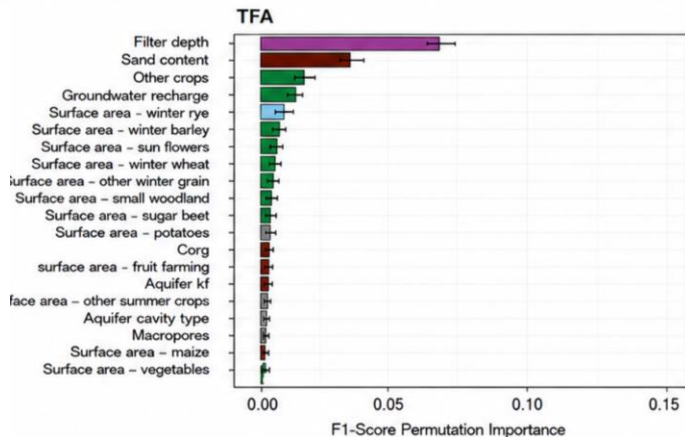
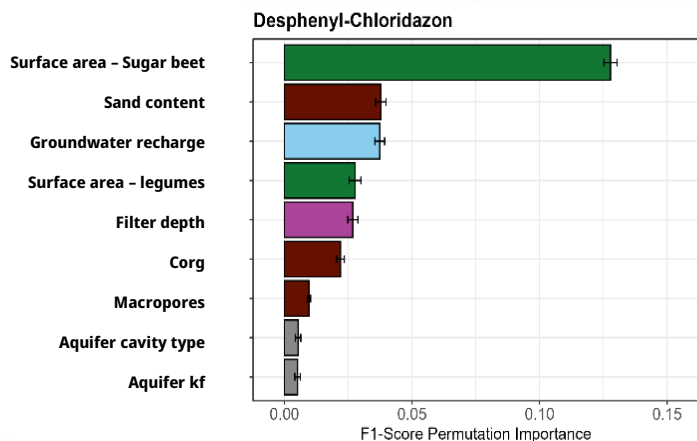
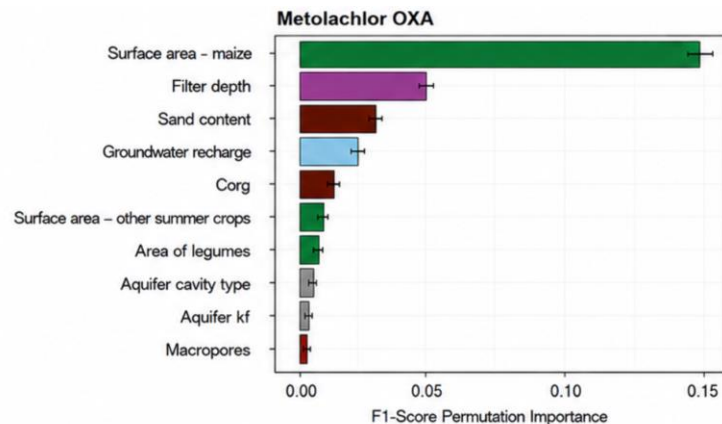
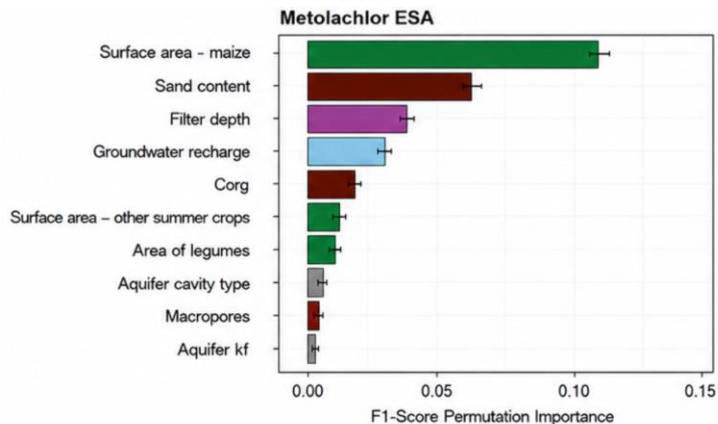
predicted class 1 ($x \leq 0.05 \mu\text{g/L}$) 2 ($0.05 \mu\text{g/L} < x$)



