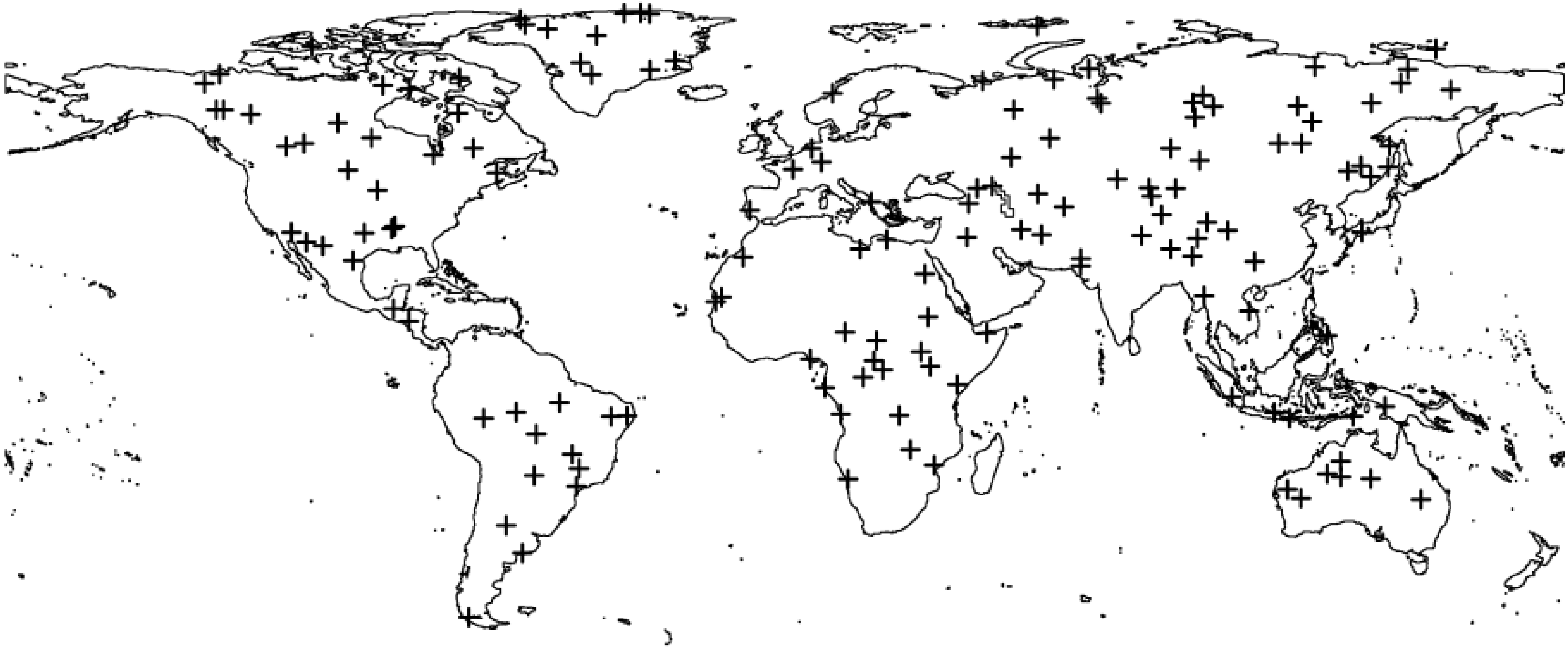
Predicting into unknown space?
Estimating the area of applicability of spatial prediction models

Hanna Meyer[1] & Edzer Pebesma[2]

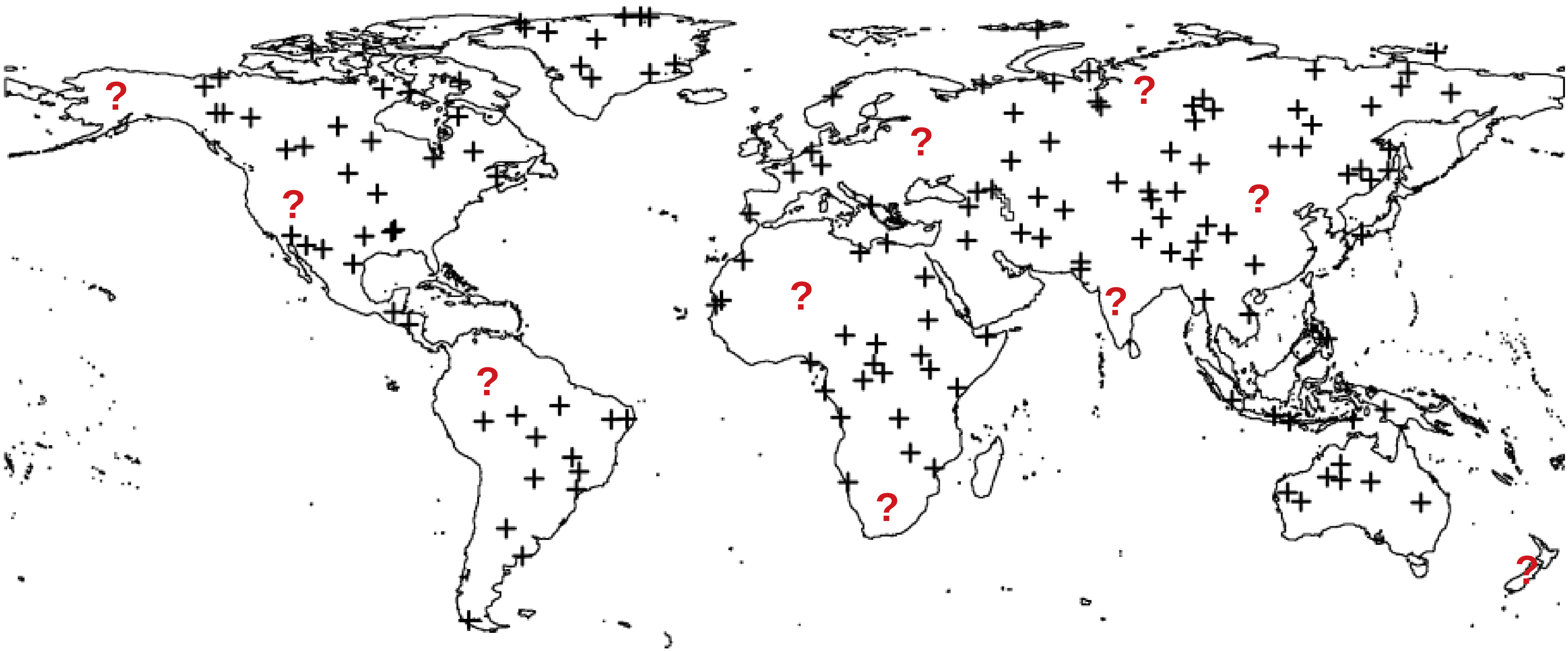
1 Institute of Landscape Ecology, WWU Münster
2 Institute for Geoinformatics, WWU Münster

Problem: Moving from field observations to maps of ecosystem variables

Problem: Moving from field observations to maps of ecosystem variables

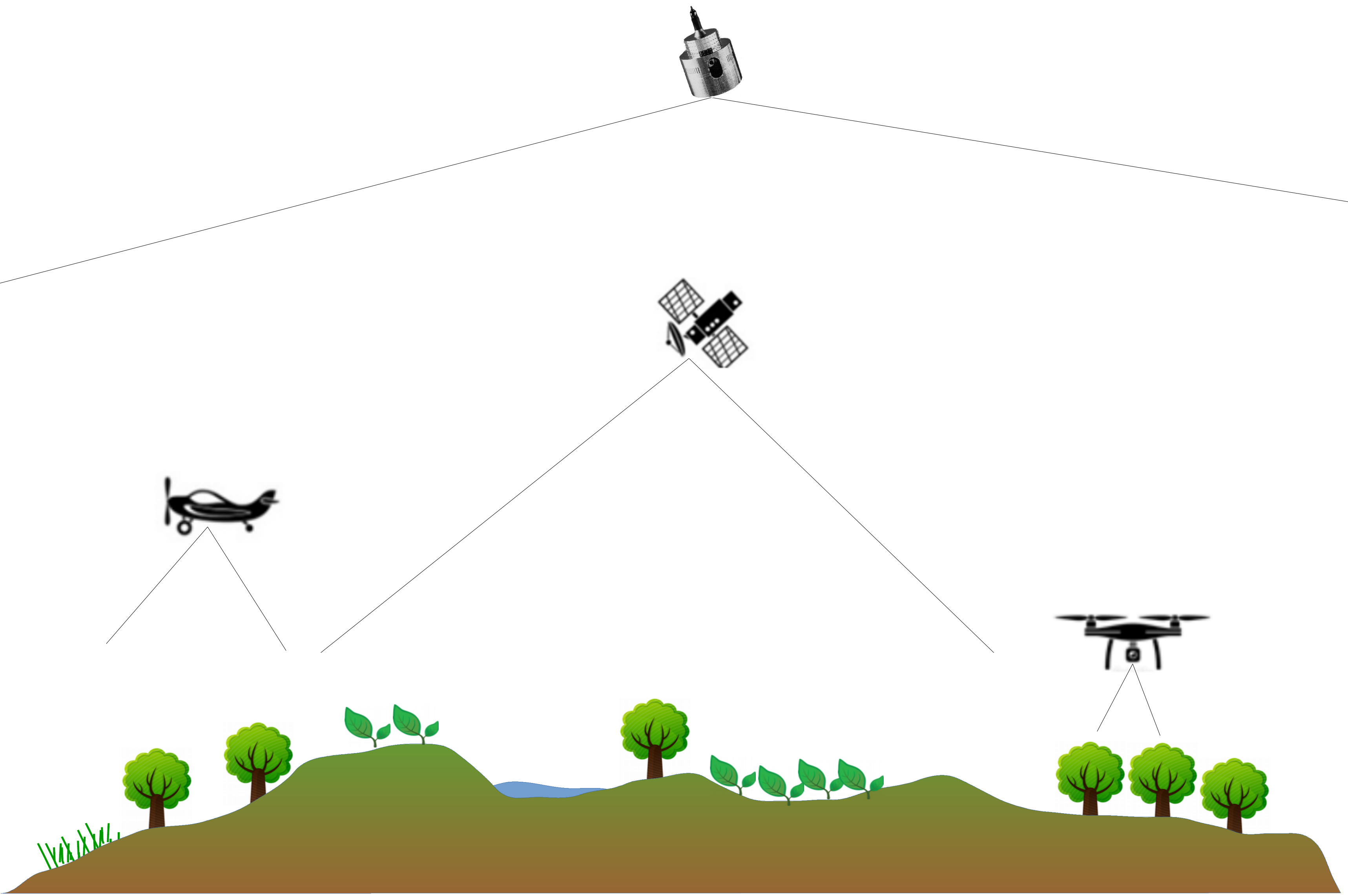


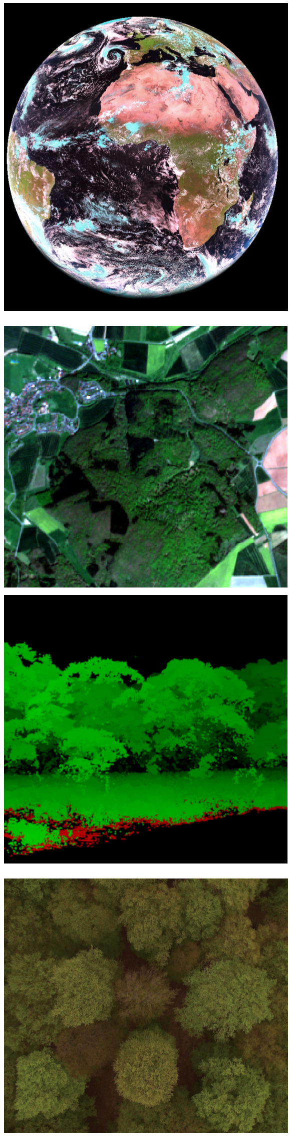
Problem: Moving from field observations to maps of ecosystem variables



