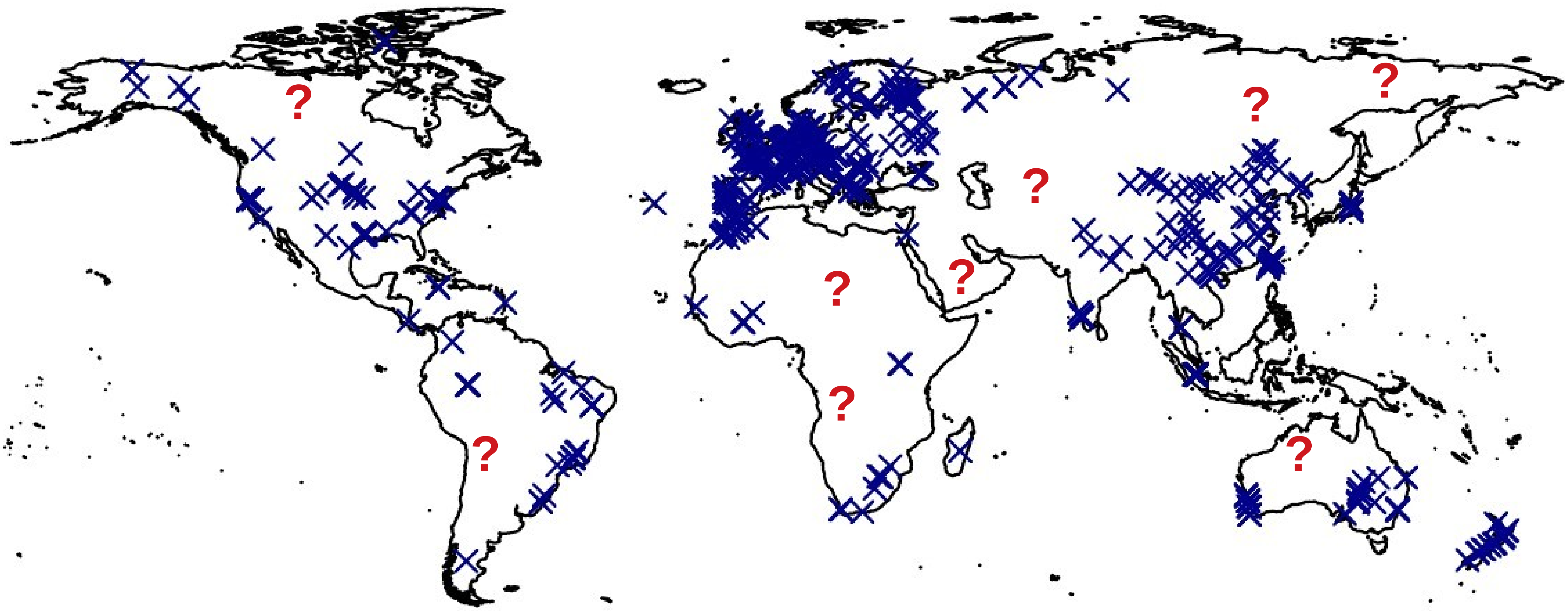
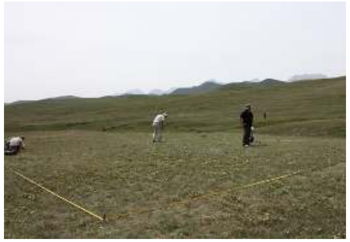
Machine learning-based maps of the environment: challenges of extrapolation and overfitting

Part 1: Introduction to the topic
Part 2: Predictive mapping and dealing with the challenges in practice

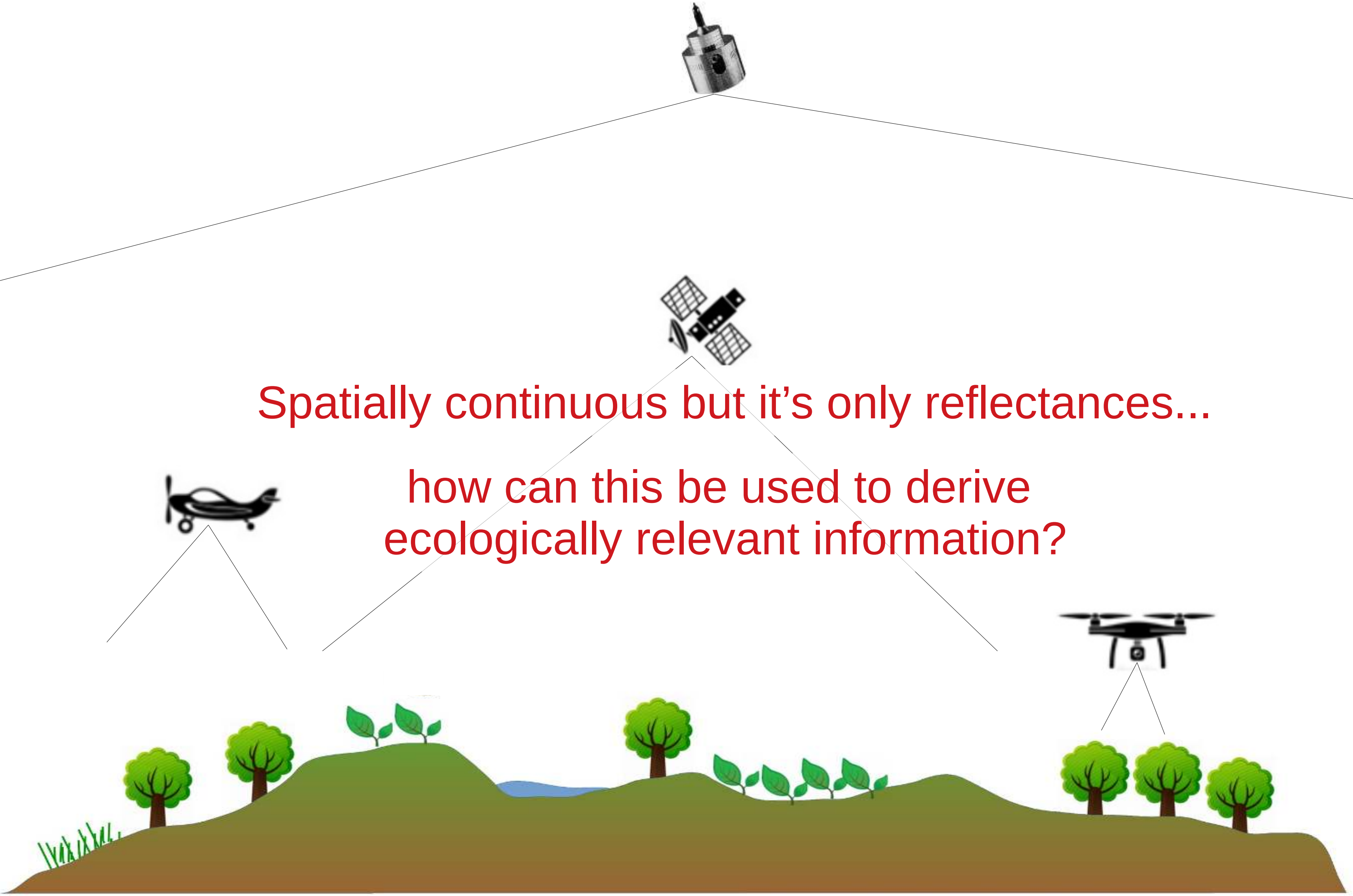
Hanna Meyer
Remote Sensing & Spatial Modelling,
Institute of Landscape Ecology, WWU Münster

Problem: We only have limited (point) information about the environment



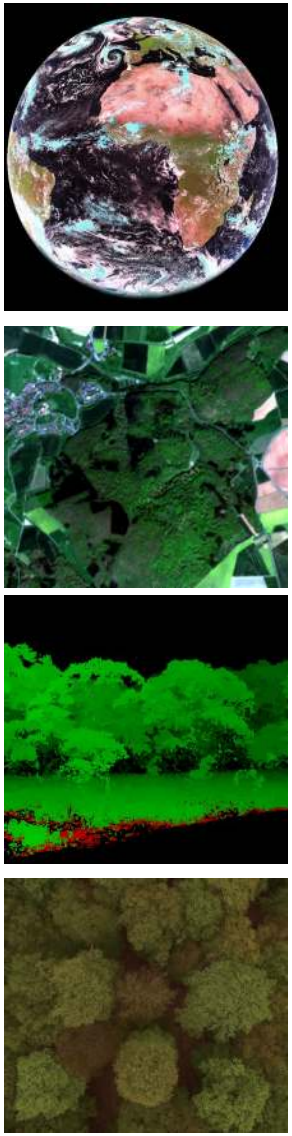


Remote Sensing to derive continuous information



Spatially continuous but it’s only reflectances...

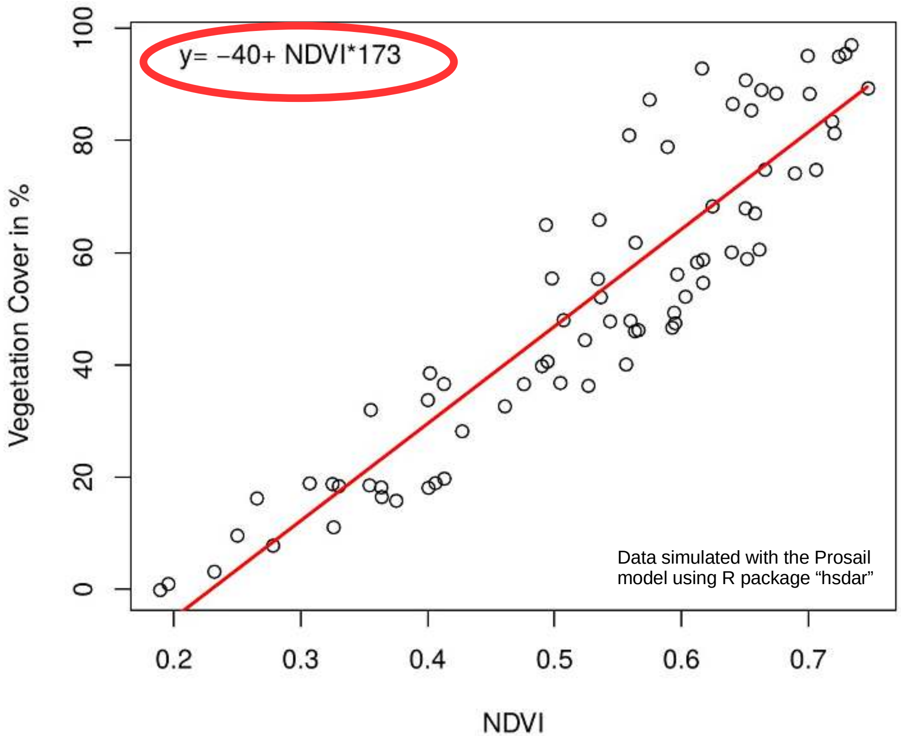
how can this be used to derive ecologically relevant information?



How can we translate the remote sensing information to the ecological variable?

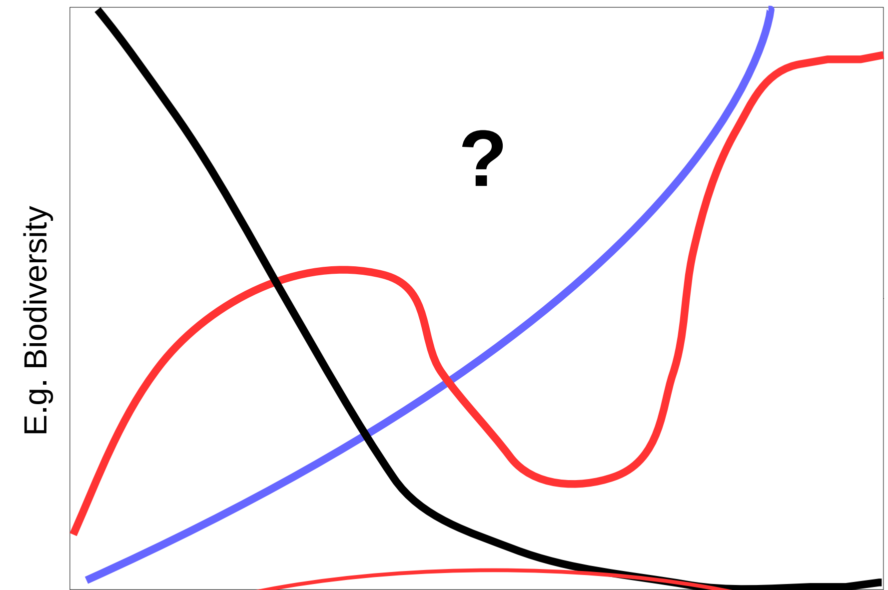
e.g. vegetation cover from satellite (VIS/NIR)

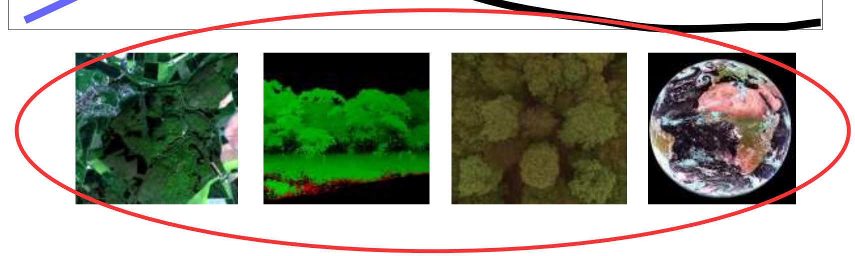
Field data



Remote sensing data

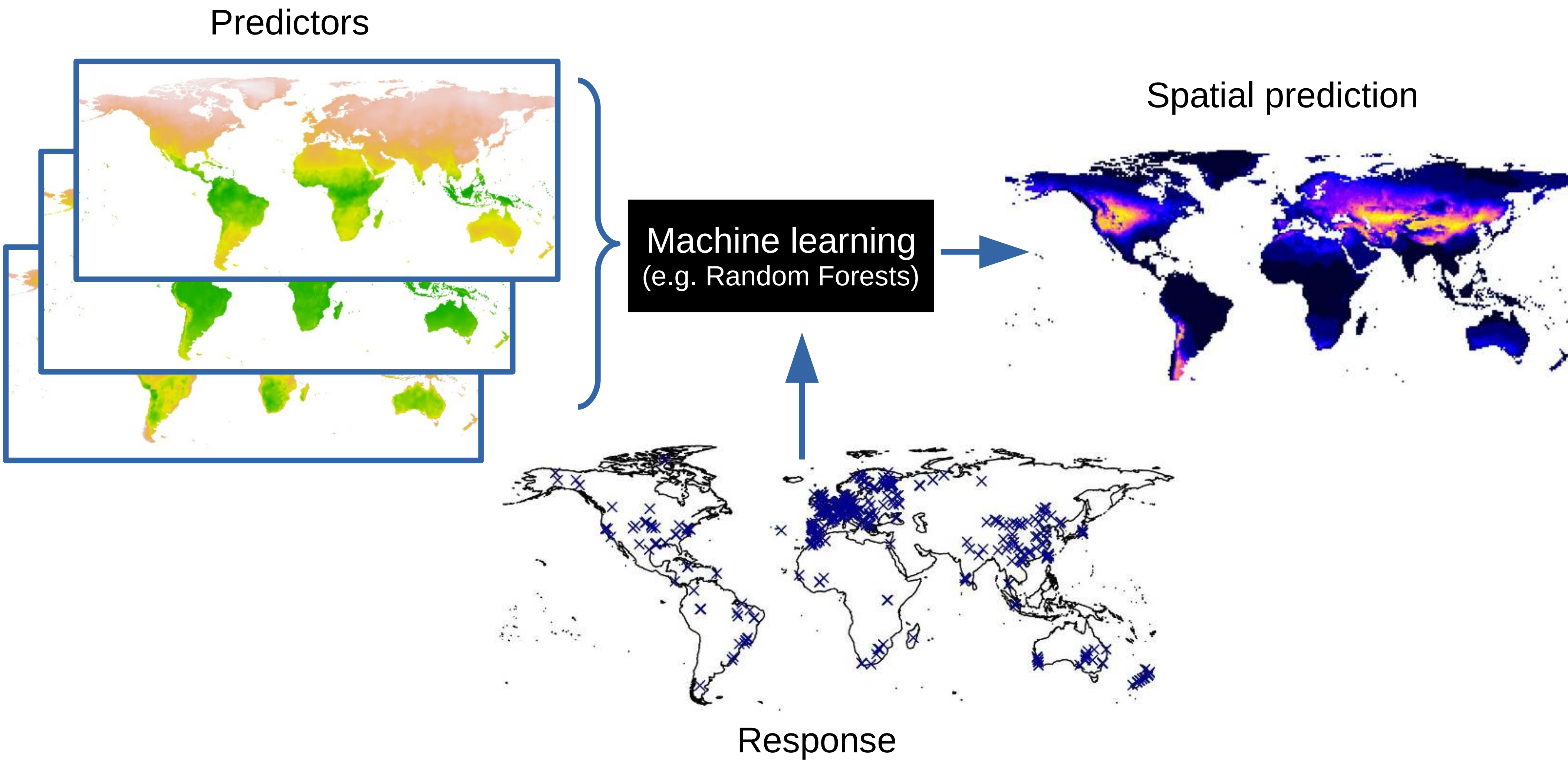
Typical ecological variables from satellite?



