

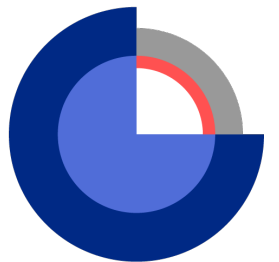
Kortlægning af terrænnære grundvandsdybder i høj rumlig opløsning ved hjælp af maskinlæring

Julian Koch¹, Helen Berger², Carsten Fjorback², Morten Revsbæk³, Mathias Rav³,
Hans Jørgen Henriksen¹, Torben Sonnenborg¹, Raphael Schneider¹, Jane
Gotfredsen¹, Lars Trolborg¹, Simon Stisen¹

¹Afdeling for Hydrologi, De Nationale Geologiske Undersøgelser For Danmark og Grønland (GEUS)

²COWI A/S

³ SCALGO ApS



G E U S

Indhold

- **Coast2Coast Climate Challenge** projektet (2018/2019), sammenarbejde med COWI og SCALGO, faciliteret af Region Midtjylland
- **HIP4PLUS** projektet (2019/2020) i FODS 6.1 regi

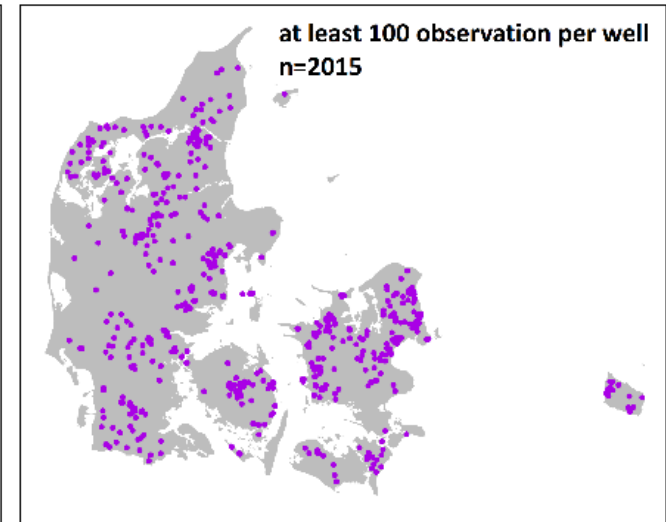
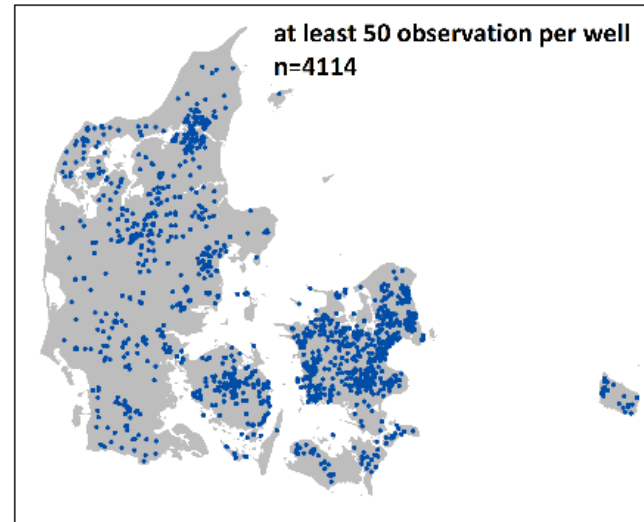
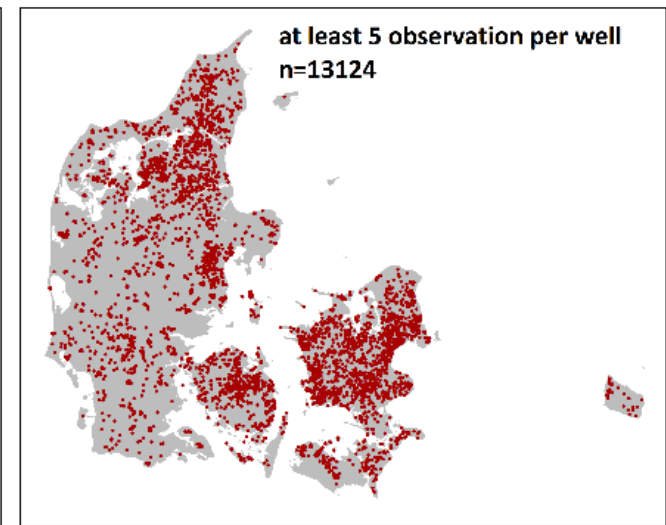
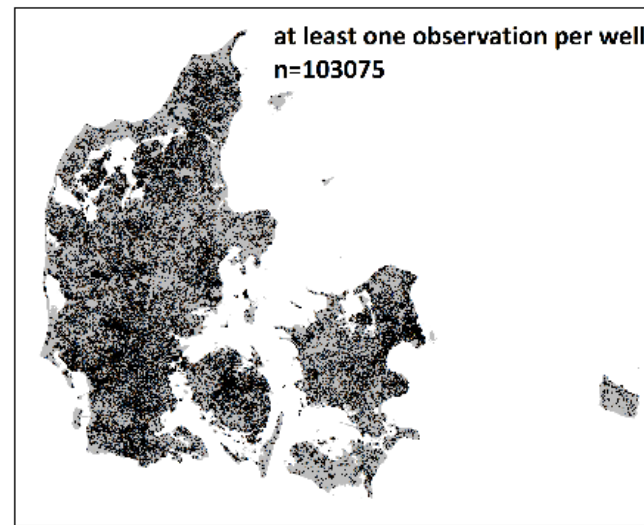
Terrænnært grundvand

- Hvorfor ML?
- Hvad kan ML?
- Hvad kan ML ikke?
- Synergi mellem ML og PB modeller



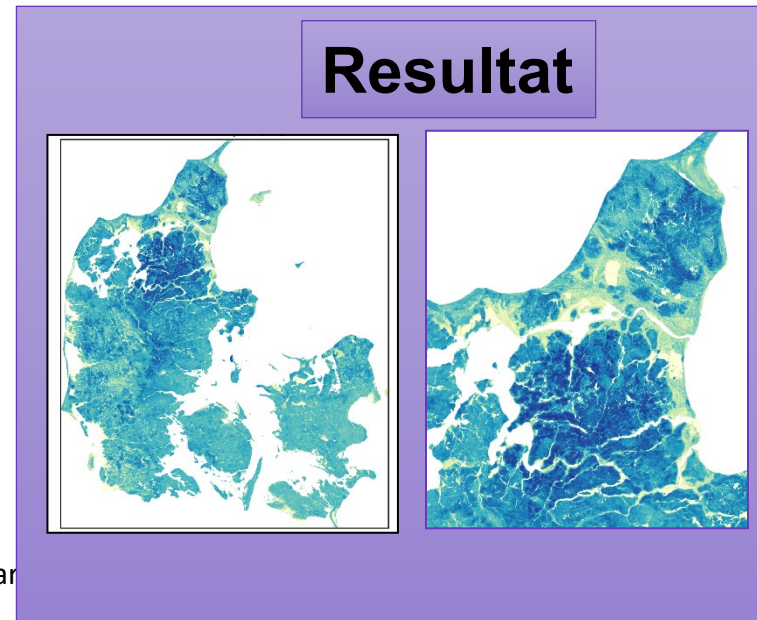
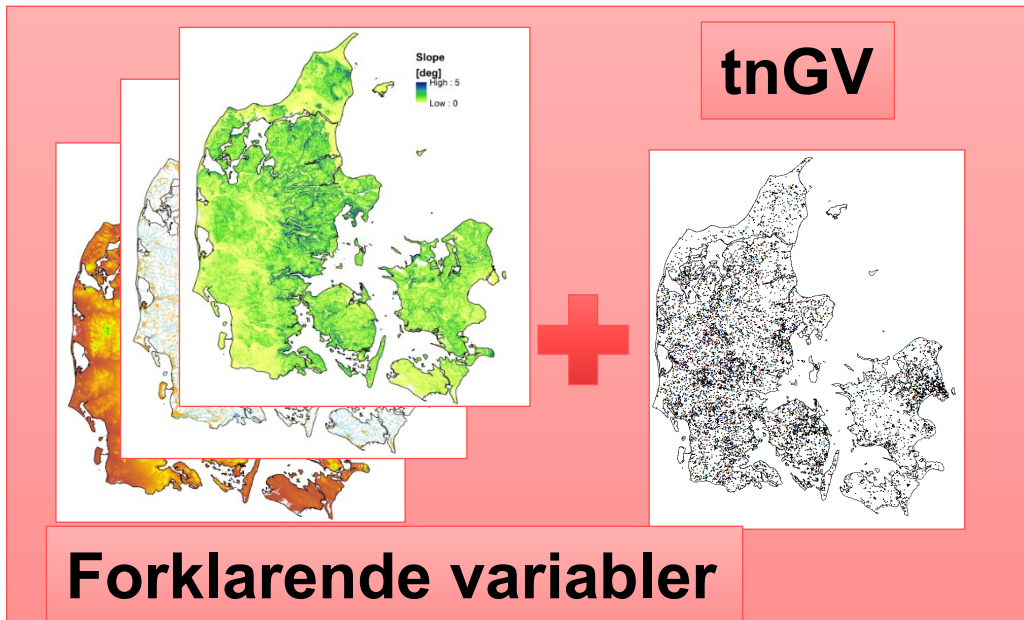
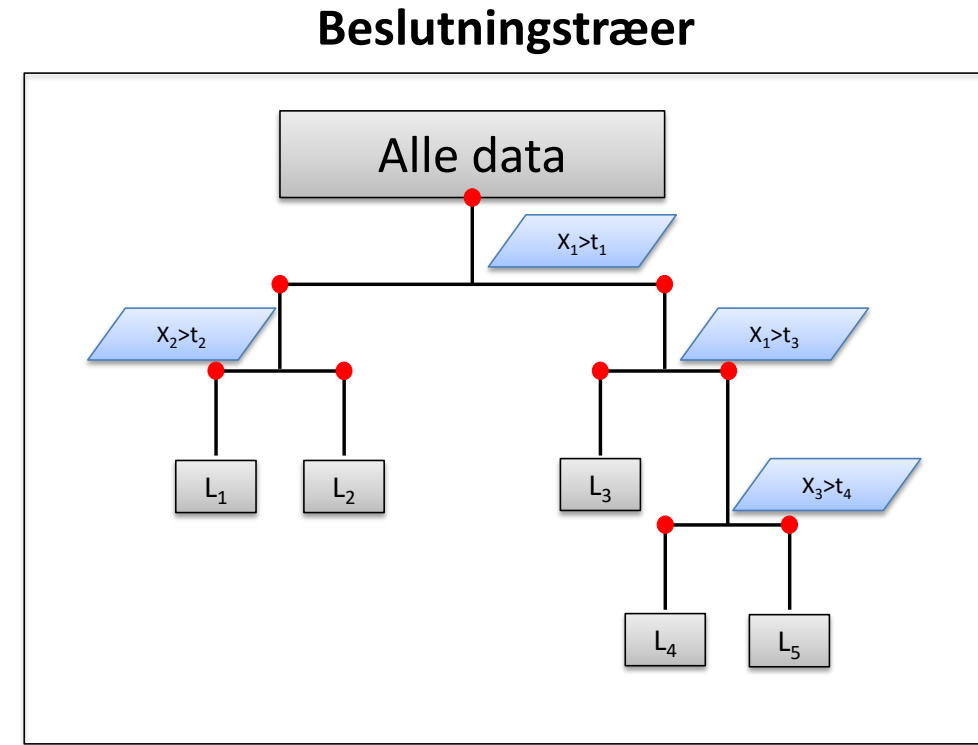
Hvorfor ML?

