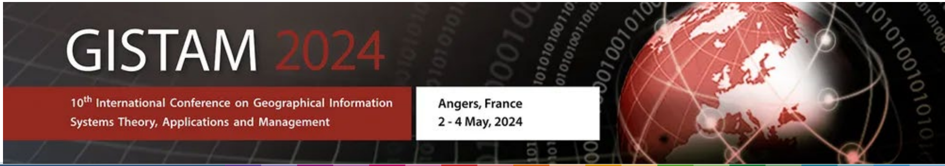
Novel Approaches to Model Assessment and Interpretation in Geospatial Machine Learning:

Addressing Spatial Dependence and High Dimensionality

Alexander Brenning

Friedrich Schiller University Jena
Department of Geography
ELLIS Unit Jena



Geospatial Machine Learning

$\mathbf{x}$

$\hat{f}$

$\hat{y}$

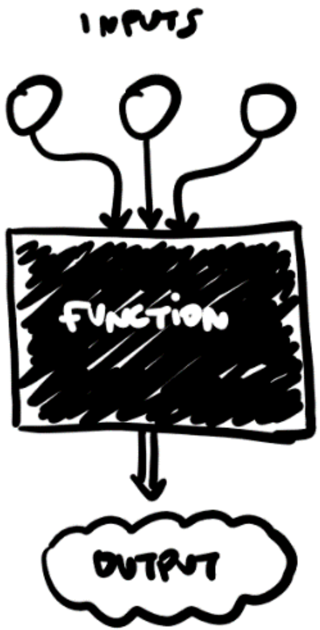


Image source: thatsoftwaredude.com

Model assessment

- Best predictive performance → reduced environmental impact, greatest societal benefit, …
- E.g. hazard assessment, regionalization of pollutants, mapping essential climate variables

Model interpretation

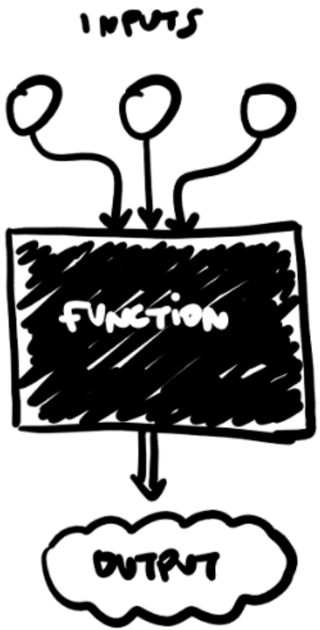
- Explainable, fair, reproducible decisions based on plausible models
- E.g. consistency with process understanding; accountability for AI-based decisions

Is This What you Want?

$\mathbf{x}$

$\hat{f}$

$\hat{y}$



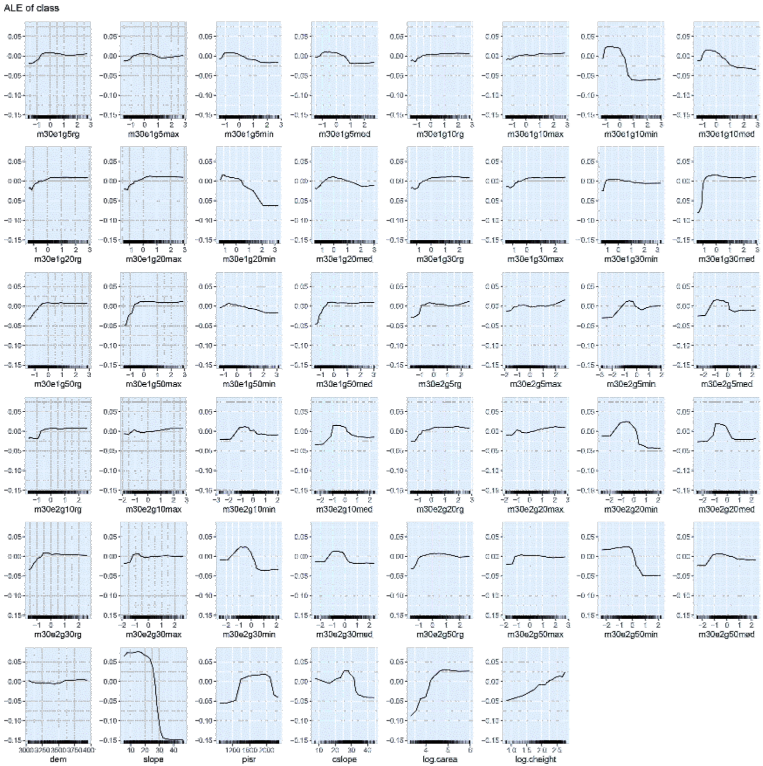


Image source: thatsoftwaredude.com

This Is What you Want!

$\mathbf{x}$

$\hat{f}$

$\hat{y}$

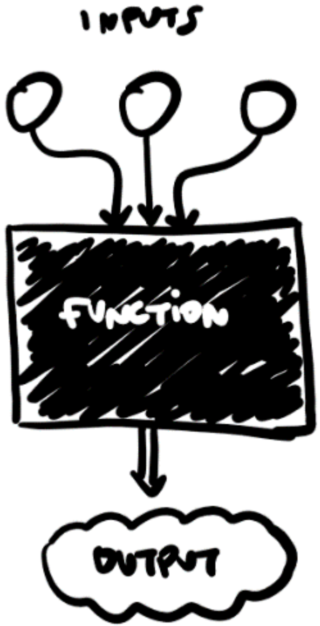
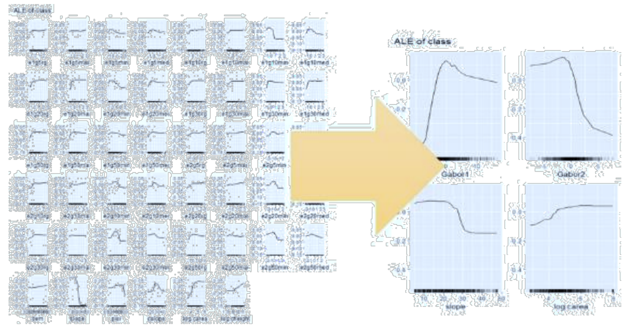