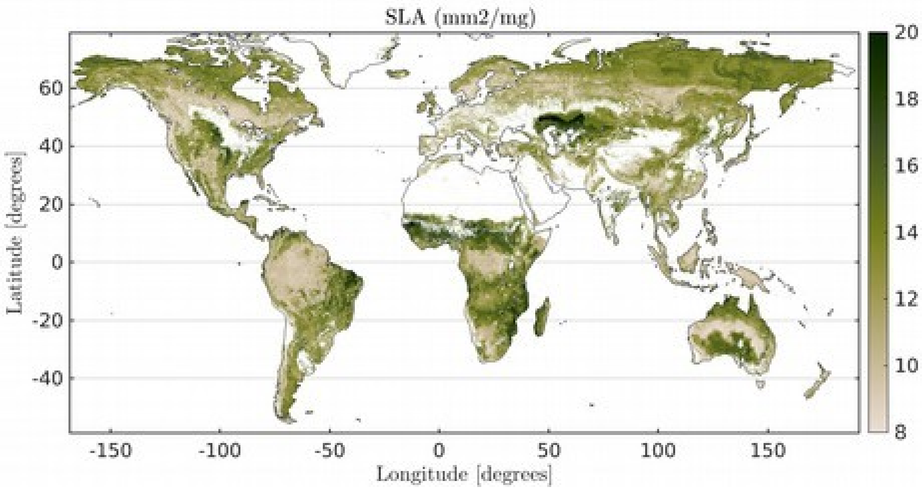


Models that can deal with complex nonlinear relationships are required!

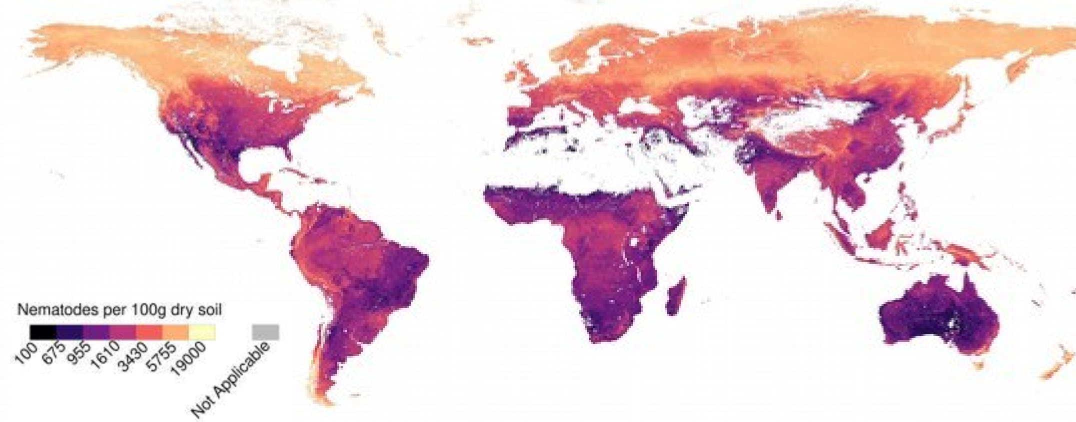
How do we get “maps” of ecosystem variables ?



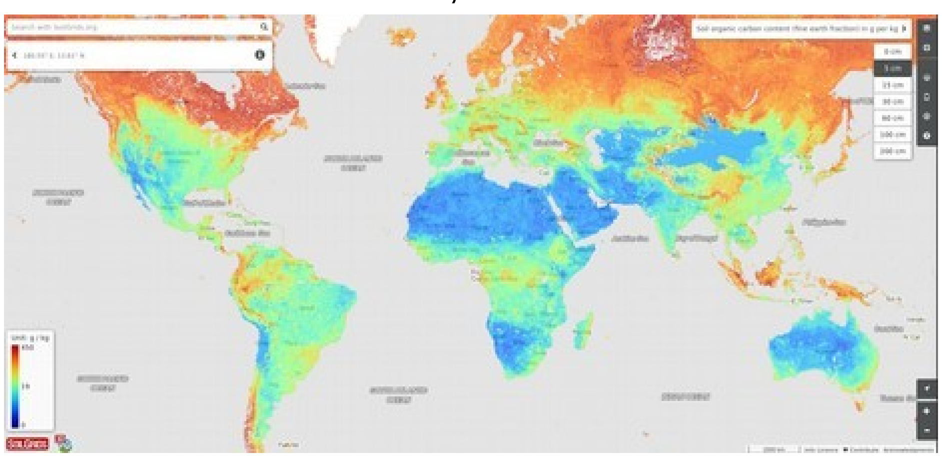
Global maps of ecosystem variables based on machine learning (a few examples)



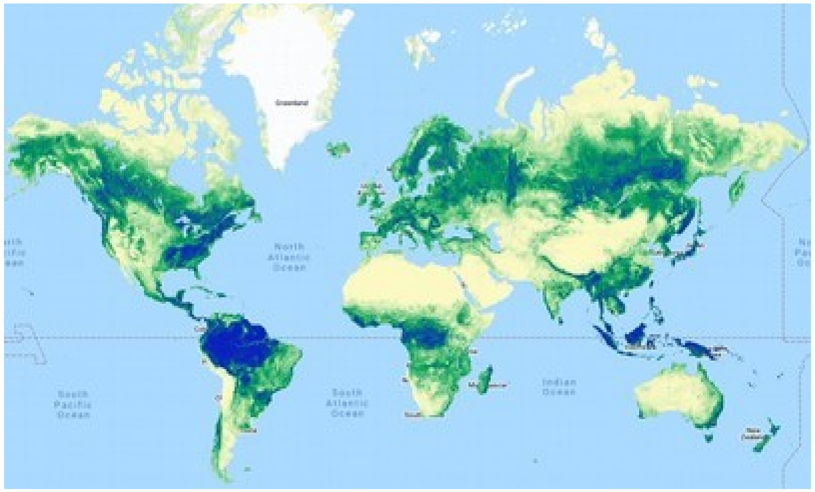
Moreno-Martínez et al., 2018



Based on van den Hoogen et al., 2019



Hengl et al., 2017



Bastin et al. 2019

Machine learning as a magic tool to map everything ?

...but there are increasingly doubts about the quality of these results

Wissenschaft

Wenn die KI daneben liegt

Welche Fehler drohen, wenn Forscher Wissenslücken per Computer schließen wollen, zeigen zwei aktuelle Klimastudien.

Von **Tin Fischer**

6. November 2019, 16:44 Uhr / Editiert am 9. November 2019, 17:42 Uhr / DIE ZEIT Nr. 46/2019, 7. November 2019 / 9 Kommentare

AUS DER ZEIT NR. 46/2019

DEEP TROUBLE FOR DEEP LEARNING

BY DOUGLAS HEAVEN

Nature 574, 163-166 (2019)

Home / News & Opinion

Researchers Find Flaws in High-Profile Study on Trees and Climate

Four independent groups say the work overestimates the c global forest restoration, but the authors insist their origina

Oct 17, 2019

KATARINA ZIMMER

Comment | Published: 23 August 2021

Conservation needs to break free from global priority mapping

Carina Wyborn & Megan C. Evans

Nature Ecology & Evolution (2021) | Cite this article

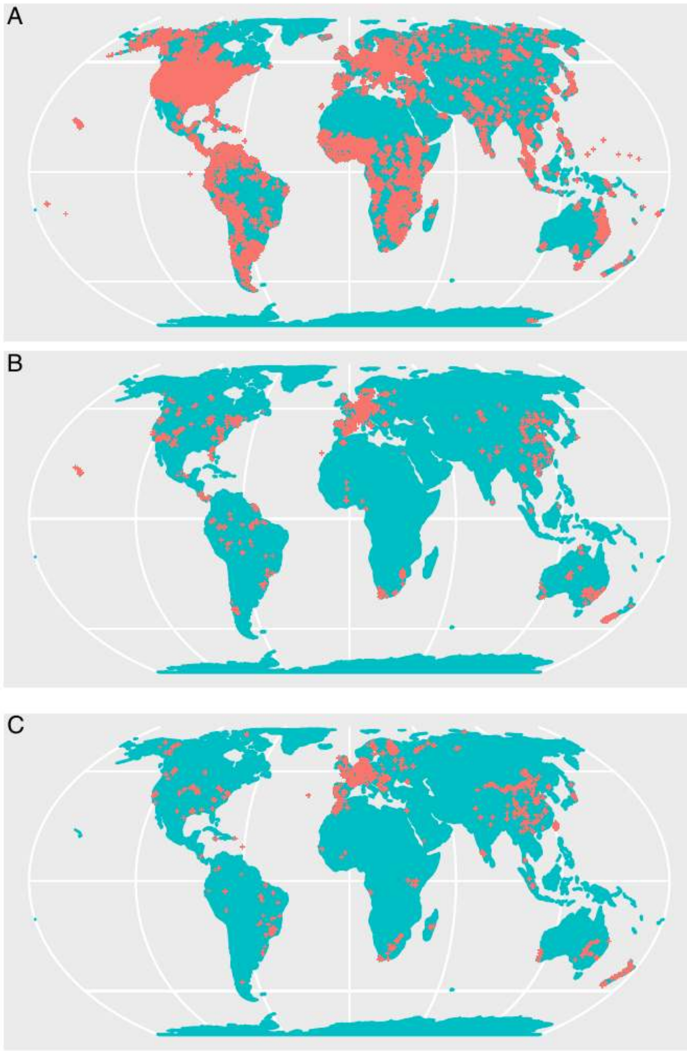
Have we been too ambitious? When and why might the models fail?

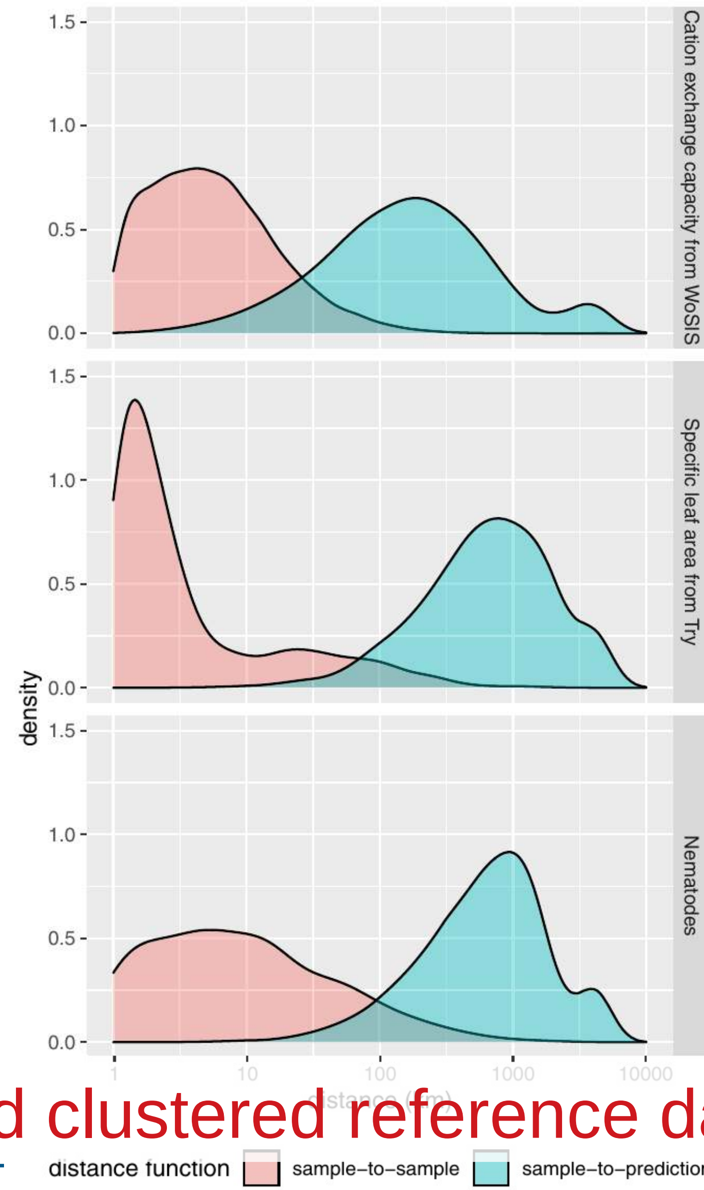
What do these applications have in common?

Soil maps

Plant traits

Nematodes

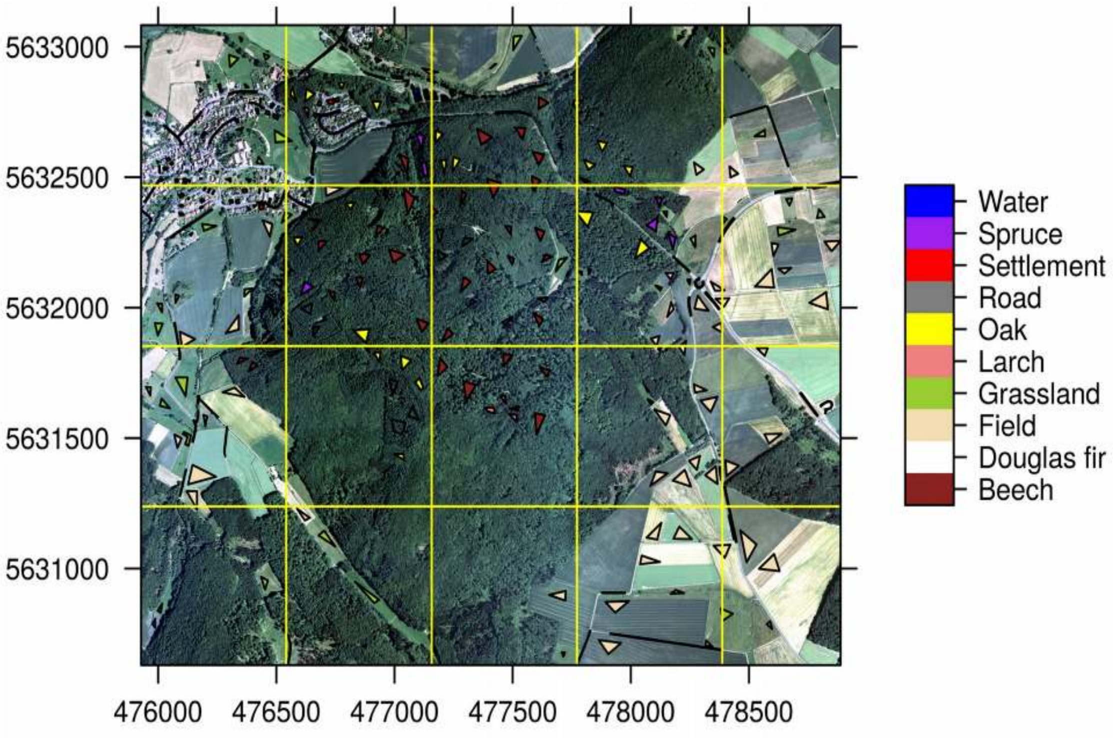


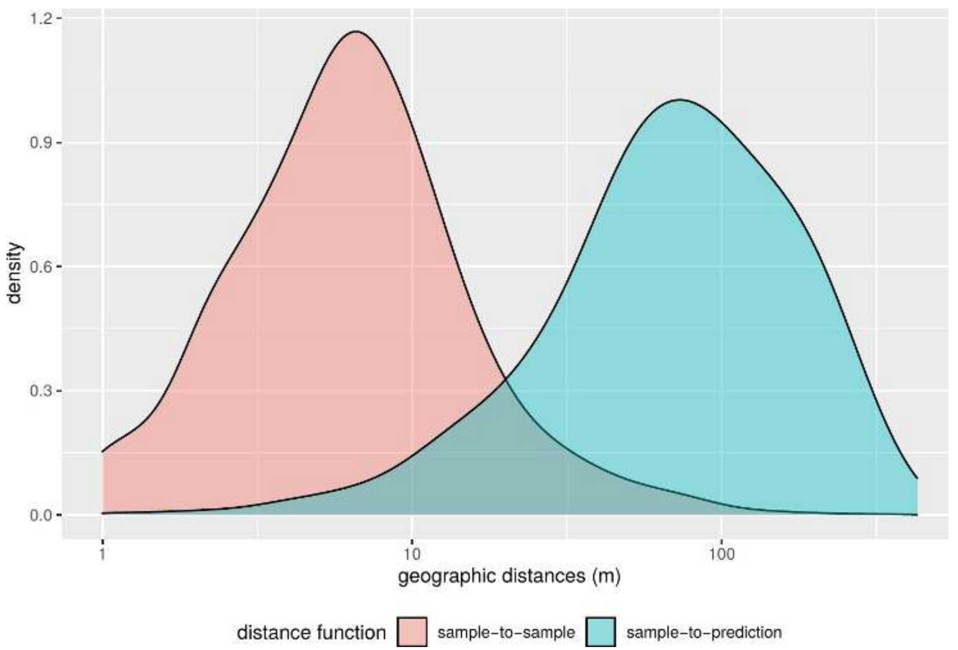


Meyer & Pebesma (2022)

Mapping requires prediction far beyond clustered reference data!