- Datagrundlaget er stort men heterogent (4,2 millioner pejlinger fra de sidste 30 år i Jupiter)
- Mange geo/miljø variabler er frit tilgængelige i høj rumlig opløsning
- Process-baseret (PB) model værktøjer opnår ofte ikke den ønskede nøjagtighed
- Den rumlige opløsning er begrænset i PB modeller



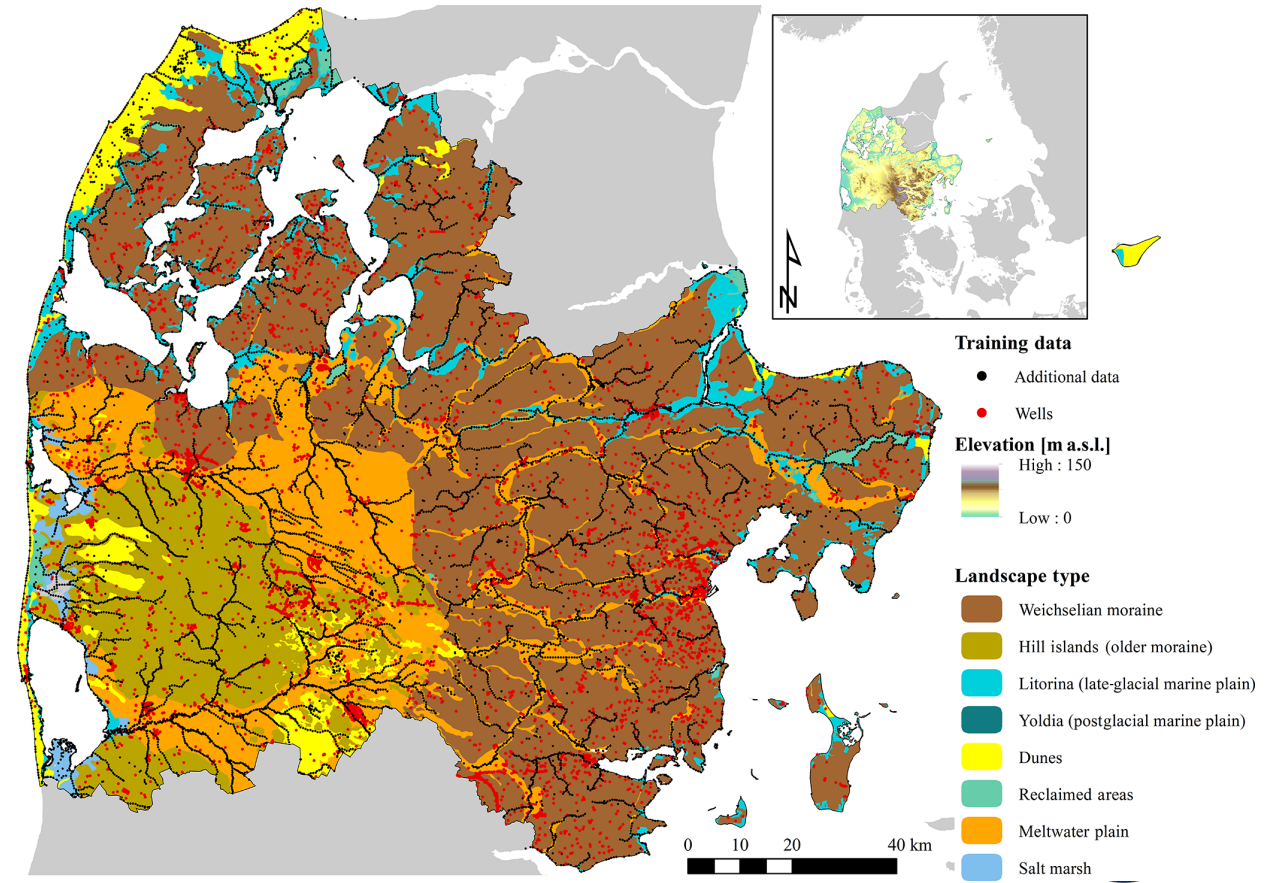
**Stort potentiale i forhold til øget nøjagtighed
og rumlig opløsning**

ML tilgang



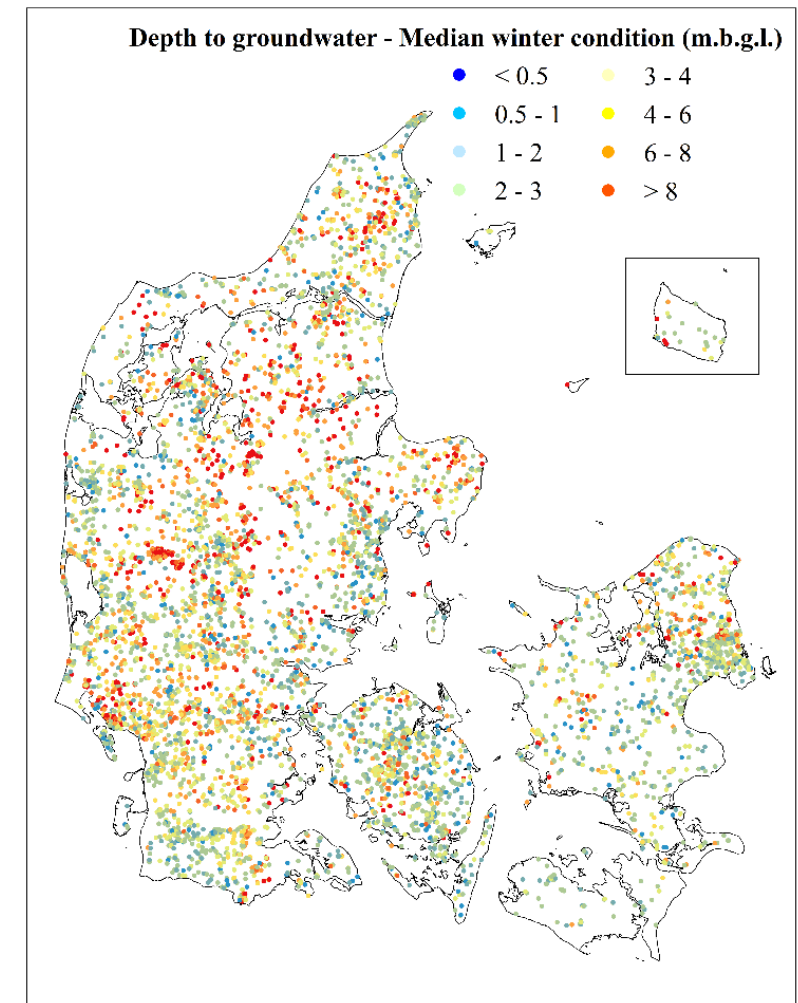
Projekt 1: C2CCC

- Modellere tnGV i 50m rumlig opløsning for et $\sim 15,000 \text{ km}^2$ domæne
- Fokus på en ekstrem vinter hændelse
- Usikkerhedskvantificering
- Effekt af klimaforandringer
- Følsomhedsanalyse
- Brugerinterface til visualisering og opdatering



Projekt 2: HIP4PLUS

- Landsdækkende modellering af tnGV i 10m rumlig opløsning (430 millioner celler)
- Fokus på en typisk **vinter** og en typisk **sommer** tilstand
- Usikkerhedskvantificering
- Sammenspillet mellem PB modellen og ML modellen



Data bearbejdning

- Træningsdata: GV
pejlinger + evt.
støttepunkter
- Forklarende variabler
- 80 % af hele projektet
med mange gentagelser
- Mangel på lange
tidsserier

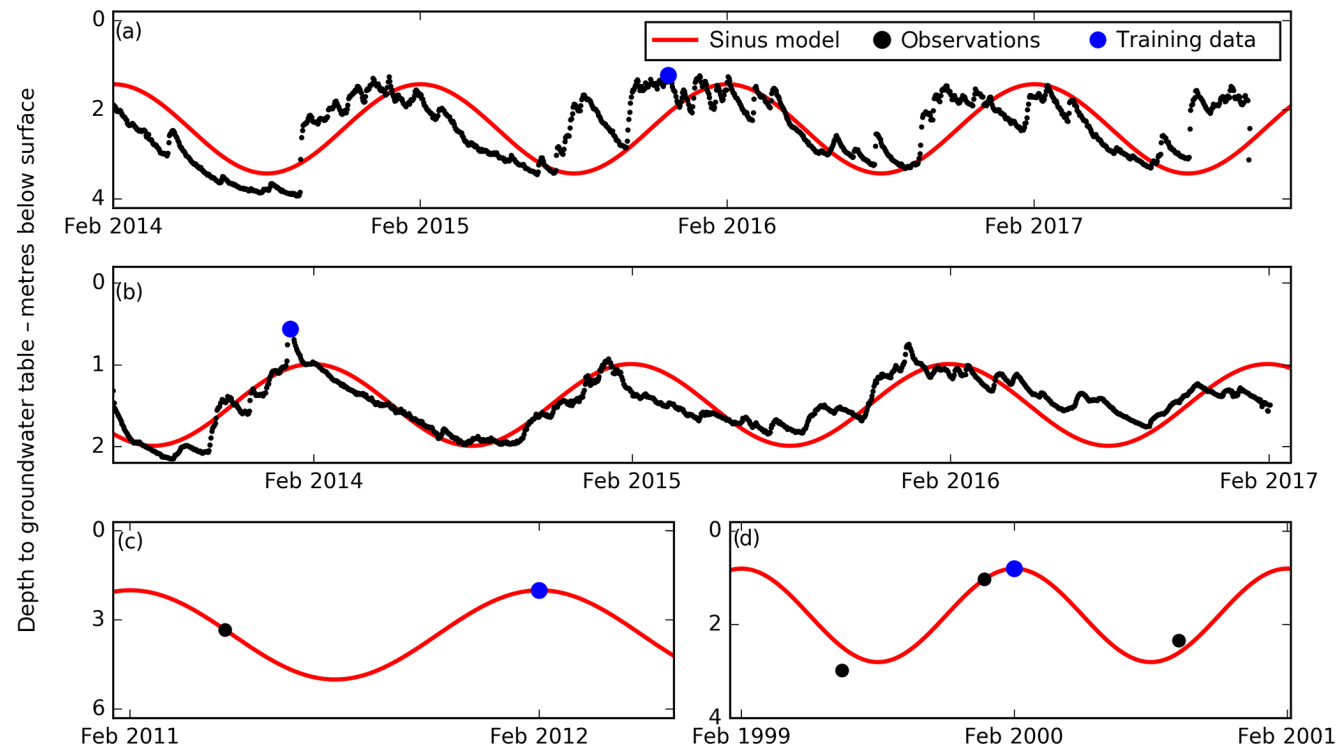
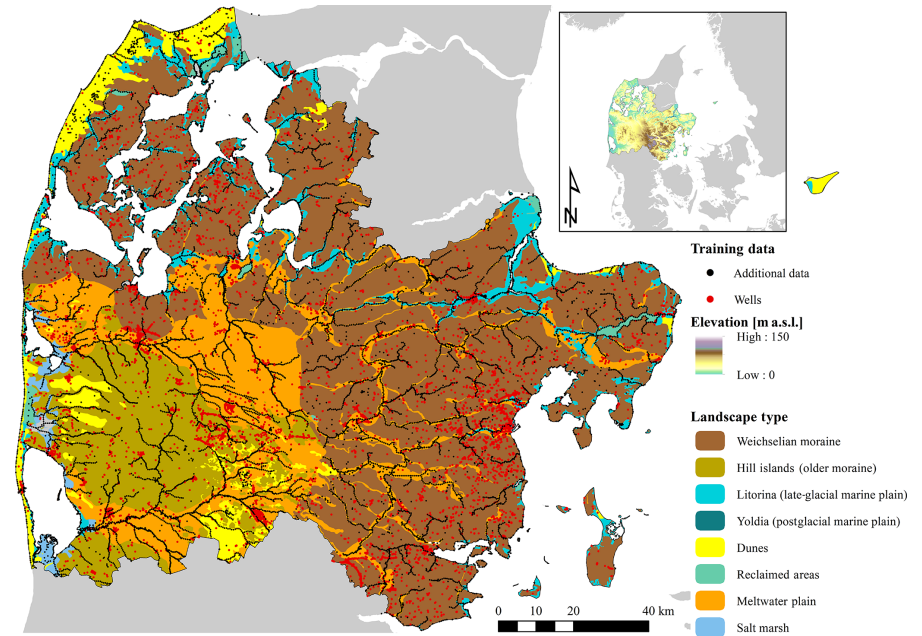
c2ccc:

15,000 pejlinger

2,000 støttepunkter

27 forklarende variabler

ATV Øst - Terrænnært Grundvand



GEUS

Data bearbejdning

- Træningsdata: GV
pejlinger + evt.
støttepunkter
- Forklarende variabler
- 80 % af hele projektet
med mange gentagelser
- Mangel på lange
tidsserier

HIP4PLUS:

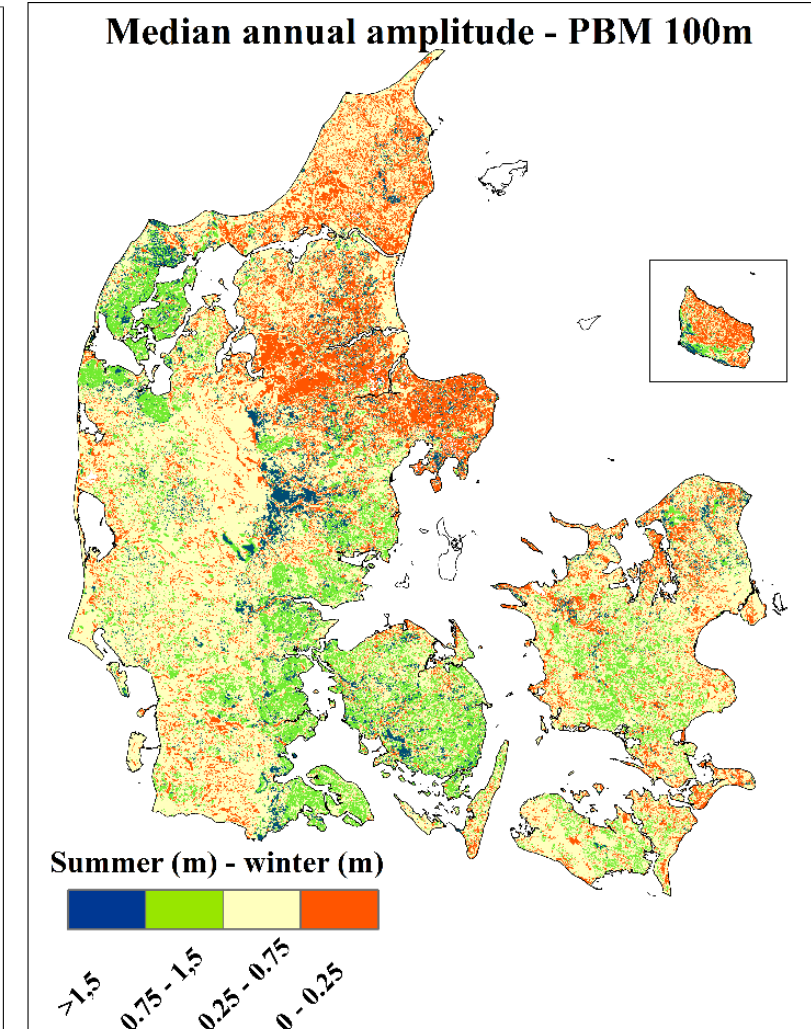
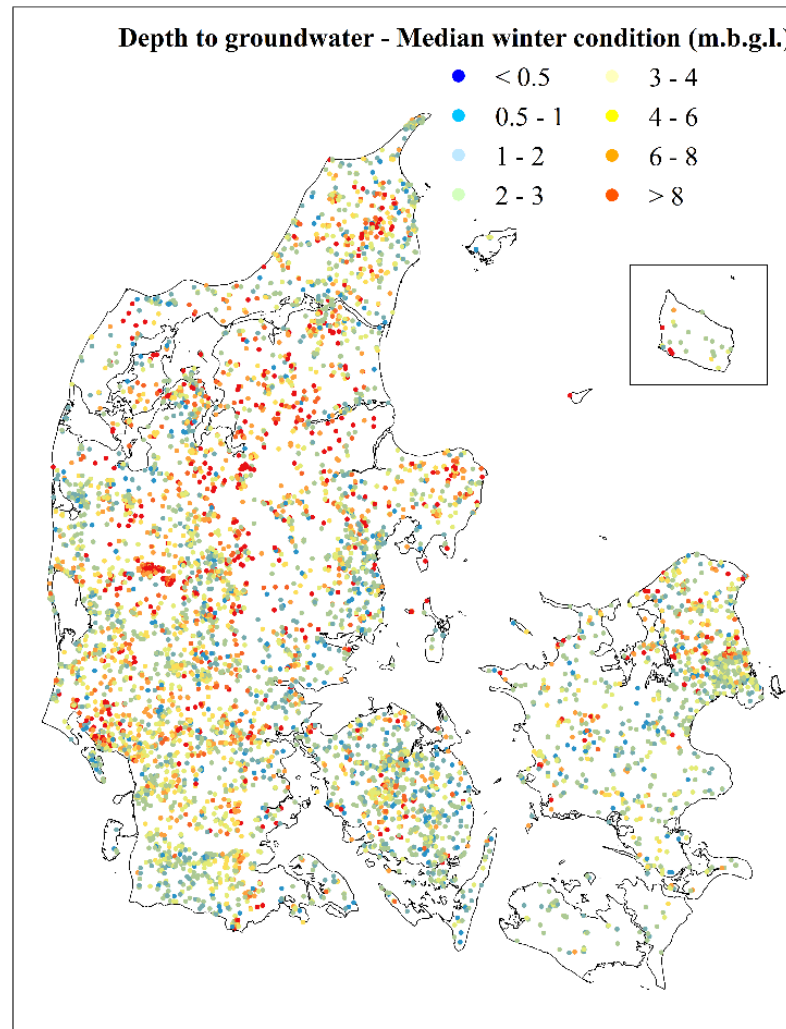
13,000 pejlinger

19,000 søer

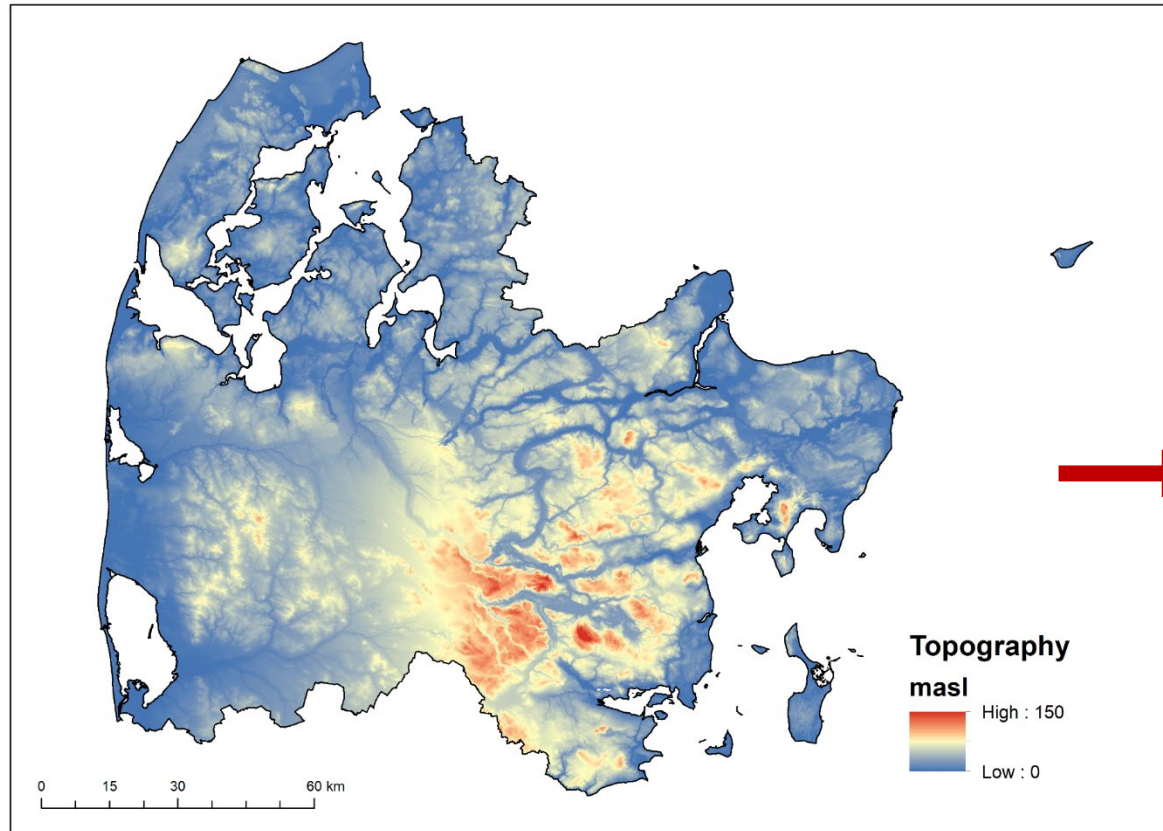
2,000 støttepunkter

14 forklarende variabler

Data type	antal
Boring (både sommer og vinter)	1378
Boring (sommer [J,J,A] gap-filled)	5620
Boring (vinter [D,J,F] gap-filled)	5989

