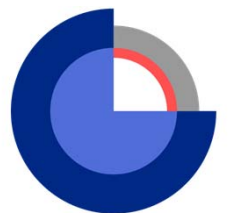


# Machine learning til kortlægning af hydrologiske variabler i høj opløsning

Julian Koch

*Hydrologisk Afdeling, GEUS*



**G E U S**

# To variabler - en metode

- Redoxgrænse og terrænnært grundvand
- Redoxgrænse: TReNDS – Innovations Fonden
- Terrænnært grundvand: C2C Projekt (Region Midtjylland, SCALGO, COWI, GEUS)
- Fokus på høj opløsning + performance + usikkerhedsanalyse + følsomhedsanalyse

Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2019-212>  
Manuscript under review for journal Hydrol. Earth Syst. Sci.  
Discussion started: 10 May 2019  
© Author(s) 2019. CC BY 4.0 License.



## Modelling of the shallow water table at high spatial resolution using Random Forests.

Julian Koch<sup>1</sup>, Helen Berger<sup>2</sup>, Hans Jørgen Henriksen<sup>1</sup>, Torben Obel Sonnenborg<sup>1</sup>

<sup>1</sup>Department of Hydrology, Geological Survey of Denmark and Greenland (GEUS), Copenhagen, 1350, Denmark

<sup>2</sup>COWI A/S, Lyngby, 2800, Denmark

## Water Resources Research





### RESEARCH ARTICLE

10.1029/2018WR023939

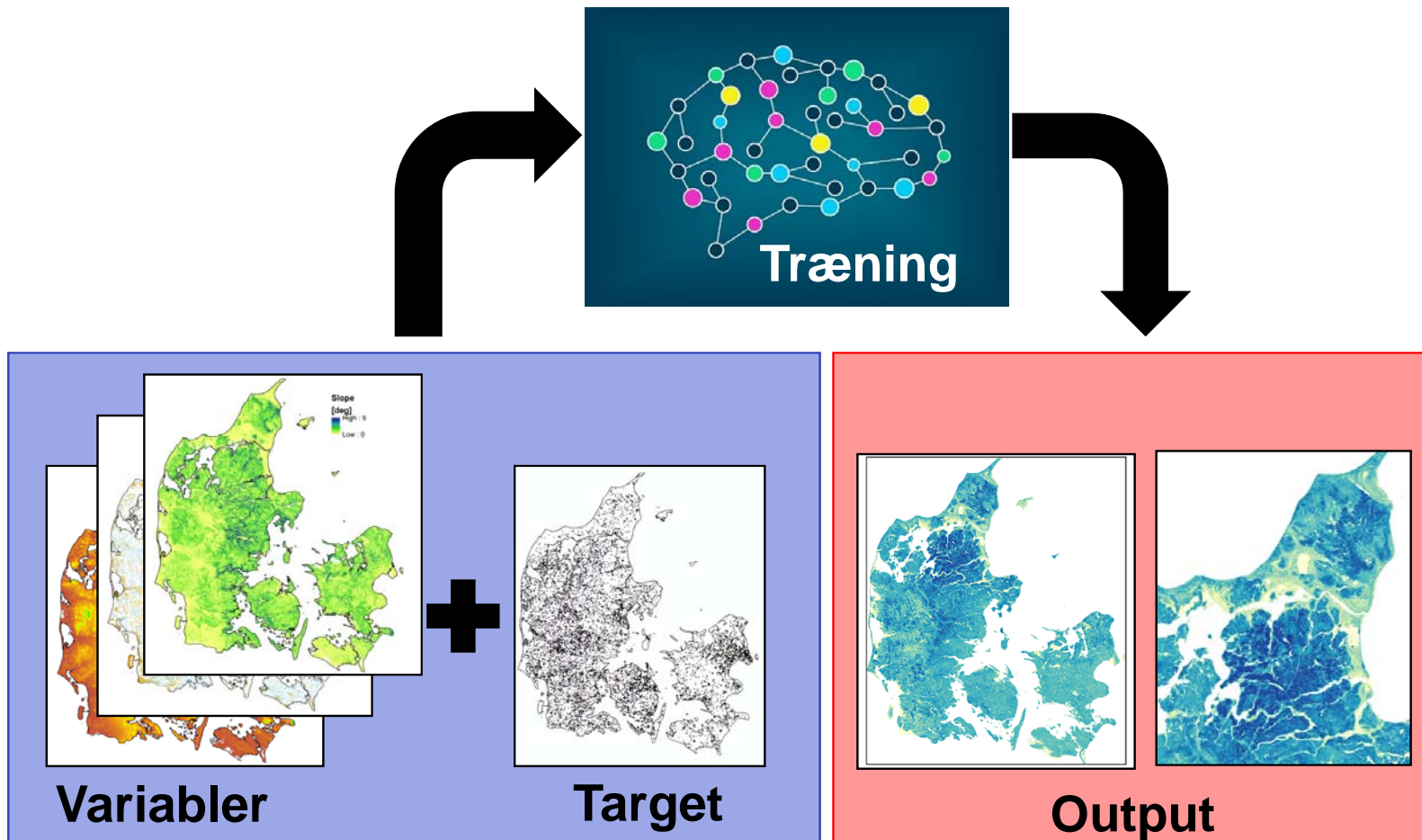
#### Special Section:

Big Data & Machine Learning  
in Water Sciences: Recent  
Progress and Their Use in  
Advancing Science