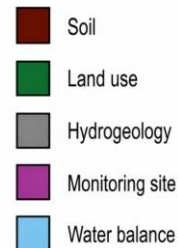
predicted class

1 ($x \leq 0.5 \mu\text{g/L}$) 2 ($0.5 \mu\text{g/L} < x$)

4. Results – Feature Importance

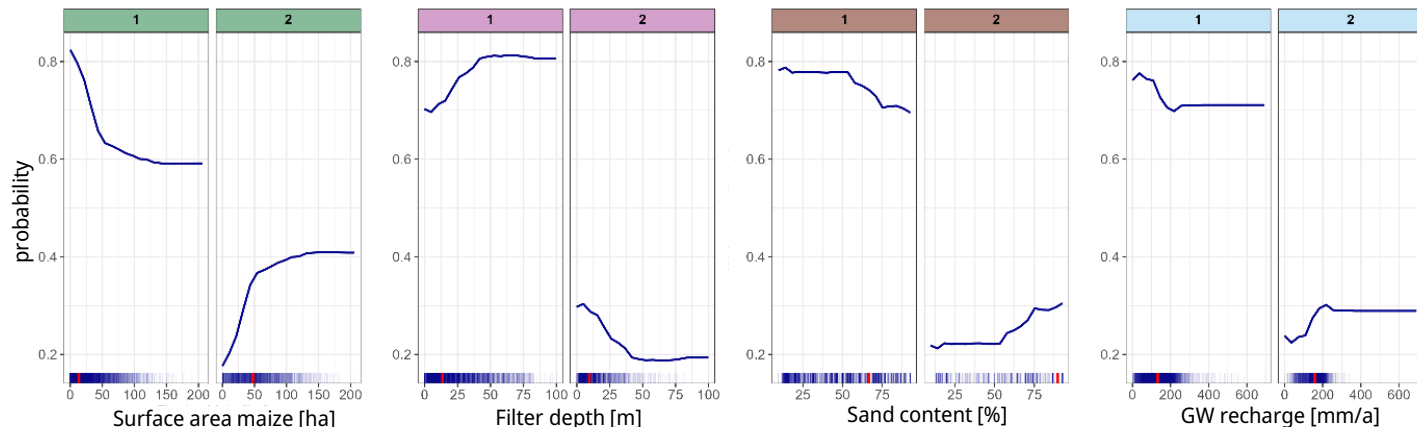


Feature category

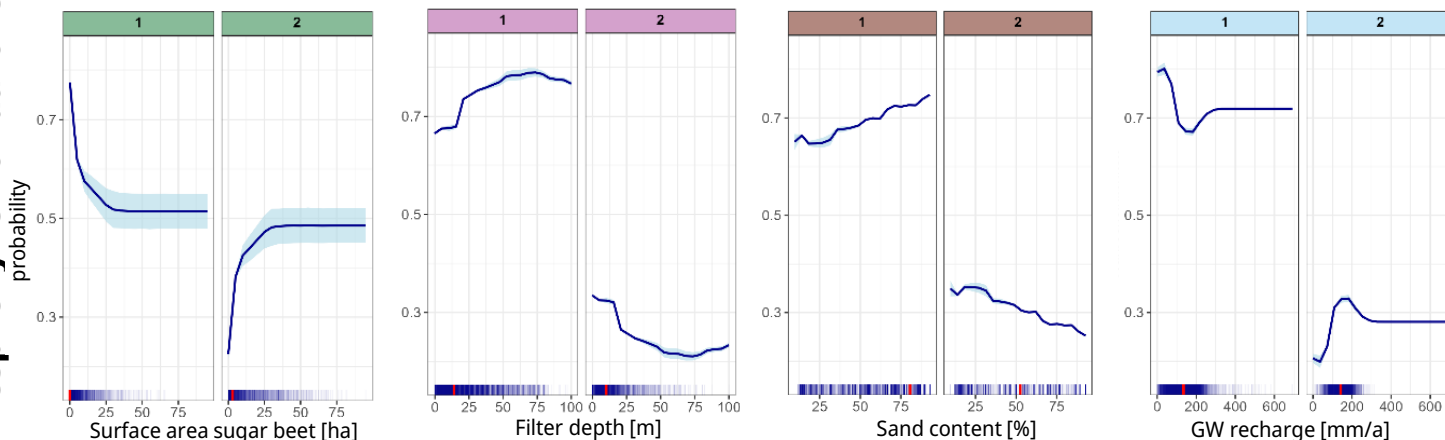


4. Results – Interpretation: Feature Effects

Metolachlor Esa



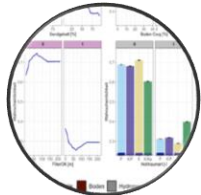
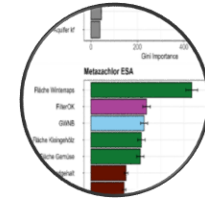
Desphenyl-Chloridazone



- Main crop: positive effects
- Well screen depth: negative effects
- Sand content (proxy soil texture) differences between metabolites
- other features show weaker effects

5. Lessons Learned – Metabolites

- Correct prediction of concentration class at majority of gw sites (F1-Score ~ 0.7)
- Area of main crop and well screen depth most important features
- Greater crop area associated with higher concentrations
- deeper well screen associated with lower concentrations
- Differences between metabolites visible: sand content, groundwater recharge, soil organic carbon
- Aquifer hydraulic conductivity, cavity type and macropore potential show weak effects
- Single compound models with acceptable model performance were only possible for a limited number of metabolites, findings cannot be directly applied to active substances or other metabolites; no development of general vulnerability index possible
- Difficult to incorporate the temporal dimension: These metabolites stem from active substance with long application history