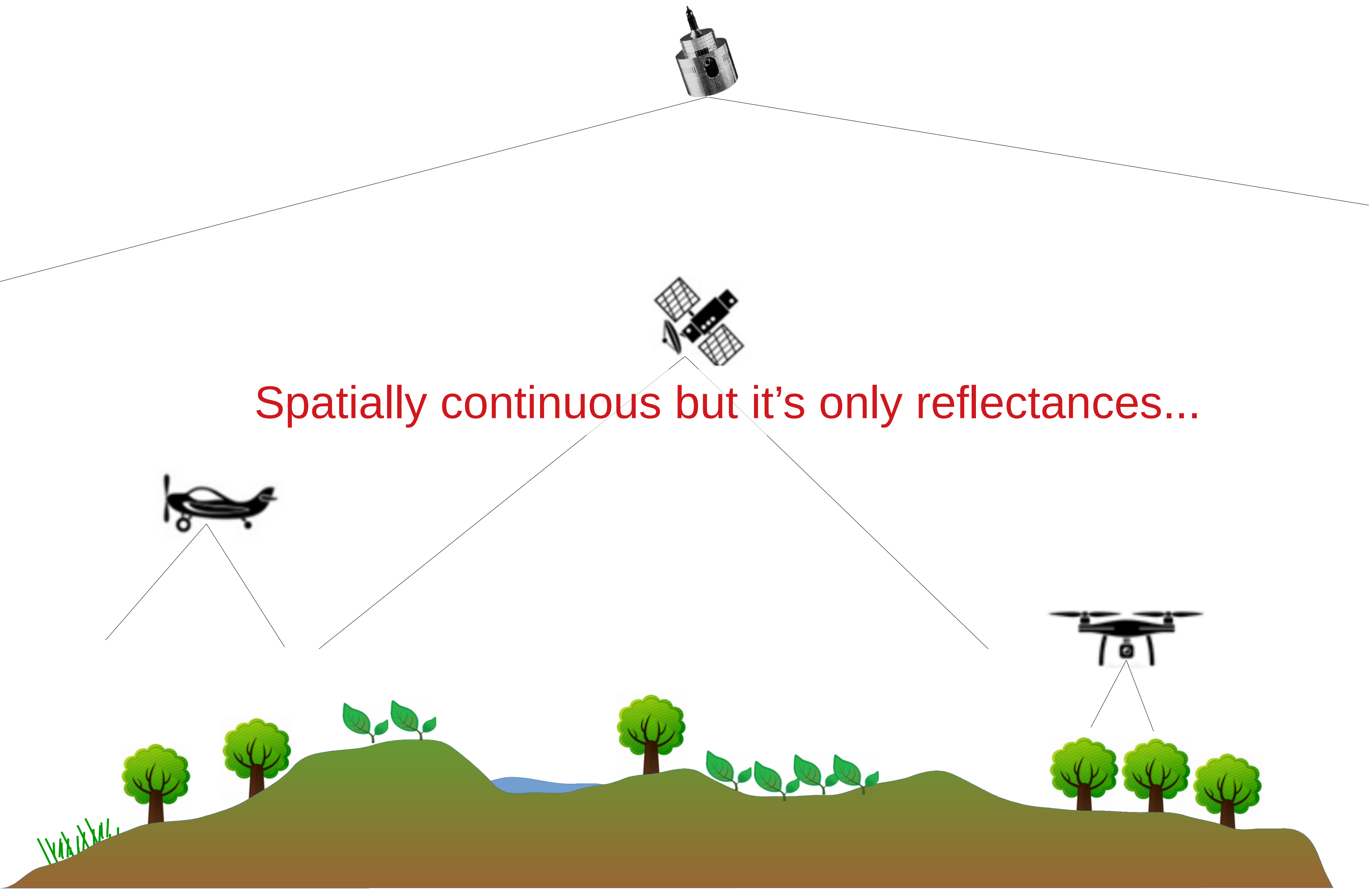
How do we fill the gaps between sampling locations?

Remote Sensing of landscapes

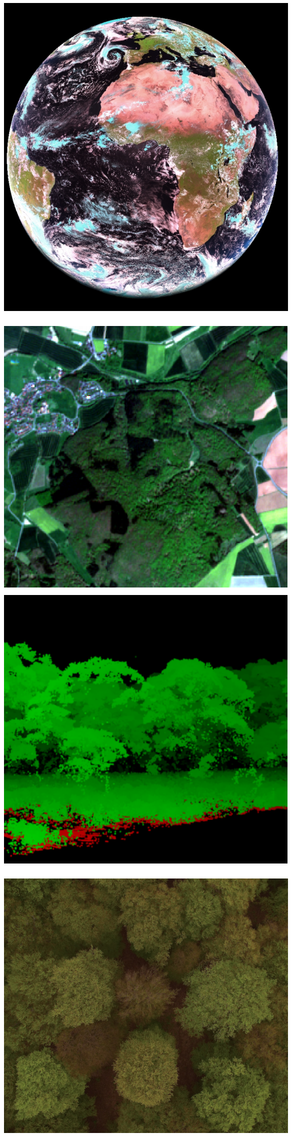




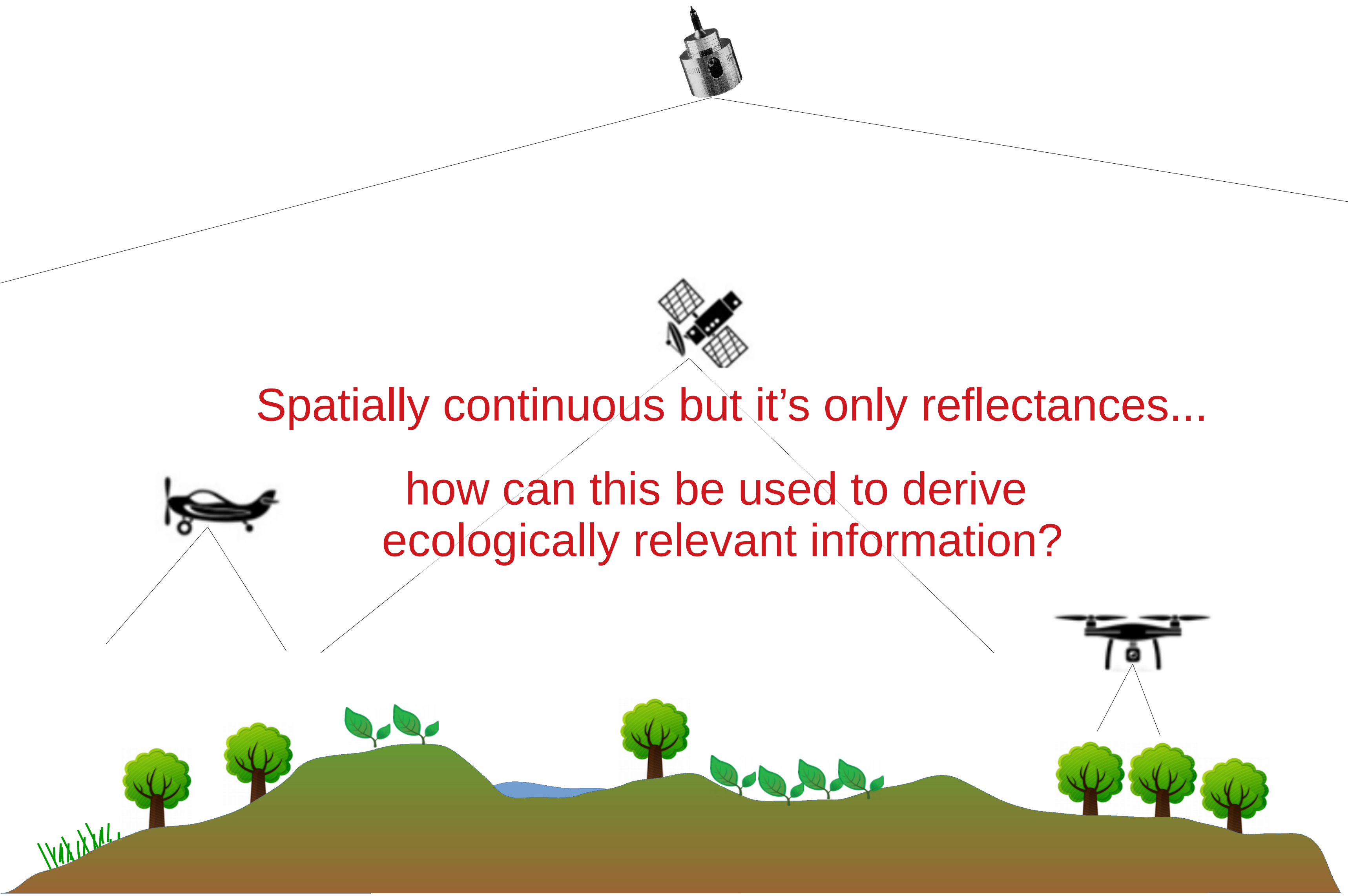
Remote Sensing of landscapes



Spatially continuous but it's only reflectances...