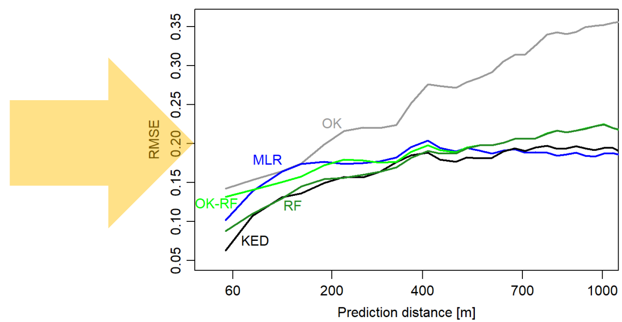
Image source: thatsoftwaredude.com

1. Interpret models in lower dimensions



2. Assess & interpret models spatially

RMSE = 0.1573



ML Model Assessment & Interpretation Challenges in Geospatial Machine Learning

$\mathbf{x}$

$\hat{f}$

$\hat{y}$

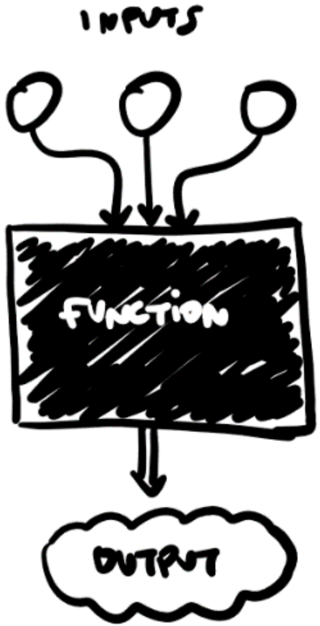


Image source: thatsoftwaredude.com

Spatial prediction

- Spatial dependence?
- Interpolation versus extrapolation skill?

Brenning (2012 in IGARSS Proc., 2023 in *IJGIS*)

High dimensionality

- Model visualization?
- Correlated features?

→ Brenning (2023) in *Machine Learning*

Model-agnostic approaches

ML Model Assessment & Interpretation
Case Studies: Pollutants in the Environment

Image credit: R. Knevels

Textbook example

- Top-soil zinc concentration
- Maas floodplain, Netherlands
- *N* = 155 with 2 predictors

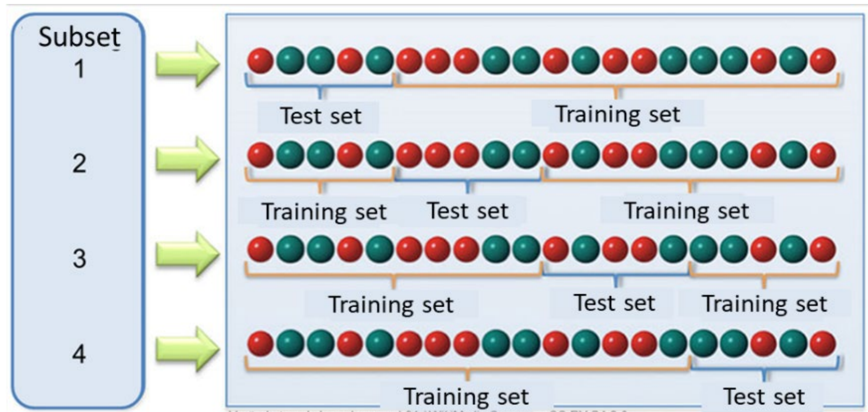
Real-world example

- Groundwater nitrate concentration (Umweltbundesamt)
- 150 km x 150 km pilot study, *N* = 471 with multiple predictors
- Countrywide analysis (*N* > 9000)

ML Model Assessment in Geospatial Machine Learning

Image credit: R. Knevels

- Design-based estimation
- Test-set estimation
- **Cross-validation**

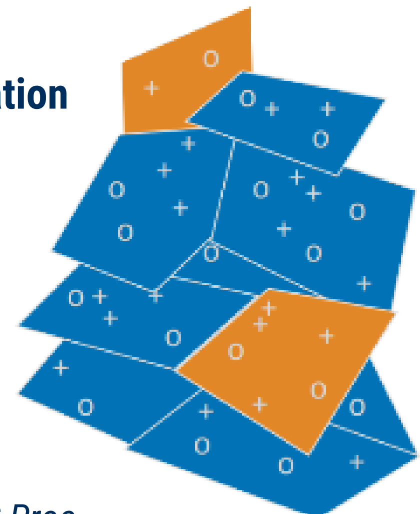


Verändert nach Joan.domenech91 / WikiMedia Commons CC-BY-SA 3.0

ML Model Assessment in Geospatial Machine Learning

Image credit: R. Knevels

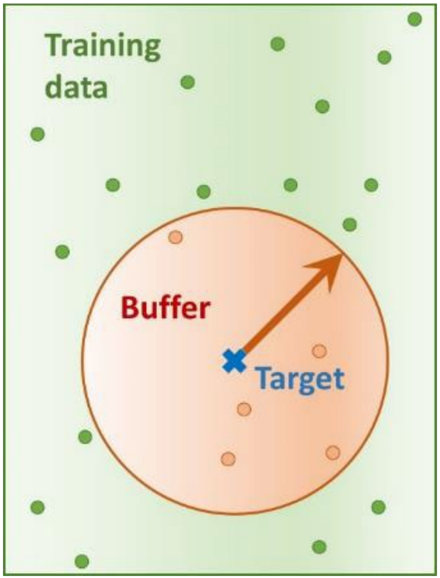
- Design-based estimation
- Test-set estimation
- Cross-validation
- **Spatial cross-validation**



Brenning (2012) in *IGARSS Proc.*
Schratz et al. (2021) in *arxiv*, in press in *J. Stat. Software*

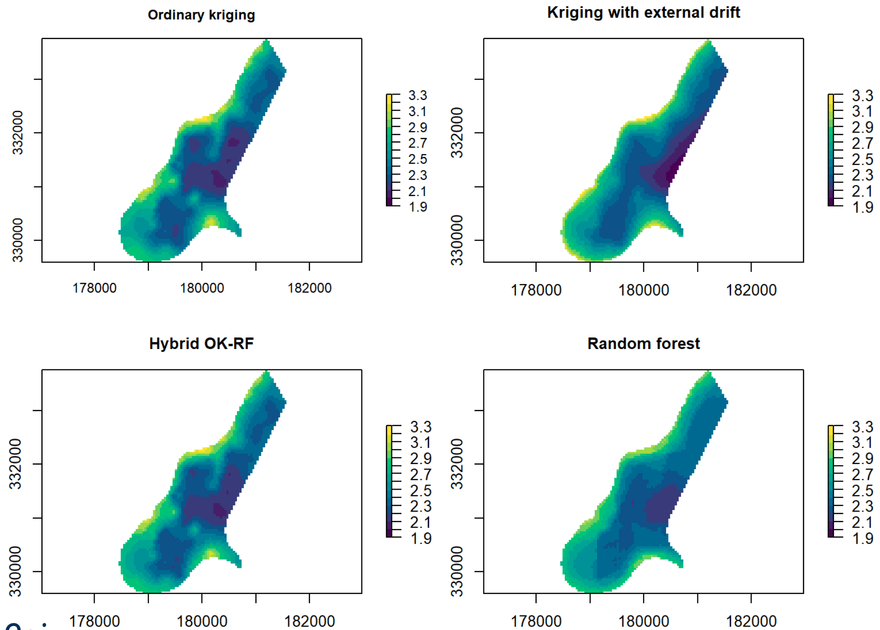
Spatial Model Assessment: *A Distance-Based Approach*