## Modeling Depth of the Redox Interface at High Resolution at National Scale Using Random Forest and Residual Gaussian Simulation

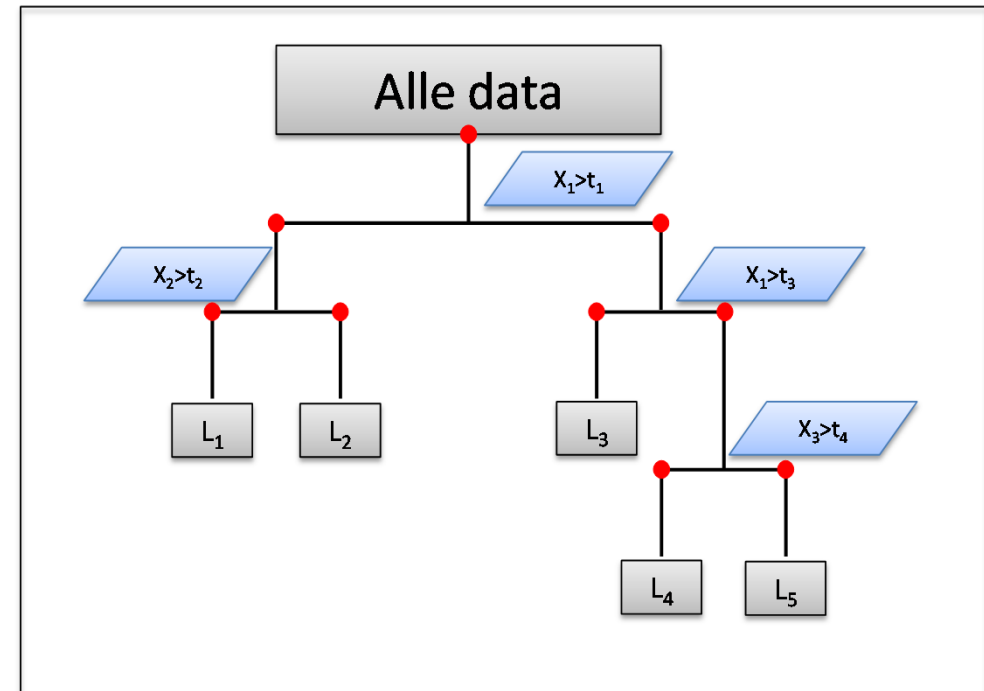
Julian Koch<sup>1</sup> , Simon Stisen<sup>1</sup> , Jens C. Refsgaard<sup>1</sup> , Vibeke Ernstsen<sup>2</sup>, Peter R. Jakobsen<sup>2</sup>, and Anker L. Højberg<sup>1</sup> 

# Machine Learning



# Random Forests

- Ensemble af beslutningstræer
- Middel af mange træer til prædiktion
- Hvert træ er bygget op på et tilfældigt udvalg af data og variabler
- Out of bag – intern cross valideringstest
- Mulighed for usikkerheds- og følsomhedsanalyse

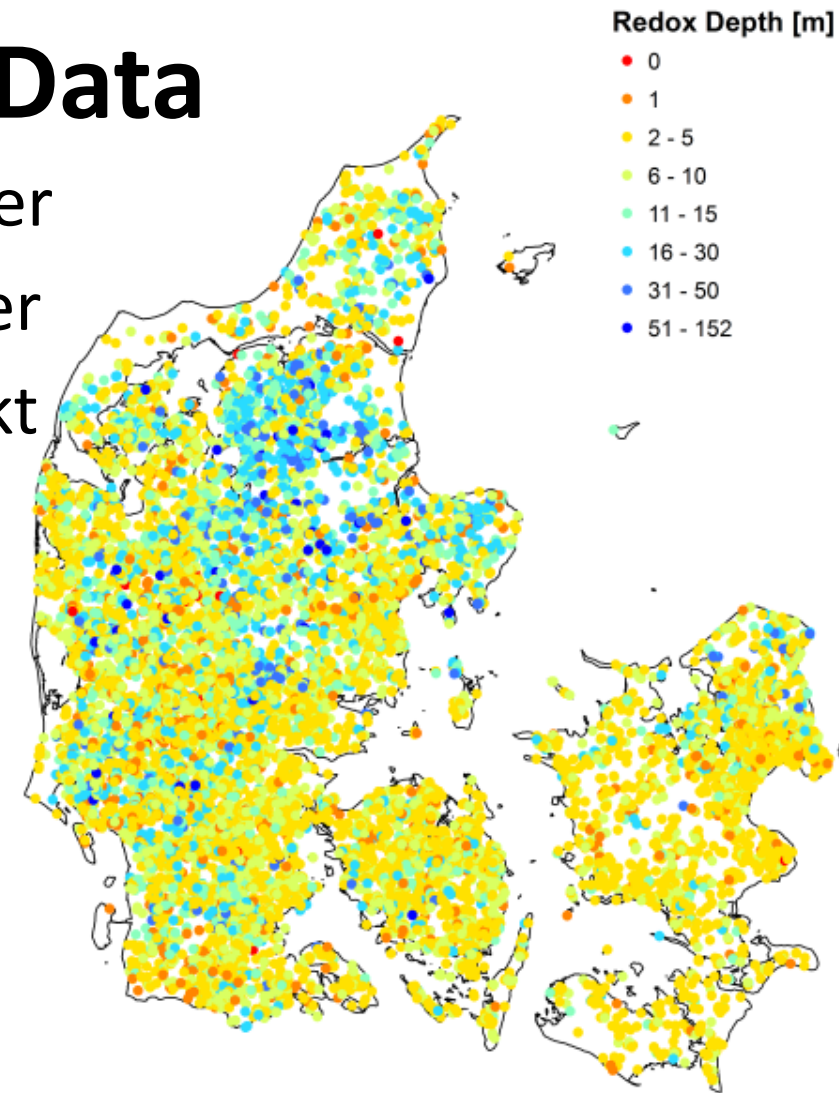


# Redoxgrænse - Data

- Oxiderede sedimenter → gule/brune farver
- Reducerede sedimenter → grå/sorte farver
- ~13000 boringer fra Jupiter + 600 støttepkt