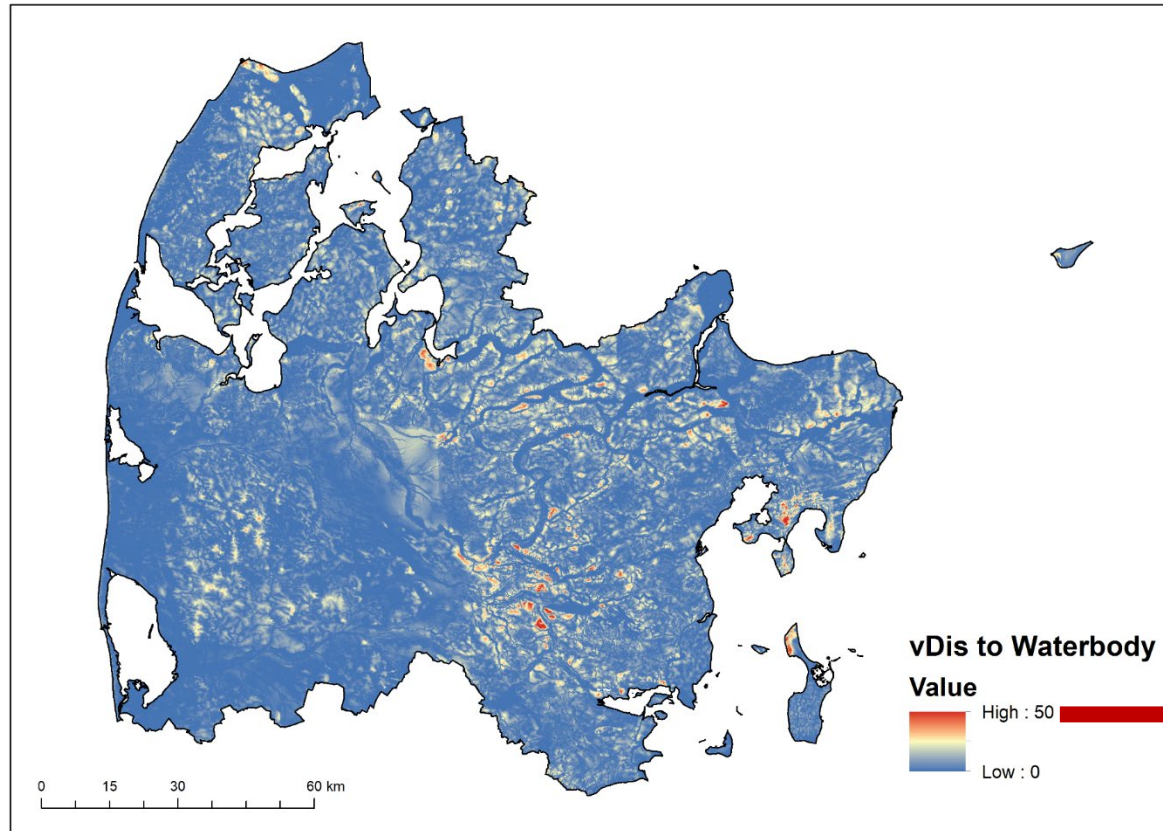


Forklarende variabler



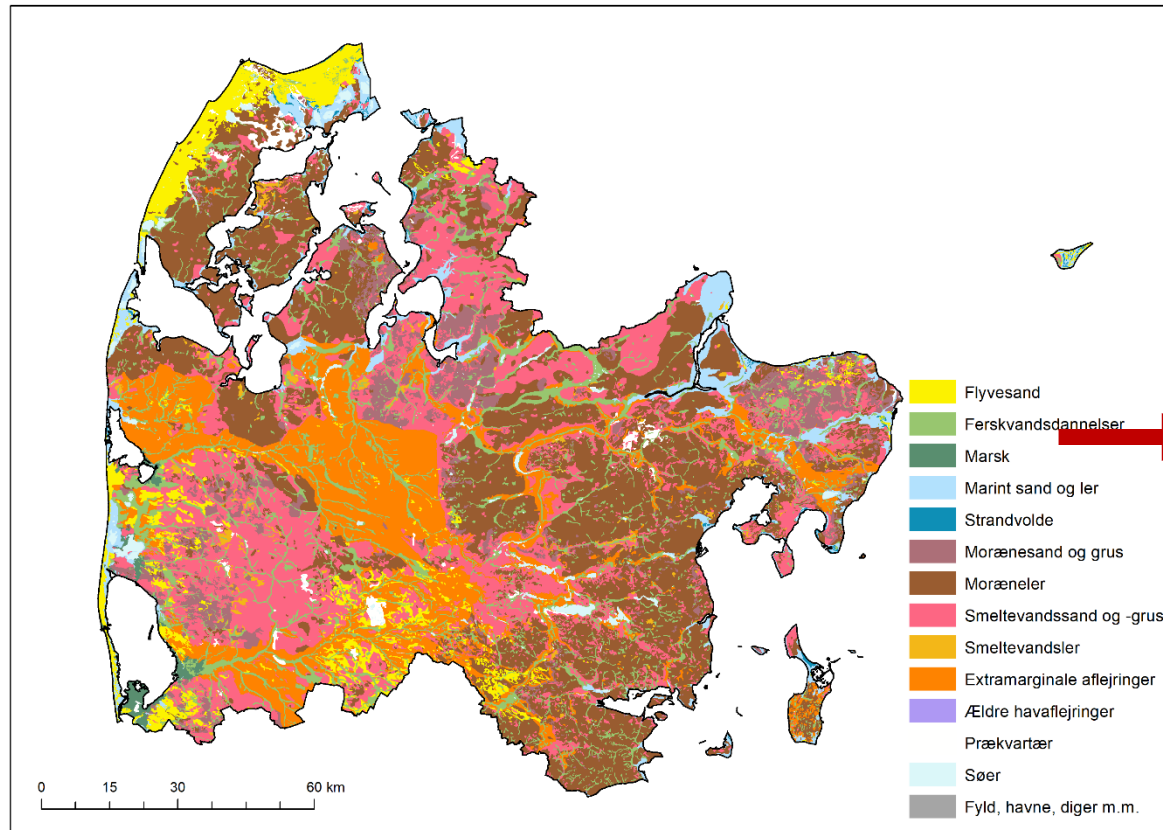
Variable	Gruppe
Ler indhold - a horisont	Geologi
Ler indhold - b horisont	
Ler indhold - c horisont	
Ler indhold - d horisont	
Kvartær lagtykkelse	
Top ler tykkelse	
Dræn sandsynlighed	
Dræn klasser	
Lavbund klassifikation	
Landskabstypologi	
Georegion	
Jordtype	Topografi
Højdemodel	
Højdemodel detrend	
Topographic Wetness Indeks	
Saga Wetness Indeks	
Opstrøms areal	
Hældning	
vertikal afstand til vandløb	Afstand til overfladvand
vertikal afstand til vandløb	
horizontal afstand til vandløb	
sø, vandløb, kyst klassifikation	Nedbør
Nedbør	
Befæstelsesgrad	Arealanvendelse
Arealanvendelse	
Koordinater (utm _x)	Koordinater
Koordinater (utm _y)	

Forklarende variabler



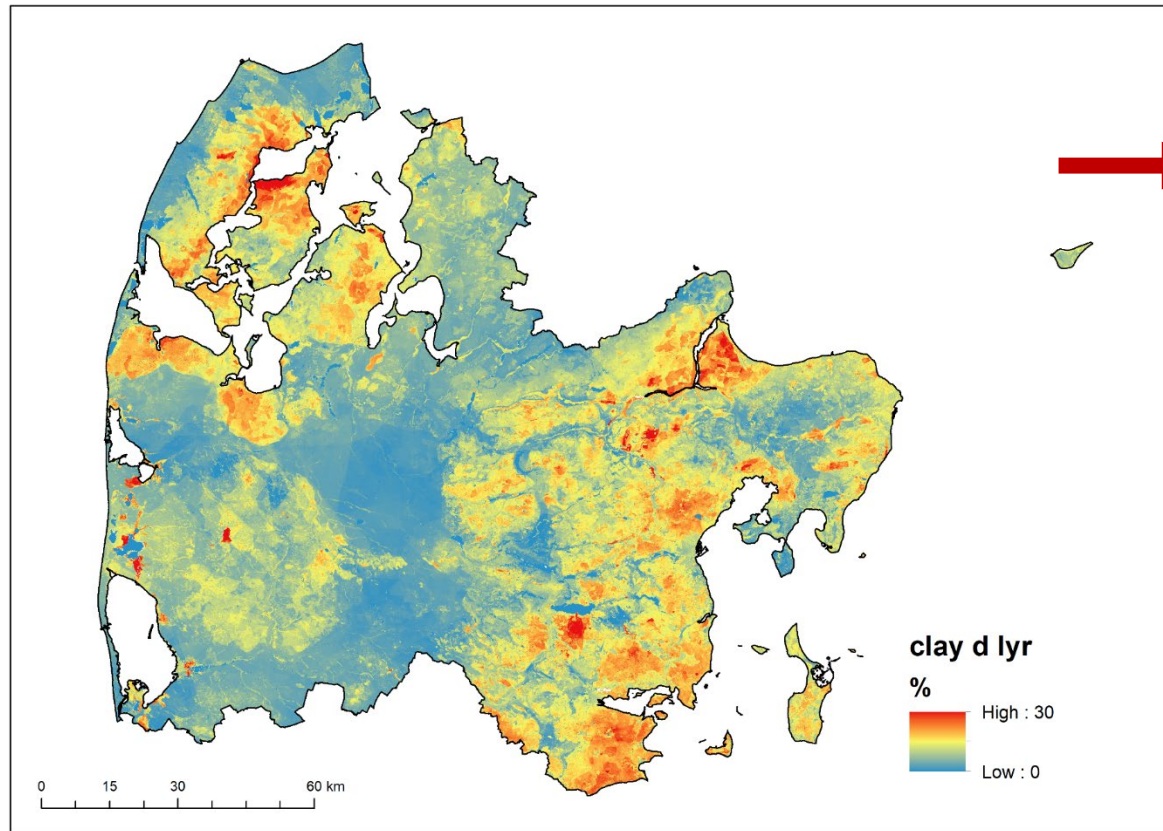
Variable	Gruppe
Ler indhold - a horisont	Geologi
Ler indhold - b horisont	
Ler indhold - c horisont	
Ler indhold - d horisont	
Kvartær lagtykkelse	
Top ler tykkelse	
Dræn sandsynlighed	
Dræn klasser	
Lavbund klassifikation	
Landskabstypologi	
Georegion	
Jordtype	Topografi
Højdemodel	
Højdemodel detrend	
Topographic Wetness Indeks	
Saga Wetness Indeks	
Opstrøms areal	
Hældning	
vertikal afstand til vandløb	Afstand til overfladvand
vertikal afstand til vandløb	
horizontal afstand til vandløb	
sø, vandløb, kyst klassifikation	Nedbør
Nedbør	
Befæstelsesgrad	Arealanvendelse
Arealanvendelse	
Koordinater (utm _x)	Koordinater
Koordinater (utm _y)	

Forklarende variabler



Variable	Gruppe
Ler indhold - a horisont	Geologi
Ler indhold - b horisont	
Ler indhold - c horisont	
Ler indhold - d horisont	
Kvartær lagtykkelse	
Top ler tykkelse	
Dræn sandsynlighed	
Dræn klasser	
Lavbund klassifikation	
Landskabstypologi	
Georegion	
Jordtype	Topografi
Højdemodel	
Højdemodel detrend	
Topographic Wetness Indeks	
Saga Wetness Indeks	
Opstrøms areal	
Hældning	
vertikal afstand til vandløb	Afstand til overfladvand
vertikal afstand til vandløb	
horizontal afstand til vandløb	
sø, vandløb, kyst klassifikation	Nedbør
Nedbør	
Befæstelsesgrad	Arealanvendelse
Arealanvendelse	
Koordinater (utm _x)	Koordinater
Koordinater (utm _y)	

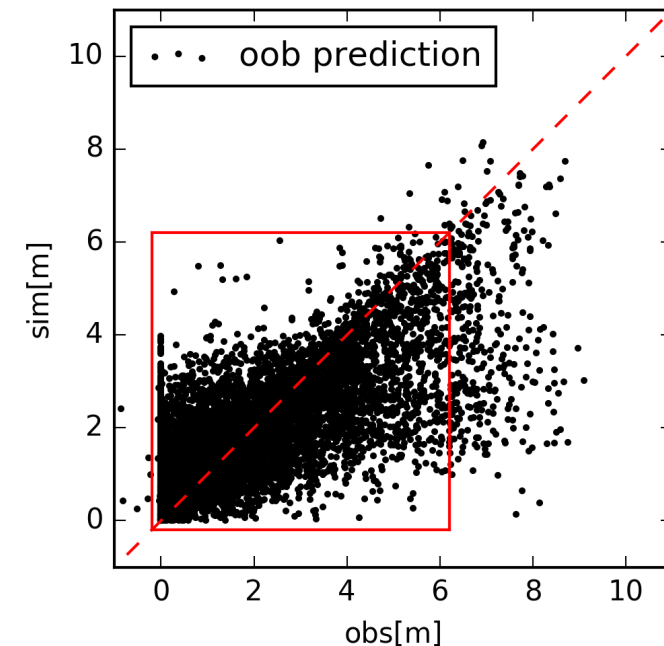
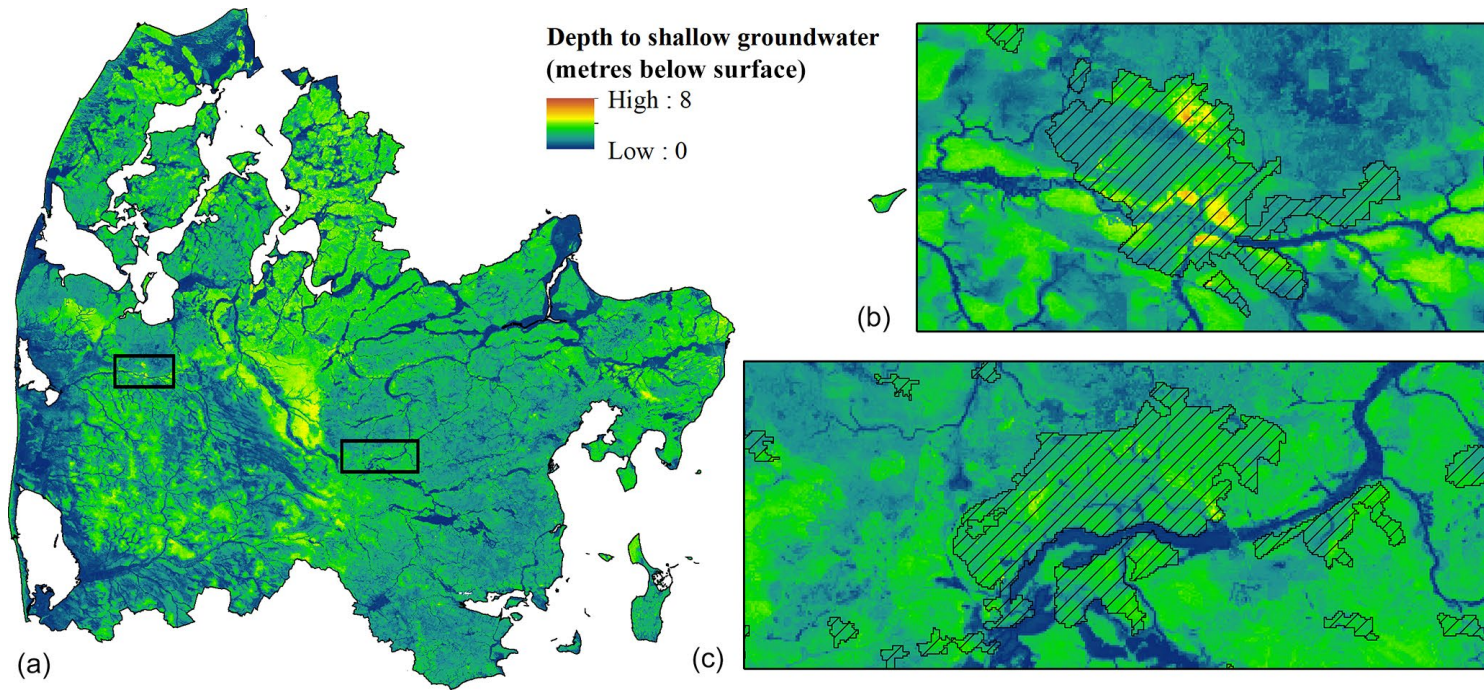
Forklarende variabler