Distance-based LOO-CV



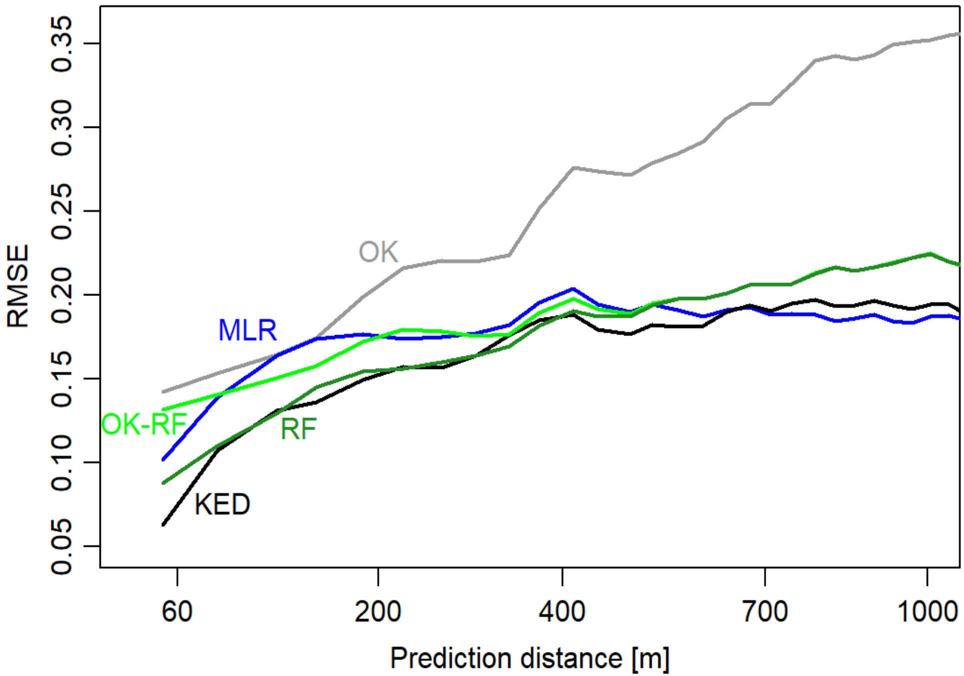
Brenning (2023) in *Int. J. Geogr. Inf. Sci.*

Sample application: Meuse data – log(Zn)



Spatial Model Assessment: *Spatial Prediction Error Profiles*

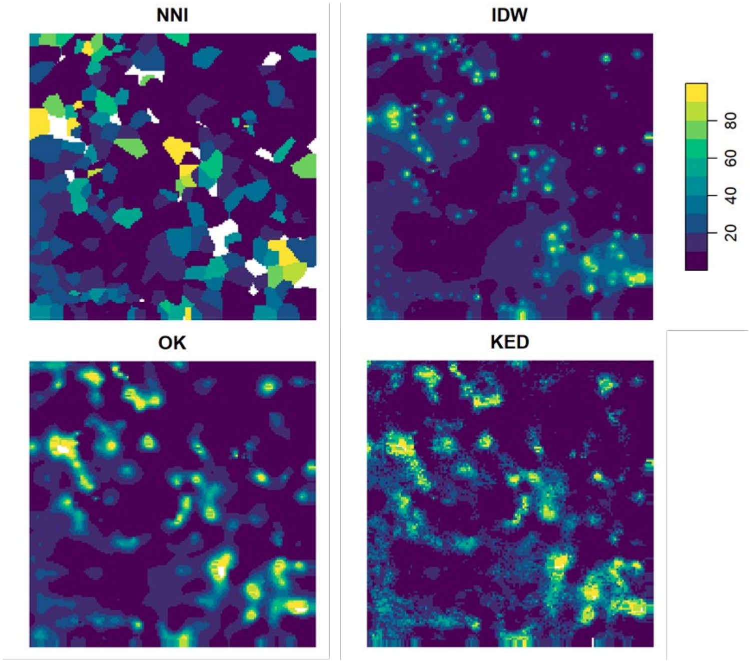
log(zinc) on the Maas Floodplain



MLR: Multiple linear regression
KED: Kriging with external drift
OK: Ordinary kriging
RF: Random forest
OK-RF: Blended OK-RF

Brenning (2023) in *IJGIS*

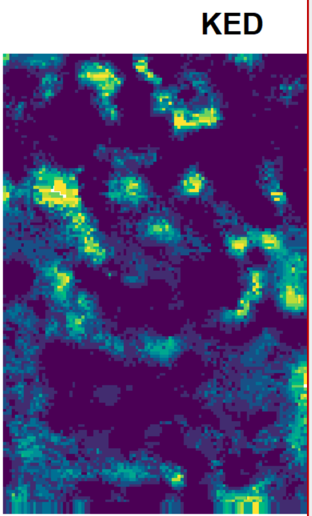
Regionalization of Nitrate in Groundwater
Deterministic and Geostatistical Interpolation Techniques



Brenning (2024) in *TEXTE*; data: Umweltbundesamt, 150 km × 150 km pilot region, N = 471

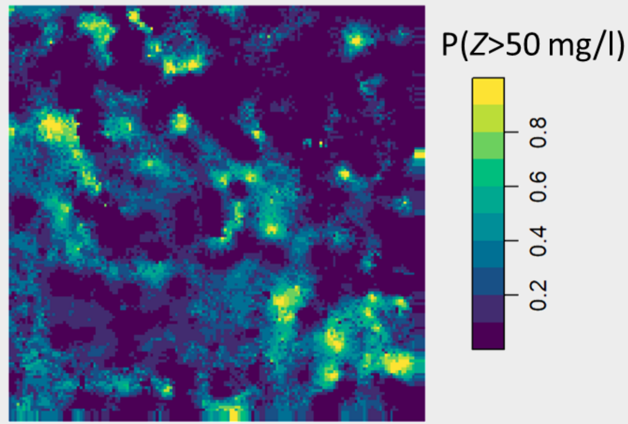
From Regionalization to Exceedence Mapping
Kriging w. External Drift & Conditional Geostat. Simulation