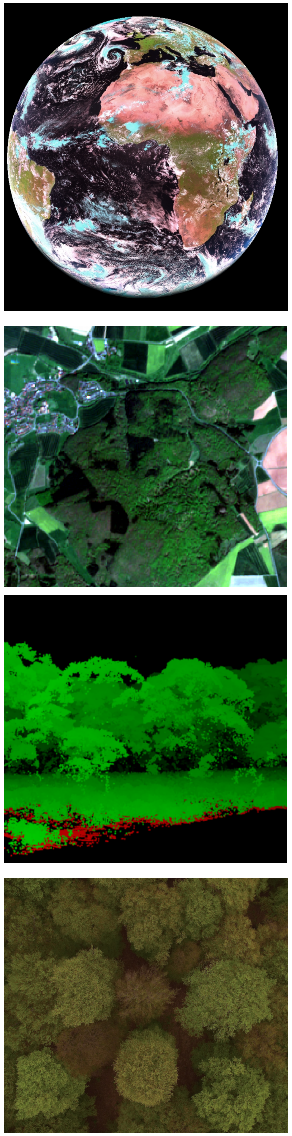


Remote Sensing of landscapes

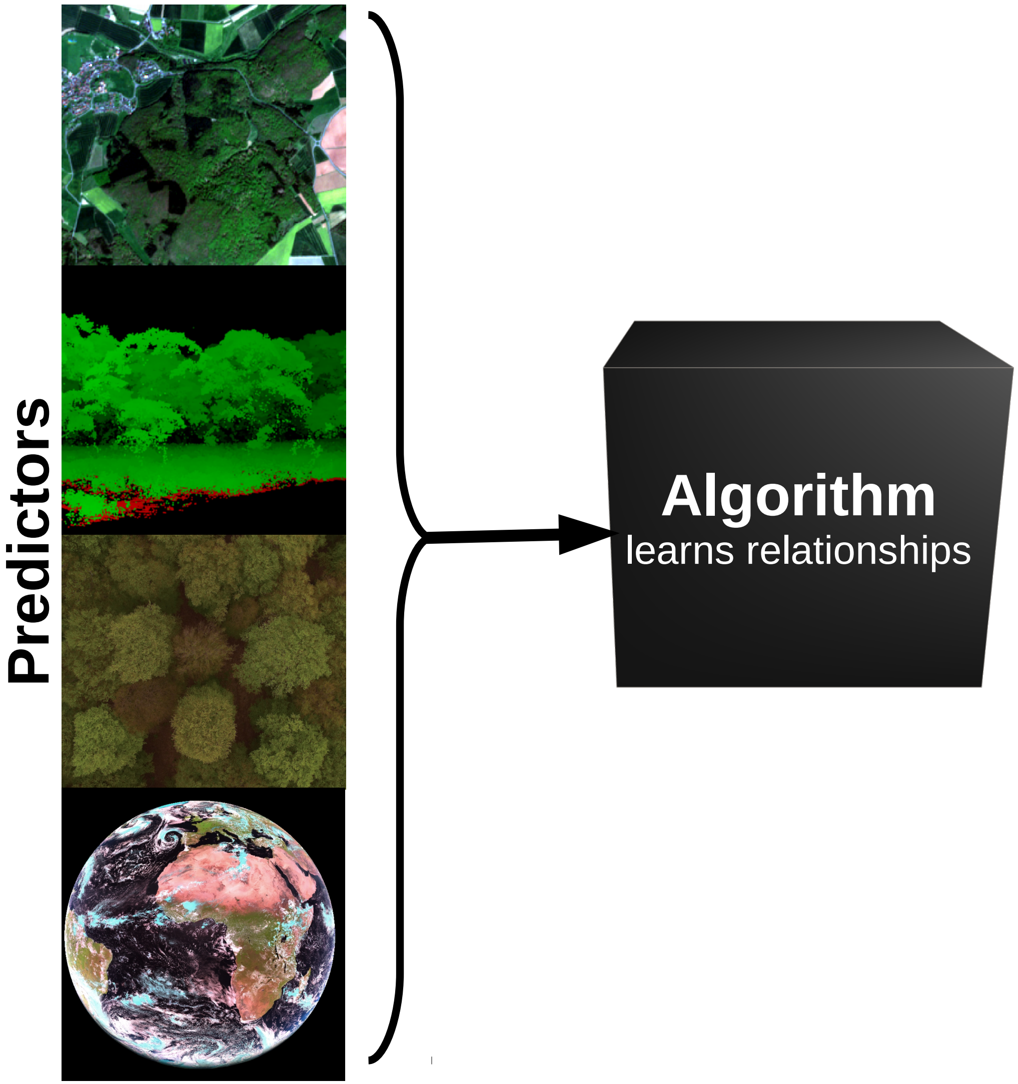


Spatially continuous but it’s only reflectances...

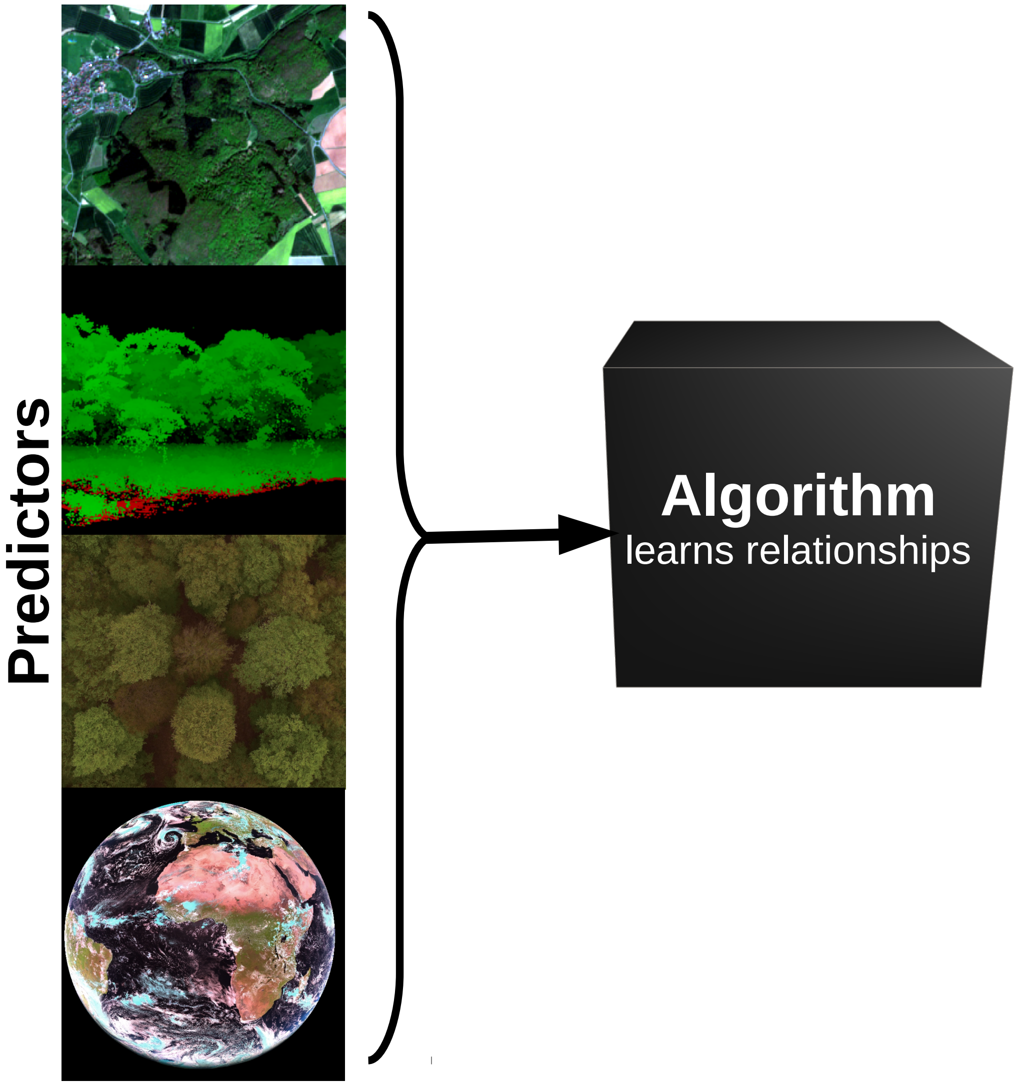
how can this be used to derive
ecologically relevant information?



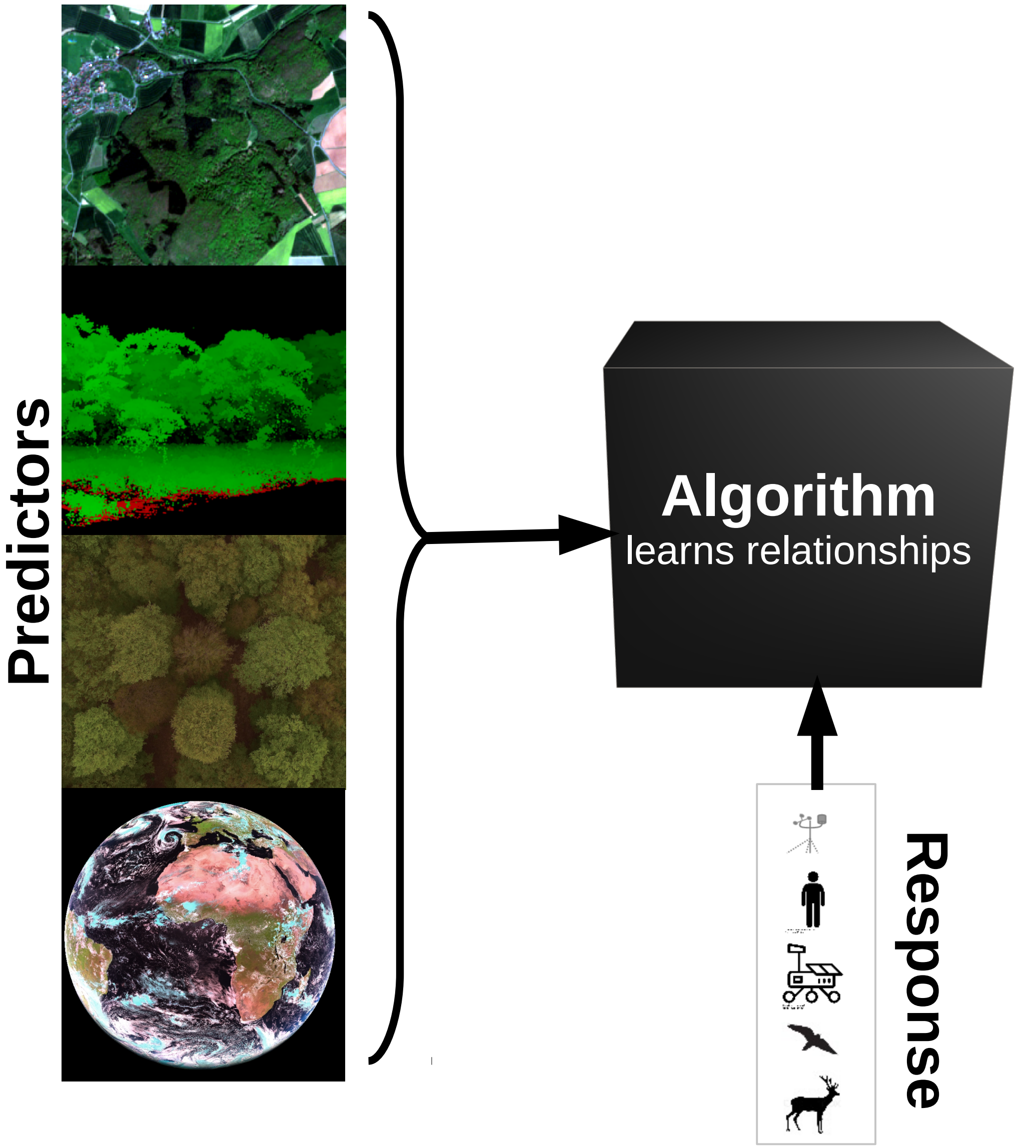
Predictive modelling of the environment: The machine learning way