Variable	Gruppe
Ler indhold - a horisont	Geologi
Ler indhold - b horisont	
Ler indhold - c horisont	
Ler indhold - d horisont	
Kvartær lagtykkelse	
Top ler tykkelse	
Dræn sandsynlighed	
Dræn klasser	
Lavbund klassifikation	
Landskabstypologi	
Georegion	Topografi
Jordtype	
Højdemodel	
Højdemodel detrend	
Topographic Wetness Indeks	
Saga Wetness Indeks	
Opstrøms areal	
Hældning	
vertikal afstand til vandløb	Afstand til overfladvand
vertikal afstand til vandløb	
horizontal afstand til vandløb	
sø, vandløb, kyst klassifikation	Nedbør
Nedbør	
Befæstelsesgrad	Arealanvendelse
Arealanvendelse	
Koordinater (utm _x)	Koordinater
Koordinater (utm _y)	

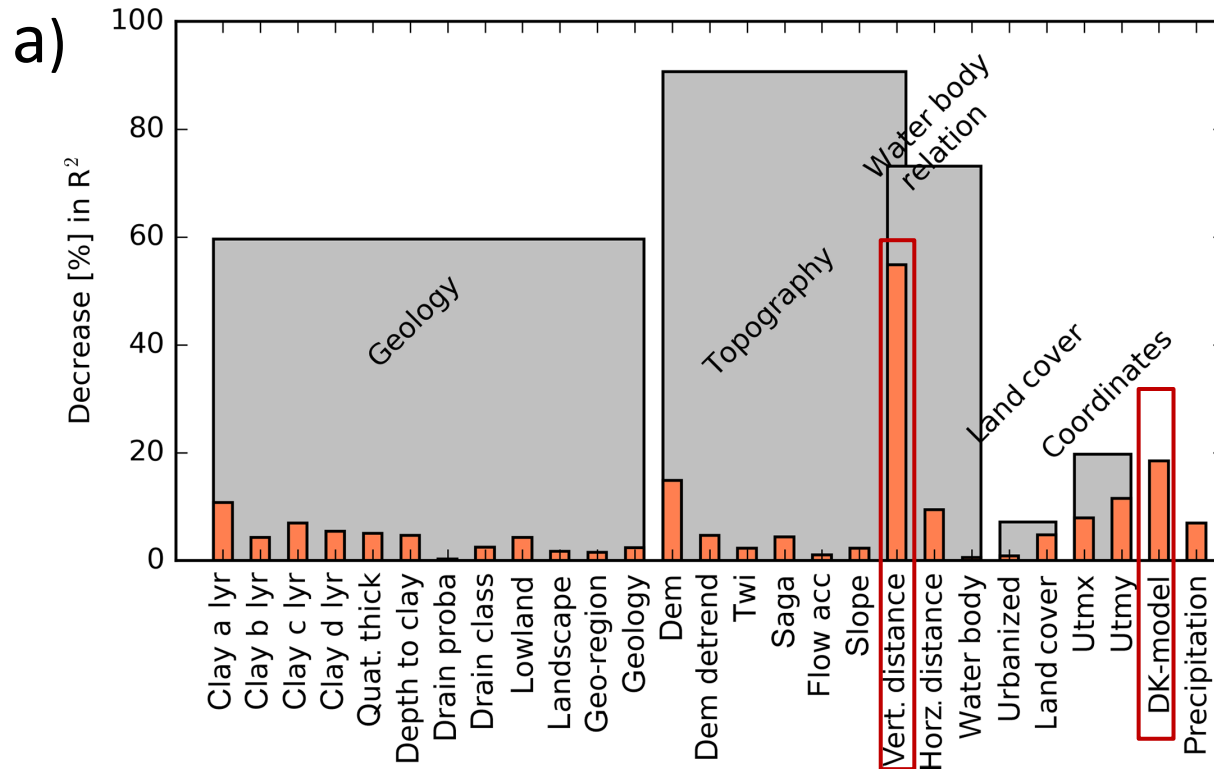
Resultater C2CCC

- 29 % of cellerne har en dybde < 1 m
- Valideringstest: MAE=0.8m, RMSE=1.1m, $R^2=0.6$



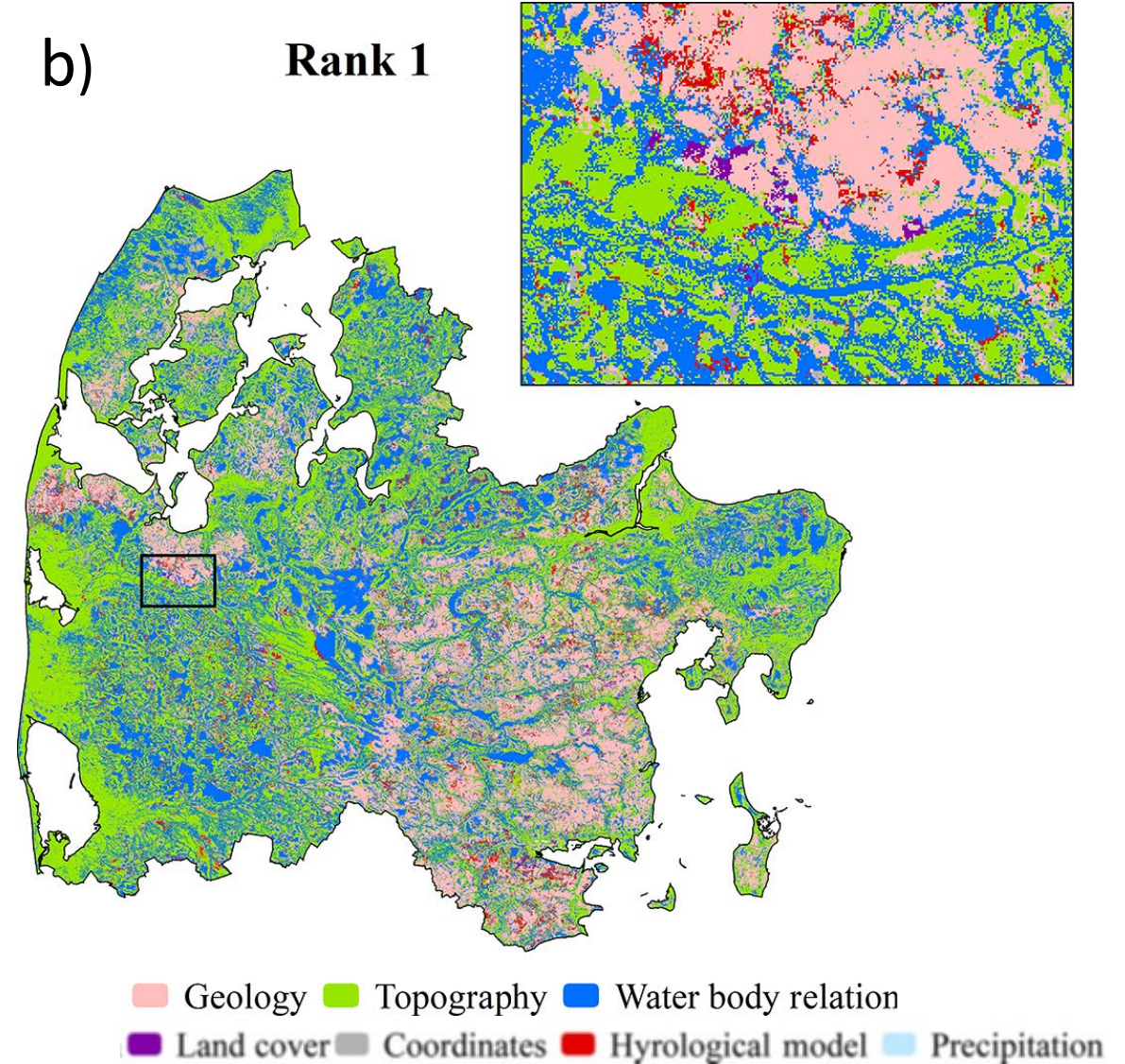
Resultater C2CCC

Følsomhedsanalyse – forklarende variable



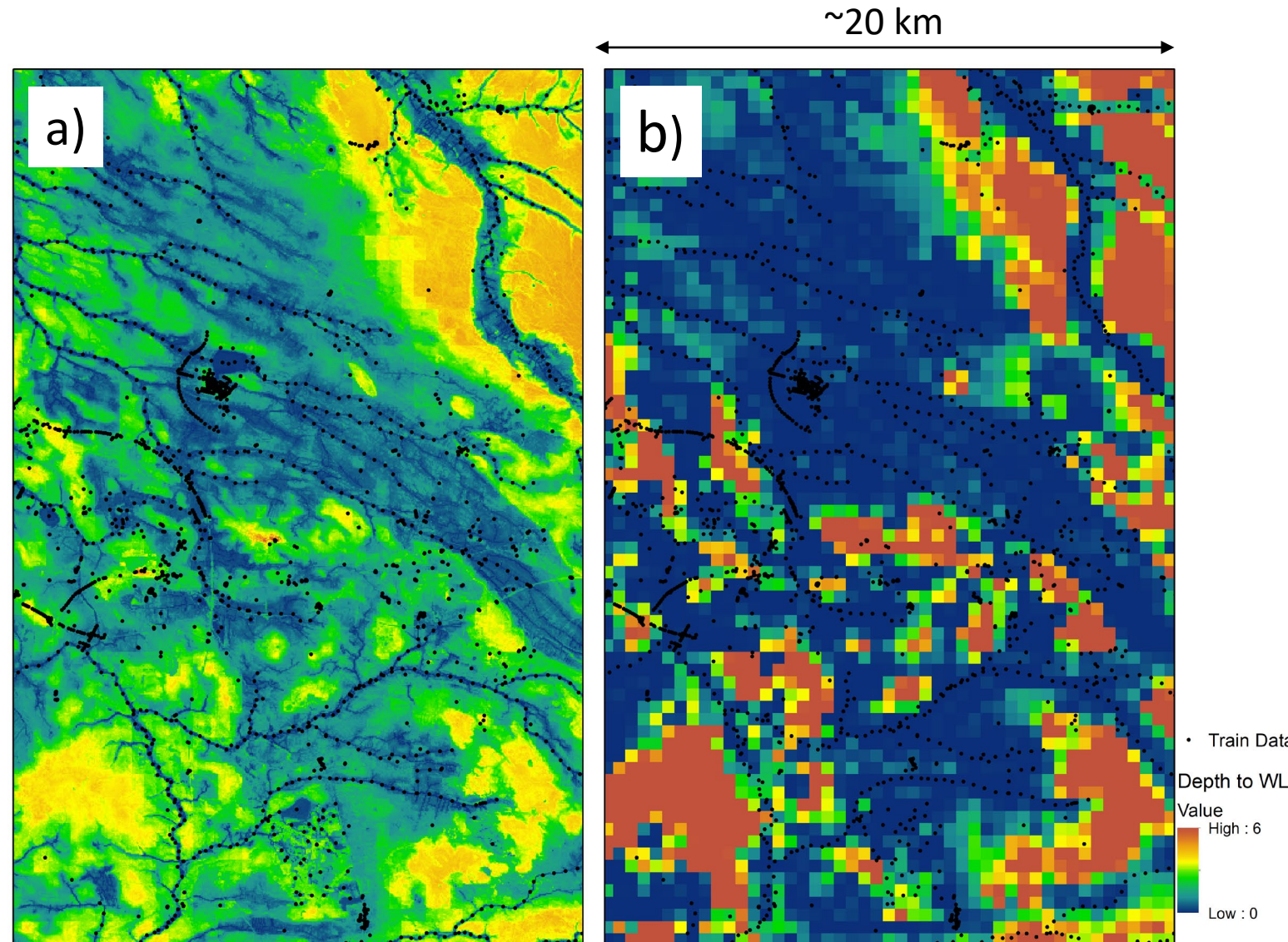
b)