Predicted nitrate co



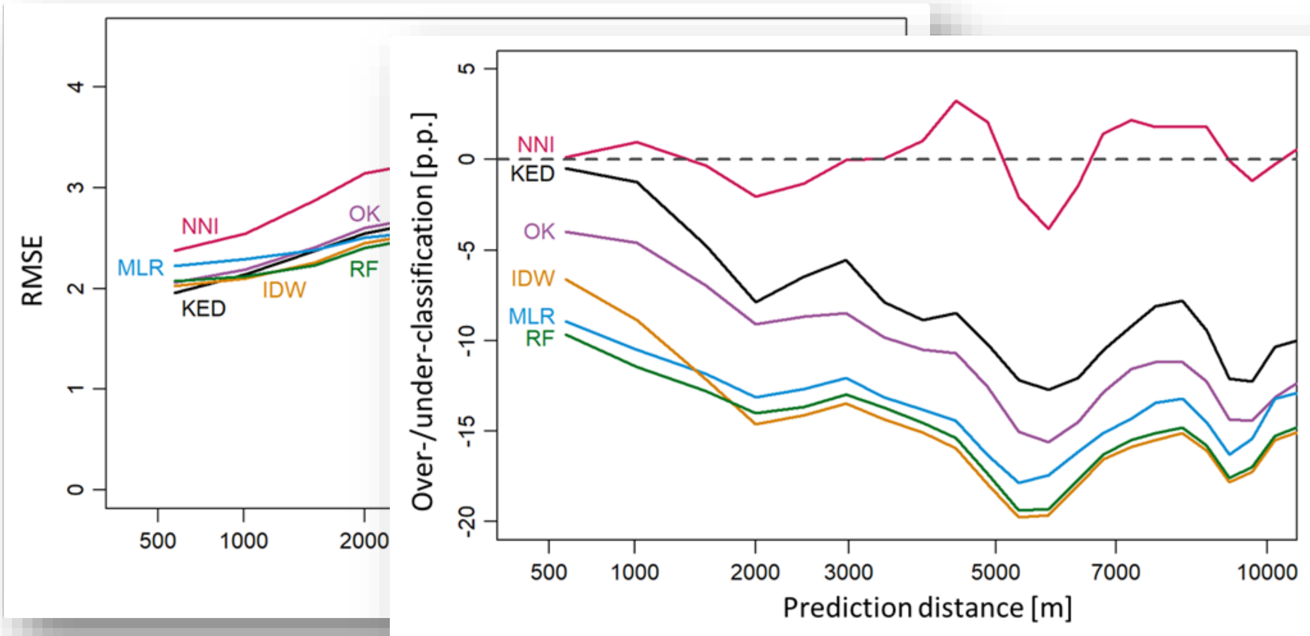
Area designation is *NOT* an interpolation task!
→ Model and estimate **exceedance regions**!

Conditional geostatistical simulation
→ 20.1% exceedance area



Brenning (2024) in *TEXTE*; data: Umweltbundesamt, 150 km × 150 km pilot region, *N* = 471

Spatial Model Assessment: ***Spatial Prediction Error Profiles***
Groundwater Nitrate Concentration in German Pilot Study



Brenning (2024) in *TEXTE*; data: Umweltbundesamt, countrywide analysis, *N* > 9000

ML Model Interpretation in Geospatial Machine Learning

$\mathbf{x}$

$\hat{f}$

$\hat{y}$

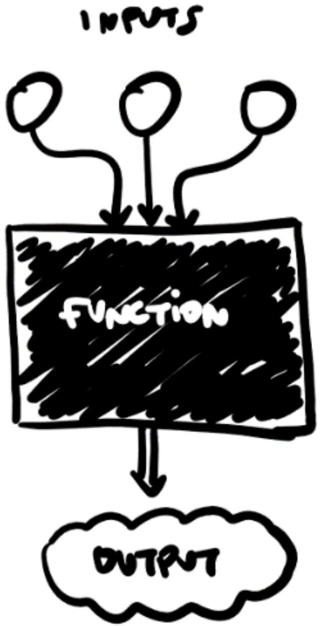
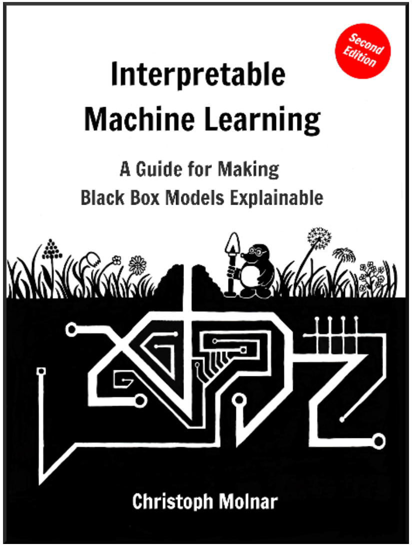


Image source: thatsoftwaredude.com

- Model-specific vs. **model-agnostic**
- Train-and-predict vs. **post-hoc**



ML Model Interpretation in Geospatial Machine Learning

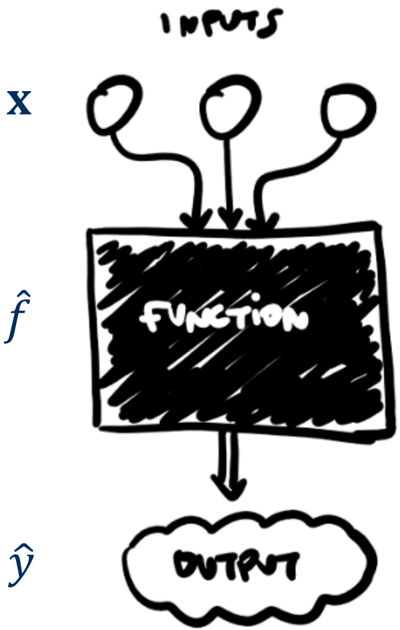


Image source: thatsoftwaredude.com

Feature summary statistics

- Permutation feature importance
- Shapley additive explanations (SHAP) feature importance

Local (i.e. instance-level) explanations

- Shapley values

Marginal effect plots

- Partial dependence
- Accumulated local effects (ALE)
- SHAP dependence

ML Model Interpretation in Geospatial Machine Learning

log(zinc), Maas Floodplain: Permutation Feature Importance

x

$\hat{f}$

$\hat{y}$

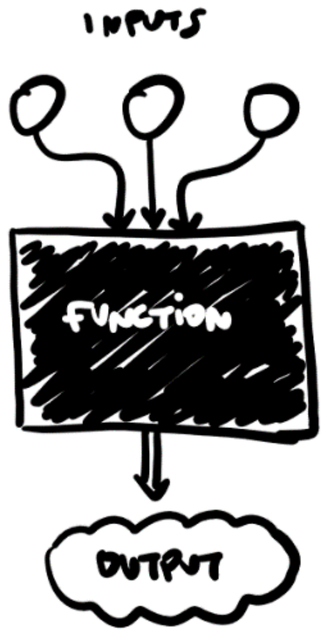
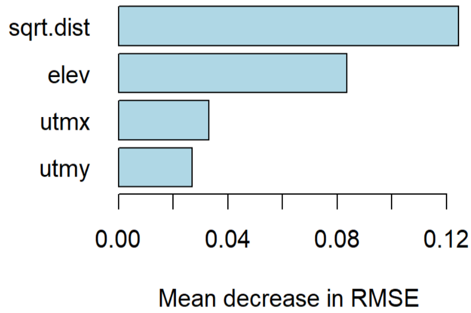
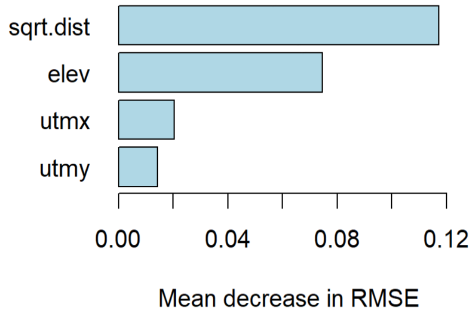