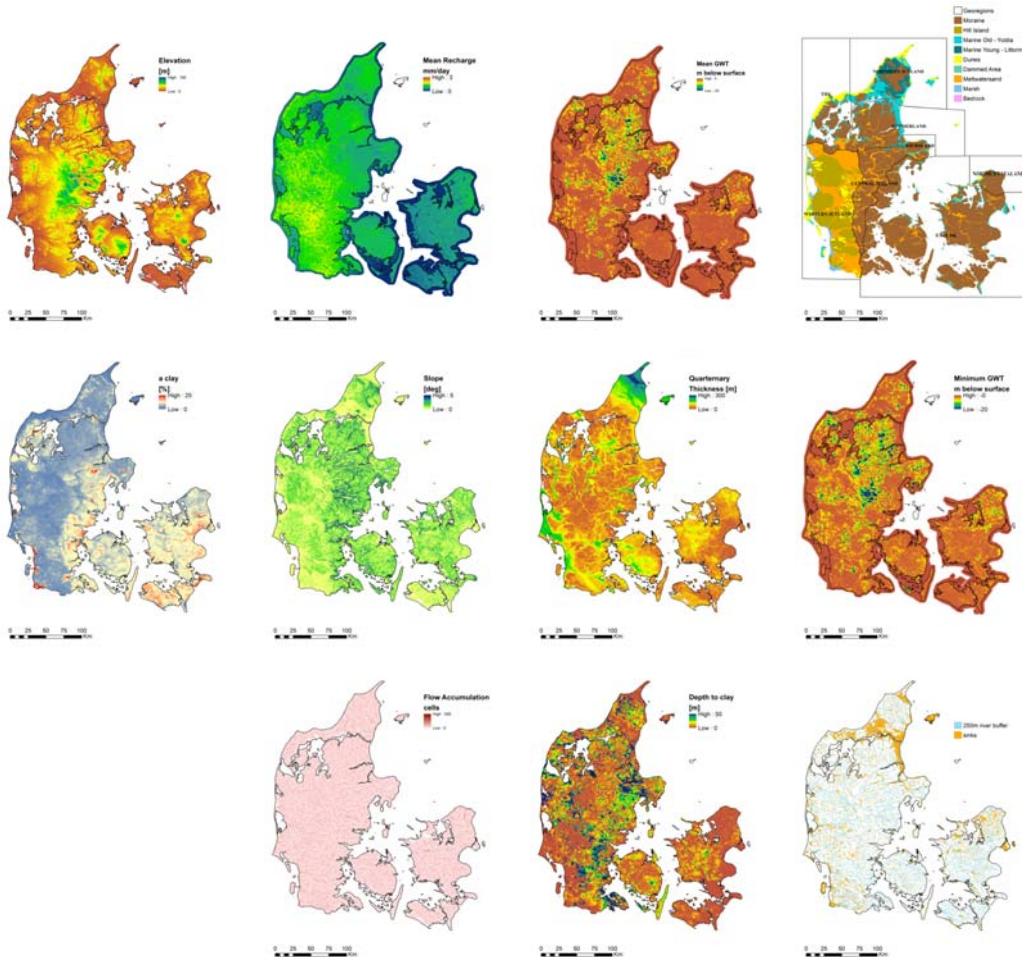


Farveskift i typisk moræne ler fra brunt til gråt. Farveskift lokaliserer redoxgrænsen. V. Ernstsén (2013)