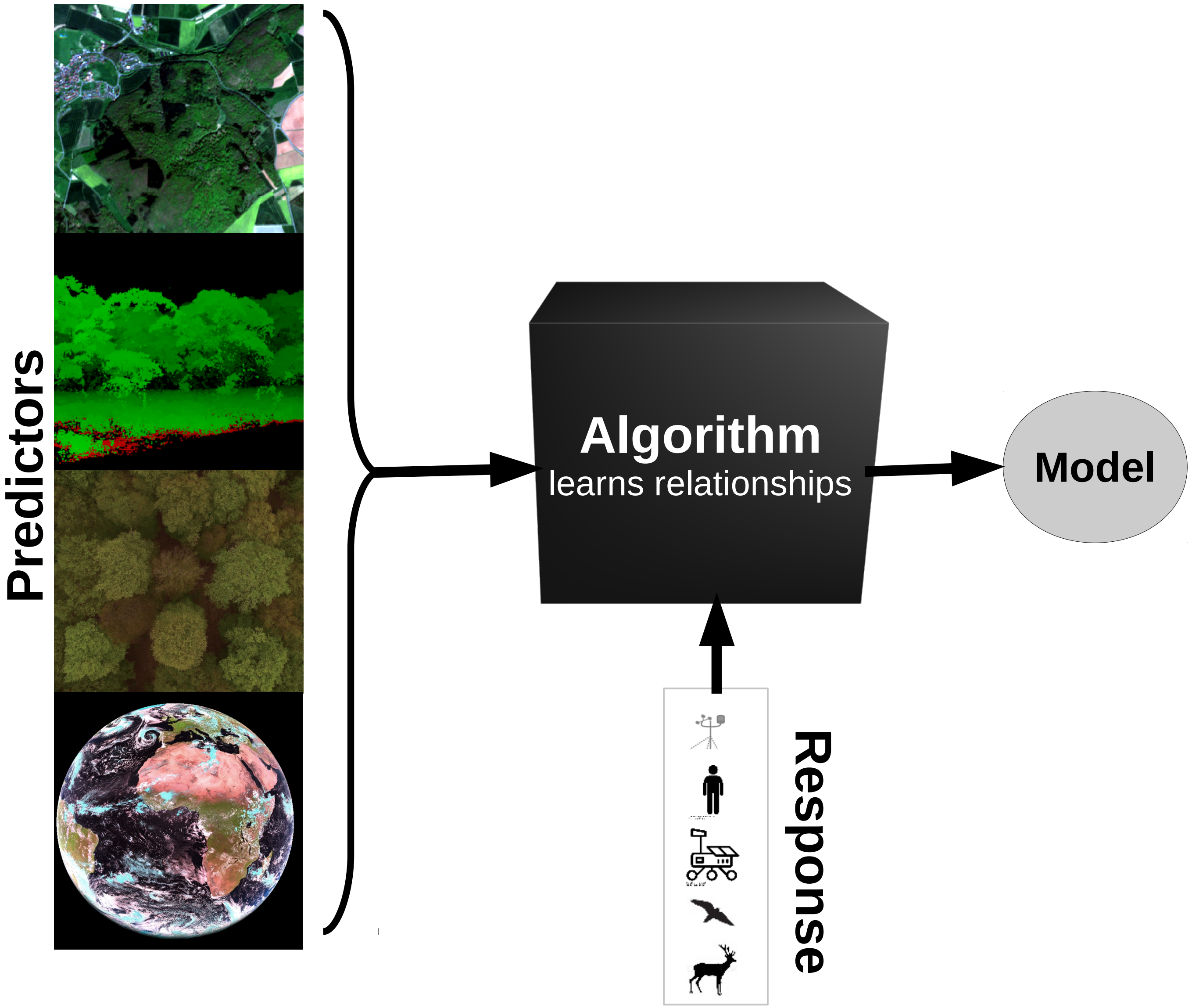
Predictive modelling of the environment: The machine learning way



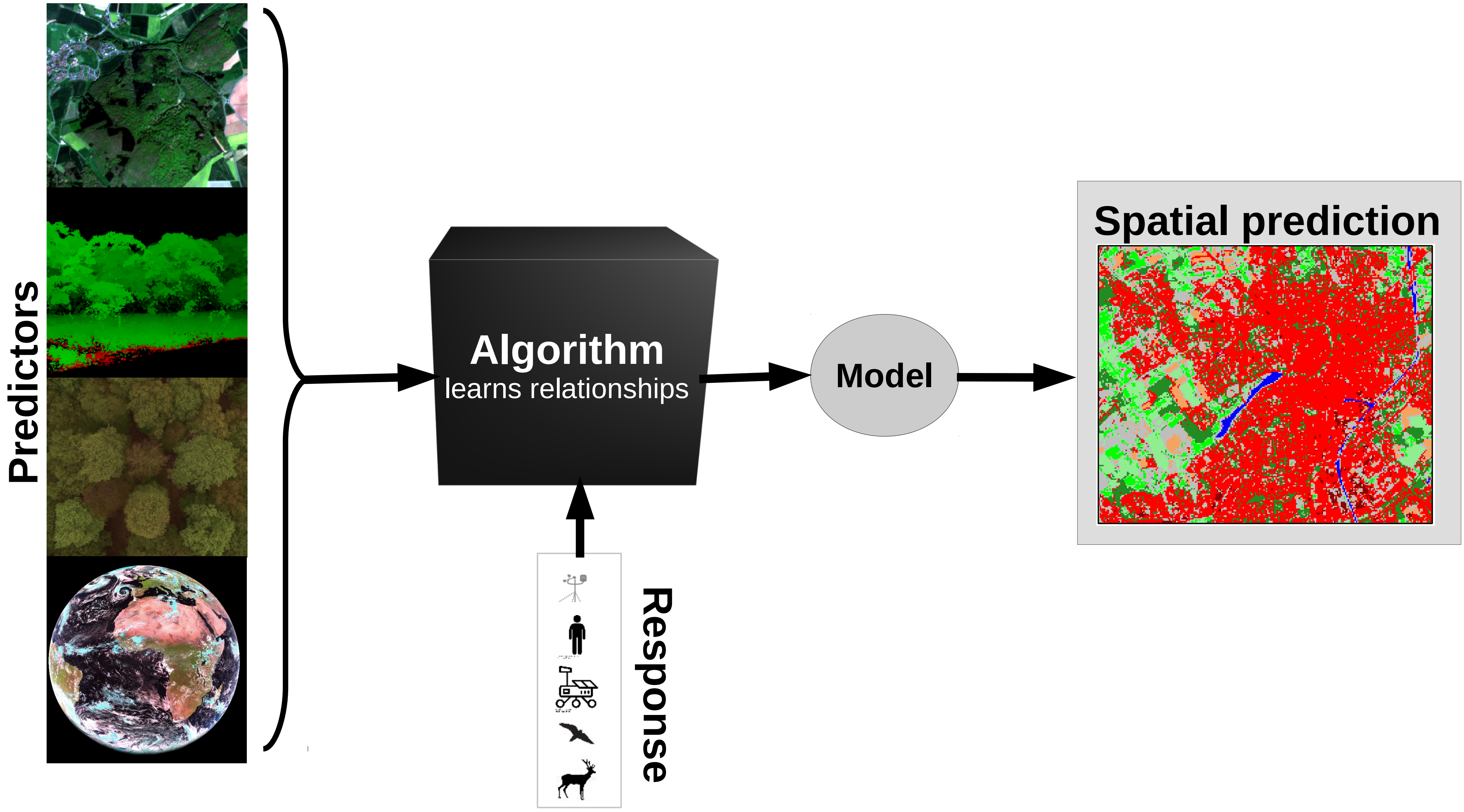
Predictive modelling of the environment: The machine learning way



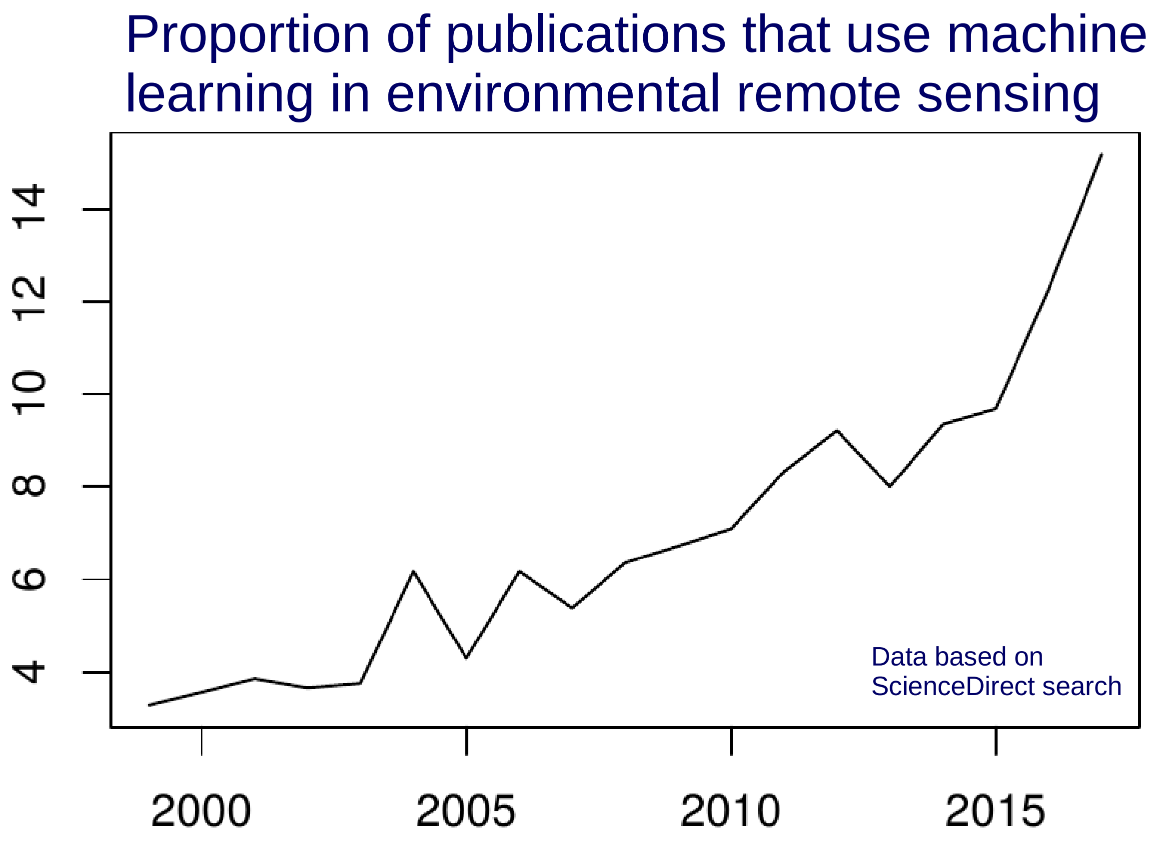
Predictive modelling of the environment: The machine learning way