Image source: thatsoftwaredude.com

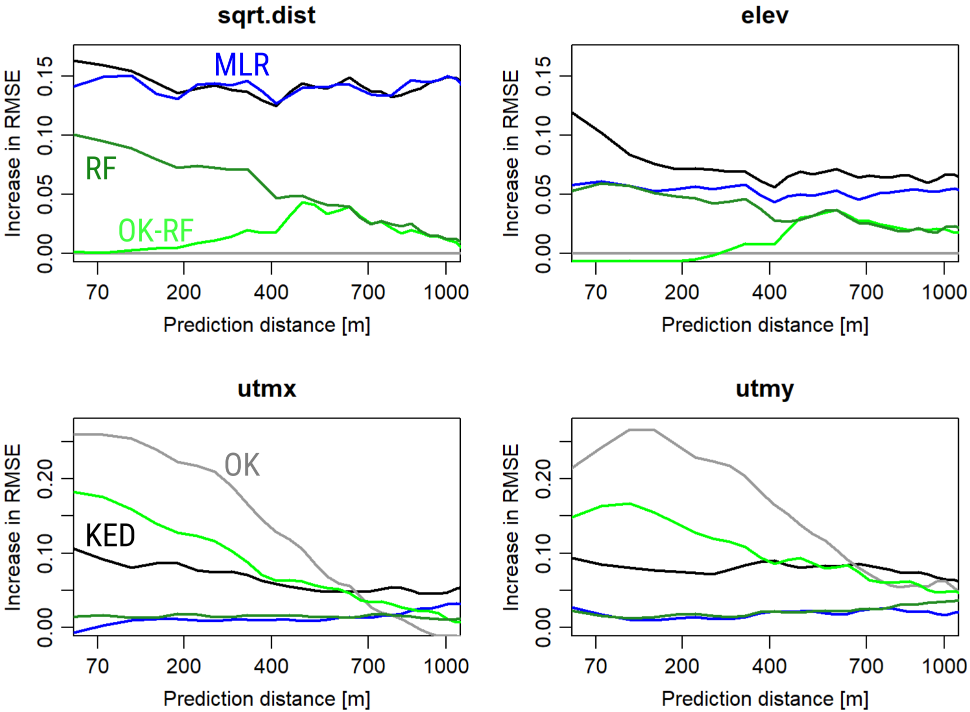
…using random cross-validation



…using *spatial* cross-validation



ML Model Interpretation: *Spatial Variable Importance Profiles*
log(zinc) on the Maas Floodplain

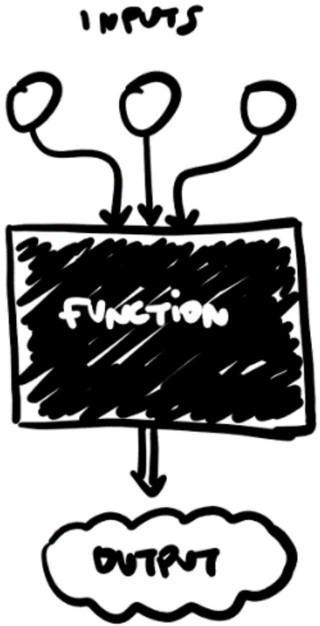


MLR: Multiple linear regression
KED: Kriging with external drift
OK: Ordinary kriging
RF: Random forest
OK-RF: Blended OK-RF

Brenning (2023) in *IJGIS*

High-Dimensional Feature Space:
Interpretation in Transformed Space

$\mathbf{x}$ $\mathbf{w}$



$\hat{f}$ $\hat{g}$

$\hat{y}$

Image source: thatsoftwaredude.com

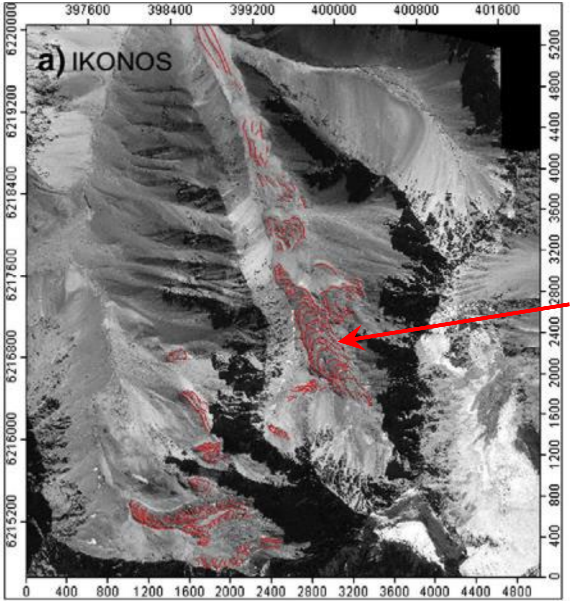
Proposal

- Use an invertible mapping $T\colon X \longmapsto W$ onto an interpretation space W
- E.g. PCA or nonlinear embedding
- Now interpret $\hat{g} := \hat{f} \circ T^{-1}$ in interpretation space W
- Does not modify the model $\hat{f}$!
- R package **`wiml`** $\longmapsto$ **`iml`**, **`DALEX`**

Brenning (2023) in *Machine Learning*

High-Dimensional Feature Space:
Mapping Rock Glaciers in the Andes

IKONOS satellite image



Small portion of study area!

Flow patterns

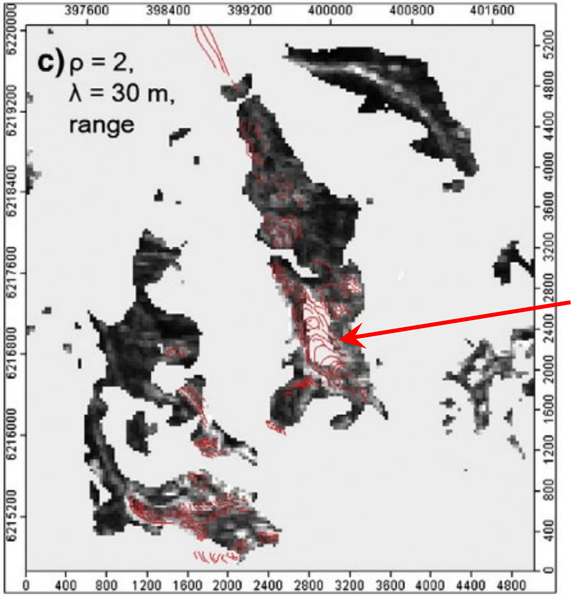
Rock glacier in the Andes….