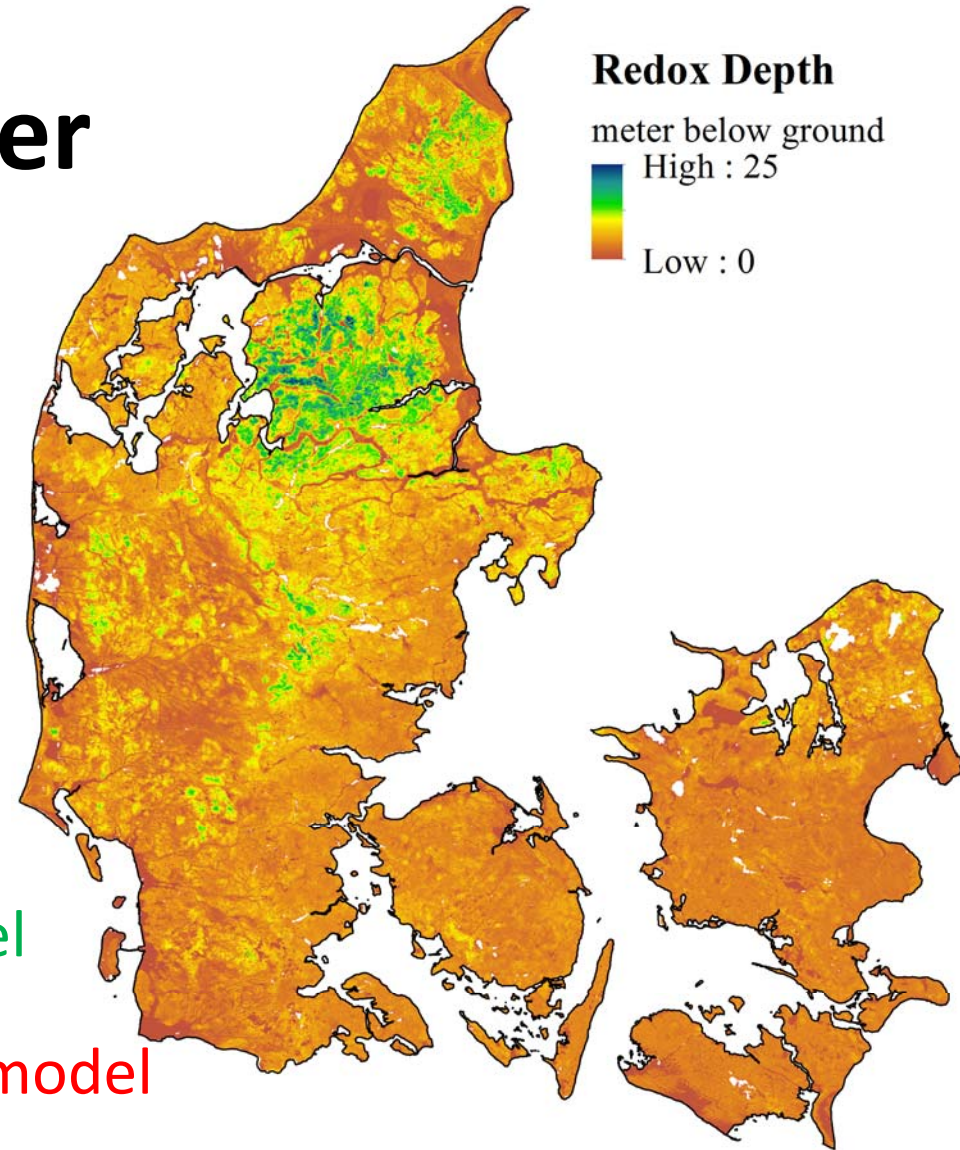
# Forklarende variabler



- Lerindhold i de øverste 2m (4 lag)
- Højdemodellen
- Opstrømsareal
- Hældning
- Afstand til vandløb
- Lavbund klassifikation (organisk/mineralsk)
- DK model - Ler tykkelsen
- Kvarter tykkelsen
- DK model – grundvandsdannelsen
- DK model – min. grundvandsspejl
- DK model – middel grundvandsspejl
- Landskabstypologi
- Georegion

# Resultater

**Redox Depth**  
meter below ground  
High : 25  
Low : 0



- 100m skala
- Validerings test →  $R^2=0.48$
- Add-on til machine learning
- Fejl interpolation med Kriging
  - Honorerer lokal data
  - Estimerer usikkerheden
  - Realisationer

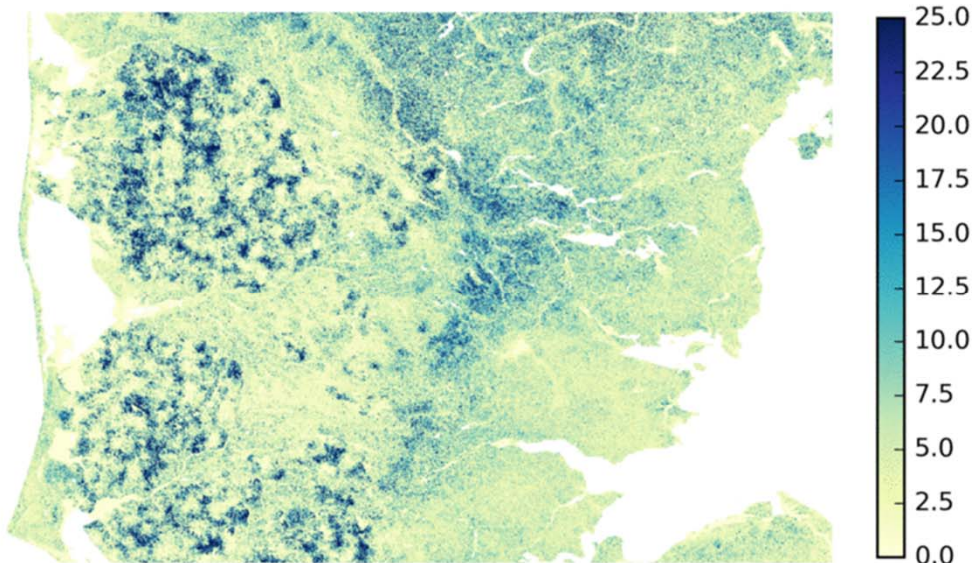
$$RD_{\text{final}} = RD_{\text{trend}} + RD_{\text{error}}$$

- Machine Learning (Random Forests)  
→ Kan forklares med data-drevet model
- Stokastisk model (sequential Gaussian)  
→ Kan ikke forklares med data-drevet model