This is not just an issue for global applications

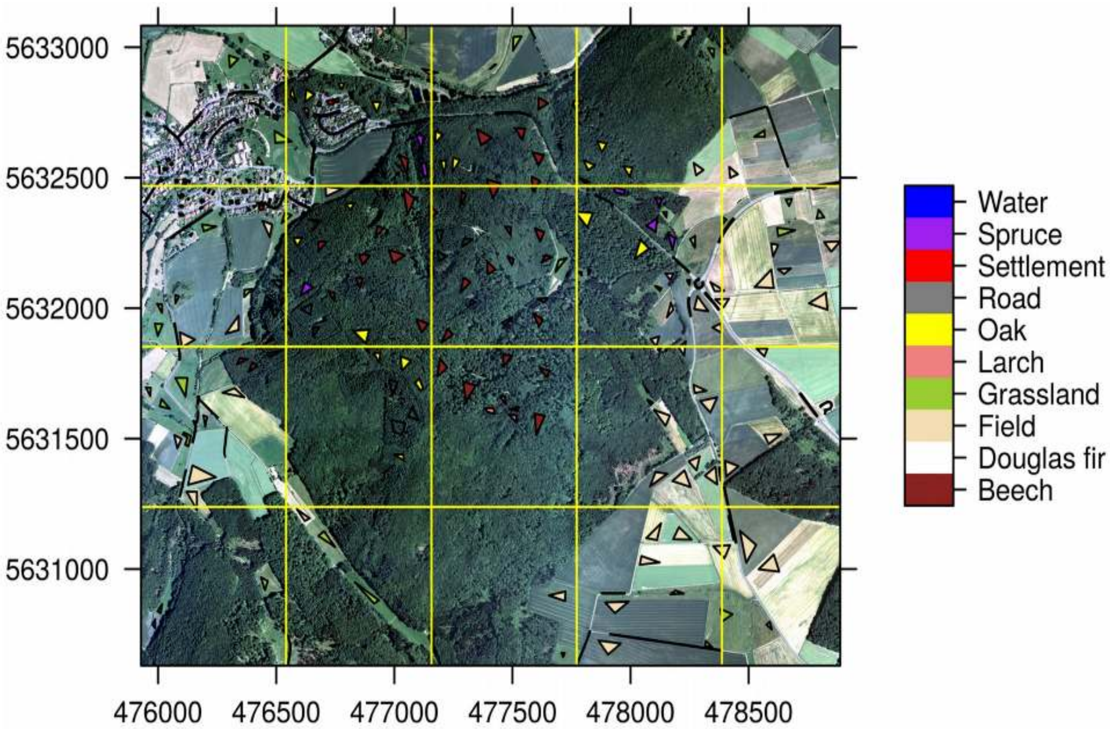




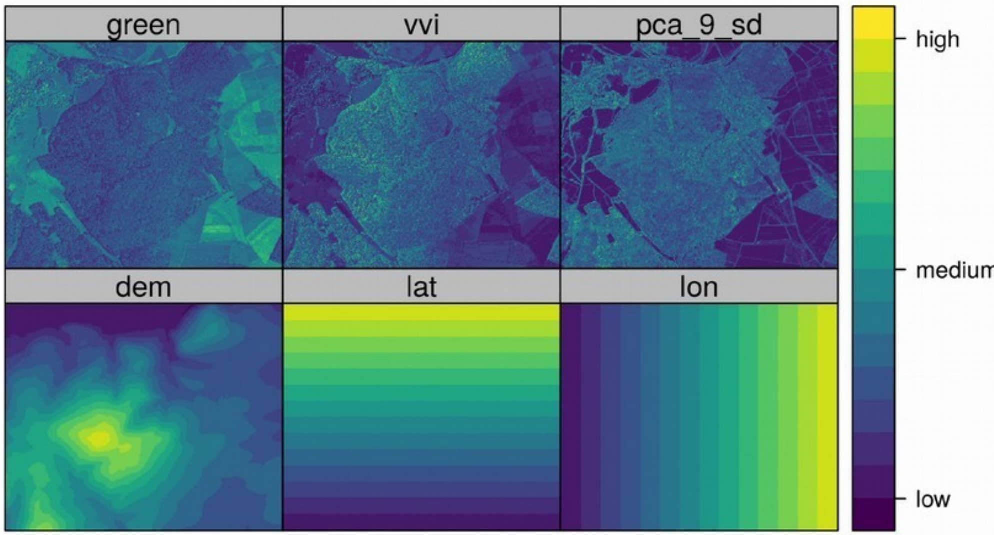
Does the clustered pattern cause problems?
Let’s explore with this case study...

Is this a problem? Example of a “classic” land cover classification

Aerial image overlayed by training sites



Example of predictors



Random Forests

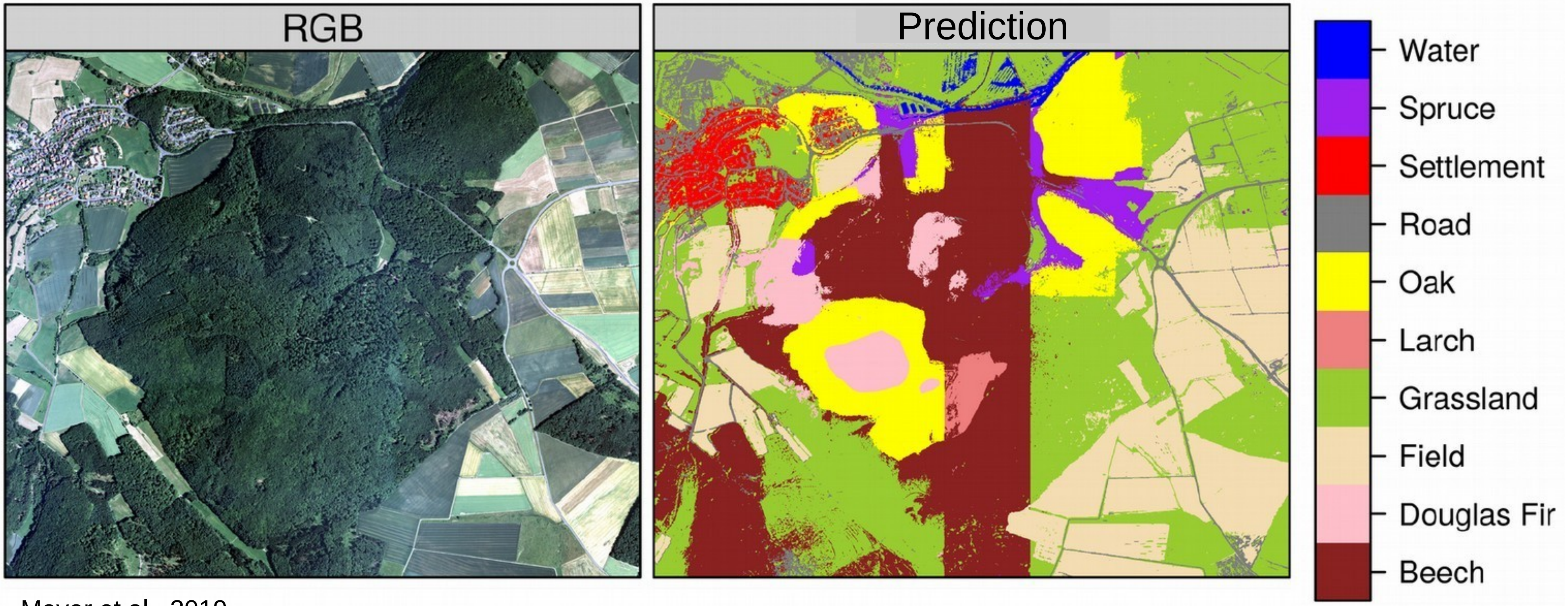
How well can we model land cover with this approach?

Performance assessment by the default validation strategy

Variables	Validation	Accuracy	Kappa
all	random	>0.99	>0.99
all	spatial	**0.68**	**0.61**

Perfect prediction?

...but it doesn’t look like a perfect prediction

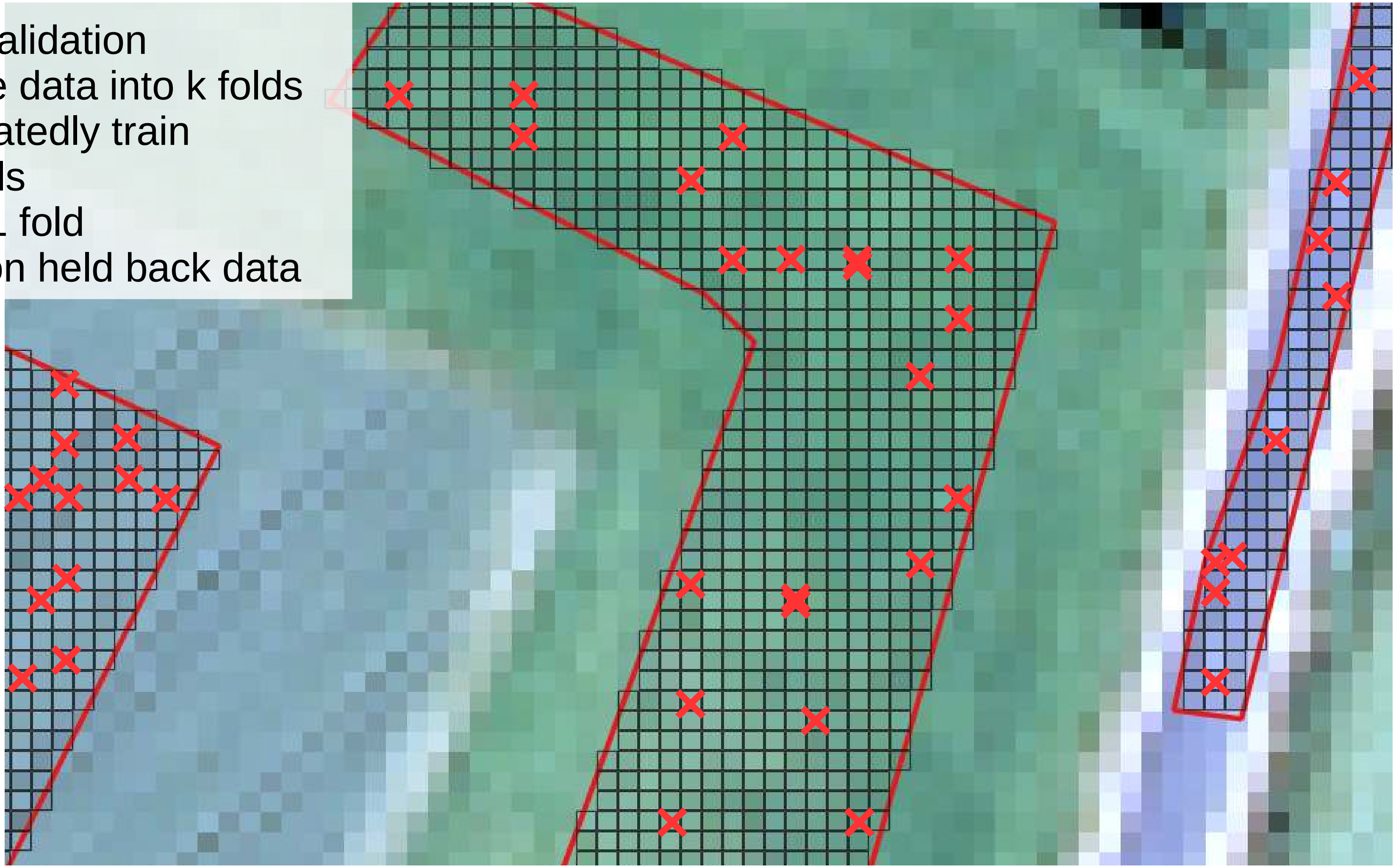


Meyer et al., 2019

But statistically it’s a perfect model.
How is this possible?

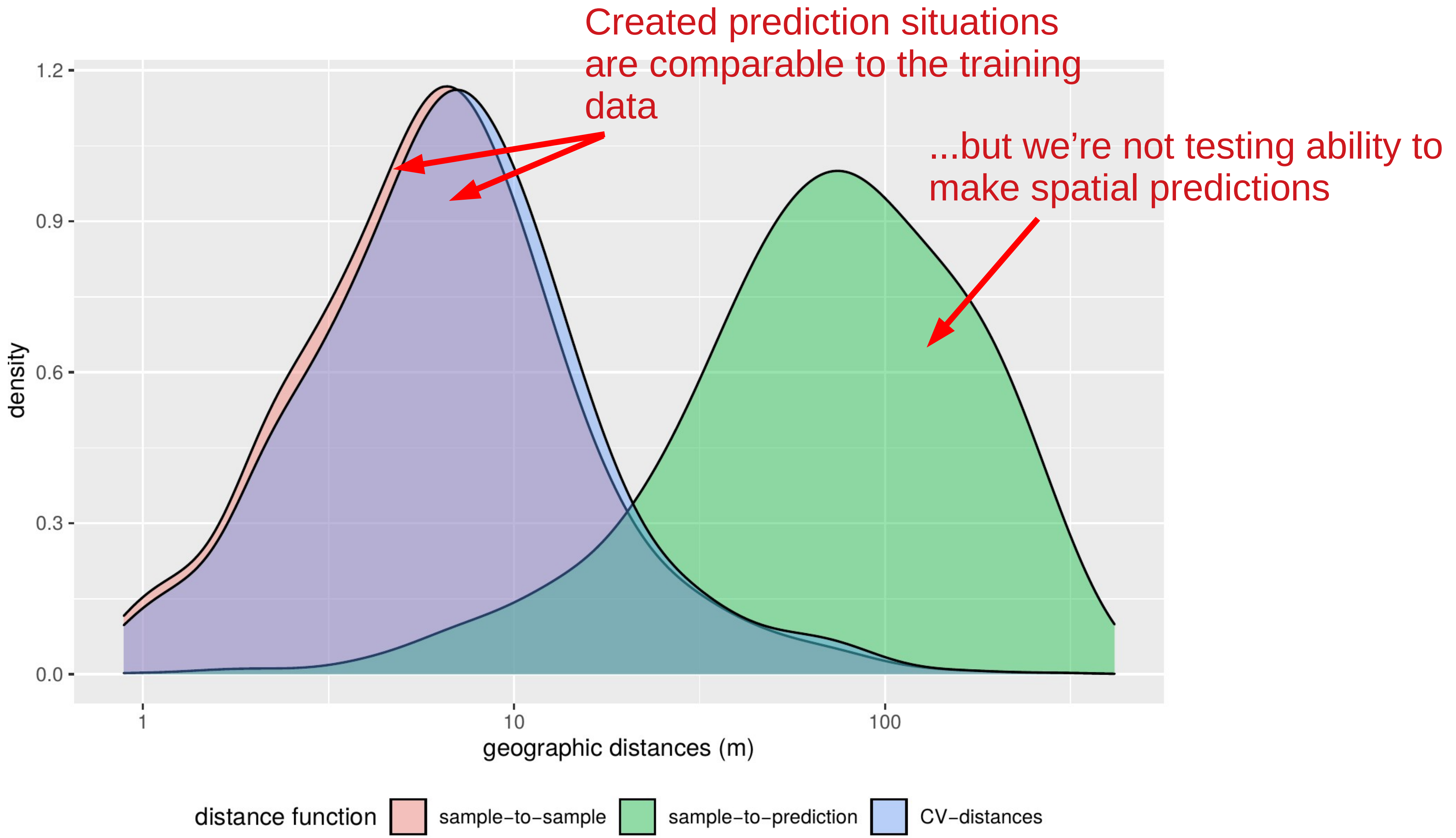
Assessment of performance by default random cross-validation

Cross-validation
- Divide data into k folds
- Repeatedly train models
  on k-1 fold
- Test on held back data



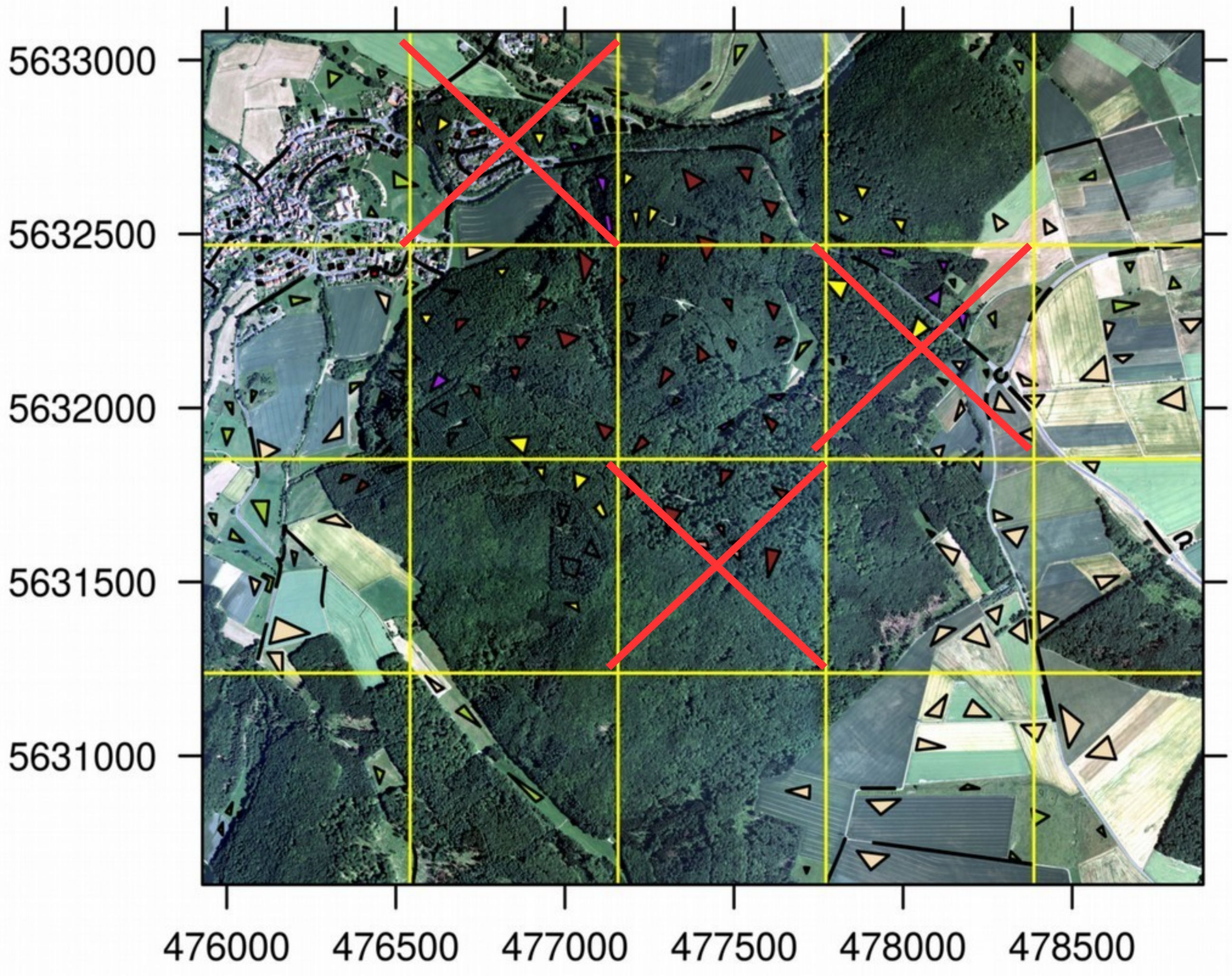
Answers question how well model performs on very similar locations