Predictive modelling of the environment: The machine learning way

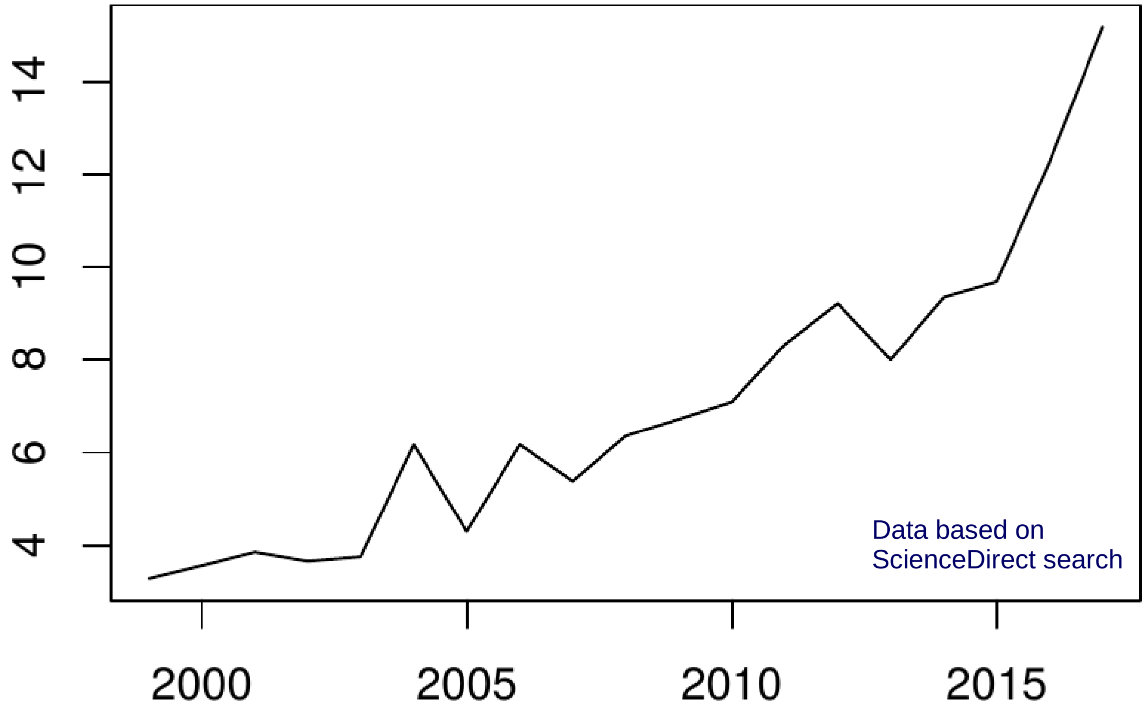


Global maps of ecosystem variables based on machine learning



Global maps of ecosystem variables based on machine learning

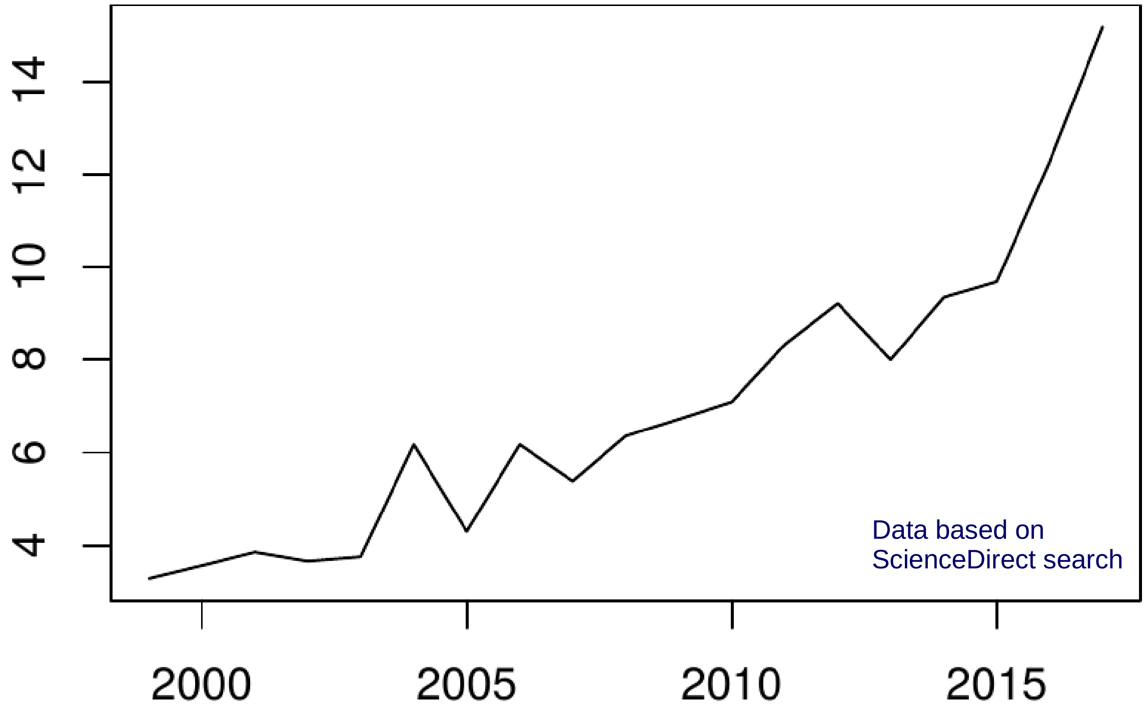
Proportion of publications that use machine learning in environmental remote sensing



Including **global** datasets on
- soil properties
- abundances of microorganisms
- Biodiversity
- tree restoration potential
- ...and many more

Global maps of ecosystem variables based on machine learning

Proportion of publications that use machine learning in environmental remote sensing



Including **global** datasets on

- soil properties
- abundances of microorganisms
- Biodiversity
- tree restoration potential
- ...and many more

Machine learning as a „magic tool“ to map basically everything ?

...but there are increasingly doubts about the methods

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 carbon-absorbing benefits of global forest restoration, but the authors insist their original estimates are accurate.

Oct 17, 2019

KATARINA ZIMMER

www.the-scientist.com

...but there are increasingly doubts about the methods

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 carbon-absorbing benefits of global forest restoration, but the authors insist their original estimates are accurate.