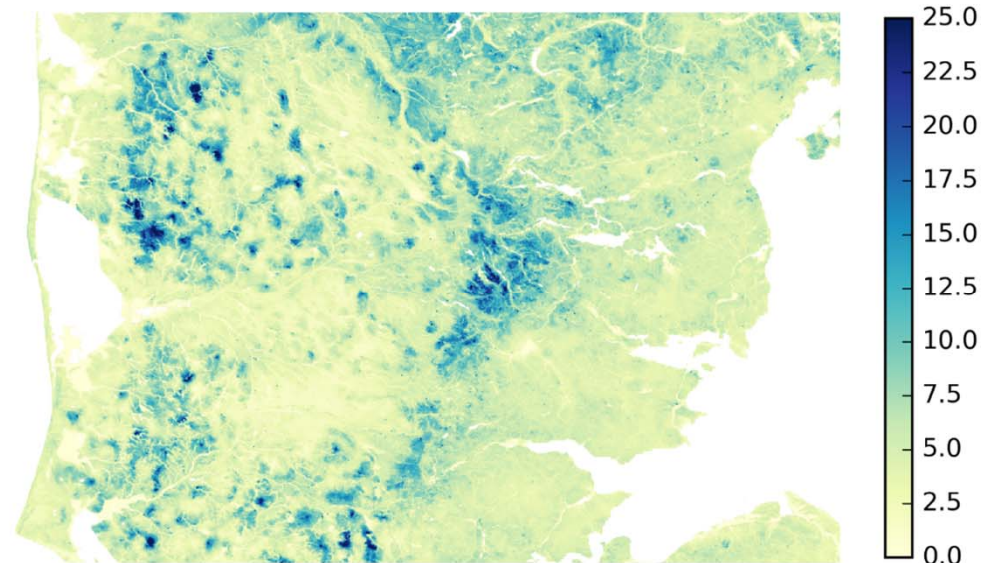
# Resultater

- Redoxgrænsen i Midtjylland

20 realisationer



Middel



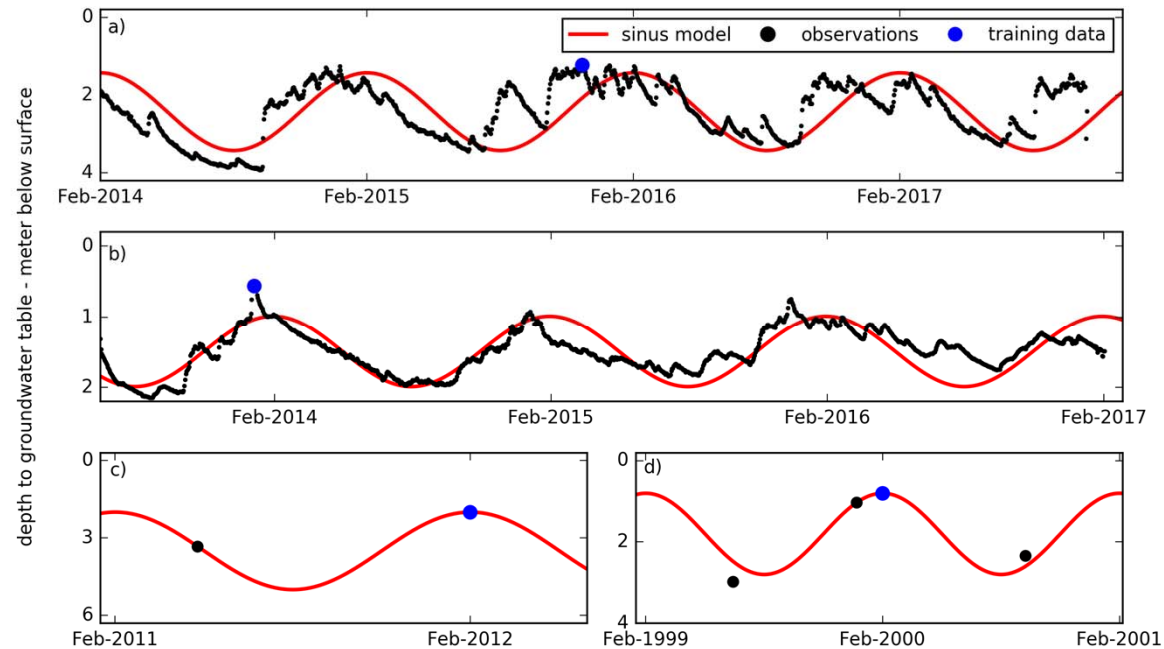


# Terrænnært grundvand

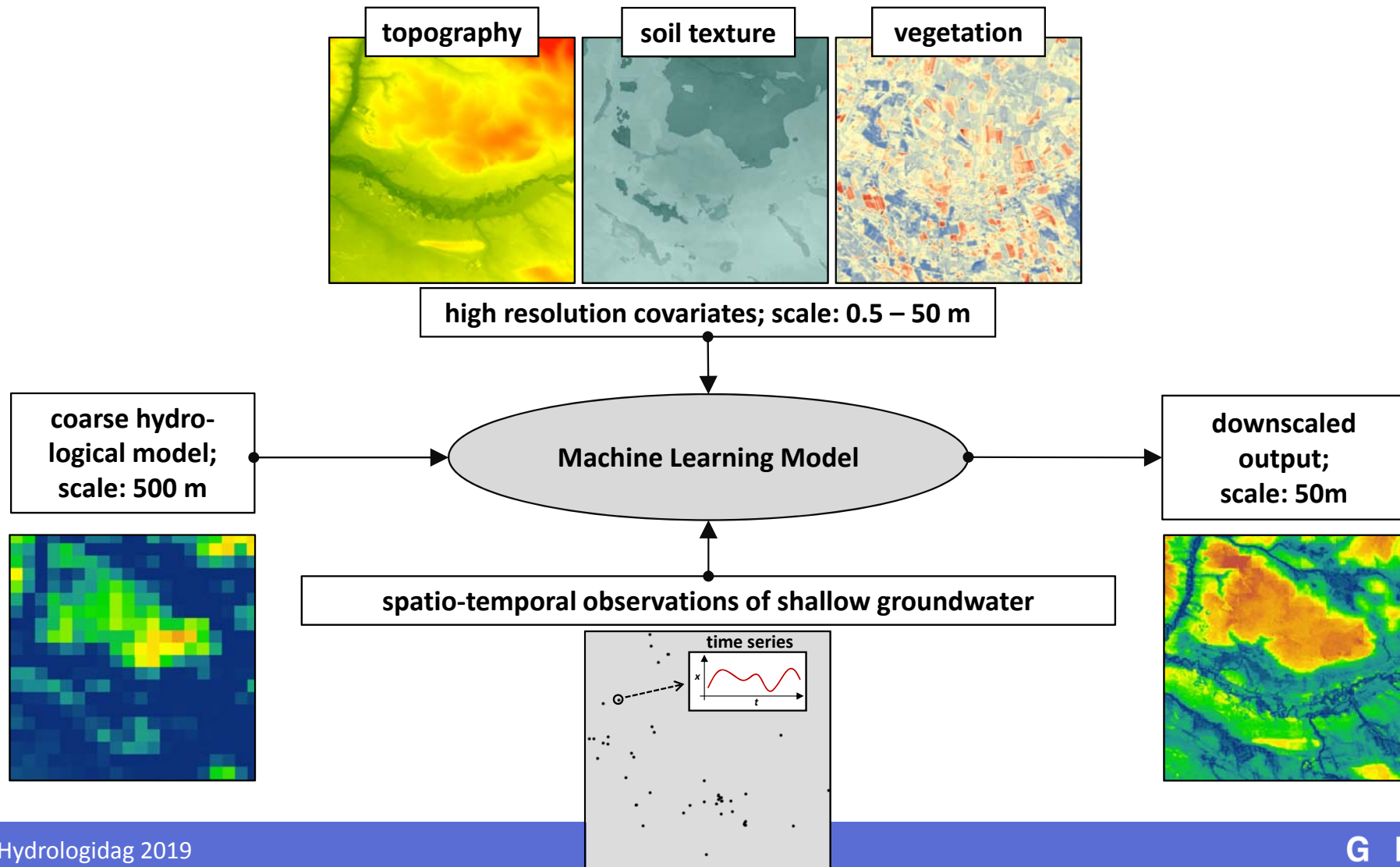
- Estimering af terrænnært grundvand i 50m
- Fokus på vinter min. dybde
- 15000 boringer + støttepunkter