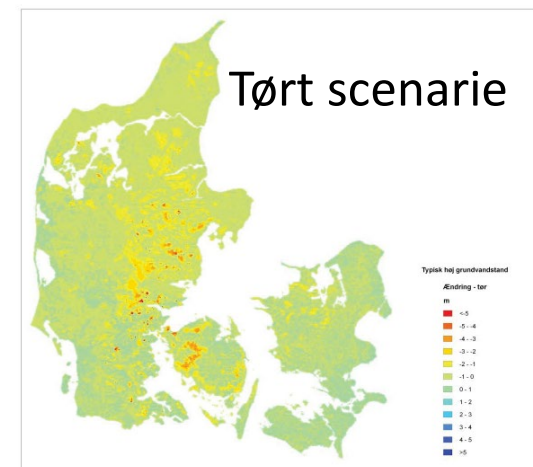
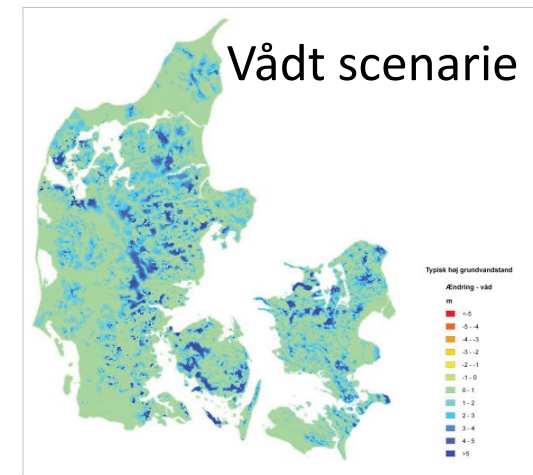
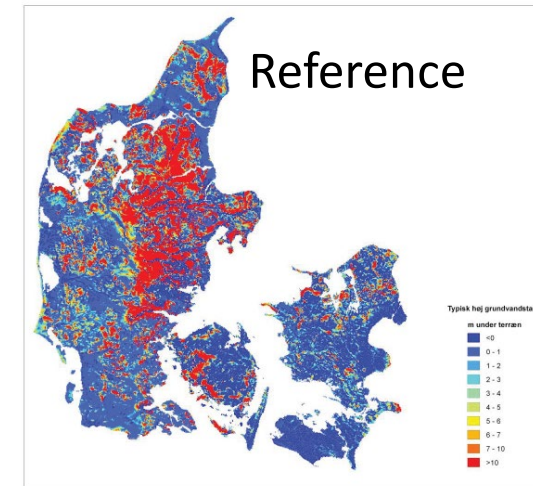
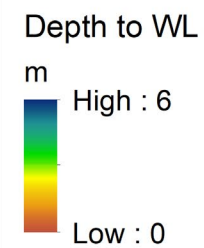
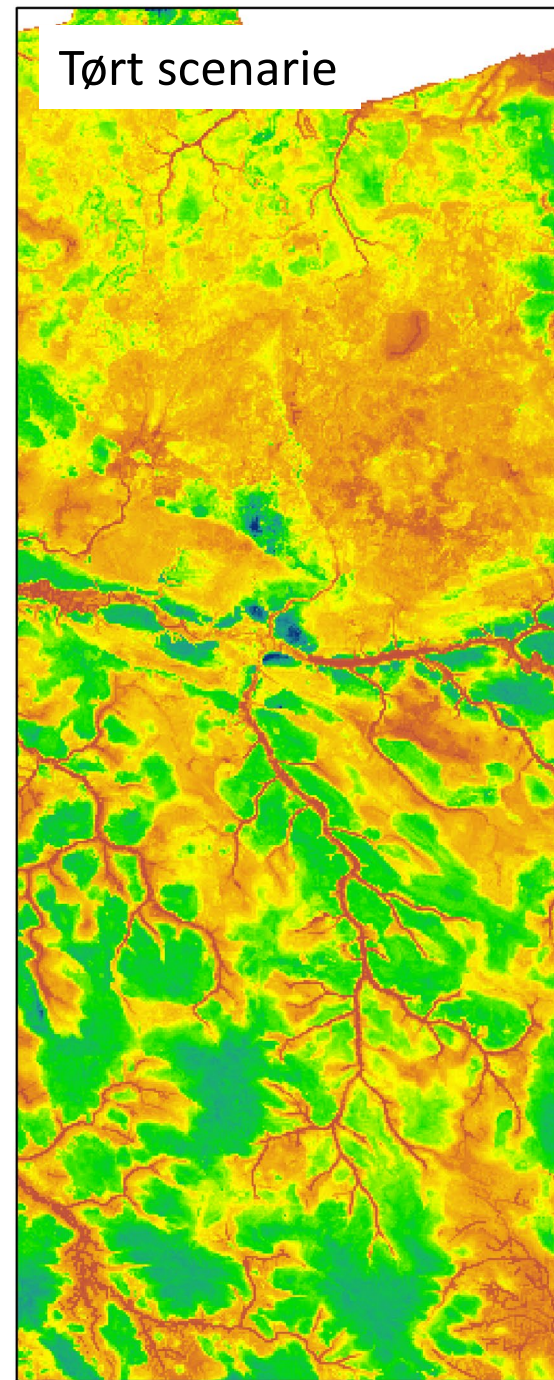
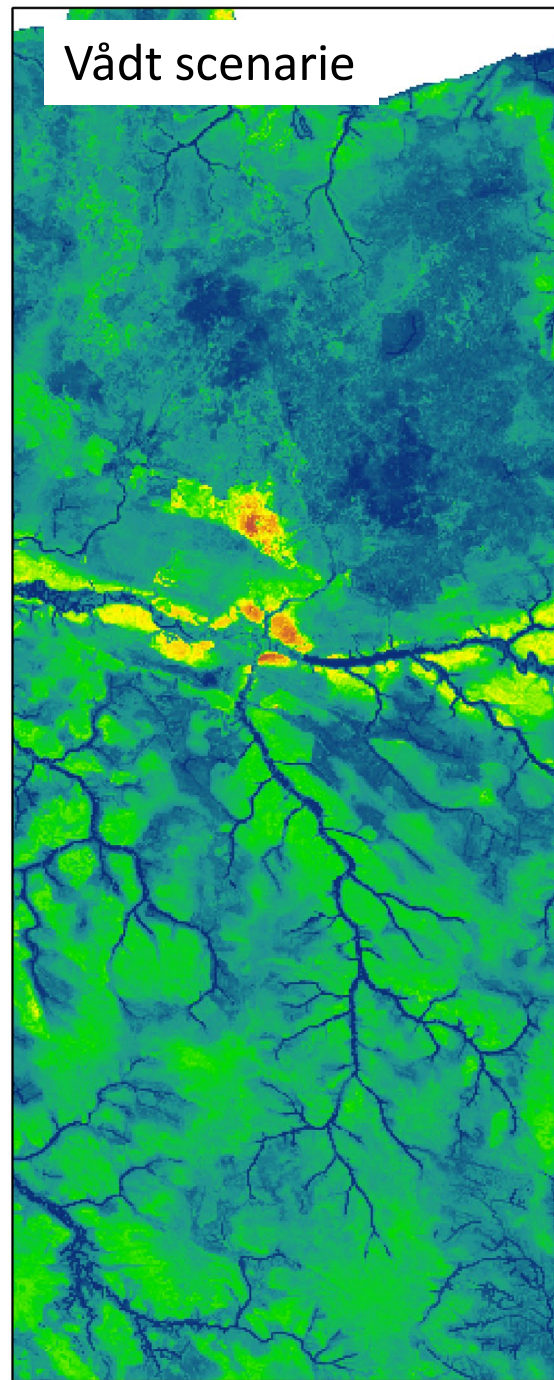
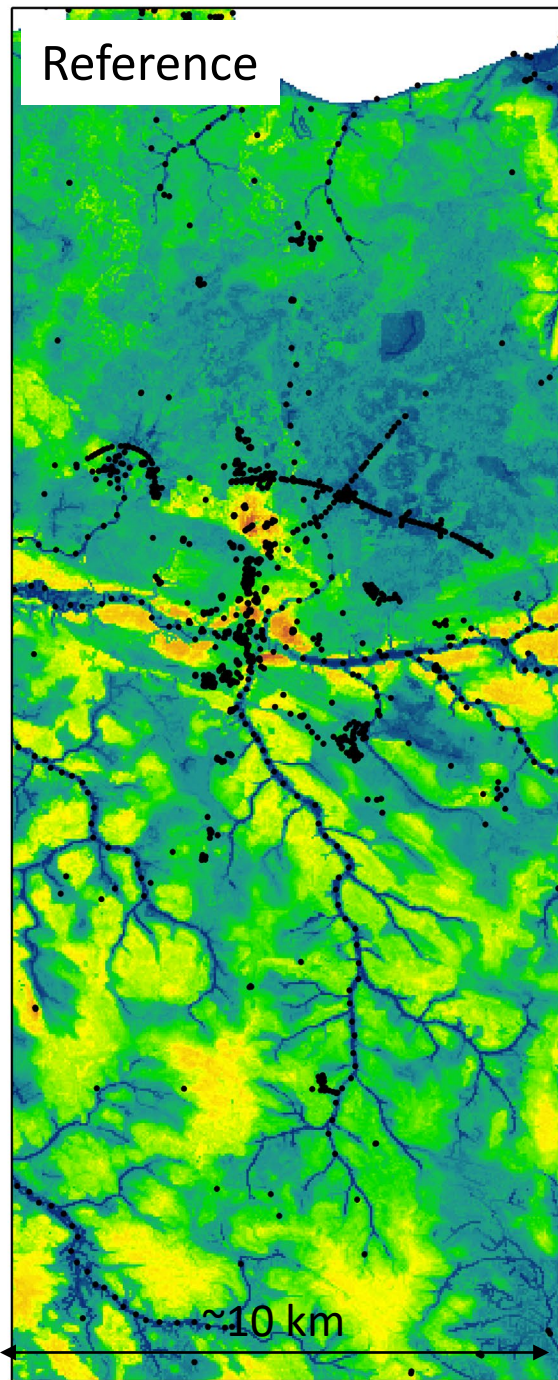
Rank 1