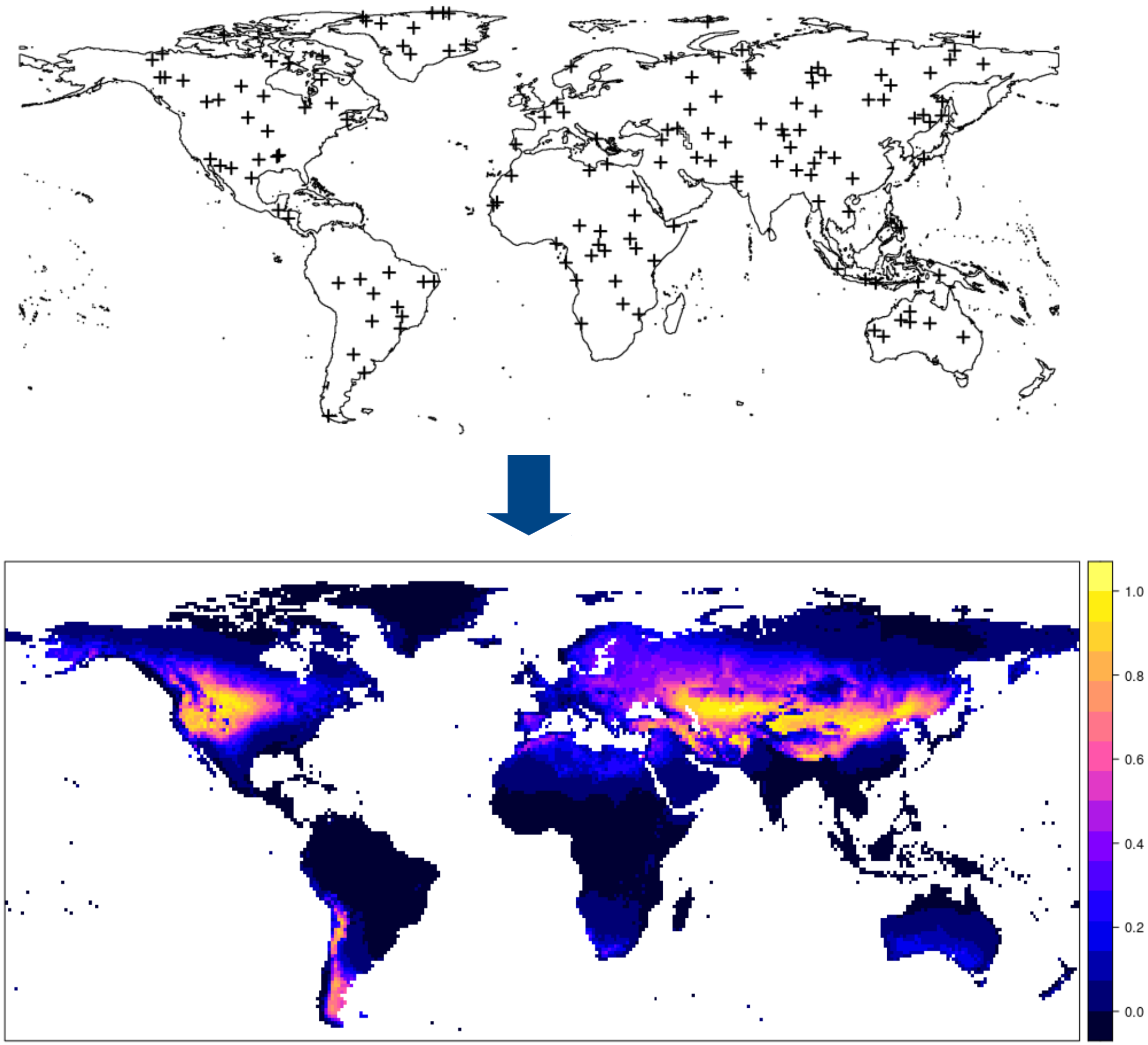
Oct 17, 2019

KATARINA ZIMMER

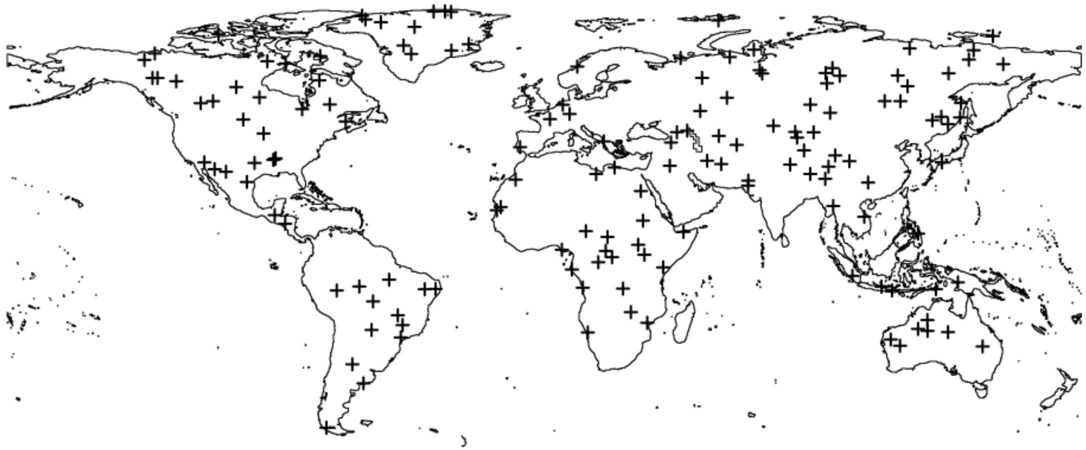
www.the-scientist.com

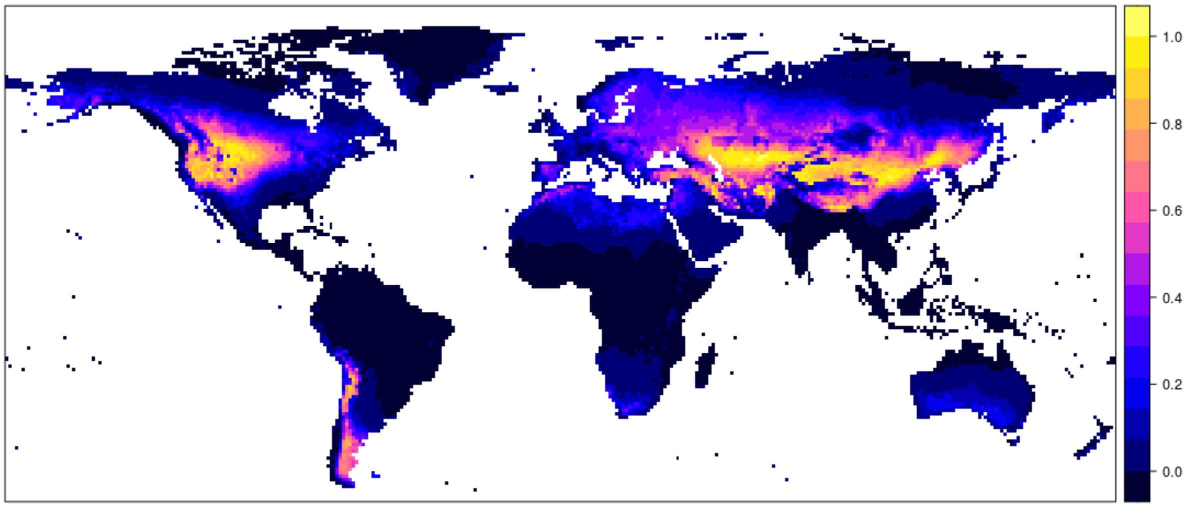
Have we been too ambitious? Why might the models fail?

Largest challenge: predictions far beyond training samples

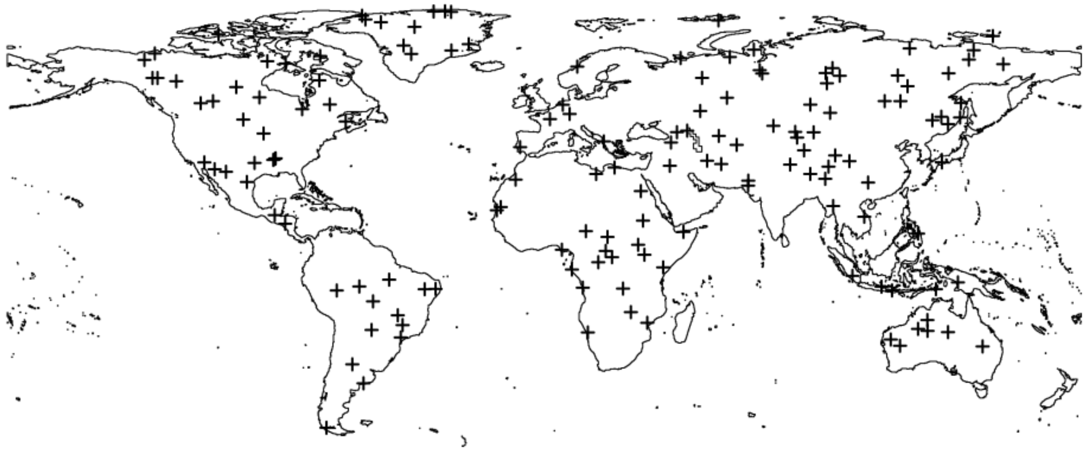


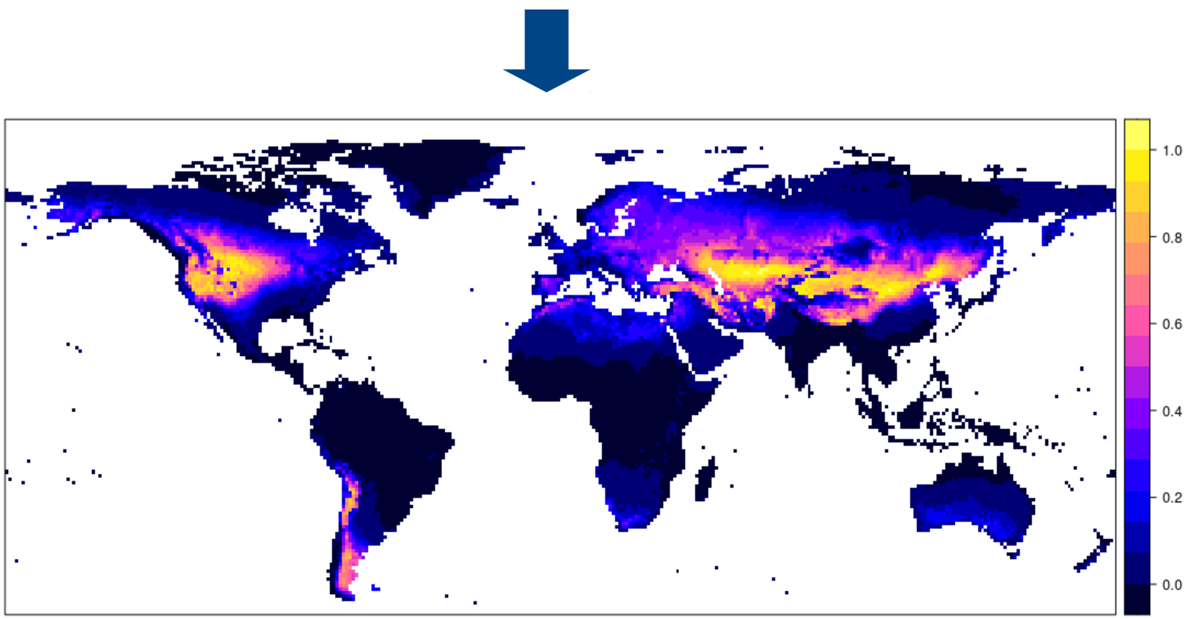
Largest challenge: predictions far beyond training samples





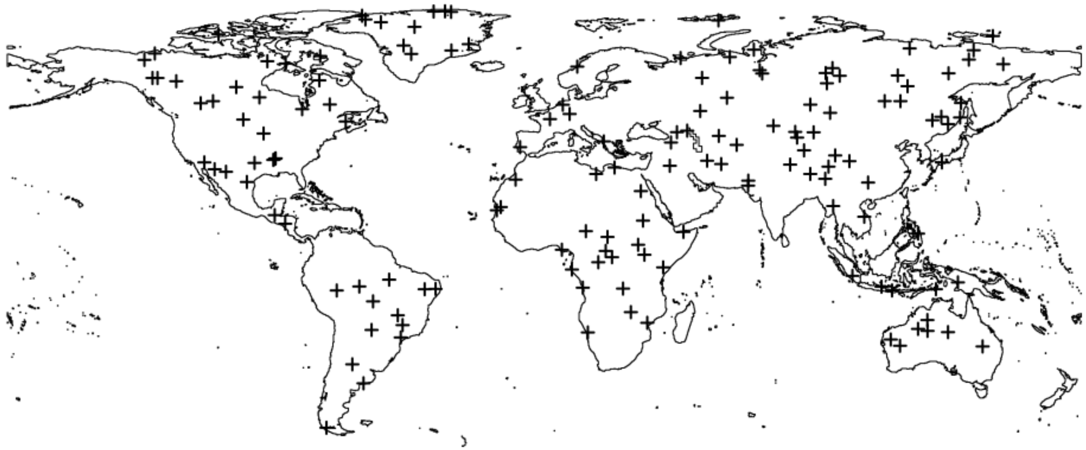
Largest challenge: predictions far beyond training samples

