Prediction of drain flow fraction at high spatial resolution by combining physically based models and machine learning

Raphael Schneider[1*], Hafsa Mahmood[2], Rasmus Rumph Frederiksen[3], Anker Lajer Højberg[1], Simon Stisen[1]

[1]*Geologic Survey of Denmark and Greenland (GEUS), Department of Hydrology, Copenhagen, Denmark*
[2]*Aarhus University, Department of Geoscience, Aarhus, Denmark*
[3]*Aarhus University, Department of Ecoscience, Silkeborg, Denmark*
*rs@geus.dk

EGU22-3694

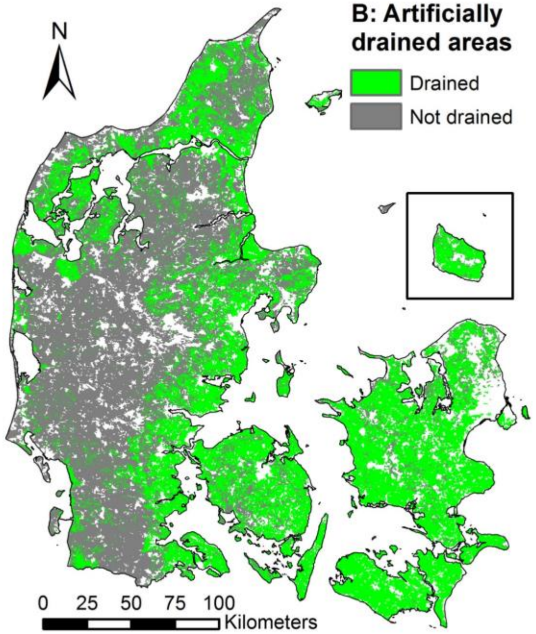
AARHUS UNIVERSITET

GEUS

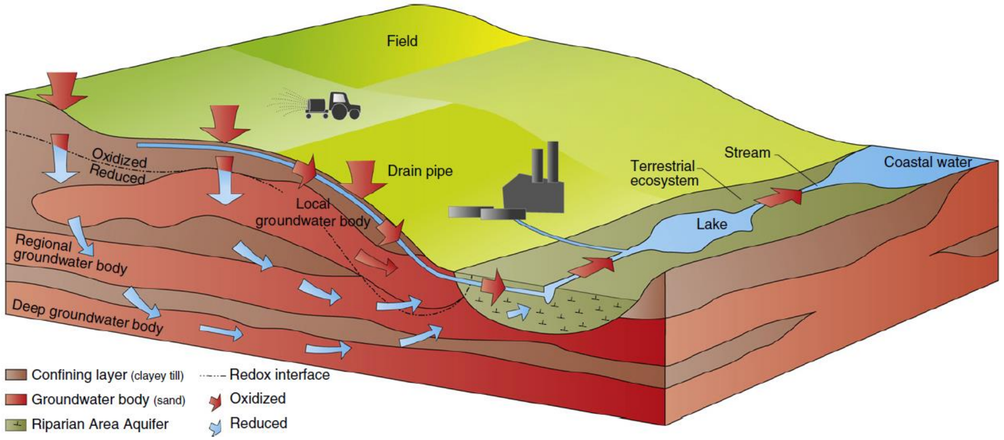
Background:
artificial drainage in Denmark

Artificial drainage in Denmark…

- is abundant: ~50% of agricultural land is artificially drained (mostly tile drains)
- has profound effect on entire hydrological cycle **(→shortcut for nitrate transport)**



Machine learning-based estimate of artificially drained areas in DK (Møller et al., 2018)

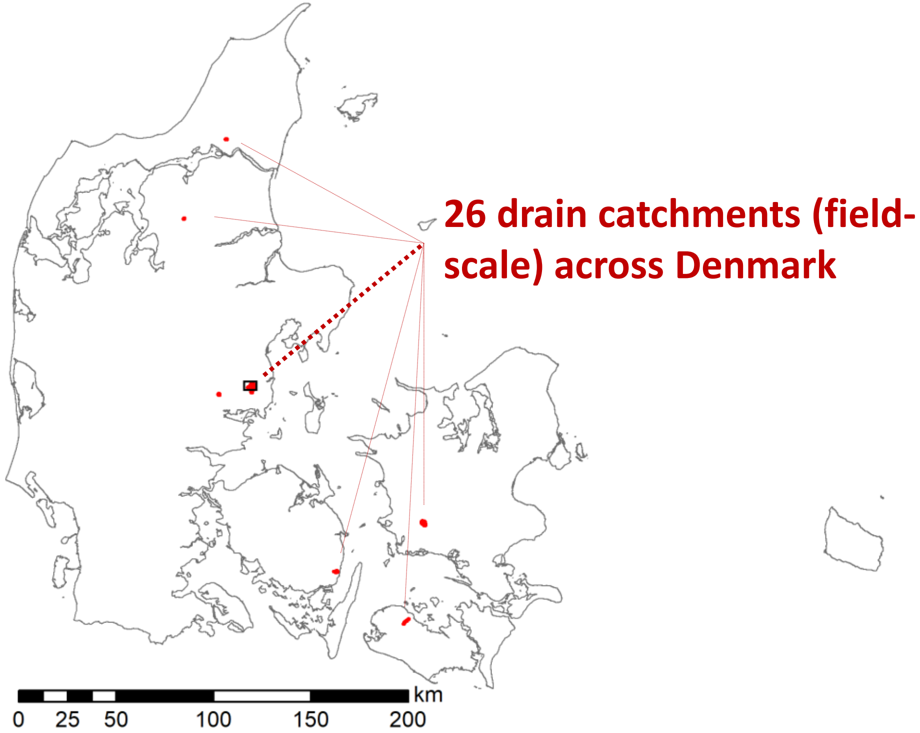


Conceptualized water flow paths and related nitrate reduction in Danish glacial till landscape (Refsgaard et al., 2014)

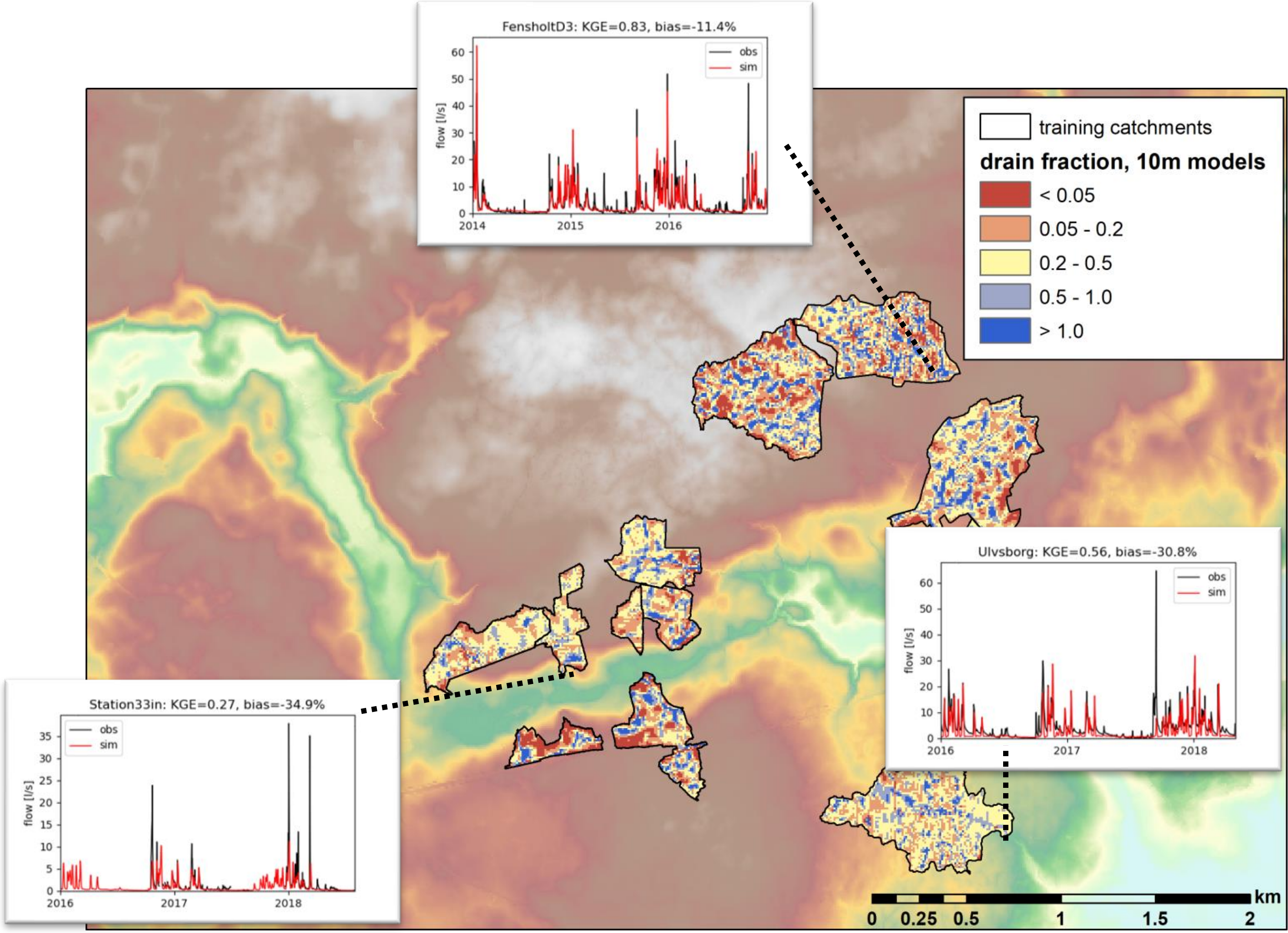
Despite this, challenging to quantify on large scale:

- insufficient data on drain infrastructure and drain flow
- highly variable in space and time (controlled by topography, geology, …)

Drain catchment modelling with physically based models



MIKE SHE hydrological models
10m horizontal resolution
26 models calibrated combined against drainflow observations



More on model calibration: EGU22-6315. Assessing the physical controls of simulated drain flow dynamics, Mahmood et al. (HS2.2.1, today 24-05-22 at 17:12 in room B)

Drain catchment modelling with physically based models

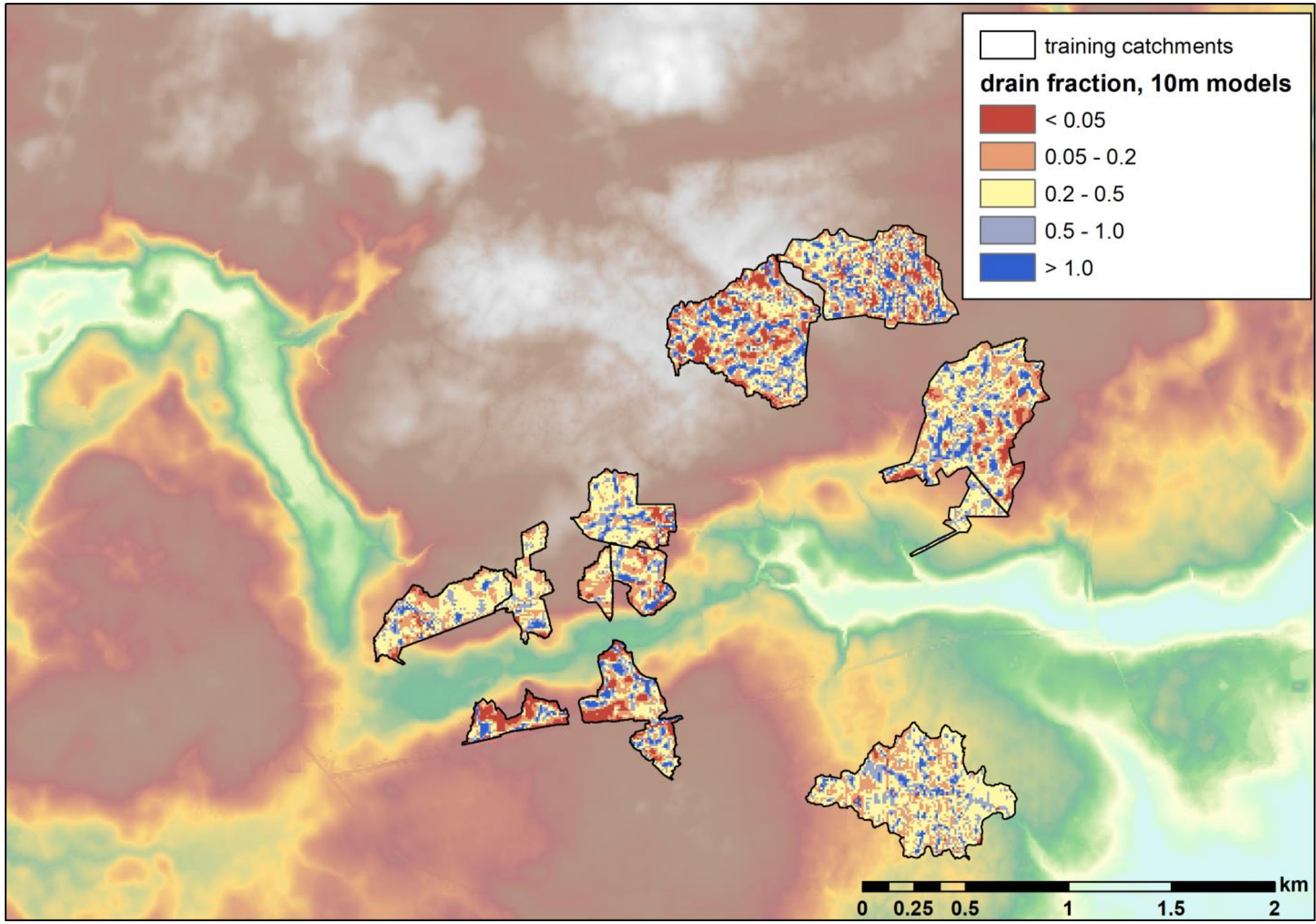
Relevant in context of nitrate transport etc:
drain fraction, either

relative to recharge $$drfr_{re} = \frac{drain}{recharge}$$

or

relative to precipitation $$drfr_P = \frac{drain}{P}$$

(examples shown for *drfr*$_P$; results for *drfr*$_{re}$ are comparable)