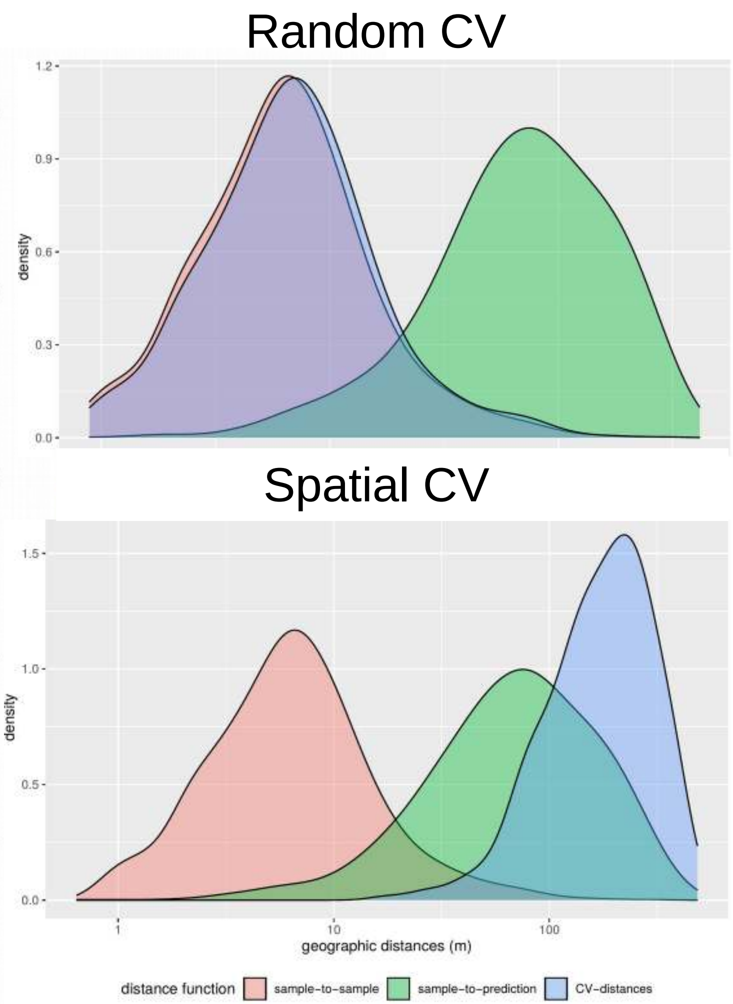
Assessment of performance by default random cross-validation



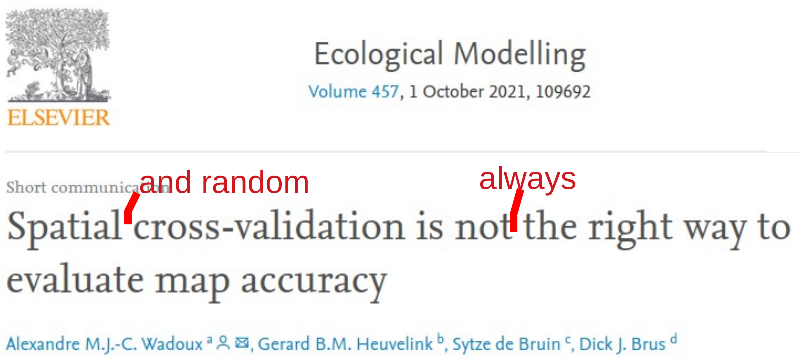
Assessment of spatial performance

...But the aim is to fill the gaps between sampling locations!
Spatial cross-validation is required





Convinced? So why is the value of spatial CV then still discussed?


Ecological Modelling
Volume 457, 1 October 2021, 109692

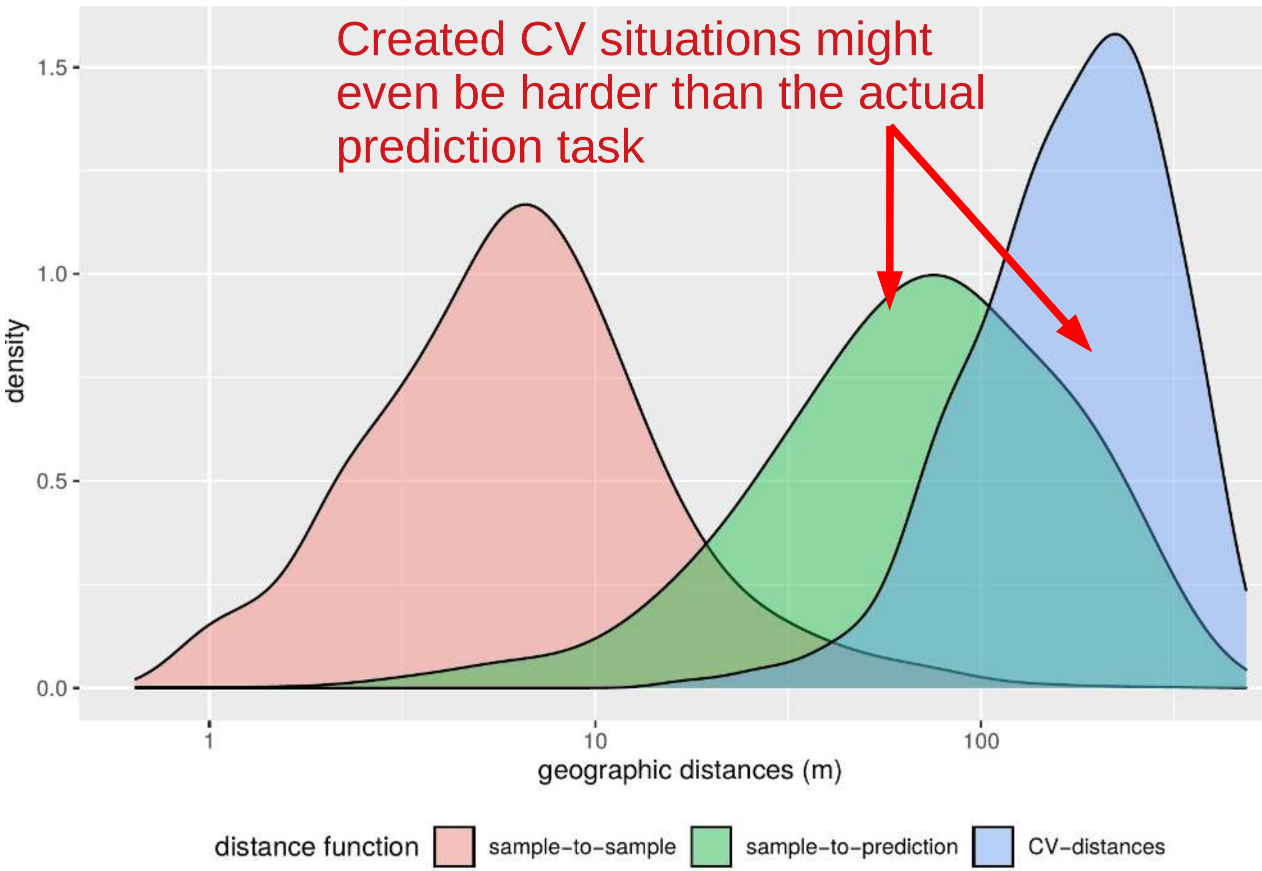
Short communication

and random   always

Spatial cross-validation is not the right way to evaluate map accuracy

Alexandre M.J.-C. Wadoux [a], Gerard B.M. Heuvelink [b], Sytze de Bruin [c], Dick J. Brus [d]

→ Our suggestion: prediction situations created during CV need to resemble those encountered while predicting the map from the reference data



We can do that the trial-and-error-way or….

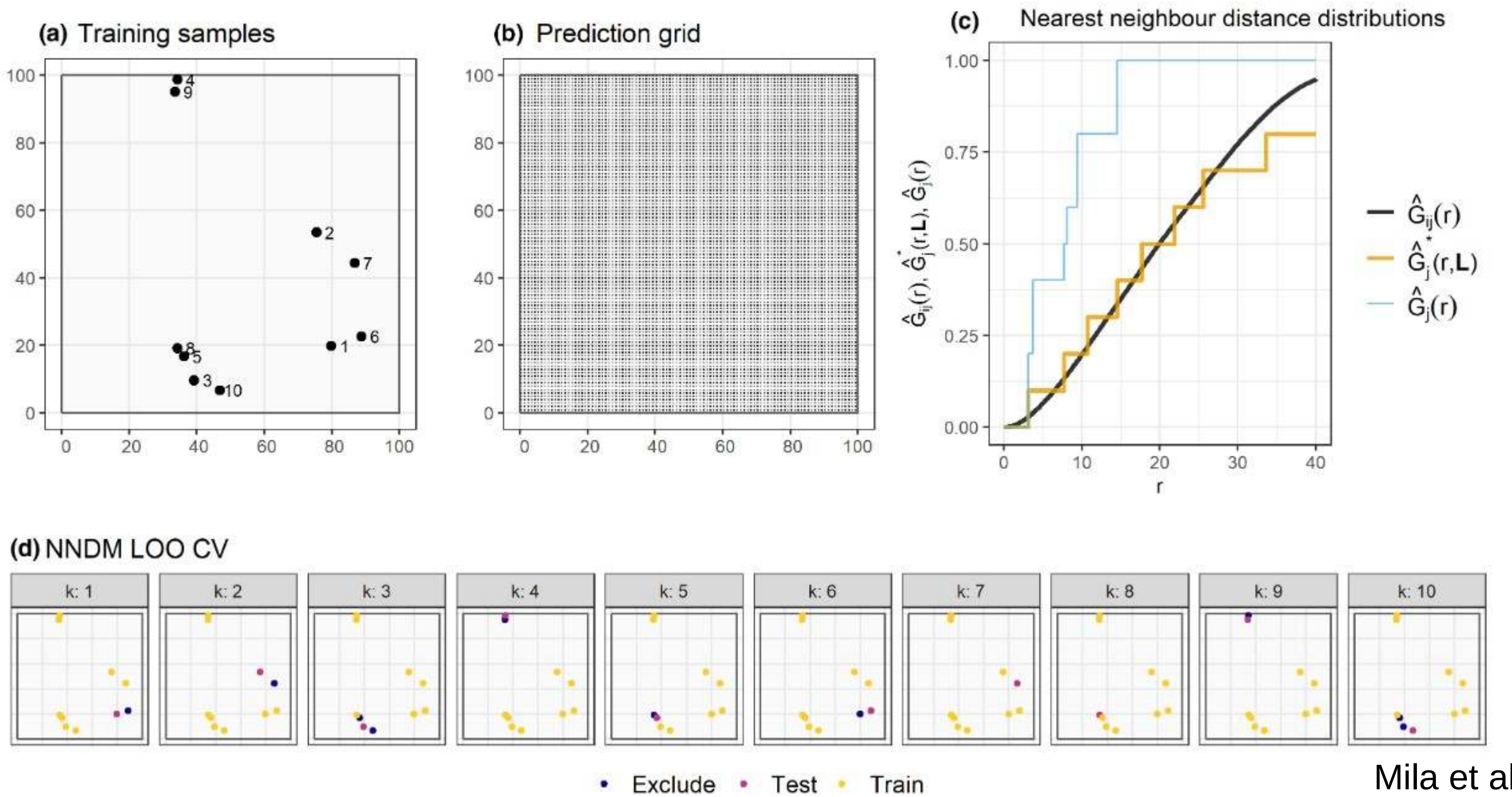
Suggestion of a nearest neighbor distance matching LOO CV

Received: 20 September 2021 | Accepted: 8 March 2022
DOI: 10.1111/2041-210X.13851
RESEARCH ARTICLE
Methods in Ecology and Evolution

Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation

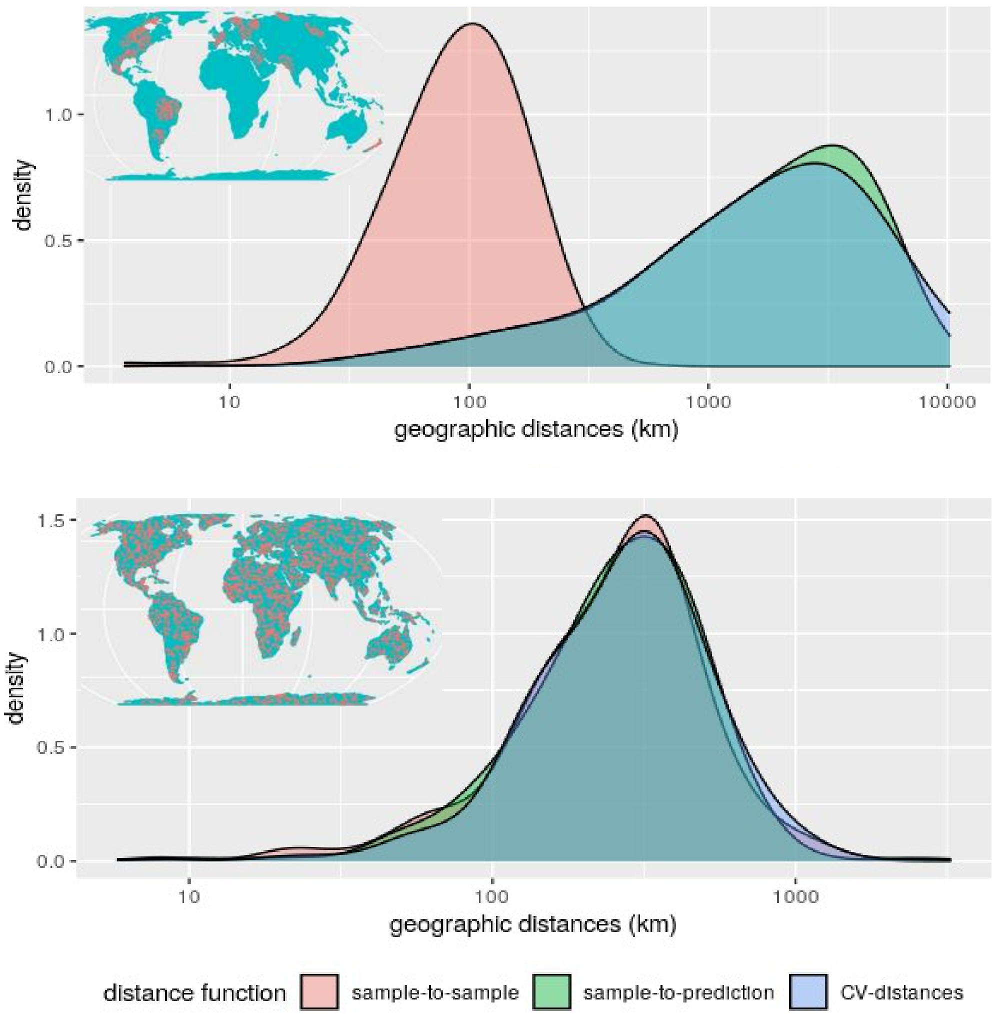
Carles Milà[1] | Jorge Mateu[2] | Edzer Pebesma[3] | Hanna Meyer[4]

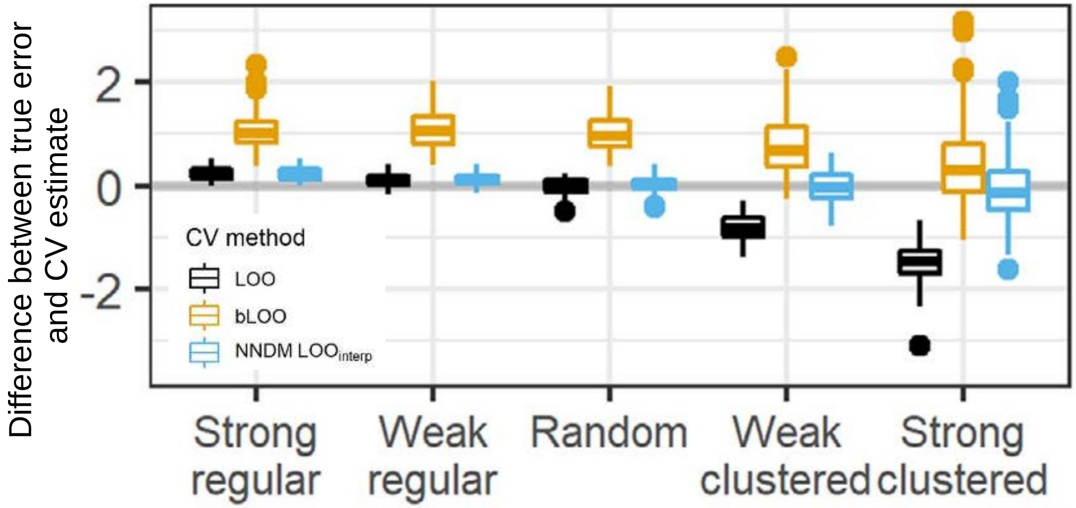
Aim: Prediction situations created during CV resemble those encountered while predicting the map



Mila et al., 2022

Suggestion of a nearest neighbor distance matching LOO CV





Mila et al., 2022

Reproduce figures: hannameyer.github.io/CAST/articles/cast04-plotgeodist.html

Coming back to our case study...

Variables	Validation	Accuracy	Kappa
all	random	>0.99	>0.99
all	spatial	0.68	0.61

Perfect prediction?
We need to assess this by a suitable CV strategy!