…and in the Alps

Brenning et al. (2012) in *Remote Sens. Env.*

High-Dimensional Feature Space:
Mapping Rock Glaciers in the Andes

Gabor texture feature (example)



Small portion of study area!

Flow patterns

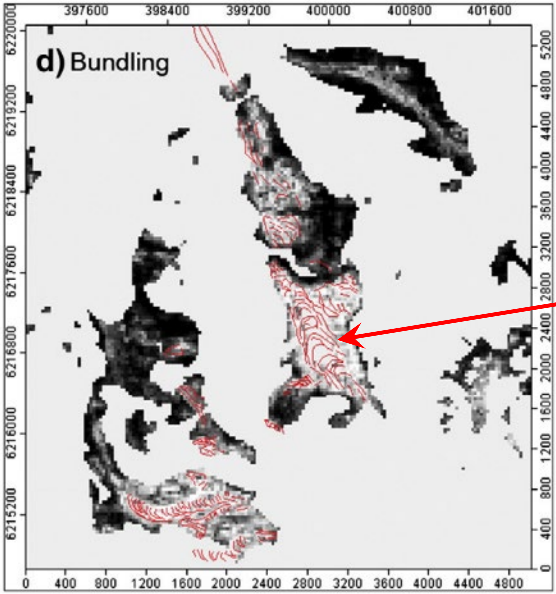
Rock glacier in the Andes….

…and in the Alps

Brenning et al. (2012) in *Remote Sens. Env.*

High-Dimensional Feature Space: Mapping Rock Glaciers in the Andes

Soft classification



Small portion of study area!

Flow patterns

Rock glacier in the Andes....

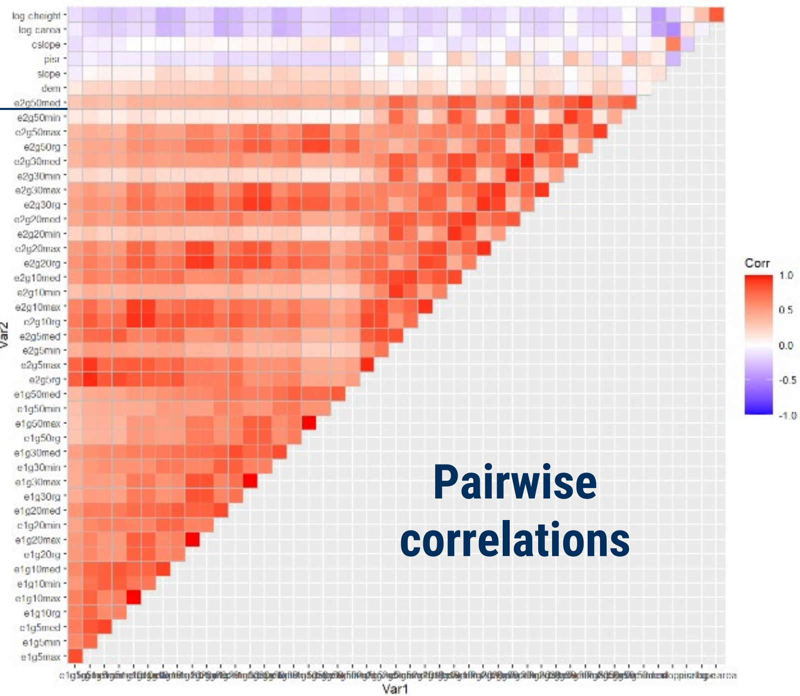
...and in the Alps

Brenning et al. (2012) in *Remote Sens. Env.*

High-Dimensional Feature Space:
Mapping Rock Glaciers in the Andes

Terrain attributes

Gabor texture features



Rock glacier in the Andes....

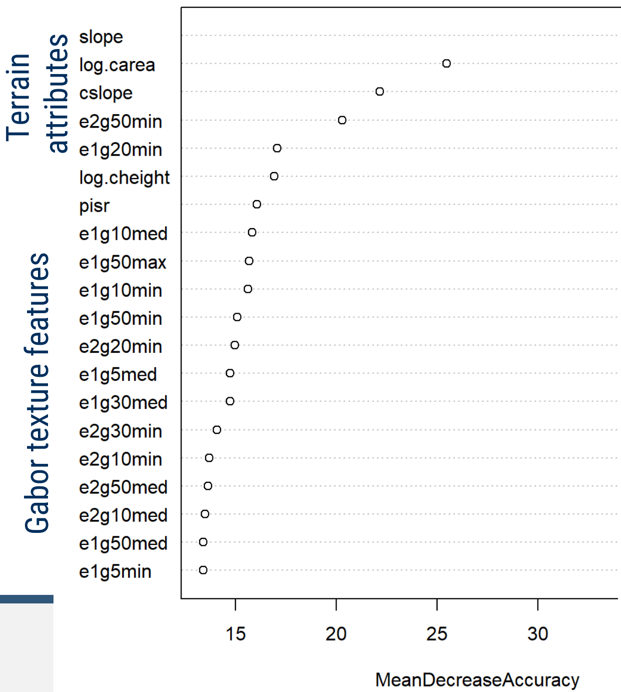
...and in the Alps

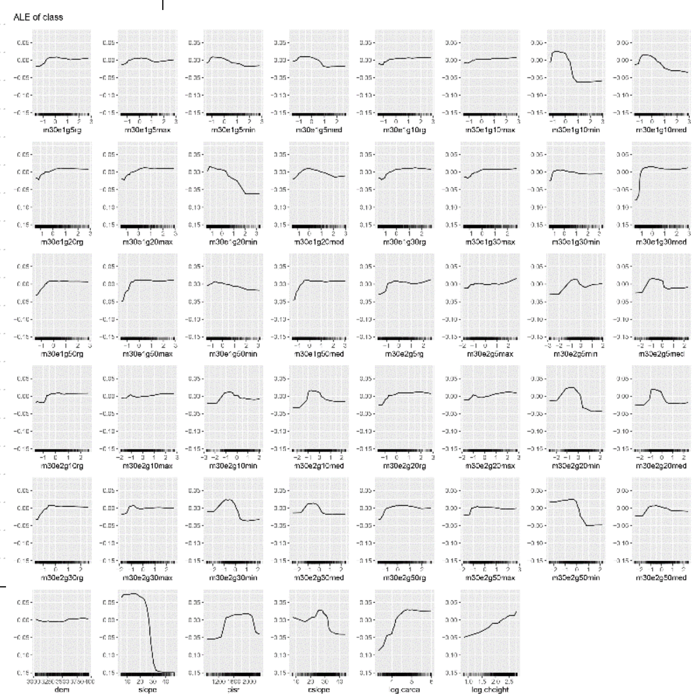
High-Dimensional Feature Space:
Interpretation in Untransformed Space
Random Forest Model

Rock glacier in the Andes

Permutation importance

ALE plots





Visualizing Marginal Effects in ML

Partial Dependence vs. Accumulated Local Effects (ALE)

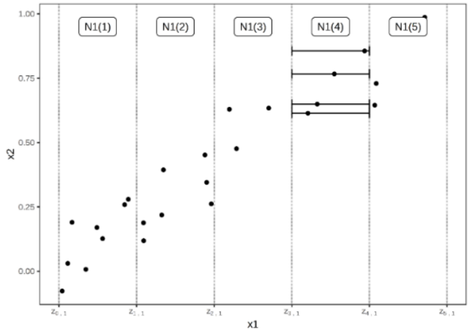
$$\hat{f}_{x_S}(x_S) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}\left(x_S, x_C^{(i)}\right)$$

How do $\hat{f}$'s predictions vary with (arbitrary values of) x_s?

x_s: selected variable, and
x_c: all other predictors

→ **Extrapolation** in feature space!