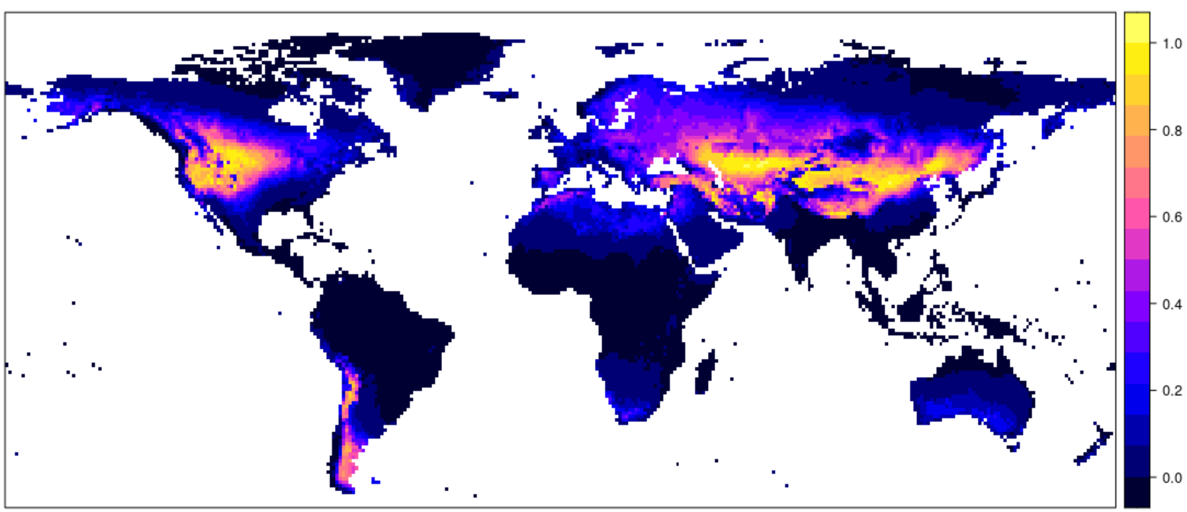




- Transfer to new space required

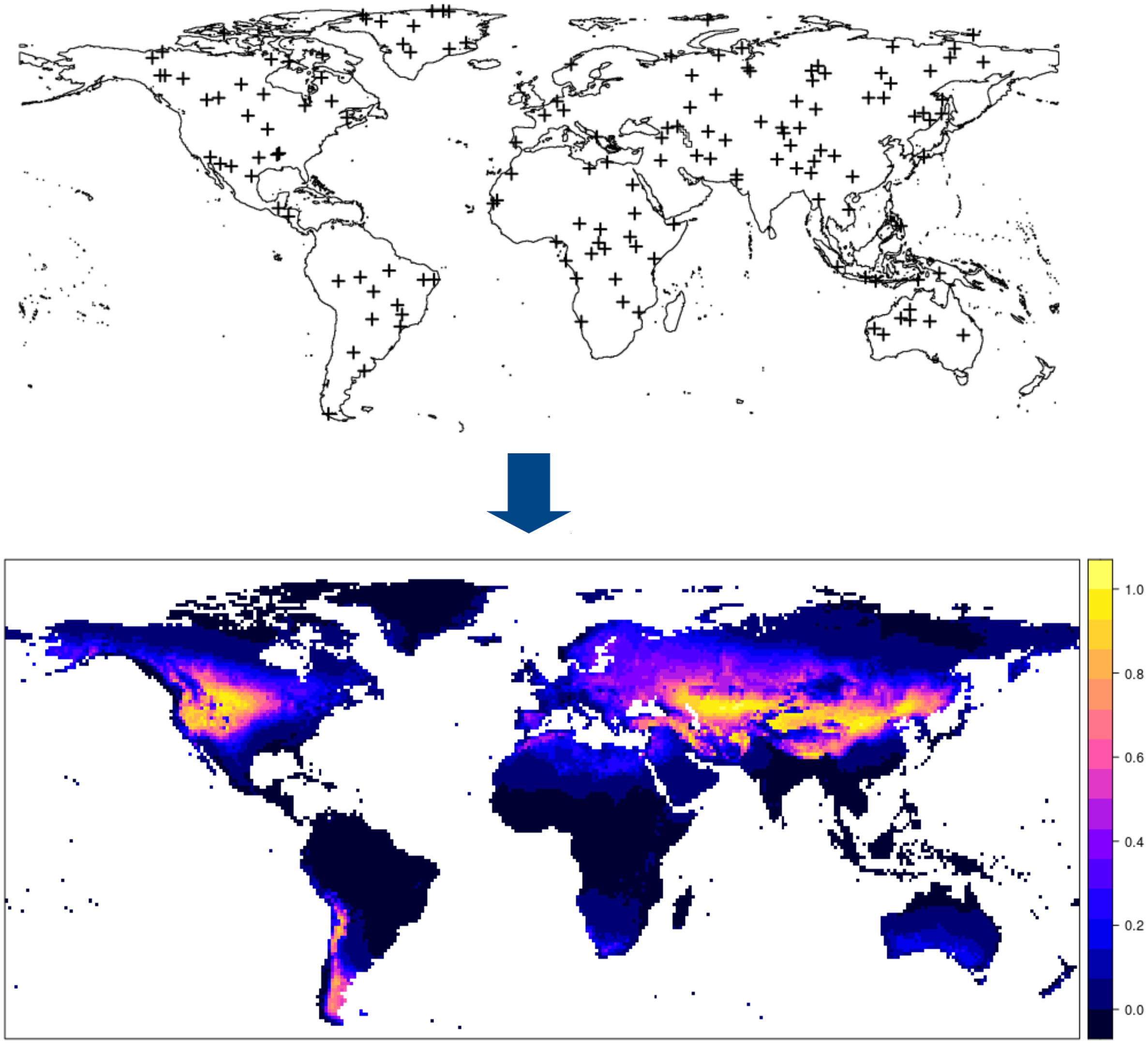
Largest challenge: predictions far beyond training samples





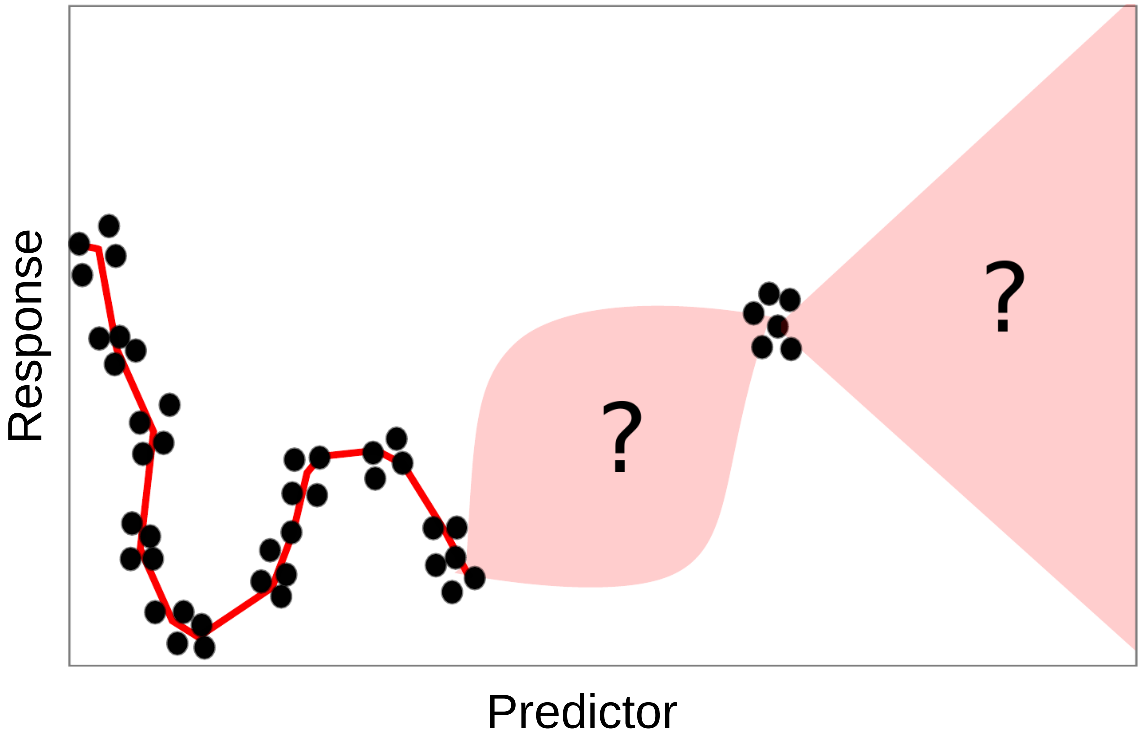
- Transfer to new space required
- New space might differ in environmental properties

Largest challenge: predictions far beyond training samples



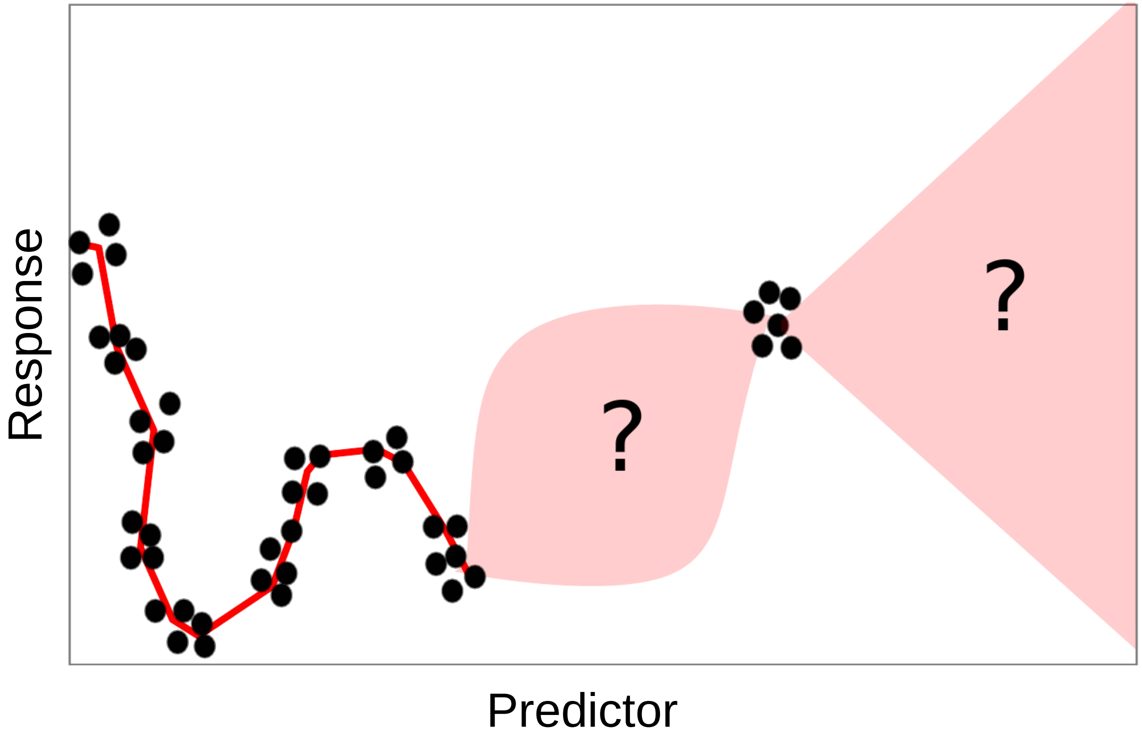
- Transfer to new space required
- New space might differ in environmental properties
- But what if the algorithm has never seen such properties?

Machine learning models are weak in extrapolations



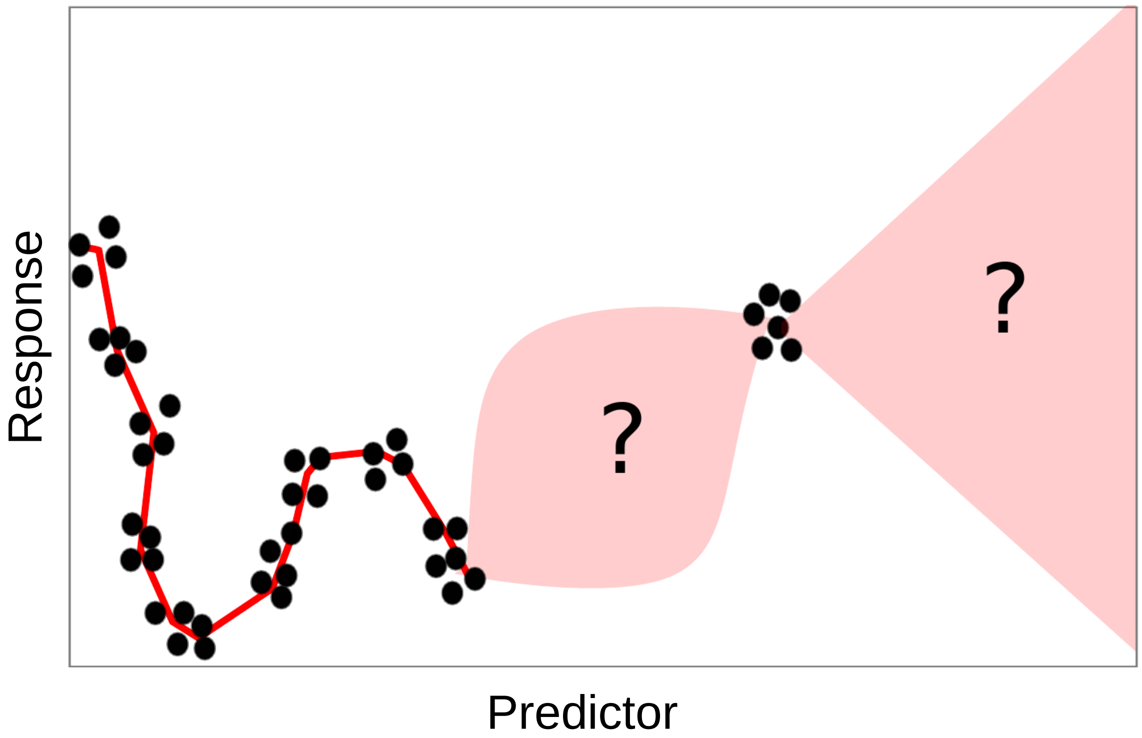
- Machine learning can fit very complex relationships.