G E U S

Machine learning predictions of drain fraction

Relevant in context of nitrate transport etc:
drain fraction

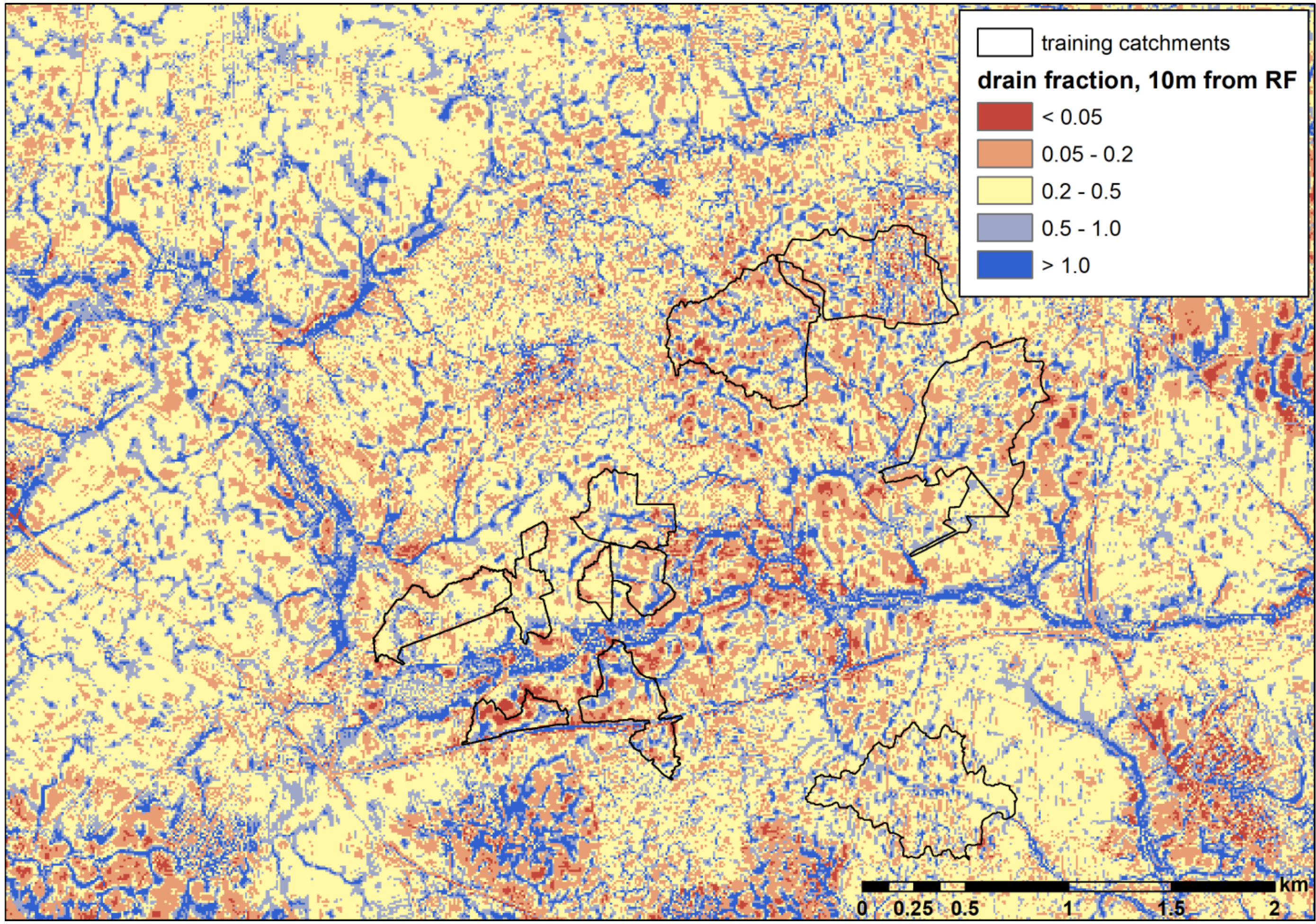
$$drfr_{re} = \frac{drain}{recharge}$$

$$drfr_P = \frac{drain}{P}$$

→**Random Forest (RF) regressor model to predict drain fraction outside of hydrological models**
Covariates:

- slope
- relative topography
- TWI
- curvature
- clay fraction
- clay thickness

Trained against drain fractions simulated by hydrological models



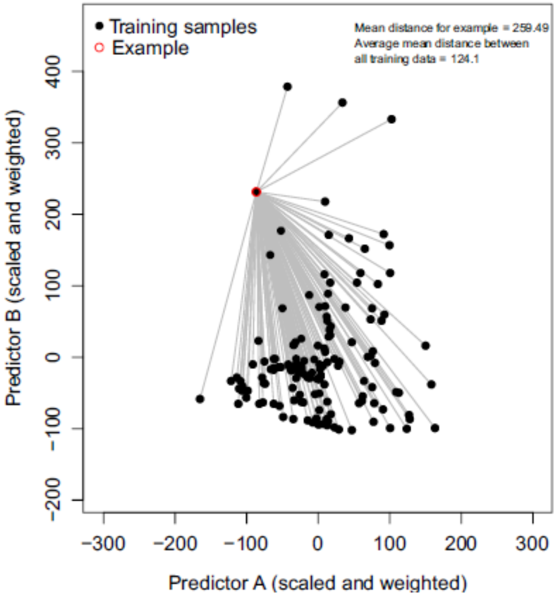
Machine learning predictions of drain fraction – spatial transferrability

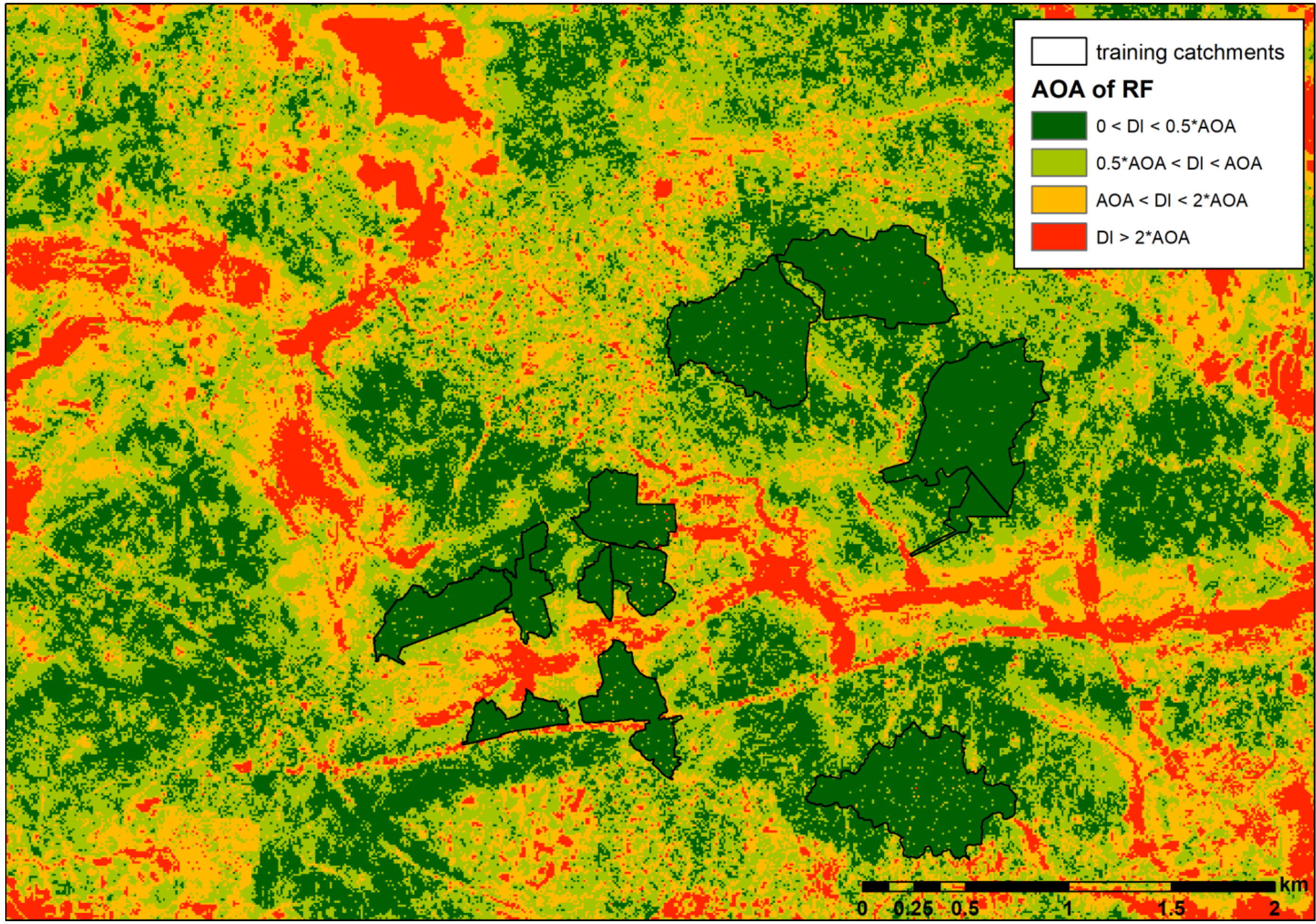
Spatial transferrability – common problem with ML methods such as Random Forest (RF)

random cross validation ⟷ **spatial cross validation**

Pearson R >0.9 | **Pearson R ~0.8 to ~0**

Evaluation of spatial transferrability with concept of **dissimilarity index (DI)** and related **area of applicability (AOA)** by Meyer et al., 2021





GEUS

Conclusions and Outlook

- Developped Random Forest regressor model to predict drain fraction at high resolution
- Applied dissimilarity index (DI) to estimate area of applicability (AOA) in a real-world example

Future work:

- Increase training data in a targeted manner, aided by DI/AOA
- Limit areas for predictions and estimate errors aided by DI/AOA
- Investigate seasonality of drain fraction

→ Develop 10m resolution **DK-wide map of drain fraction,**
limited to agricultural areas, to be used in work related to nitrate retention etc.

Acknowledgements

This work was performed within the T-REX project (Terrænnær redox og retentionskortlægning til differentieret målrettet virkemiddelsindsats indenfor ID15 oplande), funded by the Danish GUDP (Grønt Udviklings- og Demonstrationsprogram), project number 34009-18-1453.