Assessment of spatial performance

Variables	Validation	Accuracy	Kappa
all	random	>0.99	>0.99
all	spatial	**0.68**	**0.61**

- Standard validation procedures lead to an overoptimistic view on prediction performance!
- Prediction situations created during CV need to resemble those encountered while predicting the map from the reference data

...but the relevance of spatial validation is still highly underestimated

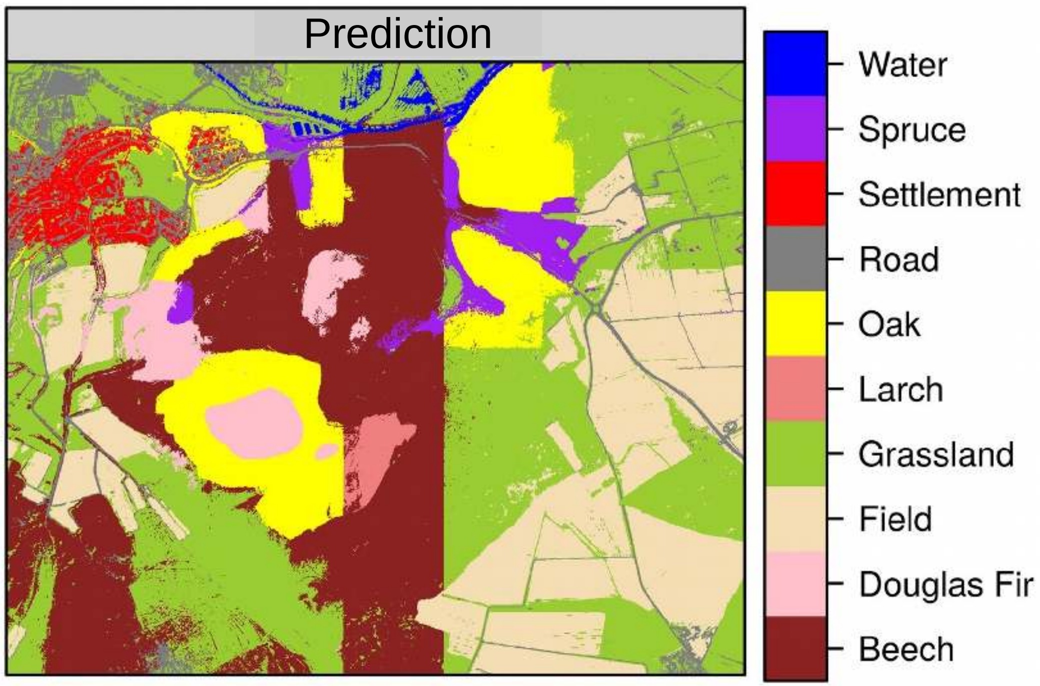
"*I am actually surprised to see the poor performance of your NN approach[...]. Typically with sufficient training data a NN approach can often reproduce the predicted variable very well even if the underlying reasons are unknown*"
(an editor from a high impact journal in the remote sensing community)

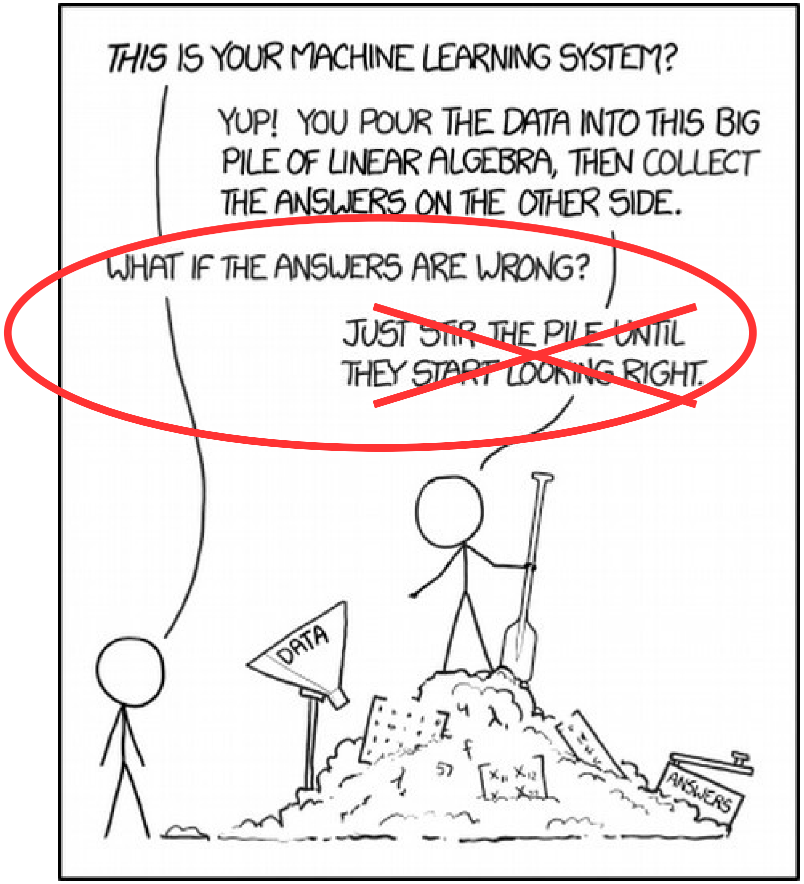
Data reproduction is not the same as data prediction!

Random
cross-validation!

Spatial
cross-validation!

Spatial performance of models needs to be improved!

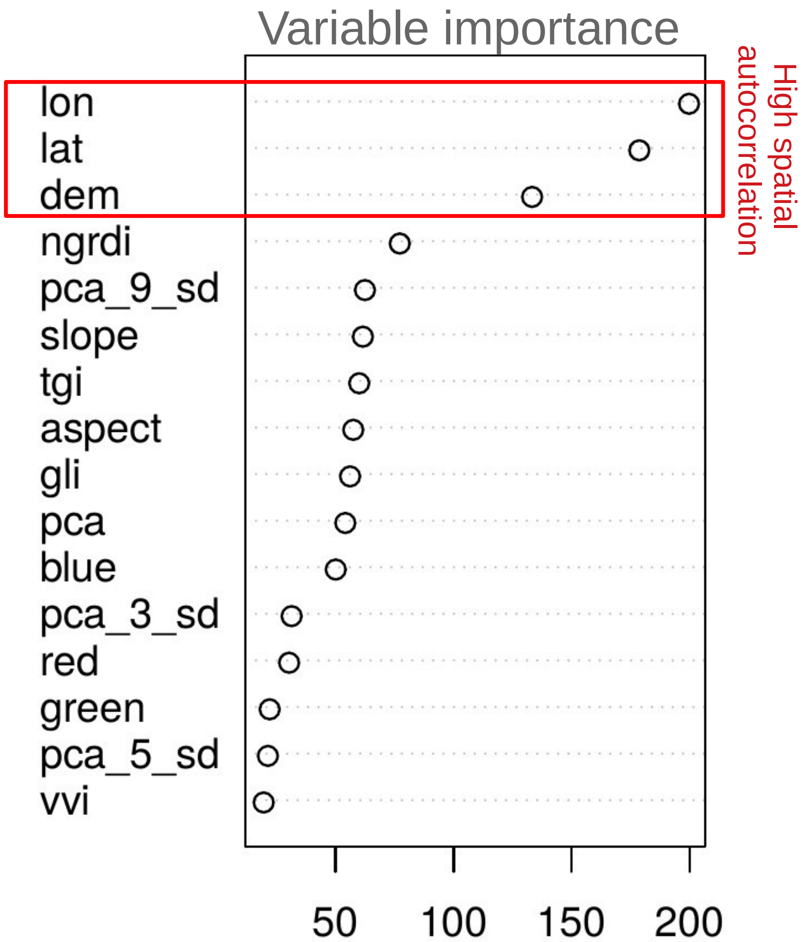




Where do these prediction patterns come from?

https://xkcd.com/1838/

An example of the “clever Hans effect” ?



Is the model behaving like the “clever Hans” ?

https://commons.wikimedia.org/wiki/File:Osten_und_Hans.jpg#/media/Datei:Osten_und_Hans.jpg

Suspicion: spatial dependencies lead to confounding variables.

→ True relationships not recognized, causing the model to fail in making predictions?

“Unmasking Clever Hans predictors and assessing what machines really learn”

(Lapuschkin et al., 2019, Nature communications)

Horse-picture from Pascal VOC data set

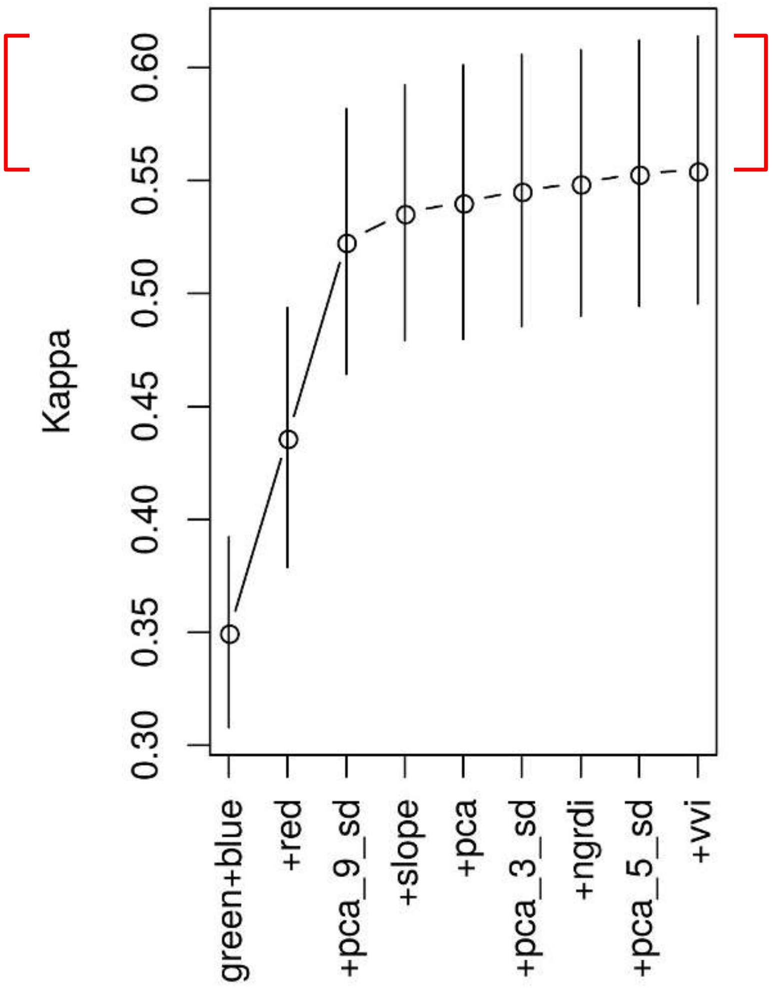
What is the information the algorithm uses to detect the horse?

Lapuschkin et al., 2019

“Right for the wrong scientific reasons” (Schramowski et al., 2020)?
If scientific reason is not right, the model won’t be able to make reliable predictions for new samples!
→ We already revealed by spatial validation that our case study model is not right…
But how to get it right?

Unmasking “clever Hans predictors” to improve the model?

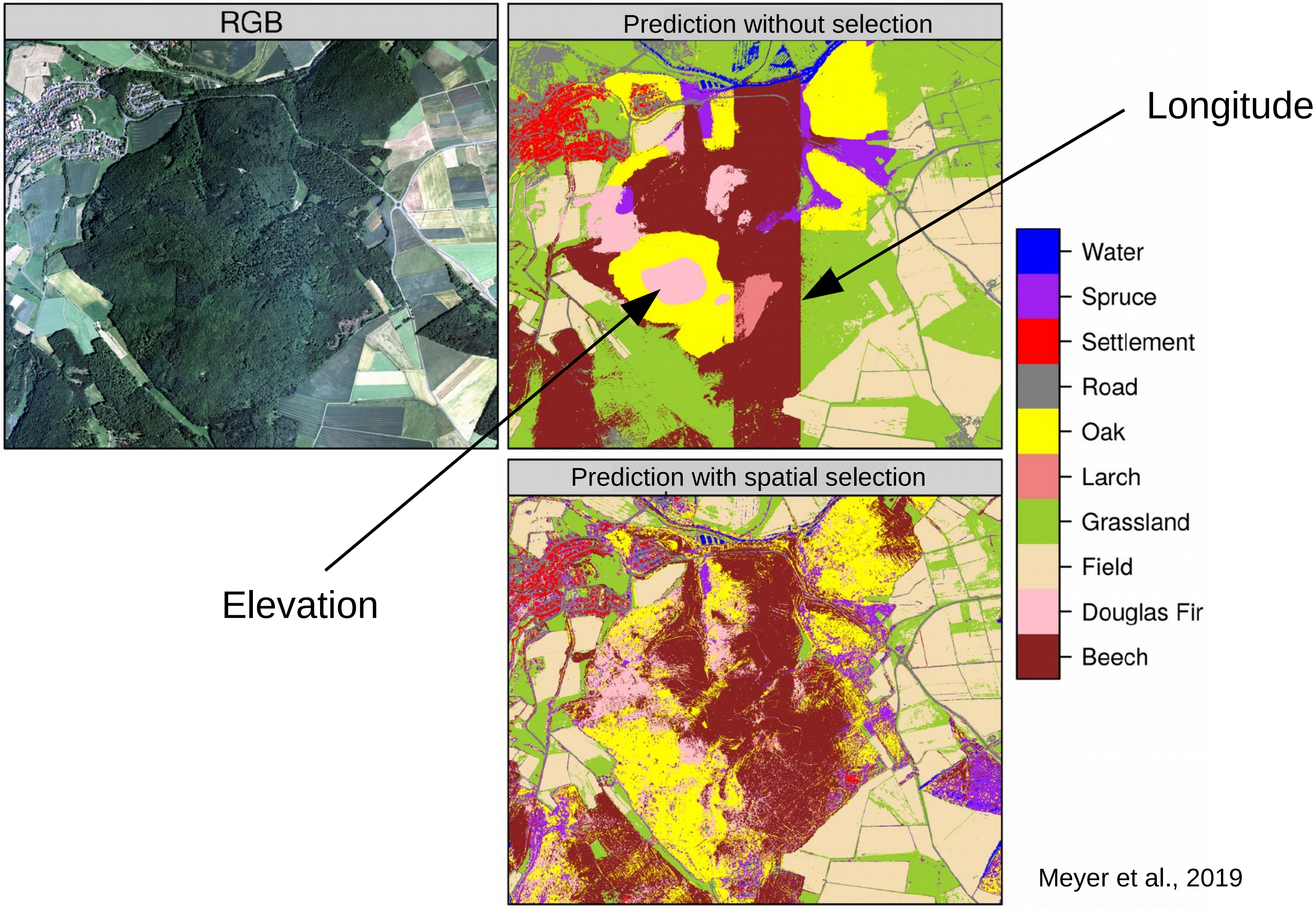
Variable importance



- Assumption: spatial autocorrelation leads to “clever Hans predictors”
- Removing those variables should improve the results
- Spatial variable selection required!

Implemented in R package “CAST”

Unmasking “clever Hans predictors” to improve the model?



Meyer et al., 2019

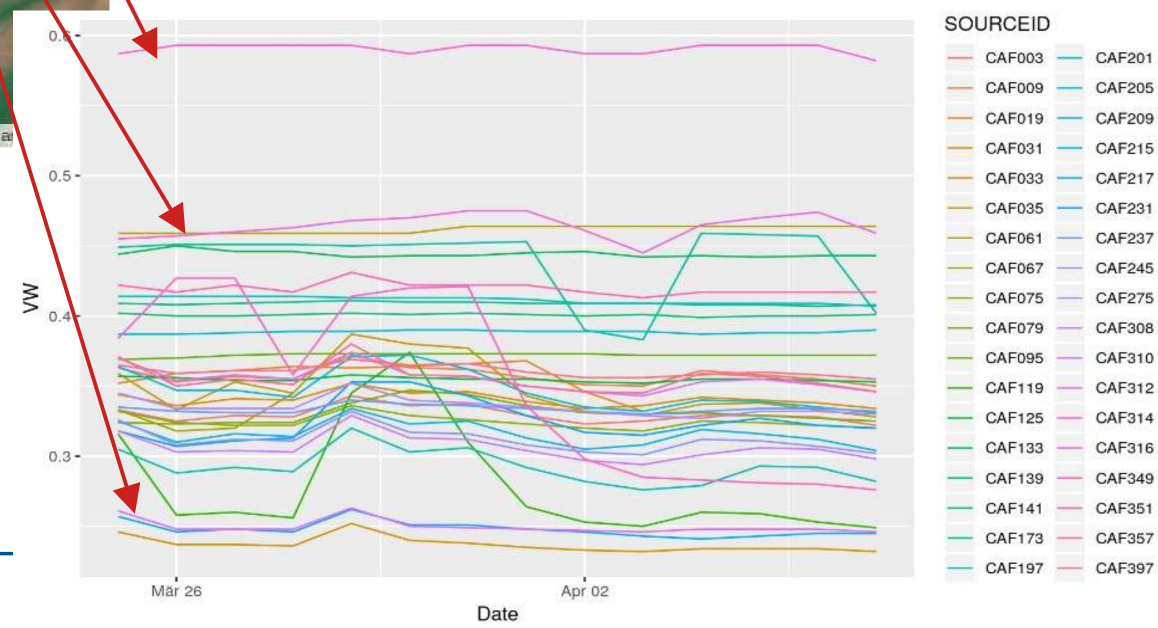
What we have learned so far...

- Cross-validation strategy affect:
  - Performance estimate
  - Selected hyperparameters
  - Variable selection
- Consequences of using an unsuitable CV:
  - Unreliable performance estimates
  - Models that can well reproduce but not necessarily predict ("clever Hans effect")
- Hence, CV strategies that fit the prediction task are required during model selection and validation!