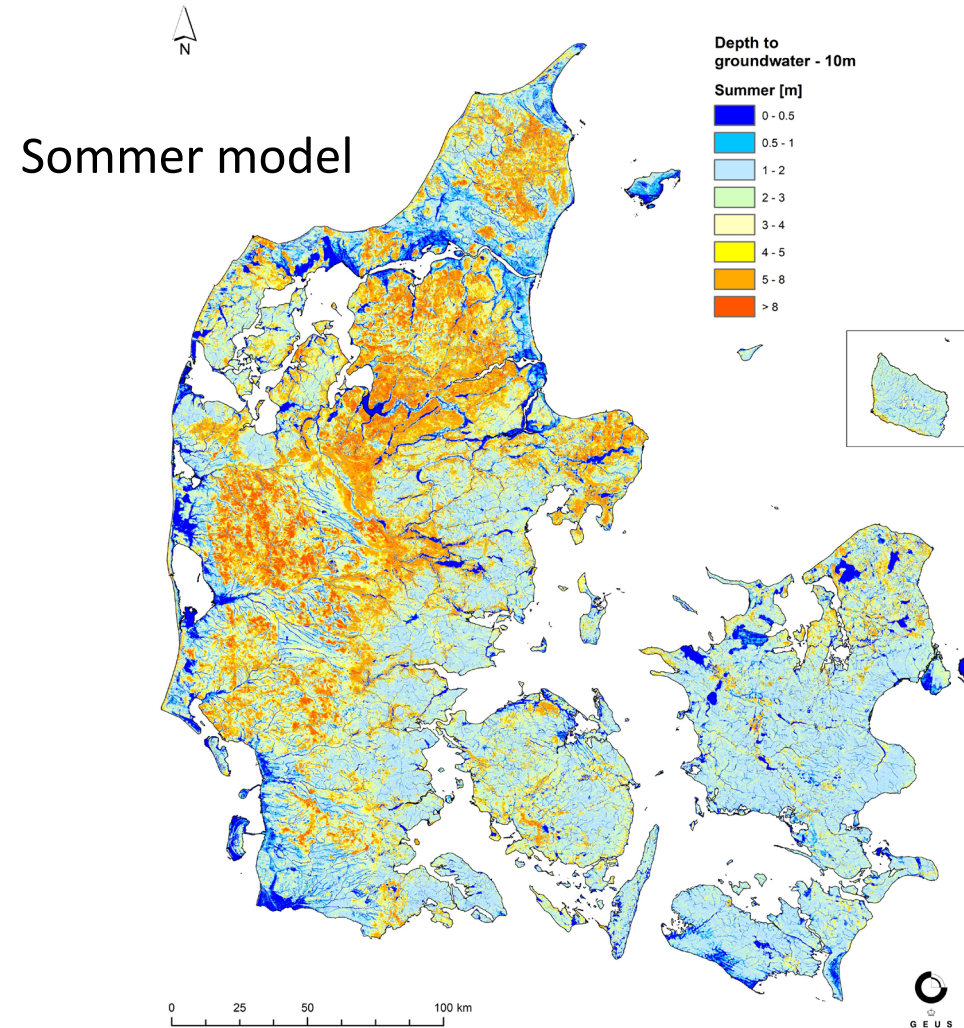
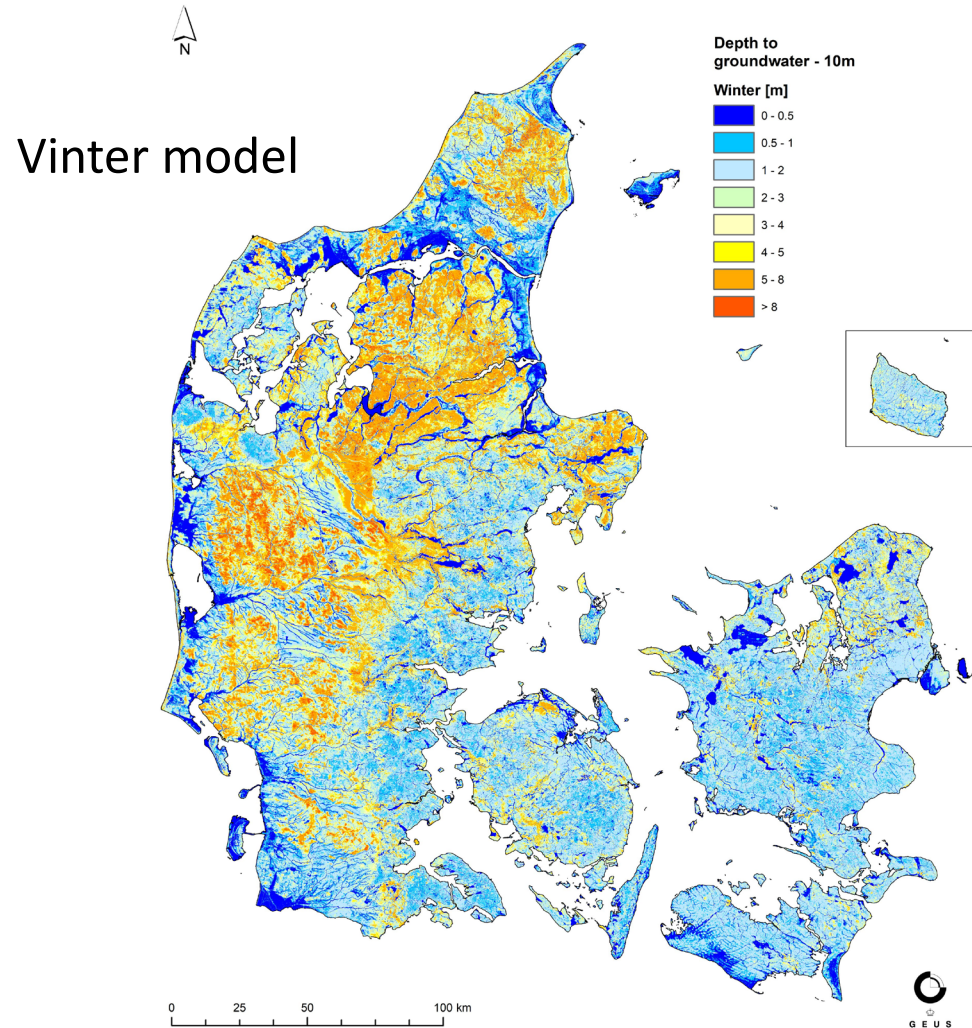
Resultater C2CCC

- a) ML 50 m opløsning
- b) DK-Model 500 m resolution



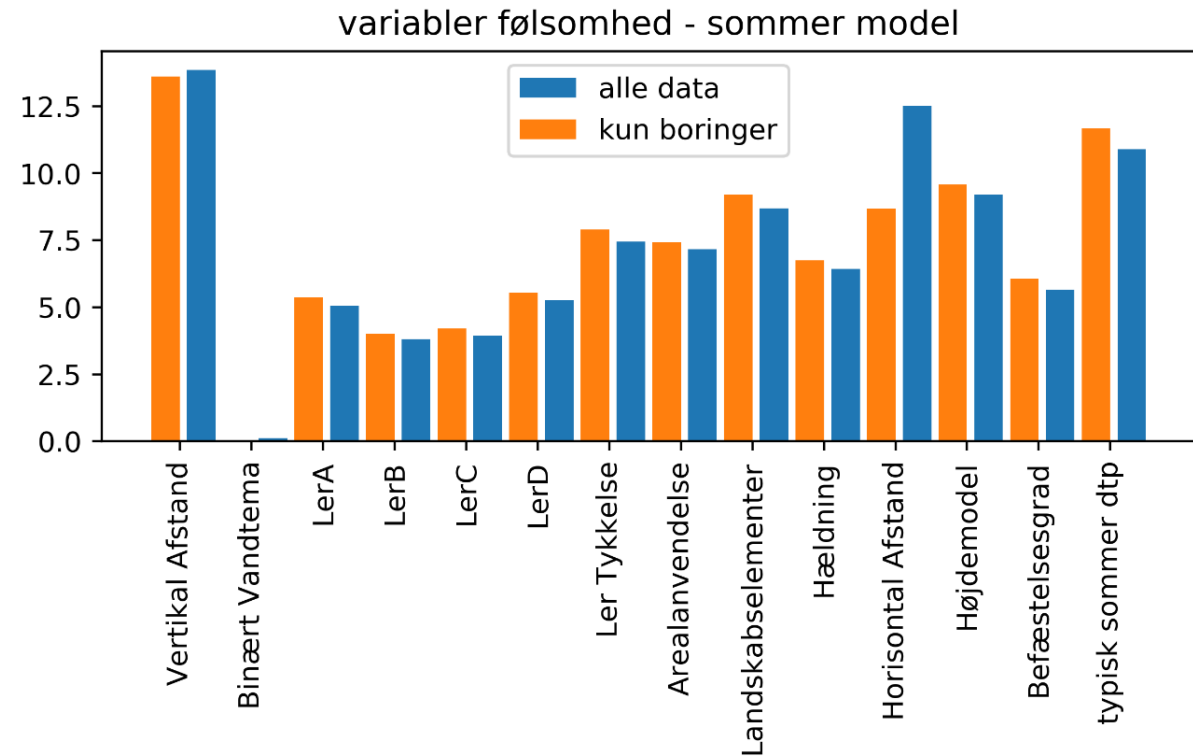


Resultater HIP4PLUS



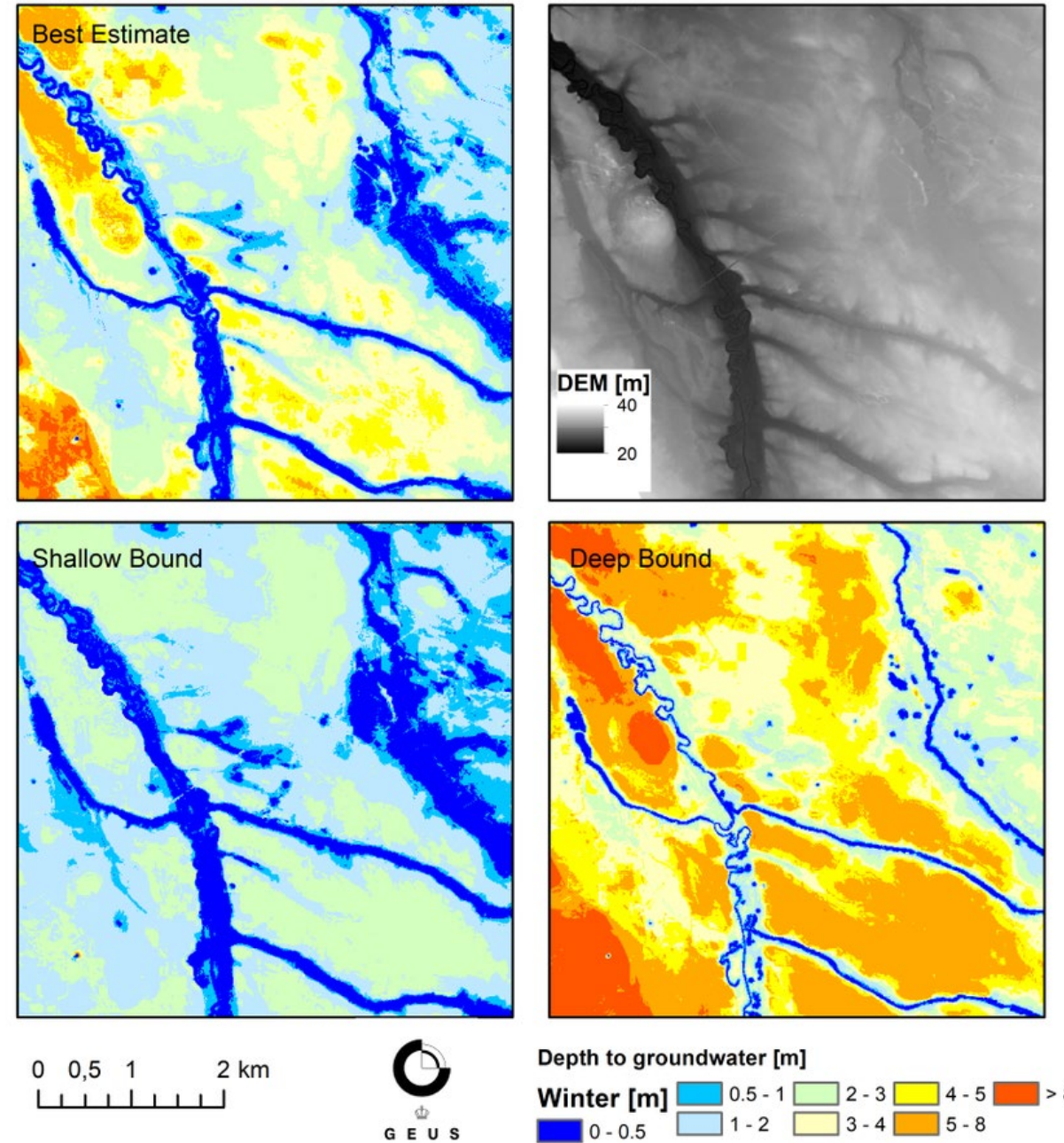
Resultater HIP4PLUS

vinter model		cv1	cv2	cv3	cv4	middel
Alle Data	MAE	0.47	0.47	0.47	0.45	0.47
	RMSE	1.23	1.21	1.21	1.18	1.20
	R2	0.72	0.72	0.71	0.73	0.72
Kun Boringer	MAE	1.16	1.17	1.18	1.13	1.16
	RMSE	1.94	1.92	1.93	1.89	1.92
	R2	0.51	0.52	0.51	0.53	0.51
Kun Søer	MAE	0.03	0.03	0.03	0.03	0.03
	RMSE	0.26	0.25	0.28	0.23	0.25
	R2					
Kun støttepunkter	MAE	0.09	0.08	0.11	0.10	0.10
	RMSE	0.27	0.22	0.28	0.27	0.26
	R2					