→For **more details on the hydrological model calibration** and physical controls on simulated drain flow, see:
Mahmood, H., Schneider, R., Frederiksen, R., Christiansen, A., & Stisen, S. (2022). Assessing the physical controls of simulated drain flow dynamics. EGU General Assembly 2022, 6315. https://doi.org/doi.org/10.5194/egusphere-egu22-6315, **presented in HS2.2.1, today 24-05-22 at 17:12 in room B**

References

Møller, A. B., Beucher, A., Iversen, B. V., & Greve, M. H. (2018). Predicting artificially drained areas by means of a selective model ensemble. Geoderma, 320, 30–42.

Refsgaard, J. C., Auken, E., Bamberg, C. A., Christensen, B. S. B., Clausen, T., Dalgaard, E., Effersø, F., Ernstsen, V., Gertz, F., Hansen, A. L., He, X., Jacobsen, B. H., Jensen, K. H., Jørgensen, F., Jørgensen, L. F., Koch, J., Nilsson, B., Petersen, C., De Schepper, G., … Viezzoli, A. (2014). Nitrate reduction in geologically heterogeneous catchments - A framework for assessing the scale of predictive capability of hydrological models. Science of the Total Environment, 468–469, 1278–1288.

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods in Ecology and Evolution, 12(9), 1620–1633.

Abstract

In Denmark, about half of the agricultural land is artificially drained. These drainage systems have a significant effect on the hydrological system. Knowledge about the spatio-temporal distribution of drain flow is crucial to understand aspects such as groundwater recharge, streamflow partitioning and nutrient transport. Still, quantification of drain flow at regional and large scale remains a major challenge: Data on the distribution of the installed subsurface drainage system are scarce, as are measurements of drain flow. Large-scale simulations of drain with physically-based hydrological models are challenged by scale, as drain flow is controlled by small-scale variations in groundwater depth often beyond the model resolution. Purely data-driven models can struggle representing the complex controls behind drain flow.

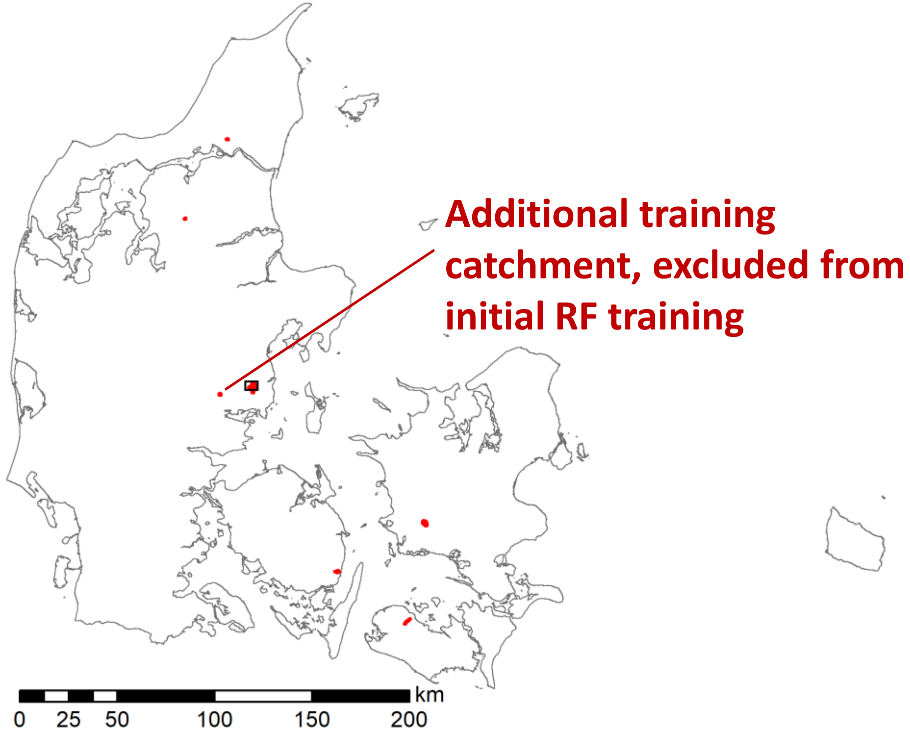
Here, we suggest a metamodel approach to obtain a more accurate estimate of generated drain flow at high spatial resolution of 10 m, combining physically-based with data-driven models. Our variable of interest is drain fraction, defined as the ratio between drain flow and recharge per grid cell, which is an indicator for flow partitioning between drain and recharge to deeper groundwater.

First, we setup distributed, integrated groundwater models at 10 m grid resolution for 28 Danish field-scale drain catchments with observations of drain flow timeseries. A joint calibration of these field-scale models against observed drain flow resulted in an average KGE of above 0.5.

Subsequently, the simulated drain fractions from the field-scale models were used to train a decision tree machine learning algorithm. This metamodel uses various mappable covariates (topography and geology-related) available at high resolution for all of Denmark. The metamodel then is used to predict drain fractions, within its limits of applicability, across relevant areas of Denmark with significant drain flow outside of the field-scale models.

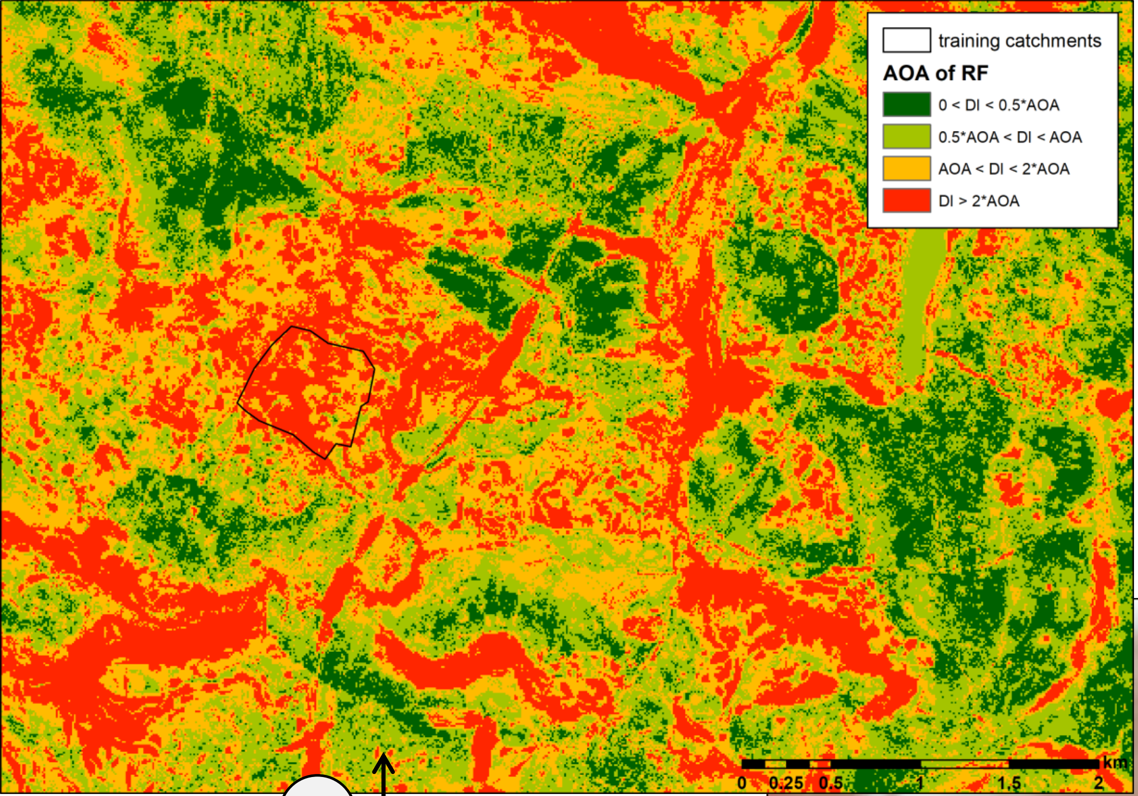
Eventually, the predicted drain fractions are intended to inform national, large-scale physically based hydrological models: An improved representation of drain can, for example, make those models more fit to improve national targeted nitrate regulation.