- 26 forklarende variabler inkl. DK-model i 500m

# Sinus fremskrivning

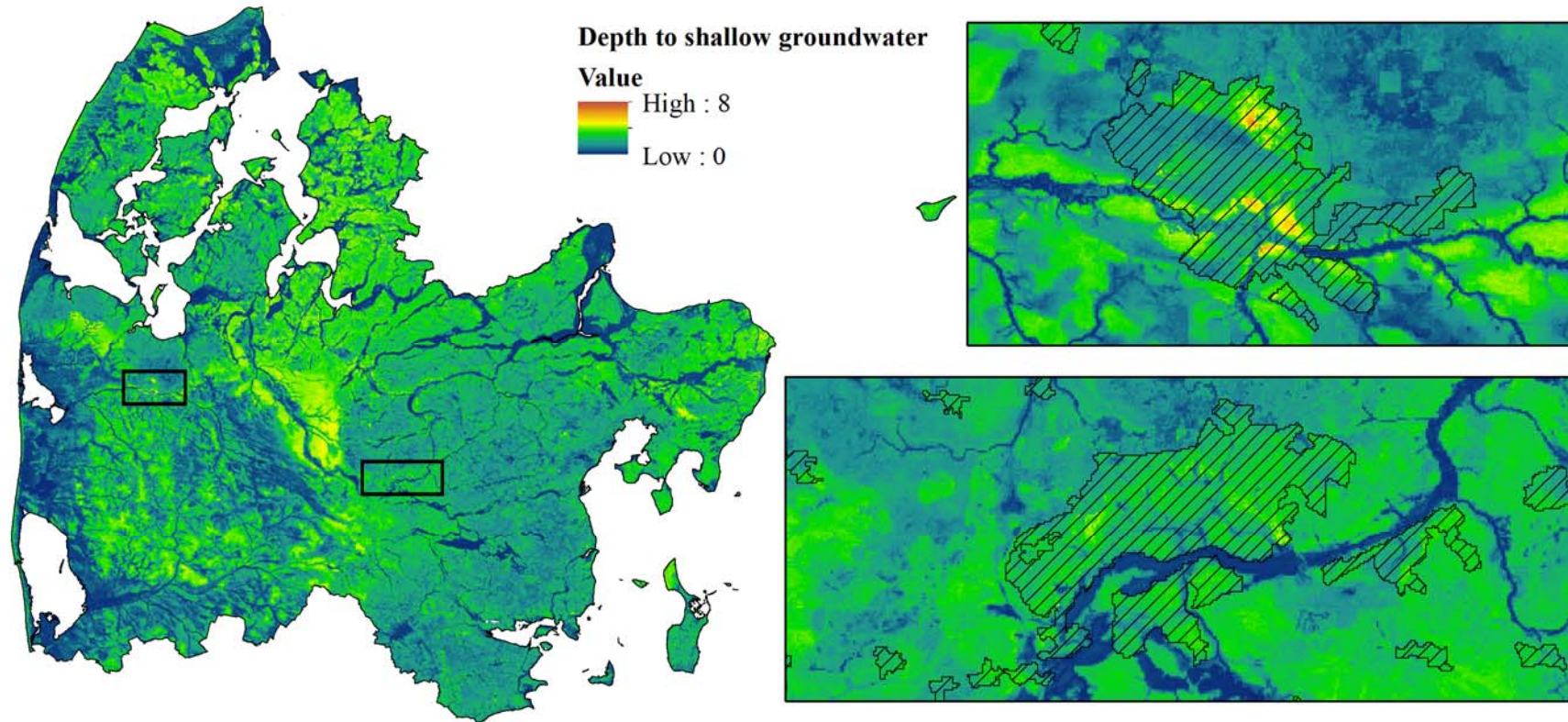
- Vinter min. grundvandsdybde
- Boringer med kort ( $n \leq 5$ ) pejleserie bliver fremskrevet
- Sinus amplituden fra boringer med lange ( $n > 5$ ) pejleserier
- Differencering i hydrogeologiske typologier