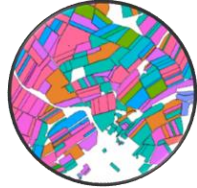


5. Lessons Learned for Vulnerability Assessment

Uncertainty possibly caused by

- Missing data on groundwater flow direction + spatial resolution of input data (geospatial maps instead of field scale data)
- Temporal aggregation of features and target variable (no short-term dynamics)

Despite these limitations the models were capable using a limited number of features to predict the concentration classes correctly in the majority of cases and to identify the main crop as the most important feature



- Process understanding or spatial effects? Models may learn spatial correlation e.g. soil properties of cultivation areas
- Feature effects of sand content as proxy for soil texture different for chloridazone metabolites, than for metazachlor or metalochlor metabolites
- Nitrate, potassium content (positive correlated) serve as a proxy for agricultural intensity

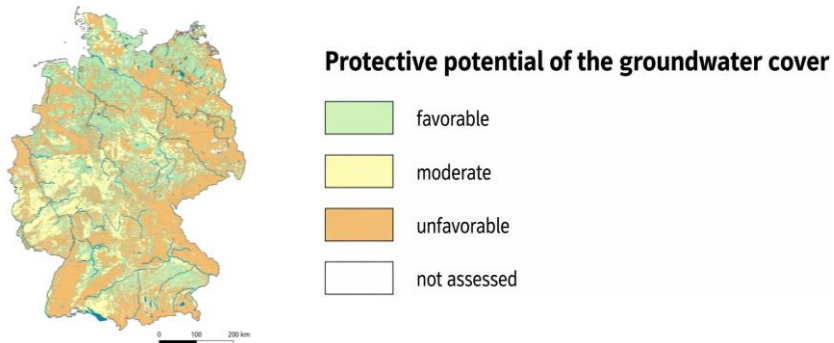
Application data required for any causal analysis + intrinsic vulnerability assessment

6. Comparison Monitoring Data with Other Approaches

1. National substance specific SDLM based on 2 PELMO versions:

- **GeoPELMO DE L:** FOCUS leaching = Matrix flow
- **GeoPELMO DE L+R+M+D:** = FOCUS Leaching + Runoff + Macropore flow + Drainage
- Simulated 80th temporal percentile concentrations were assigned to 4 **spatial percentile classes** (0-25, 25-50, 50-75, 75-100)