Additional slides: Example for extension of area of applicability



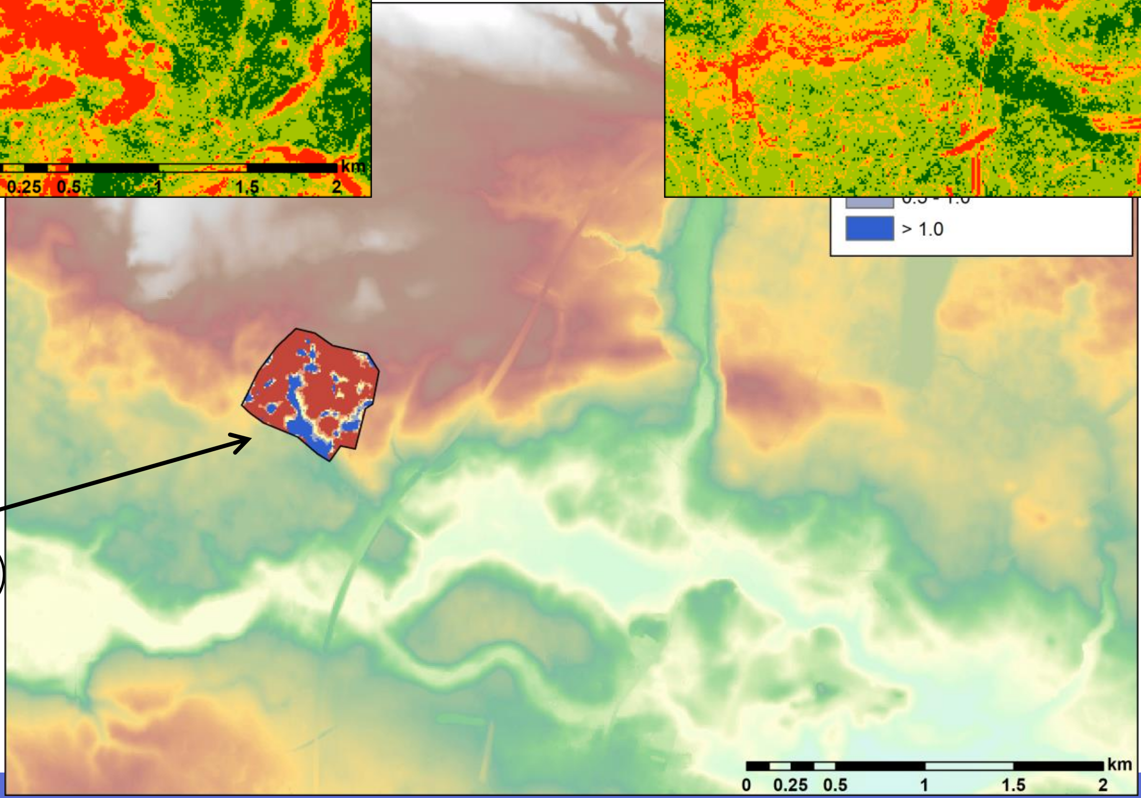
Can the area of applicability be increased by adding training data?
How does this affect the RF predictions?

Extension of area of applicability (AOA) by adding new training data



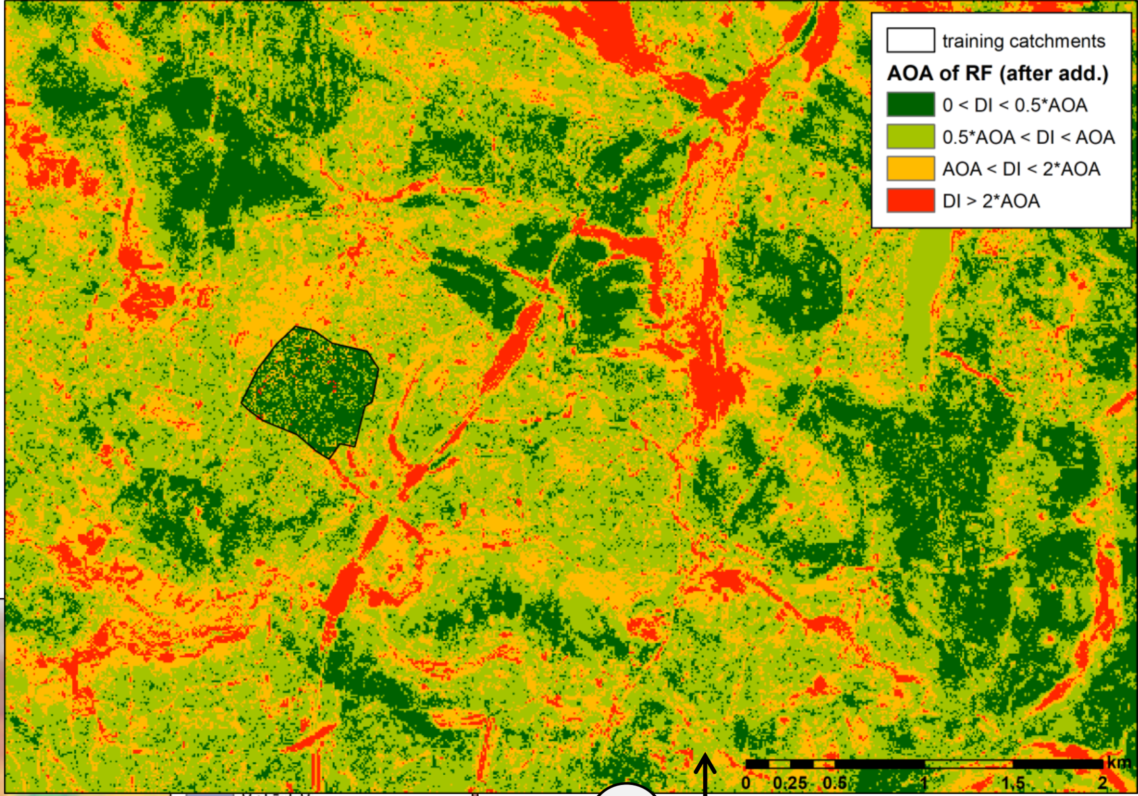
1

Initial AOA; shown for region that is mostly outside AOA
(catchment in this frame **not** included)



2

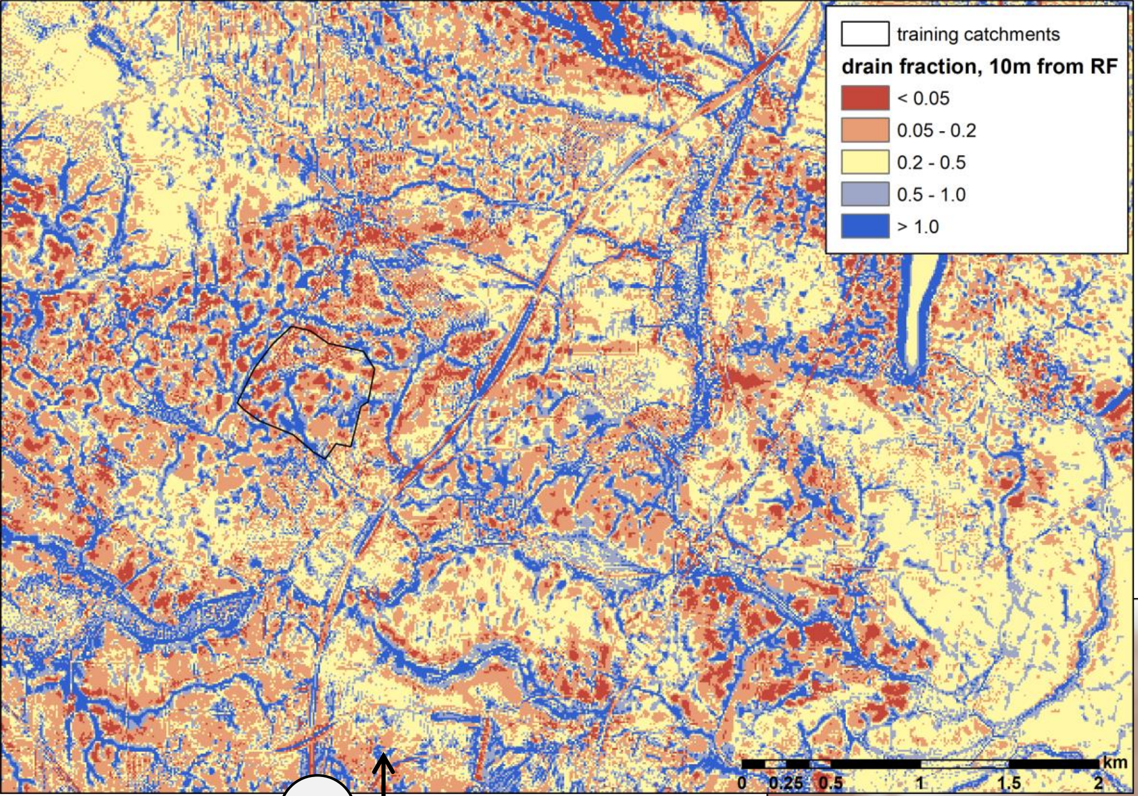
Add more training data
(drain catchment that was left out of initial RF training)