How might this help you to win the spatial prediction competition ?

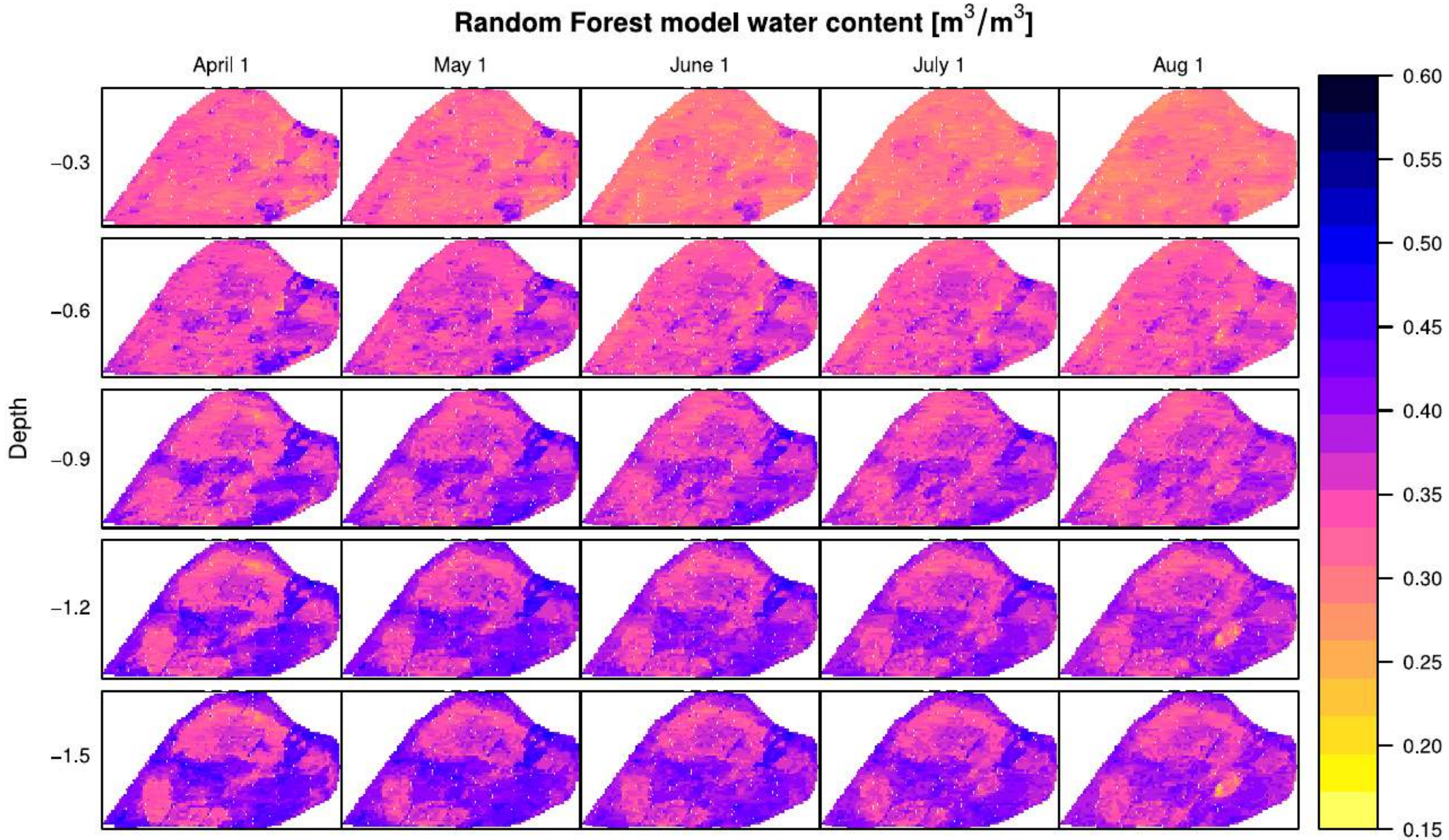
Summer school 2014 spatial prediction competition

Aim: 4D predictions of soil moisture for the Cookfarm
Response variable: Soil Moisture measured by ~40 sensors
Predictors: open (Elevation, Soil properties, rainfall etc available)



WWU MÜNSTER
ILÖK Institut für Landschaftsökologie
Remote Sensing & Spatial Modelling

Summer school 2014 spatial prediction competition - Aim



Gasch et al., 2015

Summer school 2014 spatial prediction competition

Winning modelling strategy

- Potential predictors:

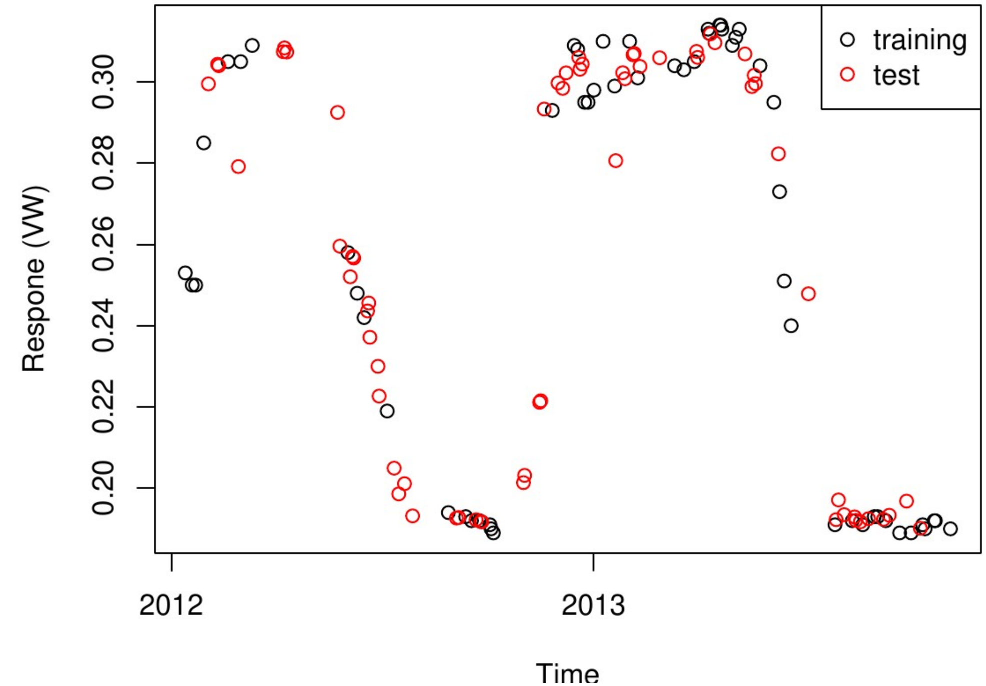
```
####################################################################
#define predictors for model
####################################################################
predictorVariables=c("altitude","year","julianDay","TAXSUSDA","latitude","longitude",
                     "DEM","TWI","bt","Precip_wrcc","MaxT_wrcc","MinT_wrcc")
```

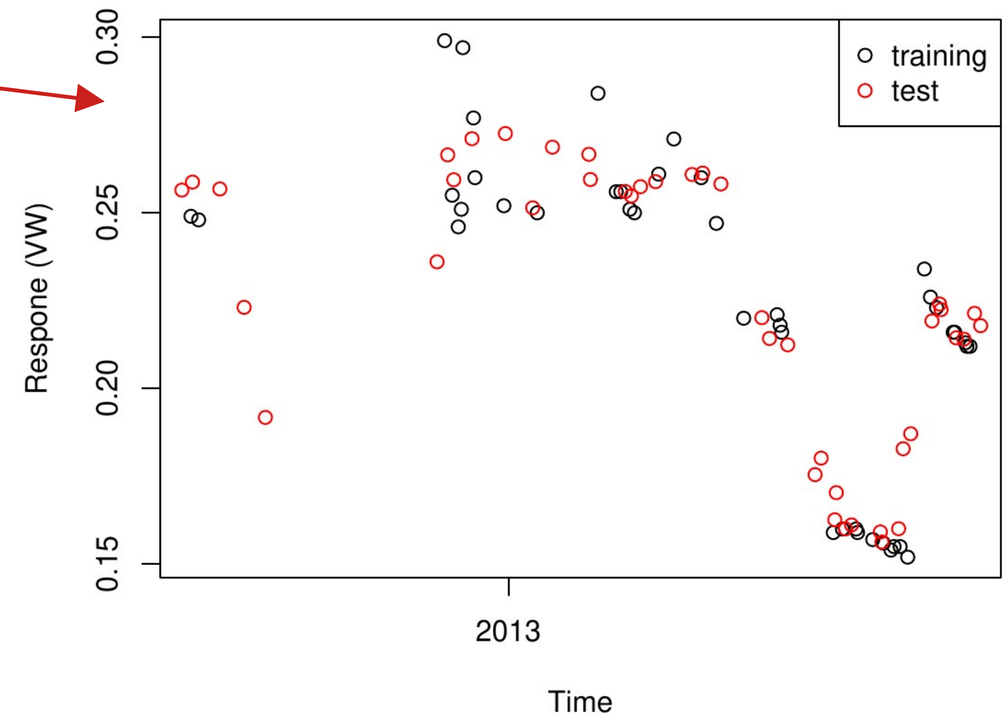
- Train random forest models with **random cross-validation**
- Make predictions for test points

- Why do you think these predictors and a random cross-validation for model selection (tuning/variable selection) could be a successful strategy ?

Summer school 2014 spatial prediction competition

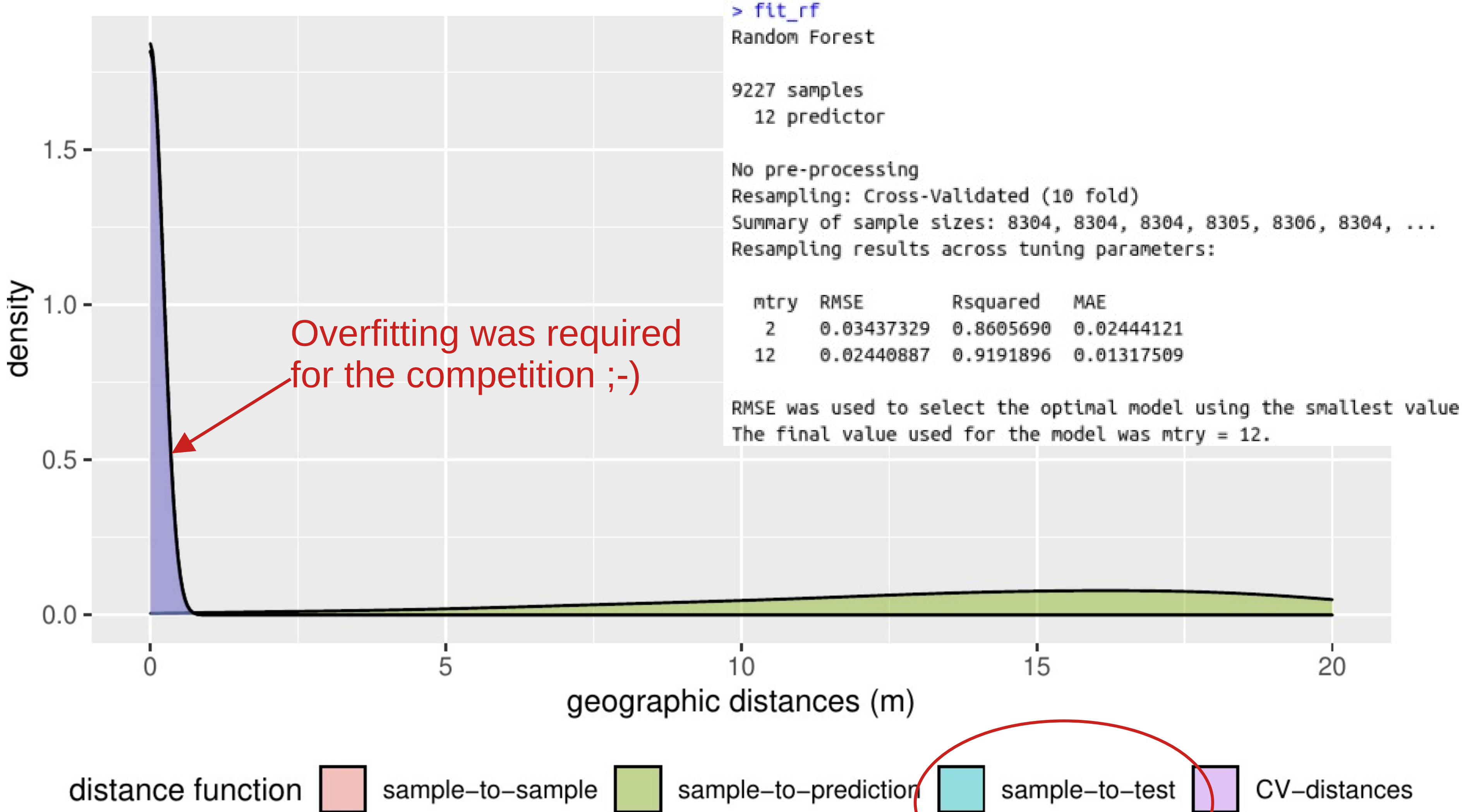
Training and test data





...

Summer school 2014 spatial prediction competition
Random cross-validation




```
> fit_rf
Random Forest

9227 samples
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 8304, 8304, 8304, 8305, 8306, 8304, ...
Resampling results across tuning parameters:

  mtry  RMSE        Rsquared   MAE
   2    0.03437329  0.8605690  0.02444121
  12    0.02440887  0.9191896  0.01317509

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 12.
```

Summer school 2014 spatial prediction competition
Spatial cross-validation

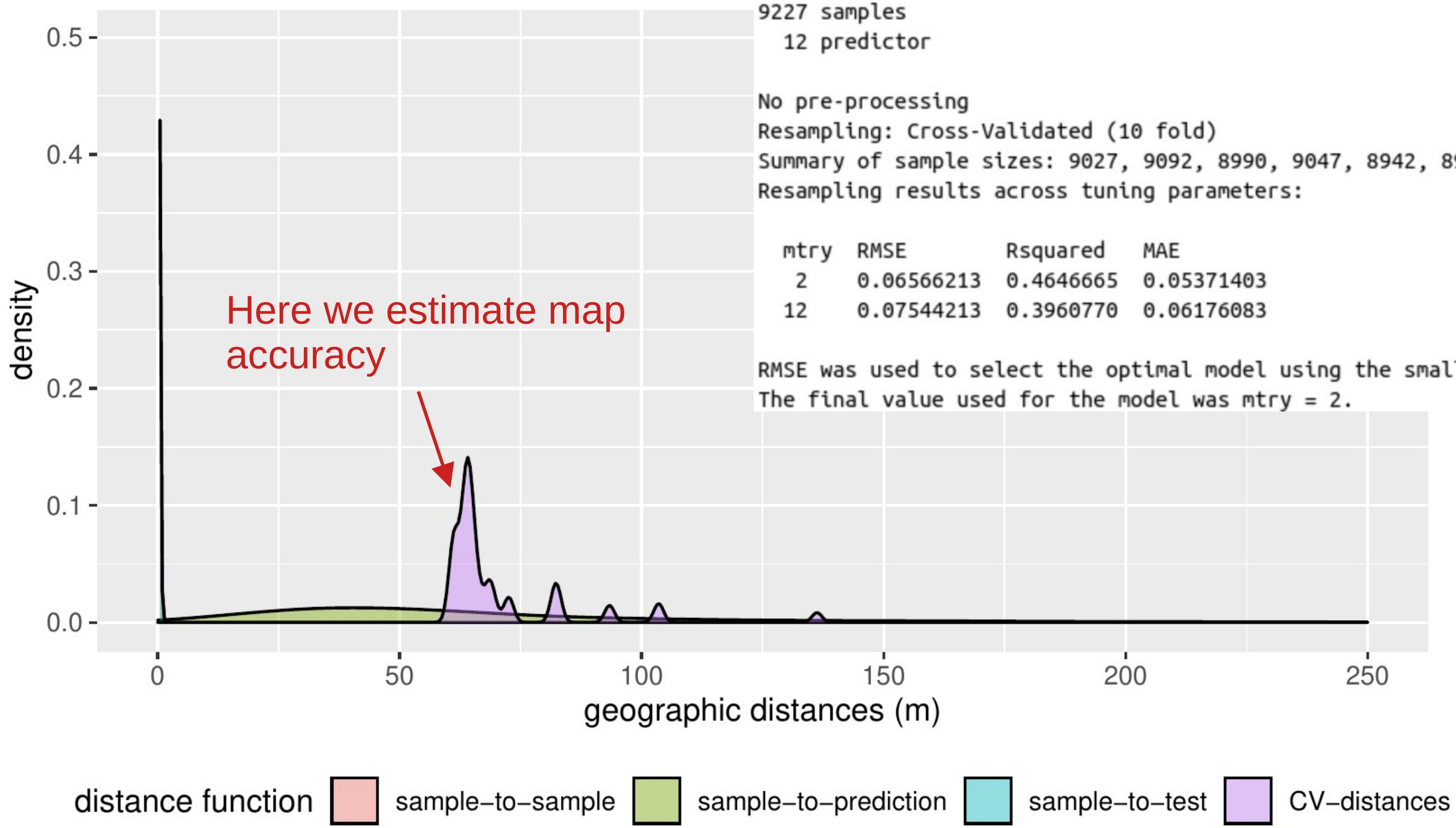
```
> fit_rf_sp
Random Forest

9227 samples
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 9027, 9092, 8990, 9047, 8942, 8920, ...
Resampling results across tuning parameters:

  mtry  RMSE        Rsquared   MAE
   2    0.06566213  0.4646665  0.05371403
  12    0.07544213  0.3960770  0.06176083

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 2.
```



Summary

- Train and validate your model with respect to the aim of the modelling approach, e.g.
  - spatial (or even 4D) mapping
  - ...or winning the competition. Strategy depends on the test data ;-)

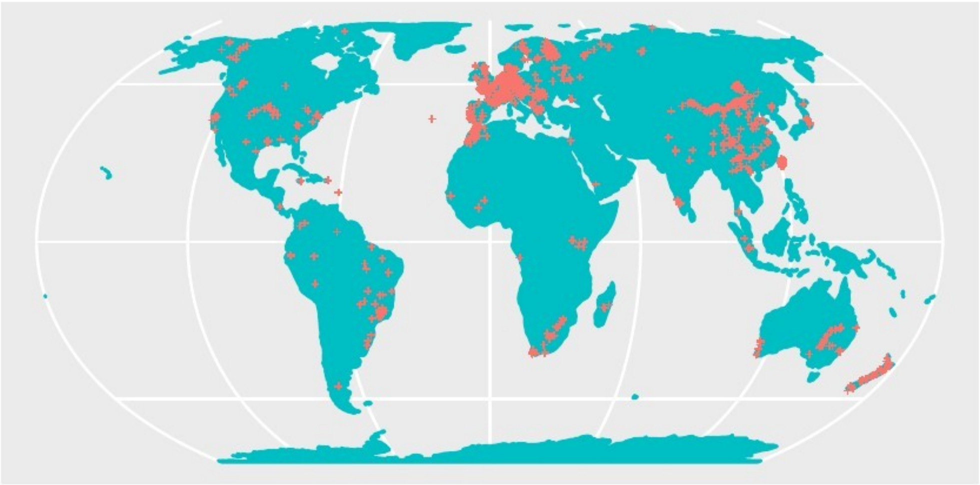
- Here latitude, longitude and day of the year as predictors allowed training models that perfectly reproduced time series (and to win the competition)...but probably not a good choice for spatial prediction → spatial variable selection!