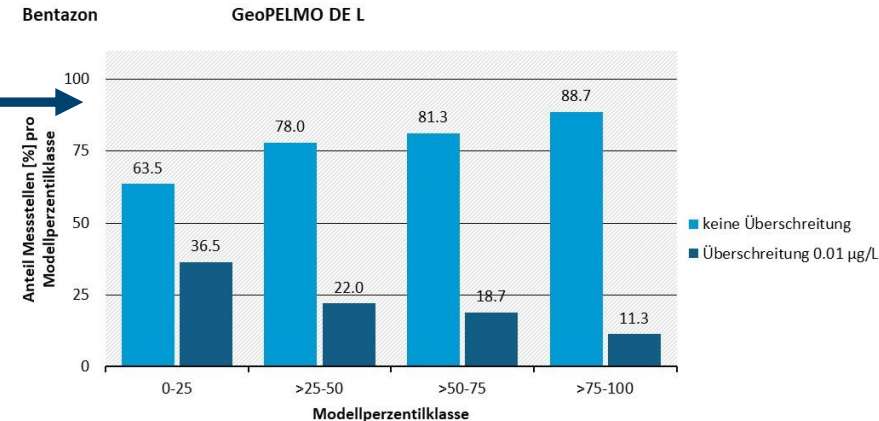
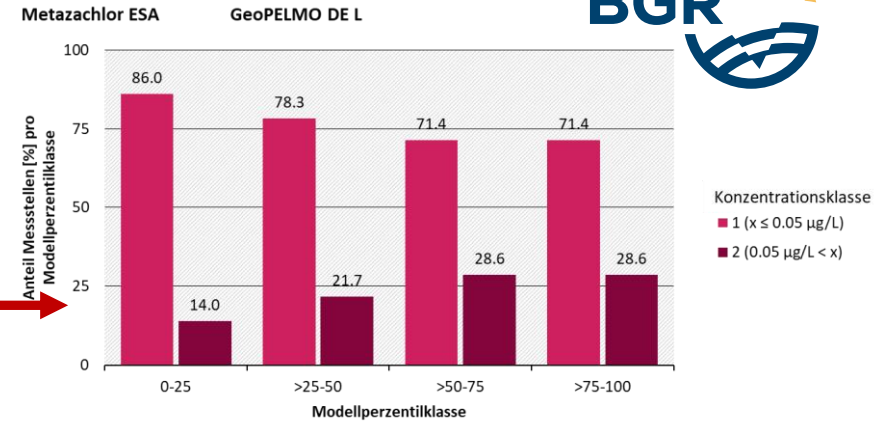
2. Intrinsic aquifer vulnerability in 3 classes (approach according to Hölting et al.)



6. Comparison – Summary

We found all cases:

- Expected relationships: more wells with detections in higher percentile classes, e.g. for **metazachlor esa** with the FOCUS version of GeoPELMO DE
- No visible correlation, e.g. for metalochlor esa and oxa
- Contrary relationship, e.g. for **bentazone**
- The comparison doesn't show any advantage of the GeoPELMO DE model with additional processes
- The same unclear relationships occur for the vulnerability classes according to Hölting et al.



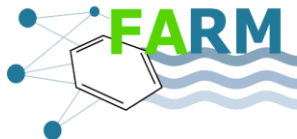
8. Key Messages



- First attempt to work with nationwide monitoring data in a data-driven approach to assess vulnerability
- Models have been trained that predict concentration classes and identify key features
- Limitations regarding data quality have been identified
- Application data is a missing link to get further with results
- Comparison between monitoring data and SDLM & Hölting et al. showed little agreement
- Recommendation for designing targeted monitoring studies: use vulnerability models not alone, combine with local data and local expert judgement

Thank you for your attention!

**Thanks to Anne-Karin
for the excellent work!**



Dr. Anne-Karin Cooke

The project "Grundwassermonitoring im ZulassungsverFAhren von PflanzenschutzMitteln" (FARM) (FKZ 3720 23 401 0) was supported by the Environmental Agency of Germany.

Full report is available: <https://doi.org/10.60810/openumwelt-8222>