3

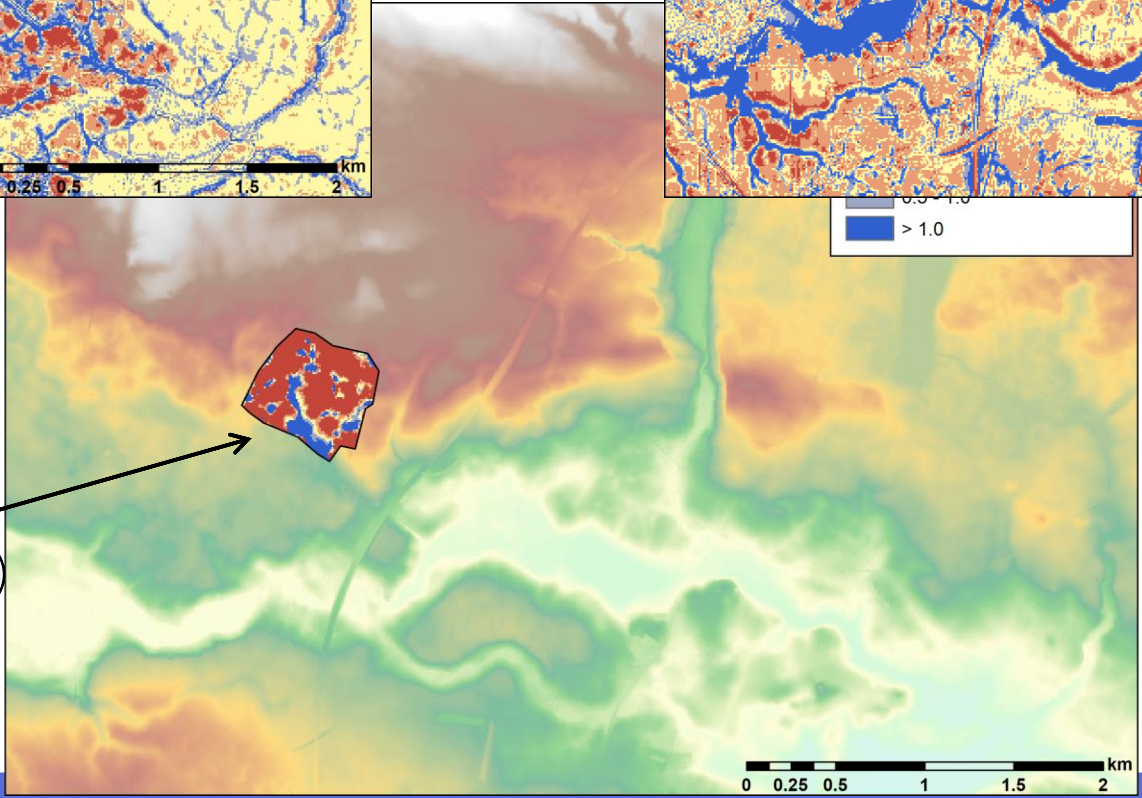
New AOA includes added training catchment
(and more of the region that covers similar covariate ranges)

Change in predictions by adding new training data



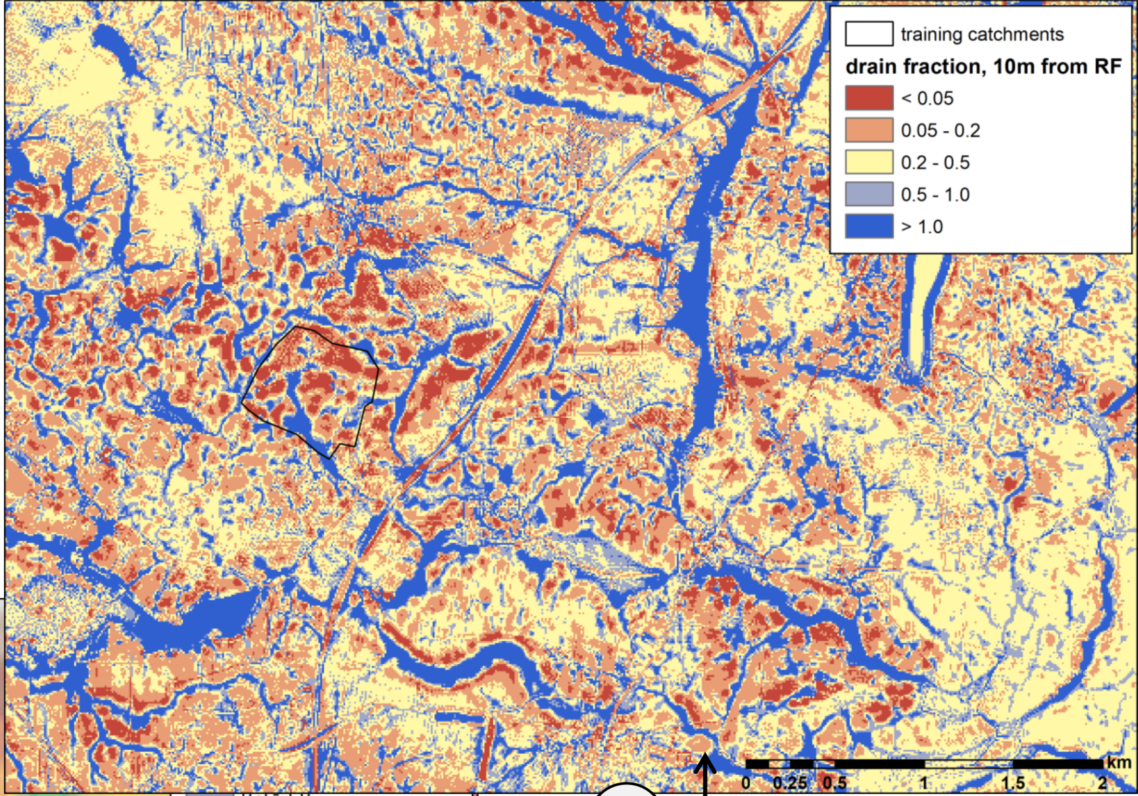
1

Initial Random Forest drain fraction predictions
(catchment in this frame **not** included)



2

Add more training data
(drain catchment that was left out of initial RF training)

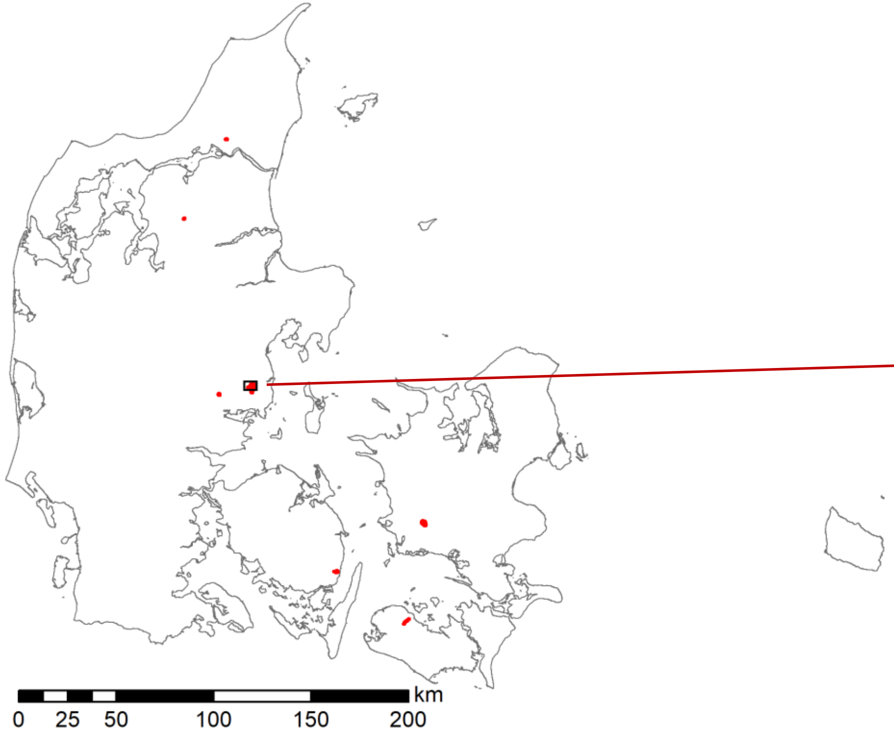


3

New Random Forest drain fraction predictions with added training catchment
closer to truth than before