# ML + PBM



# Resultater

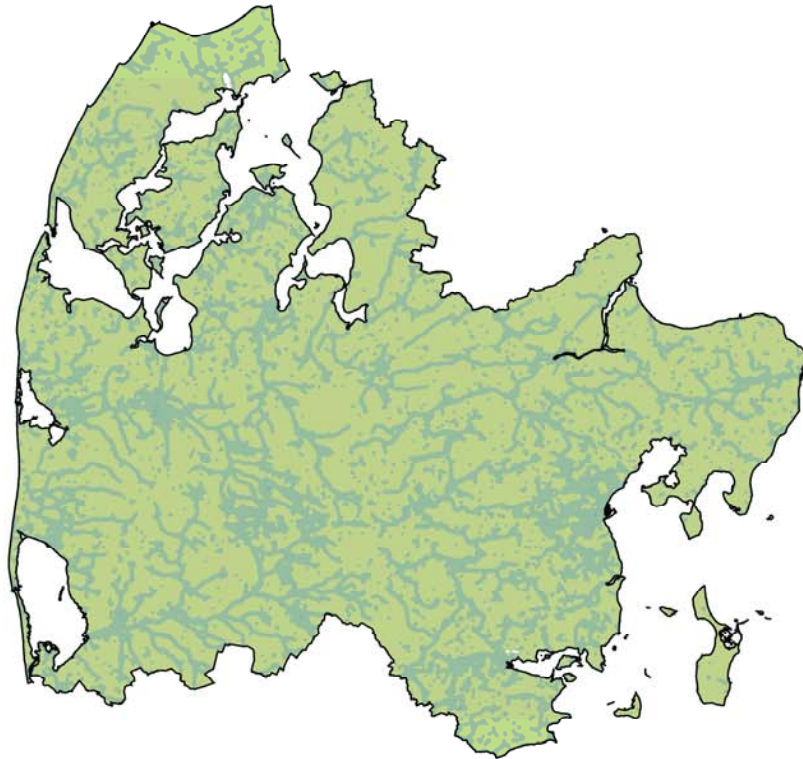


**MAE=76cm RMSE=113cm R2=0.56**



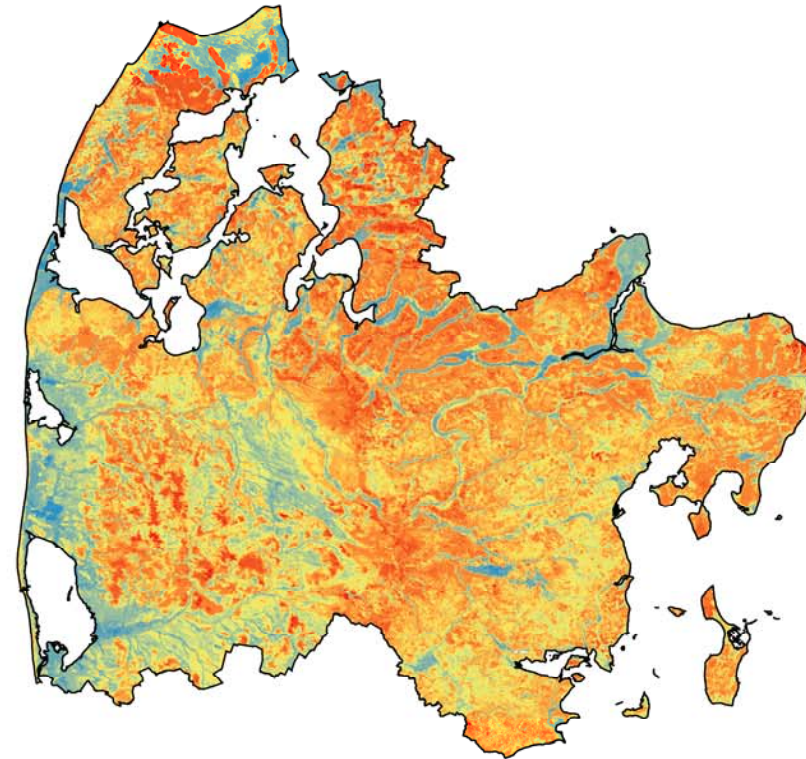
# Usikkerhedsanalyse

STD - RFRK



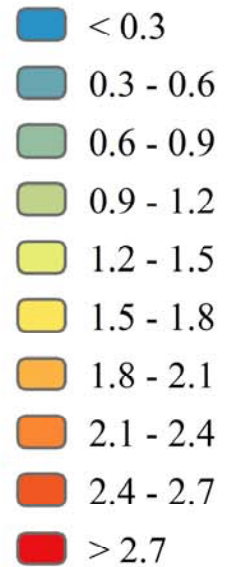
Fejl kriging

STD - QRF

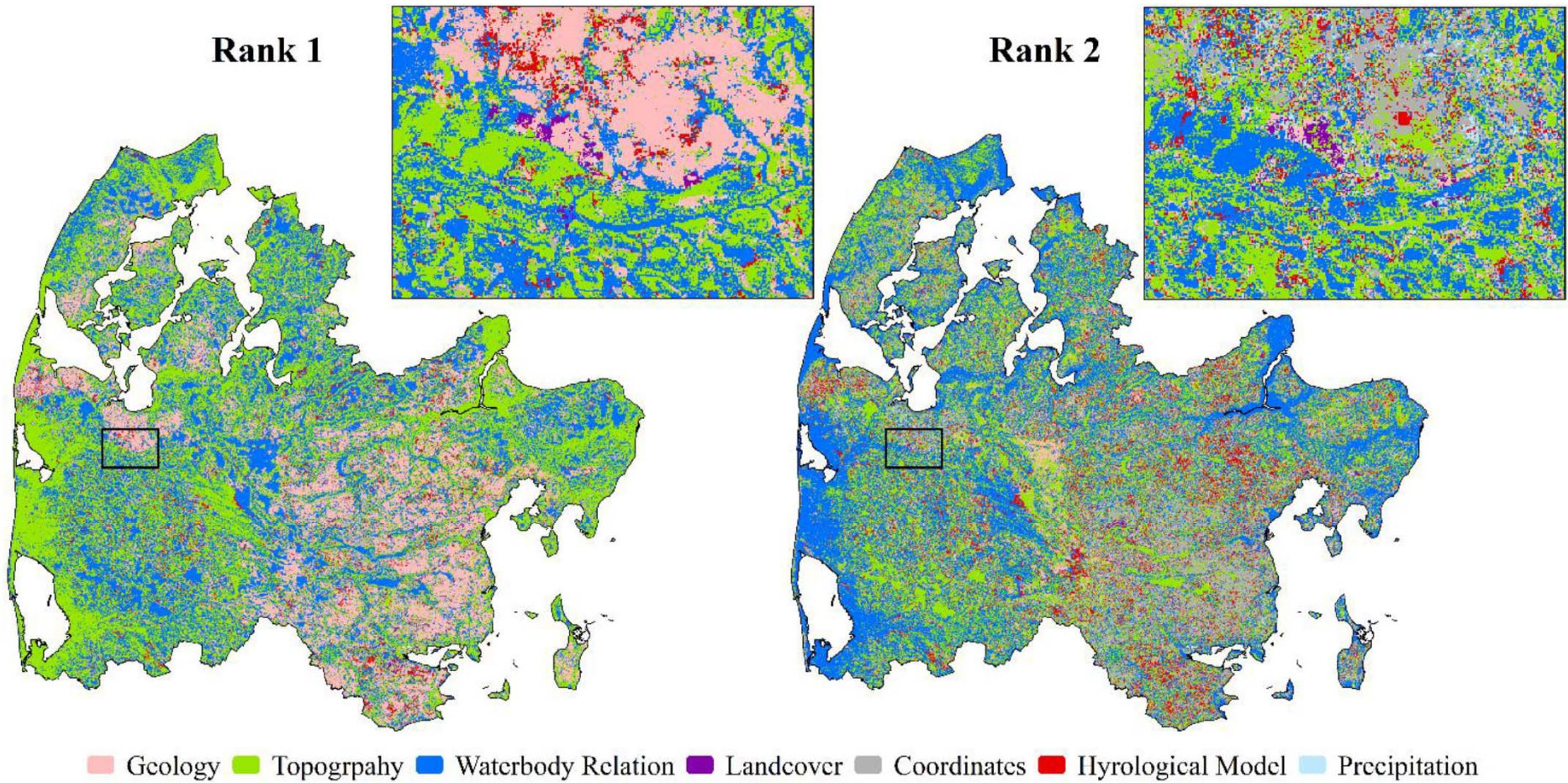


Quantile Regression Forests

Uncertainty [m]



# Følsomhedsanalyse





# Work in progress

# Konklusioner

- Støttestrukturer → understøtter træningsdatasættet ved hjælp af hydrologisk domæne viden
- Fejl kriging → nemt at tilføje ny data + stokastiske realisationer + usikkerhedsanalyse
- Forskellige metoder til at kvantificere usikkerheden – Hvad giver mest mening?
- Følsomhedsanalyse kan løse Black Box begrænsning
- Proces-baseret modeller kan guide Machine Learning