How do $\hat{f}$'s predictions change for small changes in x_s?



$$\hat{f}_{j,ALE}(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i:x_j^{(i)} \in N_j(k)} \underbrace{\left[f(z_{k,j}, x_{\setminus j}^{(i)}) - f(z_{k-1,j}, x_{\setminus j}^{(i)})\right]}_{\text{differences in predictions for different predictor values ("effects")}}$$

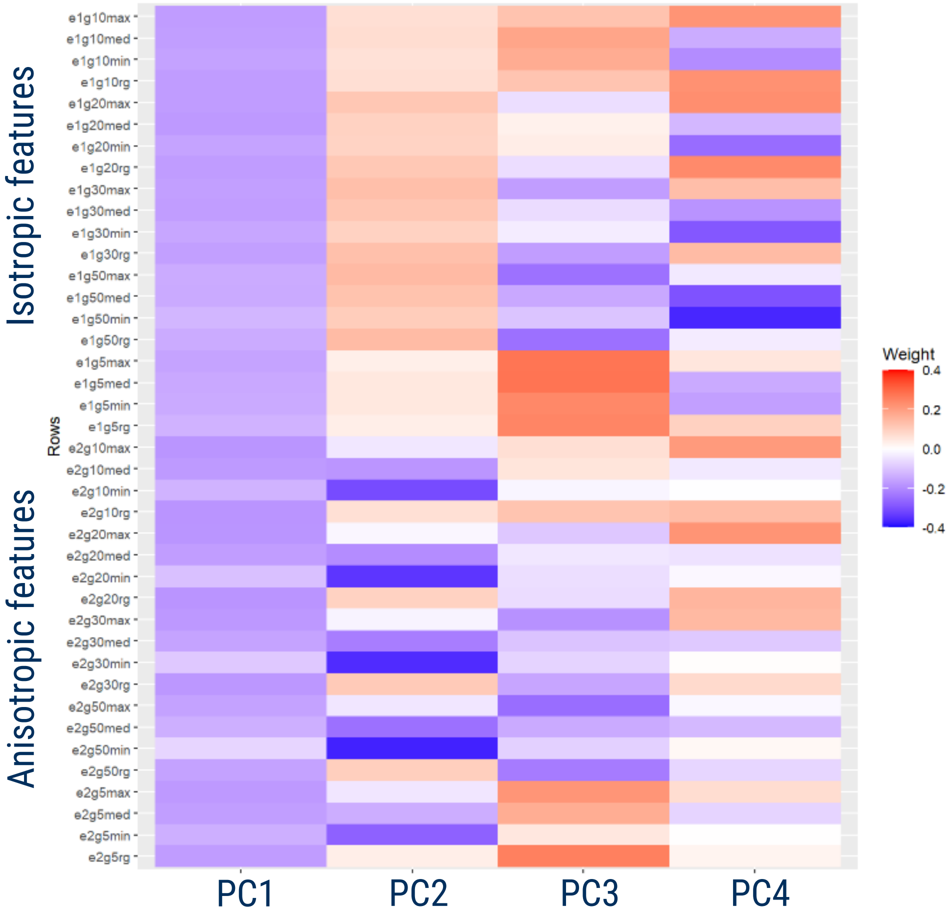
differences in predictions for different predictor values ("**effects**")

accumulate over all intervals, up to the one corresponding to x_j

sum over all features within the **local** neighborhood in the variable of interest, x_j

Molnar (2023)