(without significant trade-off for initial training catchments – see next slides)

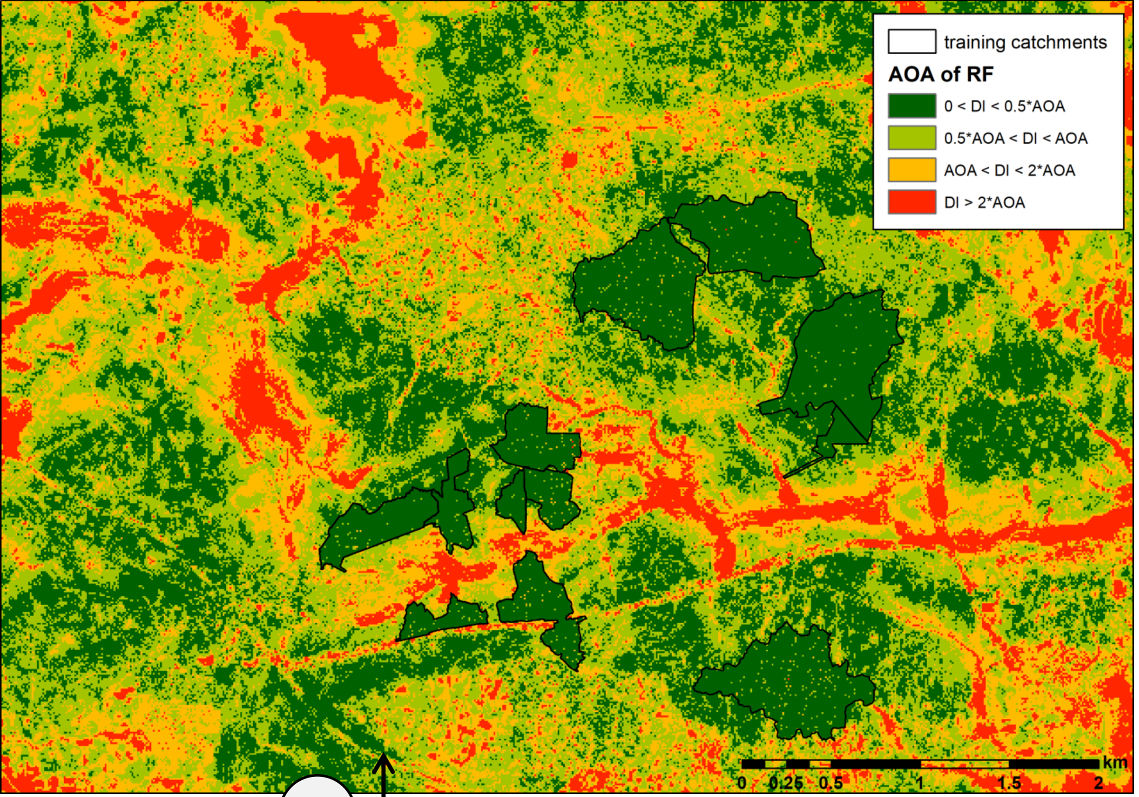
Additional slides: Example for extension of area of applicability



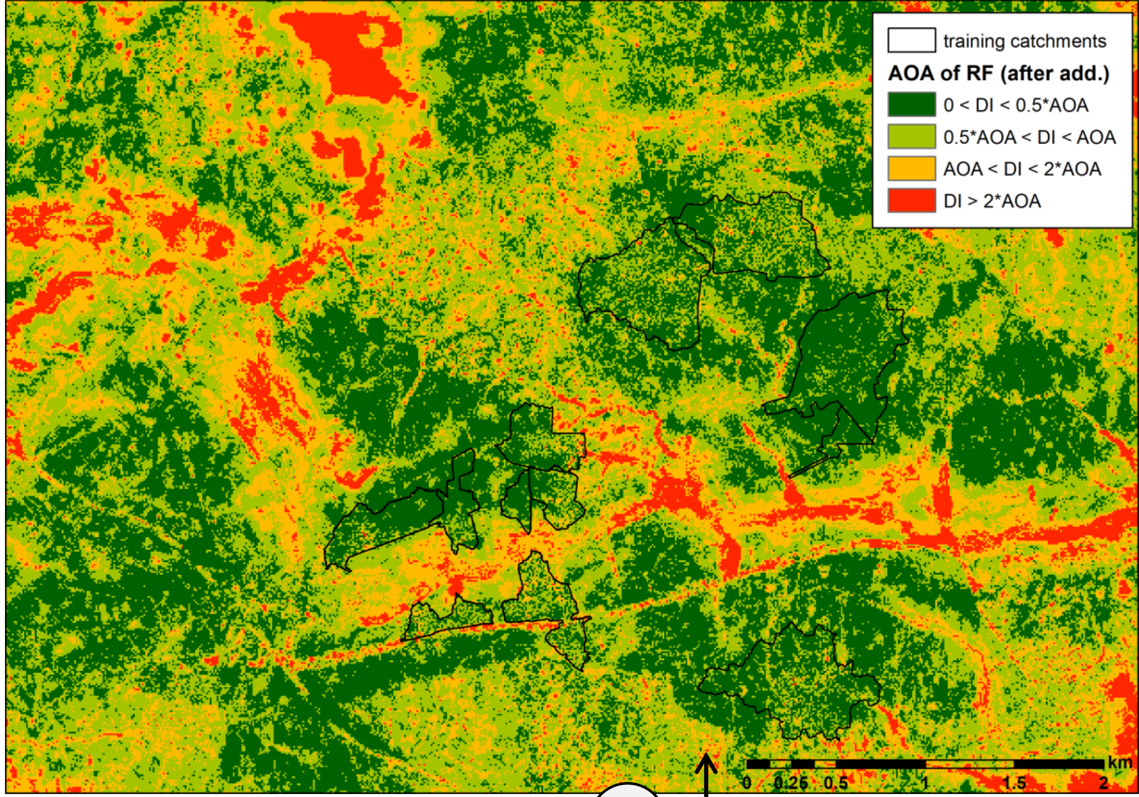
Can the area of applicability be increased by adding training data?
How does this affect the RF predictions?

How does this affect the initial training catchments?