What we have learned so far...

- Cross-validation strategy affect:
  - Performance estimate
  - Selected hyperparameters
  - Variable selection
- Consequences of using an unsuitable CV:
  - Unreliable performance estimates
  - Models that can well reproduce but not necessarily predict ("clever Hans effect")
- Hence, CV strategies that fit the prediction task are required during model selection and validation!

But is this sufficient for reliable (global) mapping ?

Limits to accuracy assessment



Based on van den Hoogen et al., 2019

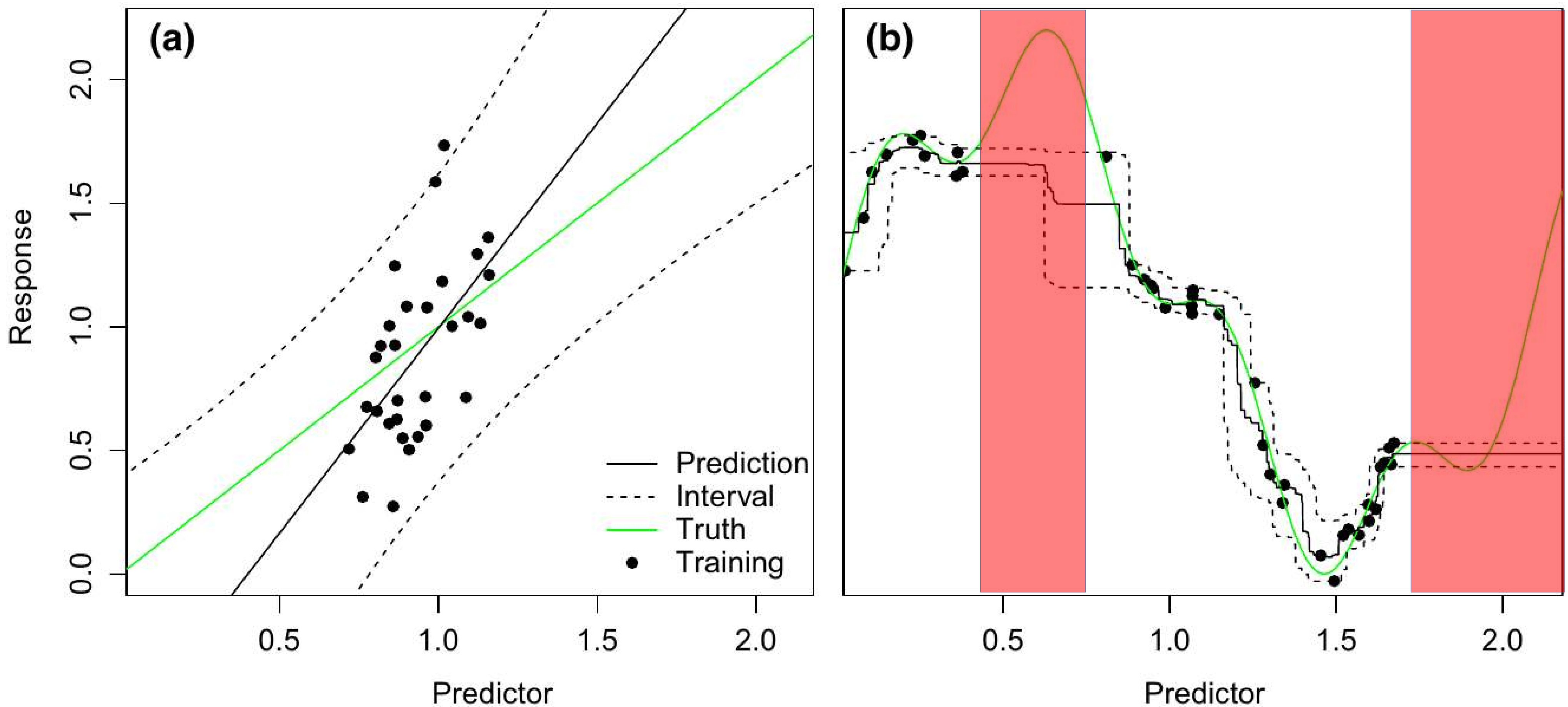
- Mapping requires prediction far beyond clustered reference data
- Transfer to new space required
- New space might differ in environmental properties

New geographic space often goes along with new predictor properties

Example of the Nematodes model by van den Hoogen et al., 2019.
Nearest neighbor distance in the predictor space:

...but what happens if the model has never “seen” such new predictor properties?

Predictions and common uncertainty measures are unreliable beyond training data



Meyer & Pebesma 2021

Shouldn't we avoid predictions into "unknown space"?

Suggestion: Area of Applicability (AOA)

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE | Open Access

Predicting into unknown space? Estimating the area of applicability of spatial prediction models

Hanna Meyer, Edzer Pebesma

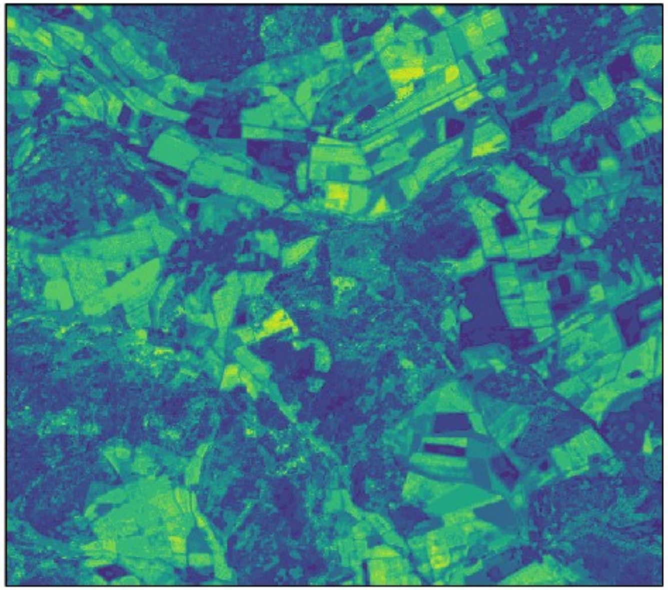
We try to derive the area...

- to which the model can be applied because it has been enabled to learn about relationships
- where the estimated performance holds

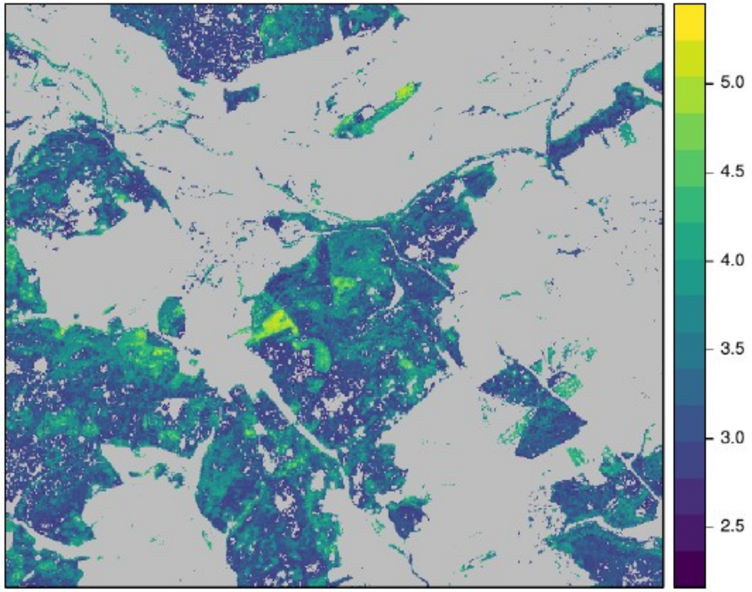
A very obvious and simple example

Sentinel-2 scene and training data points of leaf area index

Predictions

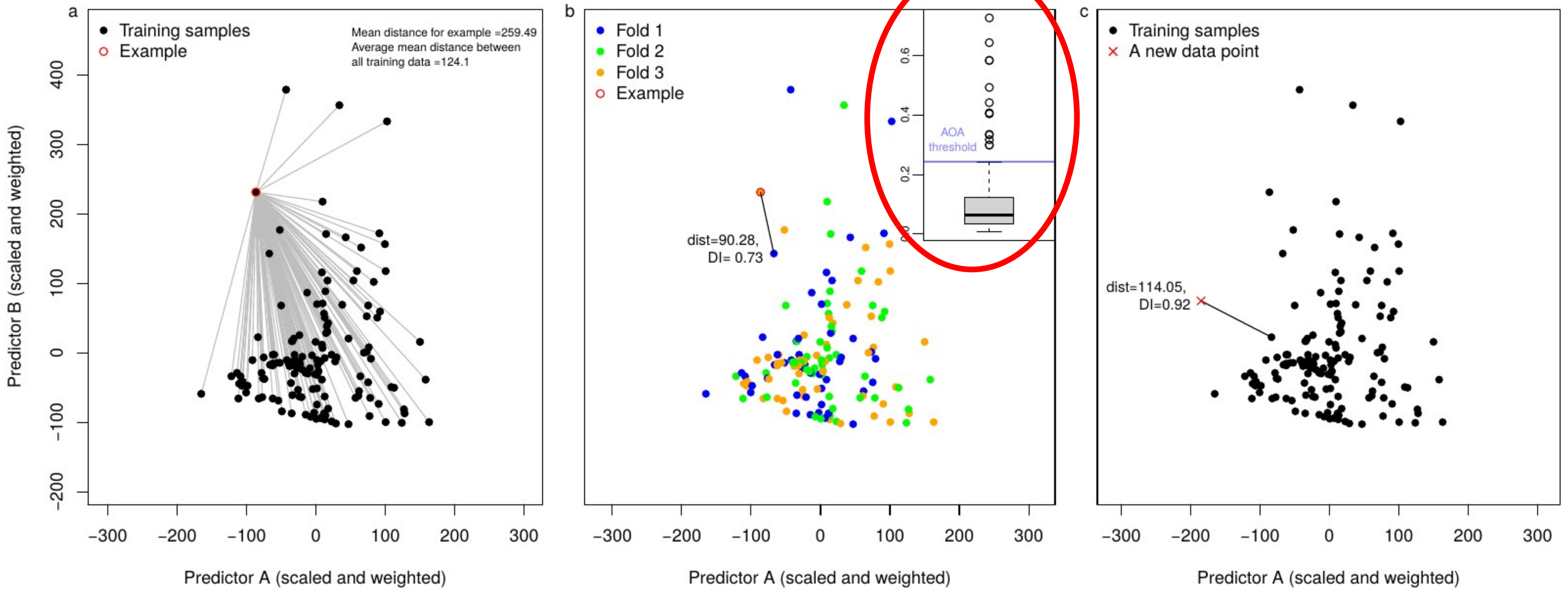


Predictions limited to the AOA



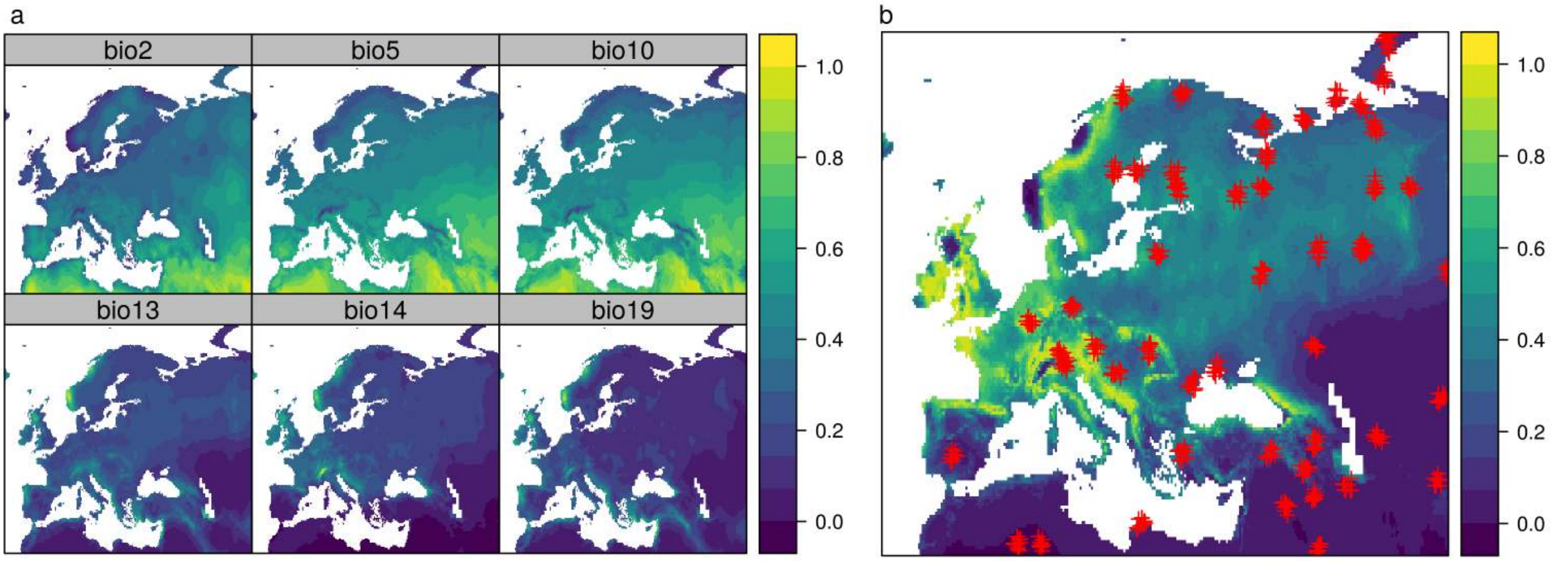
How do we derive the AOA ?