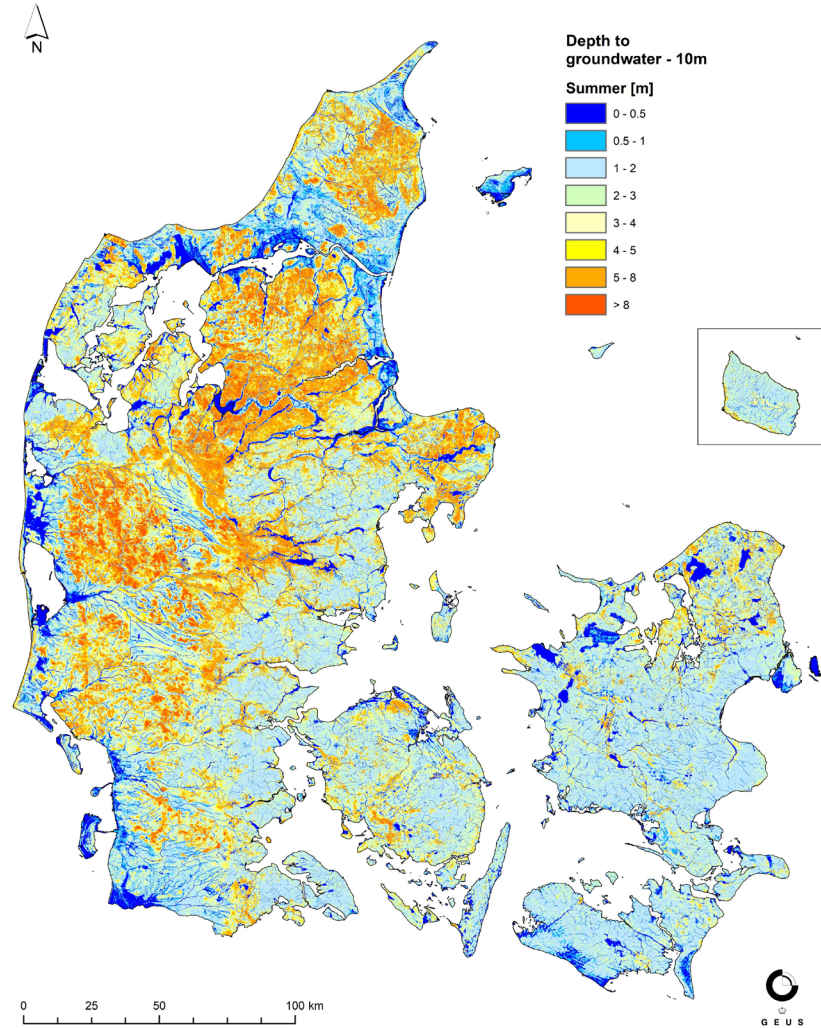


Resultater HIP4PLUS

- Usikkerhedskvantificering ved hjælp af “quantile regression”
- Angivelse af fraktil i objektiv funktionen
- 10% og 90% som svare til et 80% konfidensinterval

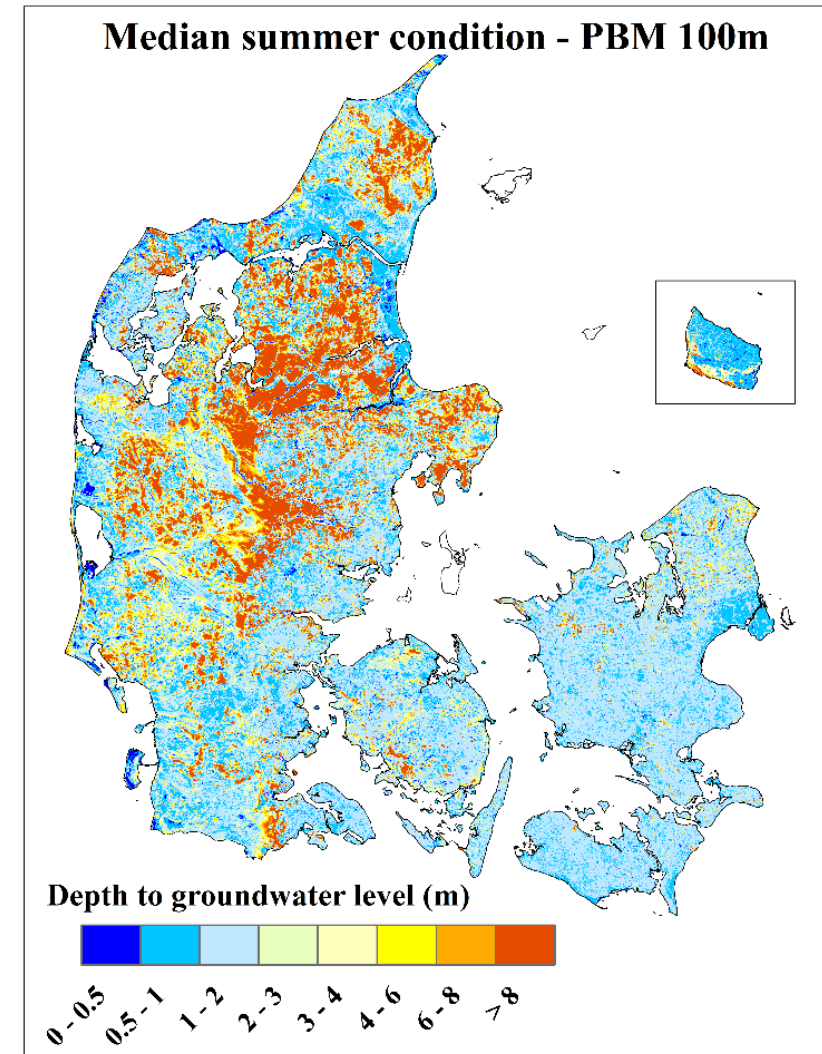


Resultater HIP4PLUS



ATV Øst - Terrænnært Grundvand

27/Januar/2021



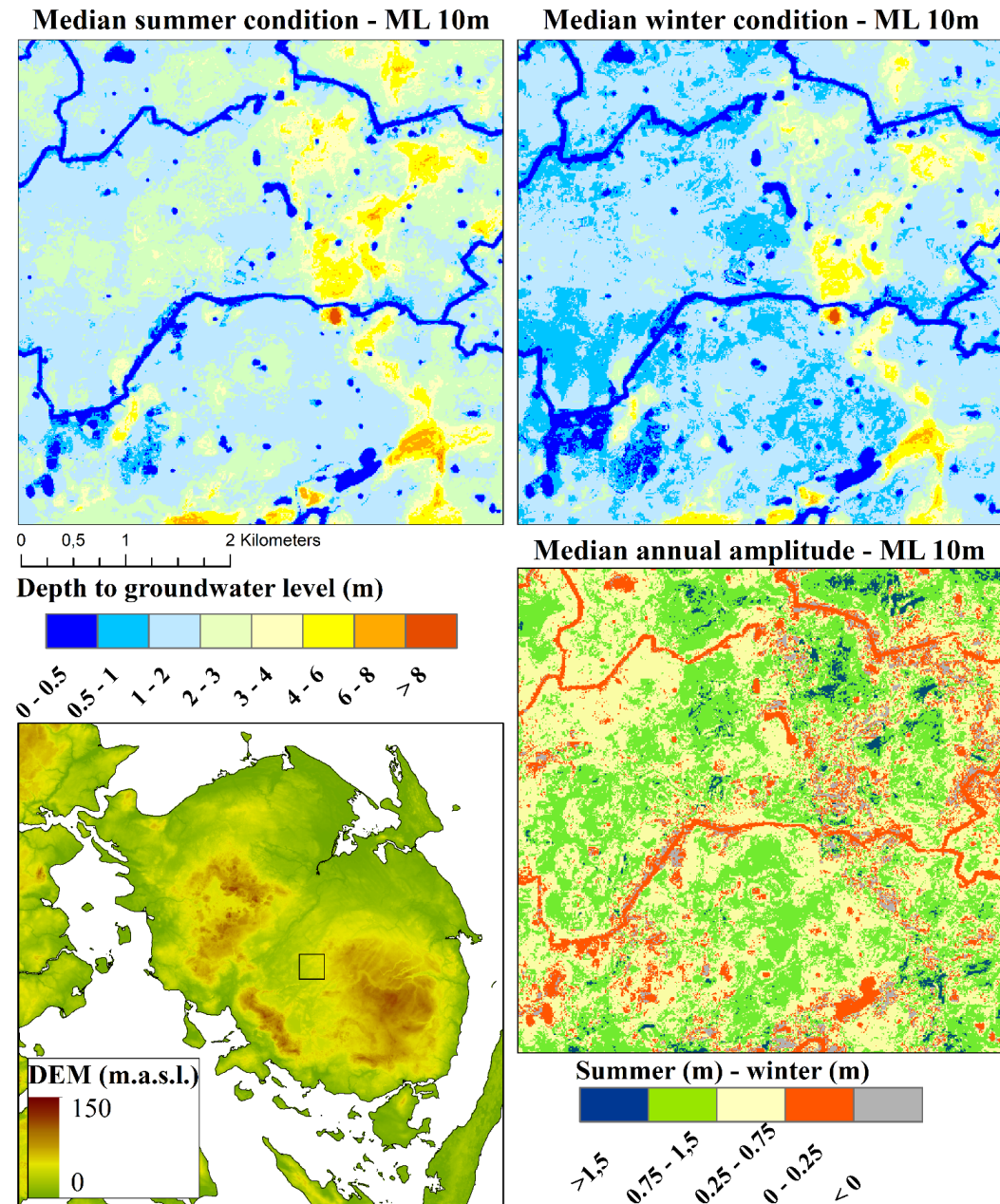
ML og PB modeller

I HIP4PLUS ML modellerne er MikeSHE simuleringer brugt til:

- 1) Udvidelse af træningsdata ved hjælp af sæsonvariationen
- 2) Model input/forklarende variabler

I C2CCC er MikeSHE simuleringer også brugt til:

- 3) Fremskrivning af pejlinger



Refleksioner

- ML kan anvendes til modellering af tnGV i høj rumlig opløsning
- ML er et fleksibelt værktøj der kan integrere alle relevante data
- Synergi mellem ML og proces-baseret modeller
- ML kan bruges til usikkerhedsanalyse, følsomhedsanalyse og klimafremskrivninger
- Tidsdimension er meget forsimplet (indtil videre)
- Resultater på screeningsniveau
- Modellen kan nemt opdateres med nye data