(The next slides show the effect of the inclusion of the additional training catchment shown in previous two slides on the AOA and predictions in and around the initial training catchments)



Extension of area of applicability (AOA) by adding new training data



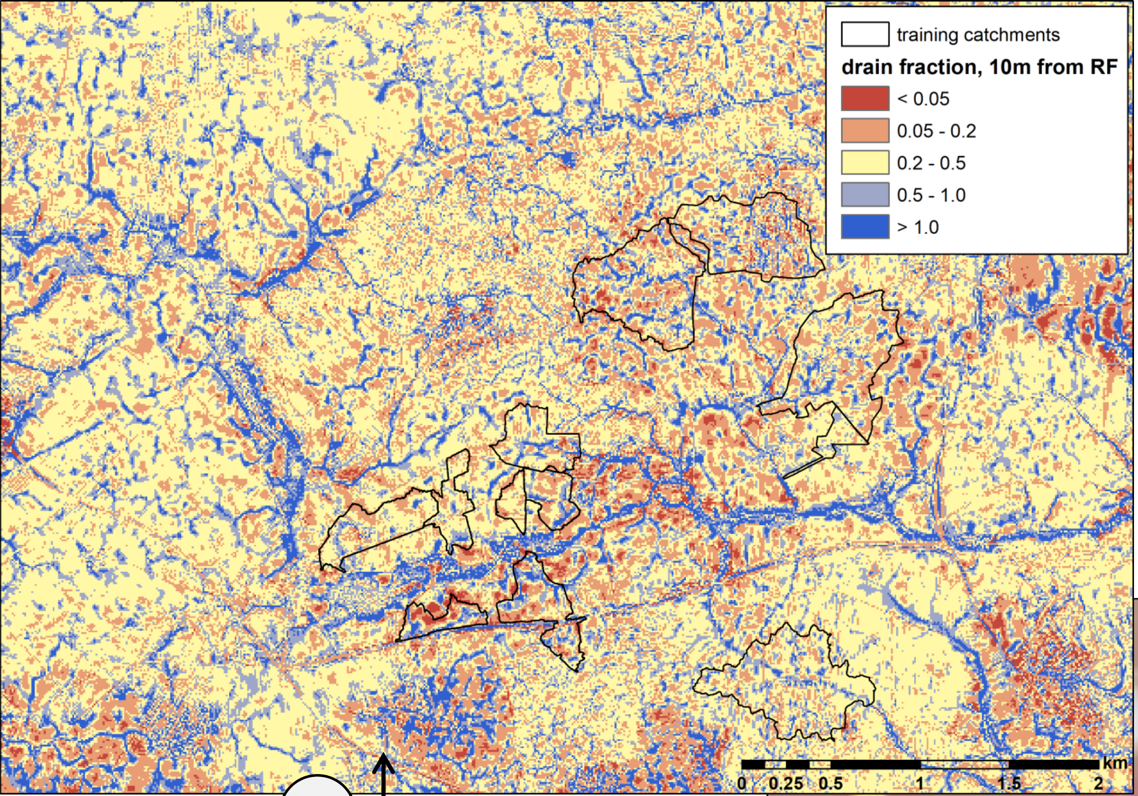
1

Initial AOA; shown for that is mostly outside initial AOA
(due to different ranges covered for important covariate – here: clay fraction)

3

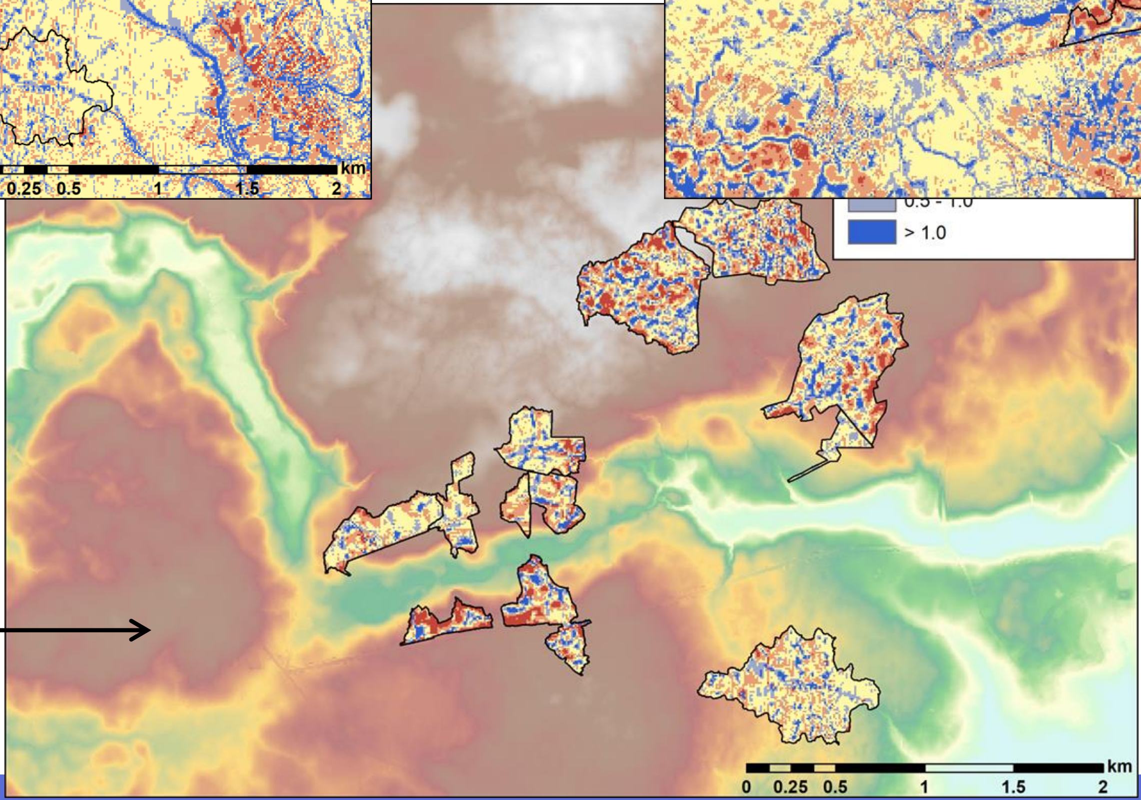
New AOA
Also increases AOA around the initial training data

Change in predictions by adding new training data – for initial training catchments

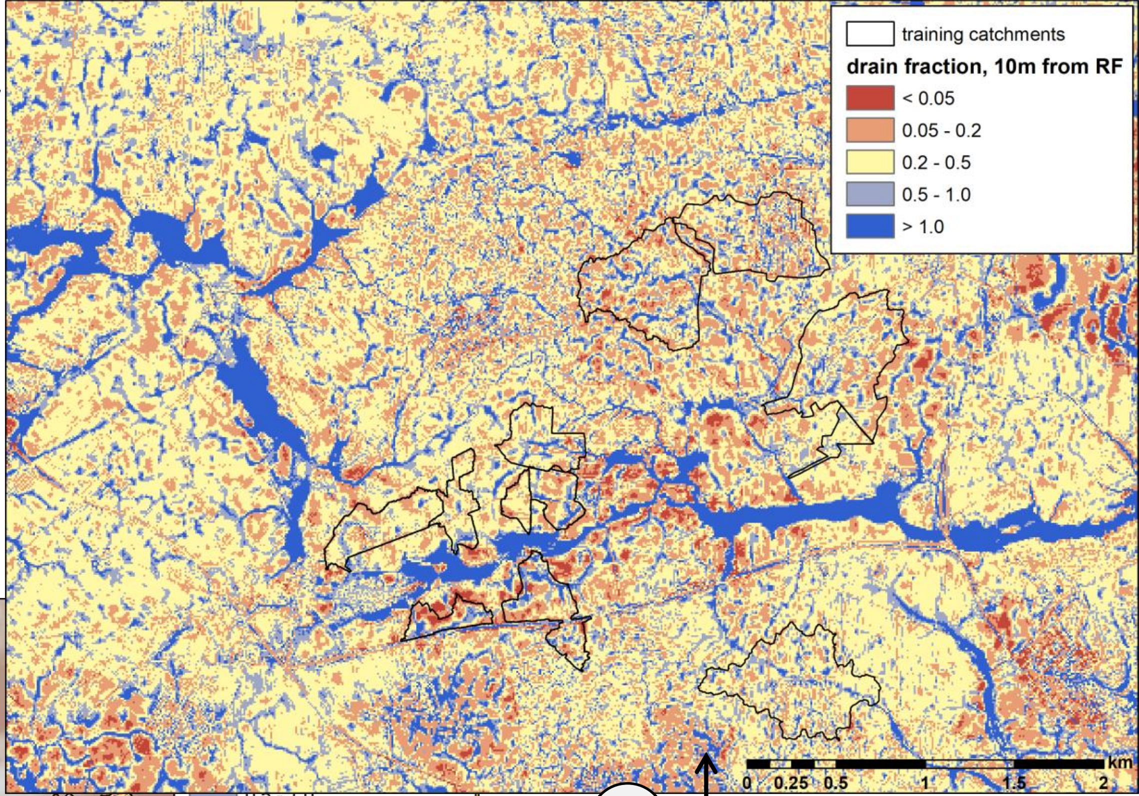


1

Initial Random Forest drain fraction predictions



Initial training data



3

New Random Forest drain fraction predictions with added training catchment

(without significant trade-off for initial training catchments; however, changes outside of these)