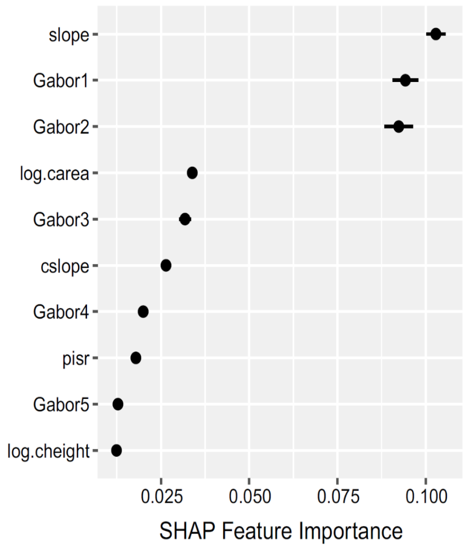
PCA of Gabor Features



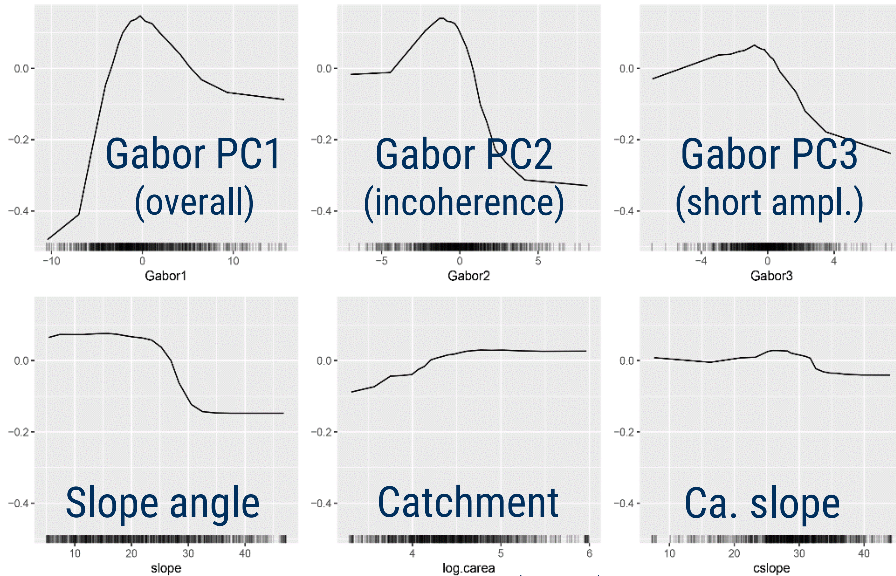
Brenning (2023)
in *Machine Learning*

High-Dimensional Feature Space:
Interpretation in Transformed Space

SHAP importance



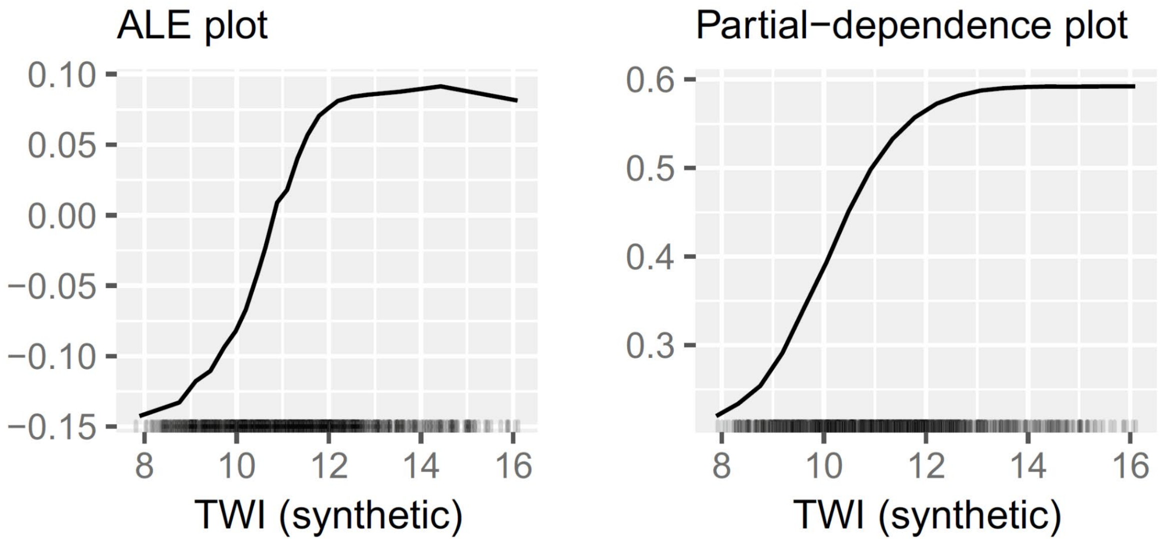
ALE plots



Brenning (2023) in *Machine Learning*

High-Dimensional Feature Space: *Interpretation Using Synthetic Features*

TWI = log(*catchment_area* / tan(*slope*))



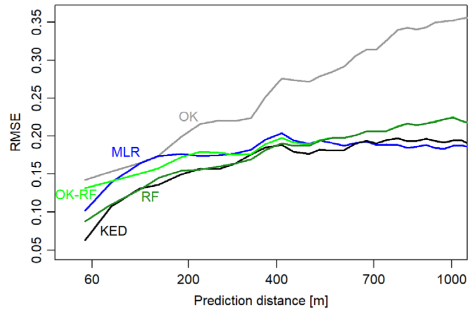
Brenning (2023) in *Machine Learning*

Lessons Learned

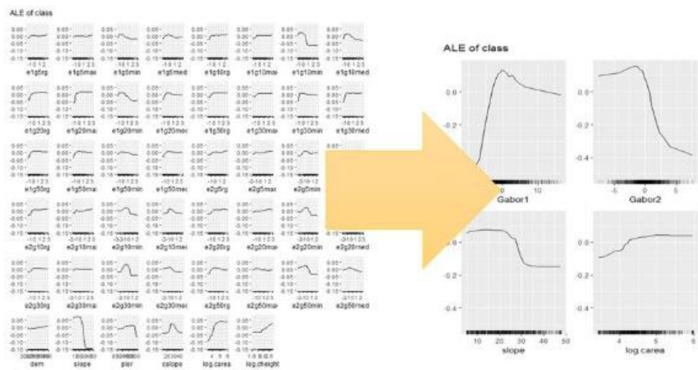
- Prediction error and feature importances depend on prediction distance.
- Suitable performance measures are needed.

- Correlated features can (and should?) be interpreted in lower-dimensional projected space.

Spatial prediction error & feature importance profiles



Model interpretation in transformed space



Lessons Learned

- Black-box model interpretation has serious limitations. (Flaws?)
- Use globally interpretable models: **Additive models**
    - Generalized additive models (Hastie & Tibshirani, 1990)
    - Model-based boosting (Hothorn *et al.*, 2010 in *JMLR*)
    - Explainable boosting machine (Nori *et al.*, 2019 in *arxiv*)

nature machine intelligence

Explore content | About the journal | Publish with us | Subscribe

nature > nature machine intelligence > perspectives > article

Perspective | Published: 13 May 2019

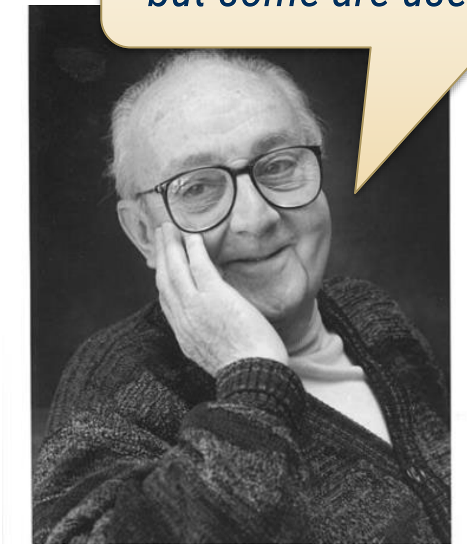
Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead

Cynthia Rudin

Lessons Learned

- Black-box model interpretation has serious limitations. (Flaws?)
- Use globally interpretable models: **Additive models**
  - Generalized additive models (Hastie & Tibshirani, 1990)
  - Model-based boosting (Hothorn *et al.*, 2010 in *JMLR*)
  - Explainable boosting machine (Nori *et al.*, 2019 in *arxiv*)

All models are wrong, but some are useful



George E. P. Box (1919-2013)

Image credit: David M.C. Eddy via Wikipedia

Thank you for your attention!

Brenning (2023) in *IJGIS*
Brenning (2023) in
Machine Learning
Brenning (2024) in *TEXTE*

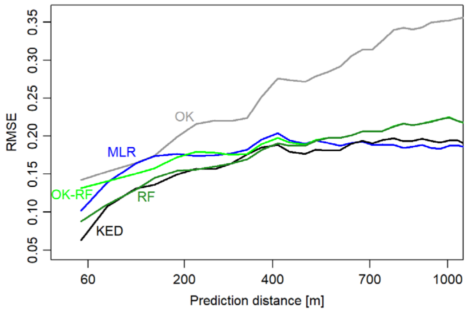
Blog: **geods.netlify.app**

@geobrenning

@alexander.brenning

Acknowledgements:
P. Fieguth, S. Long, T. Suesse, M. Fink,
Umweltbundesamt

Spatial prediction error & feature importance profiles



Model interpretation in transformed space