Calculation of a
Dissimilarity Index (DI)



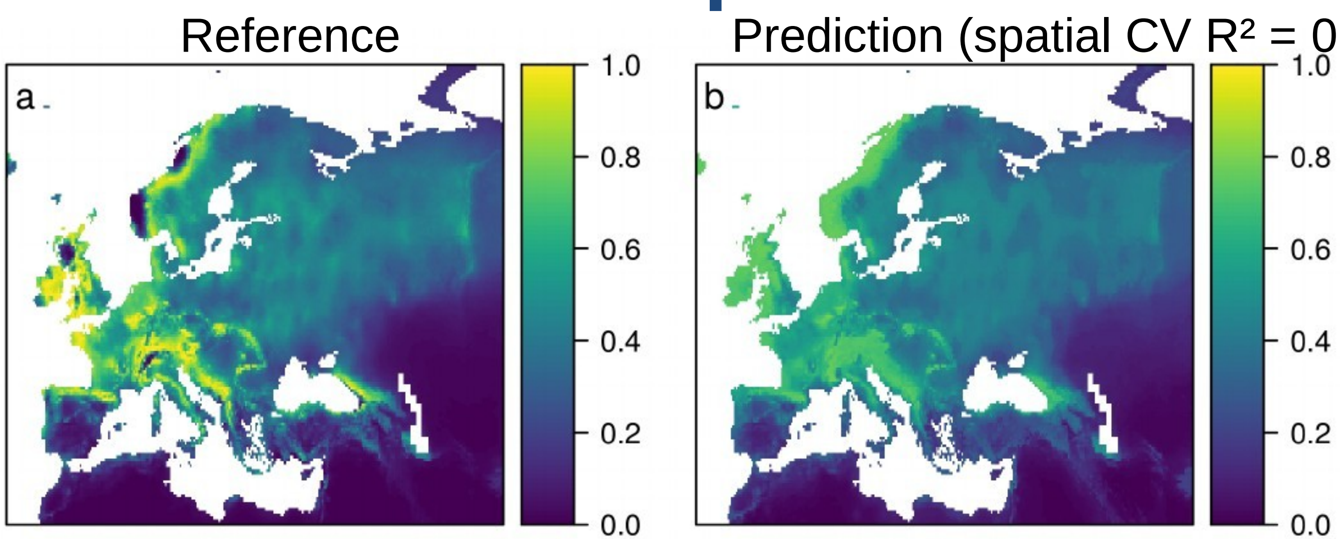
Meyer & Pebesma (2021)

Simulated example: Predictors and response

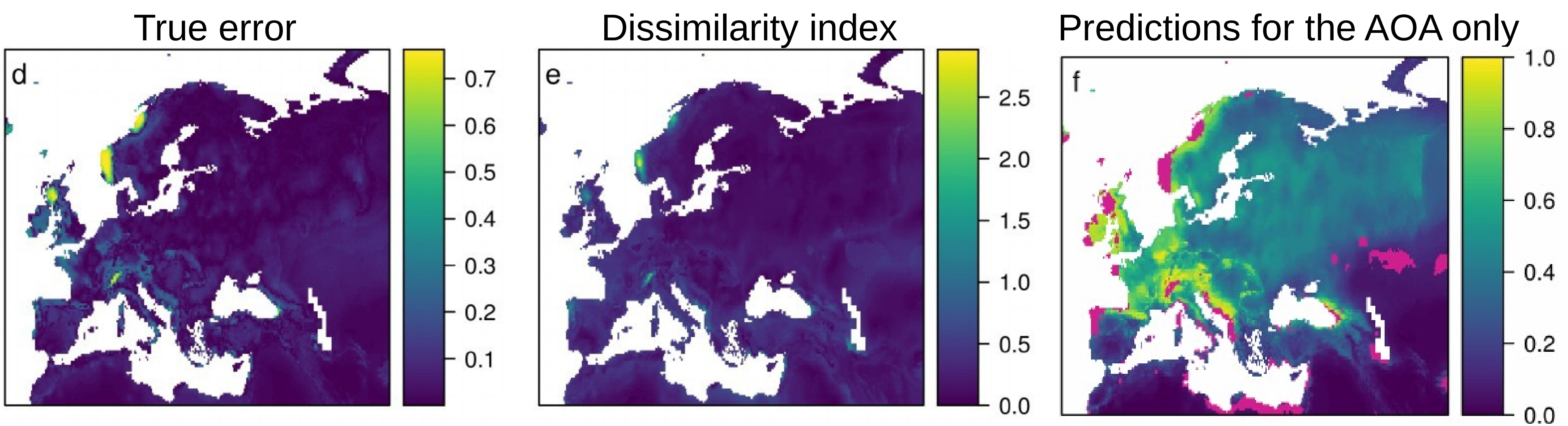


Meyer & Pebesma (2021)

Simulated example: Results



Reproduce example:
github.com/HannaMeyer/MEE_AOA

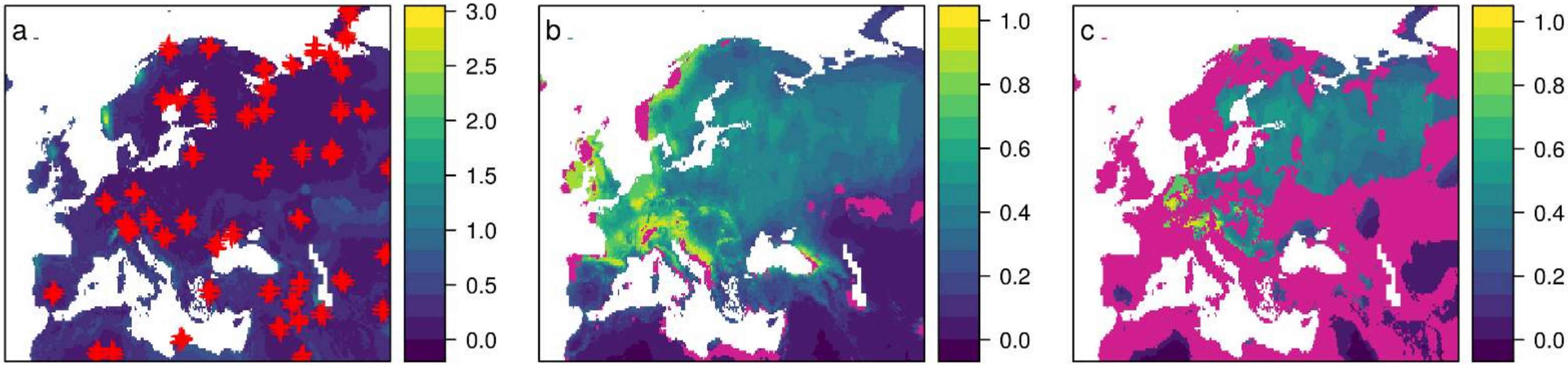


Meyer & Pebesma (2021)

Why does the AOA threshold depend on the CV strategy?

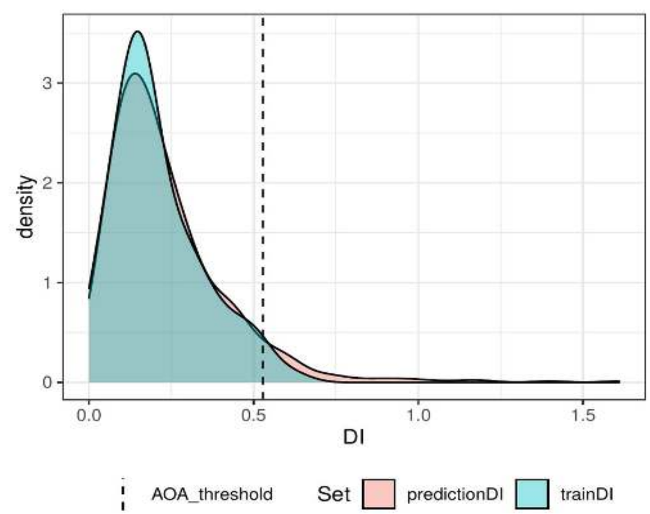
AOA estimated with **spatial CV** (R^2 = 0.79)

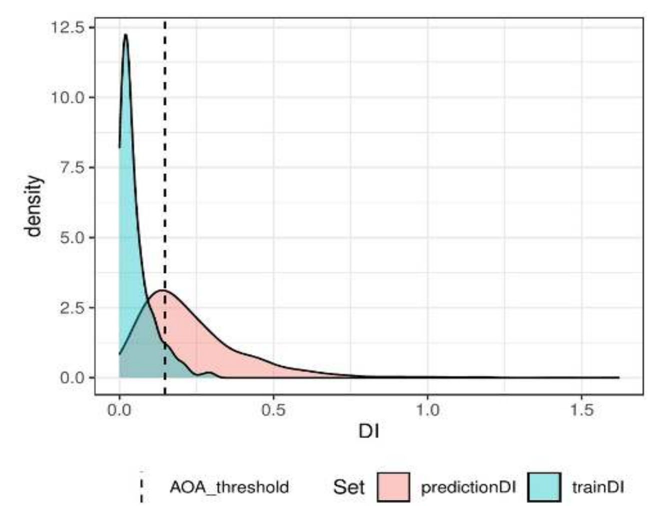
AOA estimated with **random CV** (R^2 = 0.99)



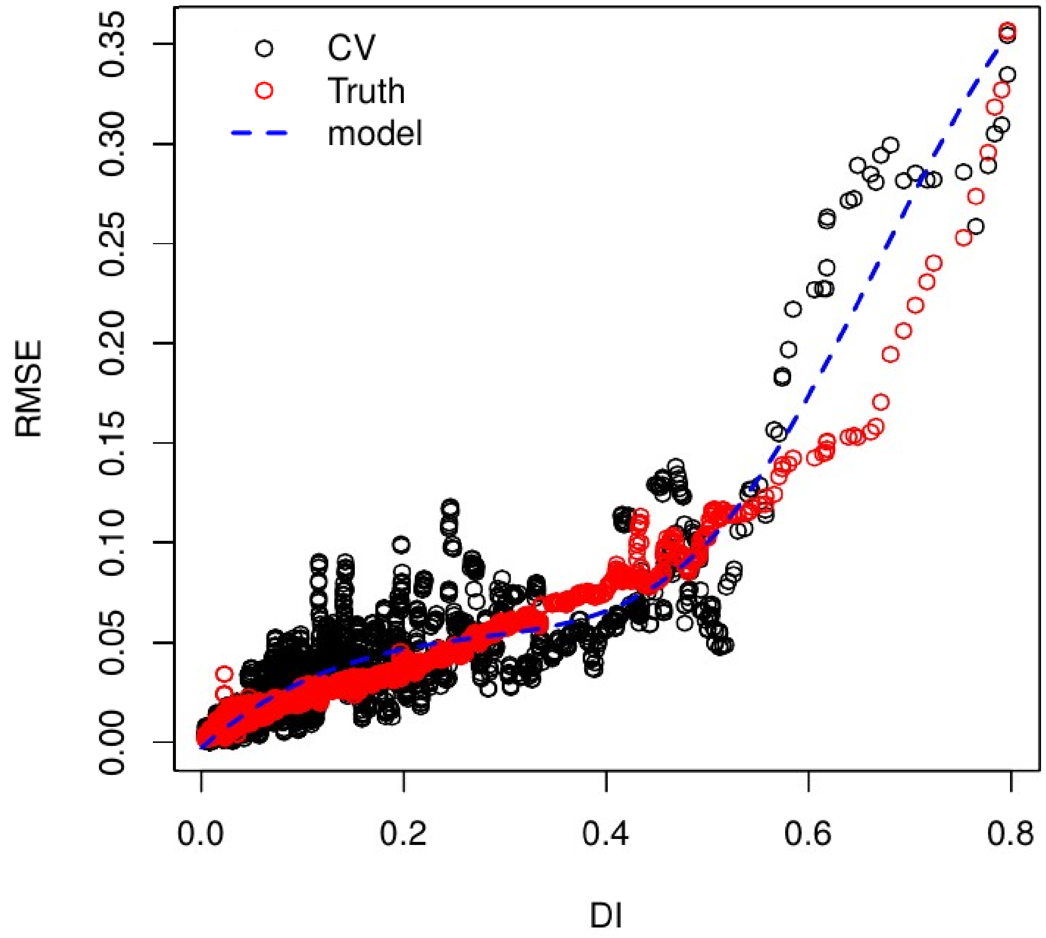
Why do we have these two different AOA’s, depending on the CV strategy ?

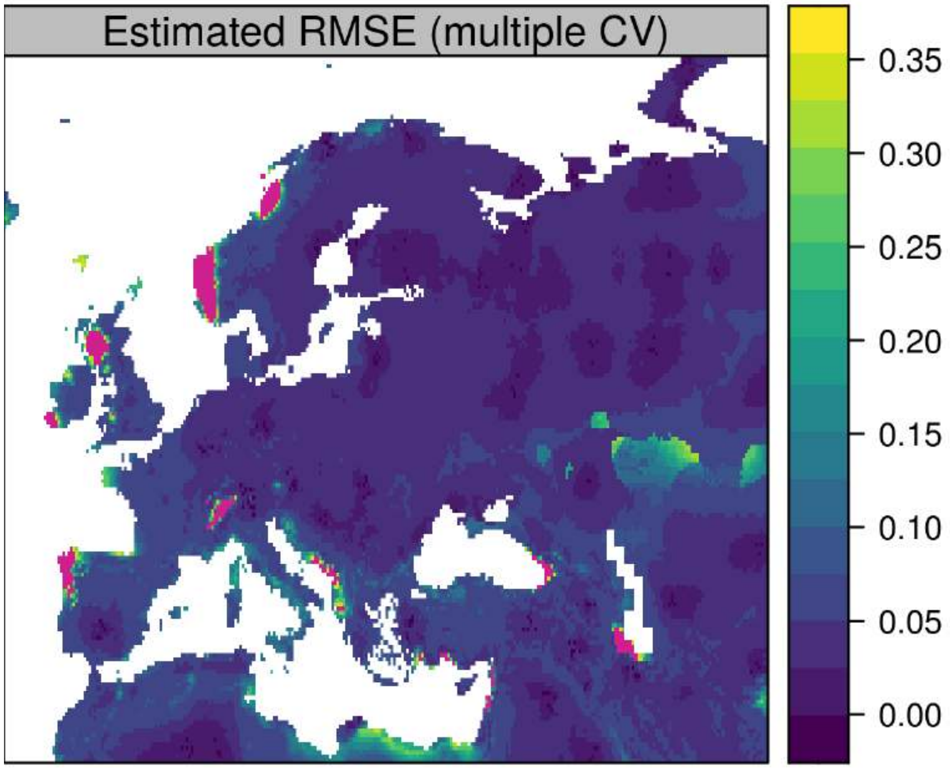
...because we never tested dissimilar prediction situations during random CV





Threshold depends on the CV…
Could we instead use the DI to map estimated performance ?





Why is it relevant to map "unknown space"?

Results are not just nice maps but used for...

- subsequent modeling
- nature conservation
- risk assessment
- …

nature COMMUNICATIONS

COMMENT

https://doi.org/10.1038/s41467-022-29838-9 OPEN

Machine learning-based global maps of ecological variables and the challenge of assessing them

Hanna Meyer [1] & Edzer Pebesma [2]

Our opinion: predictions should only be presented for the area of applicability to avoid error propagation or misplanning

Conclusions

This is not an argument against machine learning but an appeal to keep in mind that applying models to map the entire world based on limited field samples is challenging

- Accuracy assessment: prediction situations created during CV resemble those encountered during prediction
- ...otherwise risk of "clever Hans effect" and low ability of the model to predict beyond training samples
- Predictions should only be made for the AOA (accept gaps!?)
- We (= producers of the maps) are responsible for clearly indicating usage of maps, don't leave it to the user.
- There is still a lot to do...

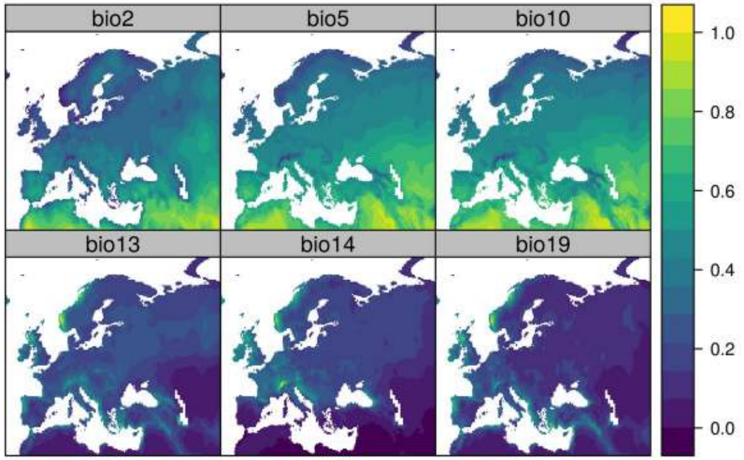
Practice (github.com/HannaMeyer/OpenGeoHub_2022)

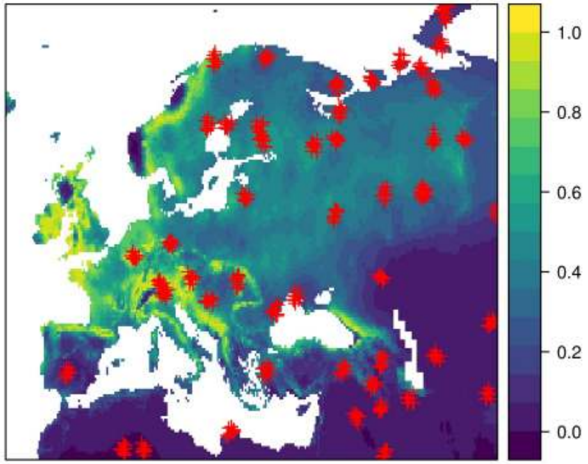
We will go through a typical prediction task and learn how to

- ...explore the suitability of cross-validation strategies
- ...use CV during training and tuning (including variable selection) and explore its effects
- ...assess the AOA of a prediction model

To do this, we will use the R package CAST, which is a wrapper around caret. We will also see how the central developments like NNDM CV and the AOA can be used with mlr3 as well

Example: Suitability for a virtual species in Europe (simulated)





References

- Bastin et al. 2019: The global tree restoration potential. Science. Vol. 365, Issue 6448, pp. 76-79.
- Batjes, N. H., Ribeiro, E. & van Oostrum, A. Standardised soil profile datasupport global mapping and modelling (wosis snapshot 2019). Earth Syst. Sci. Data 12, 299–320 (2020).
- Hengl et al. (2017): SoilGrids250m: Global gridded soil information based on machine learning. PloS one 12(2): e0169748.
- Kattge, J. et al. TRY plant trait database – enhanced coverage and open access. Glob. Change Biol. 26, 119–188 (2020).
- Lapuschkin et al (2019): Unmasking Clever Hans predictors and assessing what machines really learn. Nature Communications volume 10.
- Meyer H, Pebesma E. 2022. 'Machine learning-based global maps of ecological variables and the challenge of assessing them.' Nature Communications 13.
- Meyer H, Pebesma E (2021): Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods in Ecology and Evolution.
- Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T. (2019): Importance of spatial predictor variable selection in machine learning applications - Moving from data reproduction to spatial prediction. Ecological Modelling. 411, 108815.
- Milà, C., Mateu, J., Pebesma, E. & Meyer, H. Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation. Methods in Ecology and Evolution. 00, 1–13 (2022).
- Moreno-Martinez, A. et al. A methodology to derive global maps of leaf traits using remote sensing and climate data. Remote Sens. Environ. 218, 69–88 (2018).
- Wadoux, A. M.-C., Heuvelink, G. B., de Bruin, S. & Brus, D. J. Spatial cross-validation is not the right way to evaluate map accuracy. Ecol. Modell. 457, 109692 (2021).
- Schramowski, P., Stammer, W., Teso, S. et al. (2020): Making deep neural networks right for the right scientific reasons by interacting with their explanations. Nat Mach Intell 2, 476–486.
- Van den Hoogen, J., Geisen, S., Routh, D. et al. (2019): Soil nematode abundance and functional group composition at a global scale. Nature 572, 194–198.