Machine learning models are weak in extrapolations



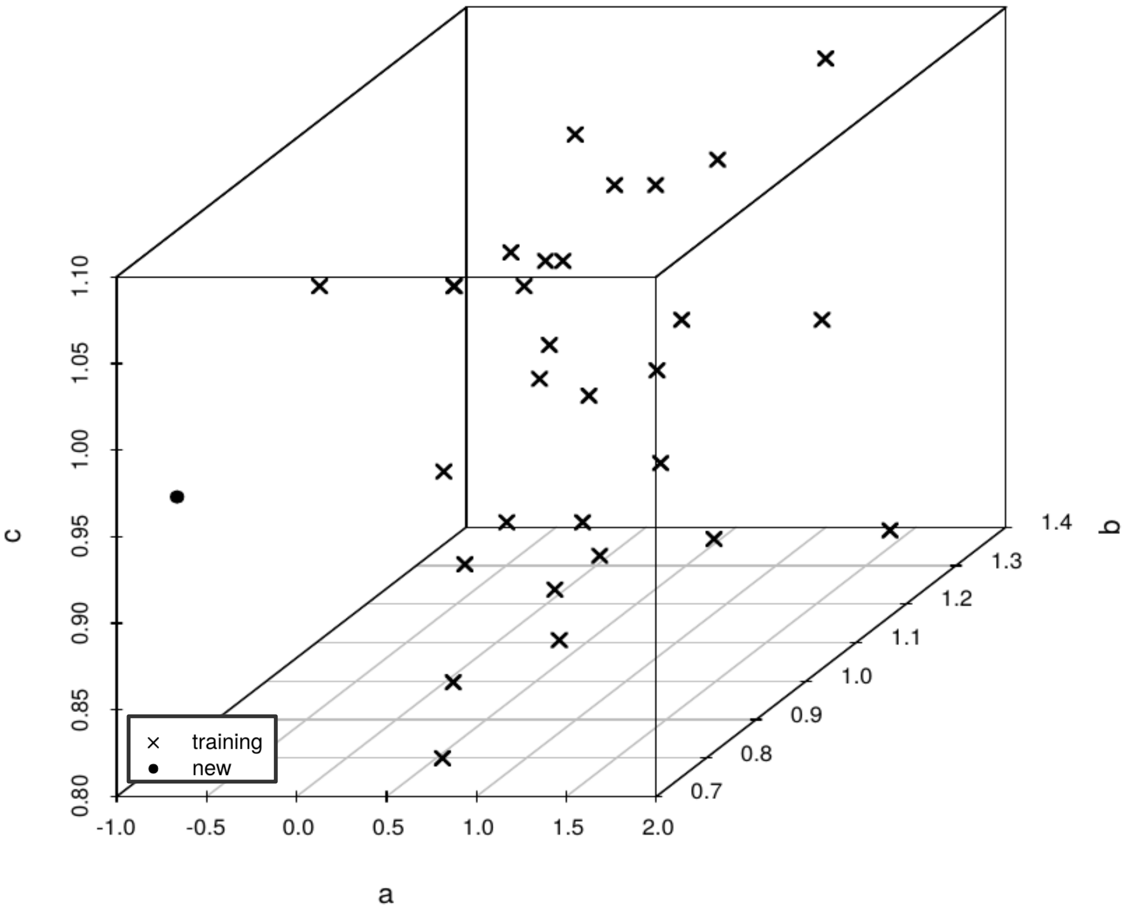
- Machine learning can fit very complex relationships.
- But gaps in predictor space are problematic (the model has no knowledge about these areas!)

Machine learning models are weak in extrapolations



- Machine learning can fit very complex relationships.
- But gaps in predictor space are problematic (the model has no knowledge about these areas!)
- A measure for “unknown space” is needed

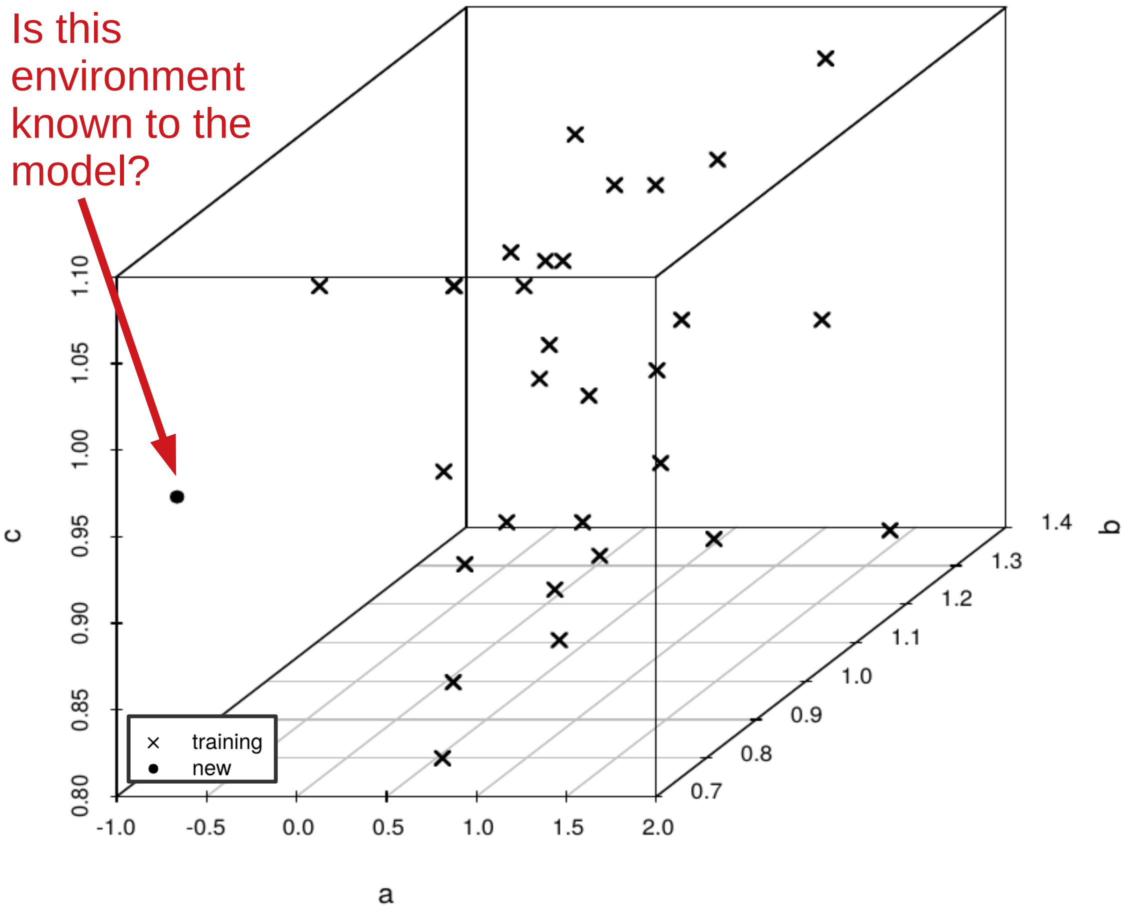
Distances in feature space as a measure for "unknown space"



- Unknown space: Environmental conditions that are very different from the training locations

*More details: https://arxiv.org/abs/2005.07939

Distances in feature space as a measure for “unknown space”

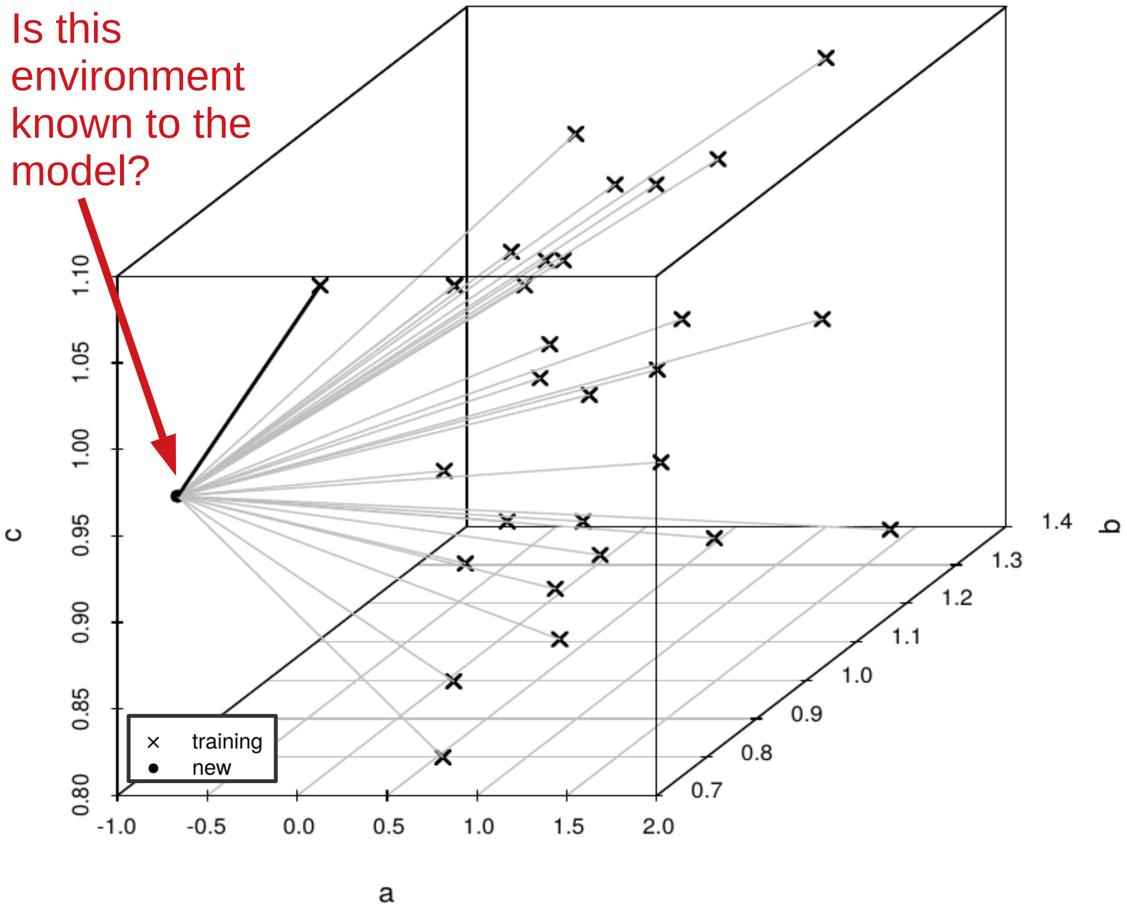


- Unknown space: Environmental conditions that are very different from the training locations

*More details: https://arxiv.org/abs/2005.07939

Distances in feature space as a measure for “unknown space”

Is this environment known to the model?

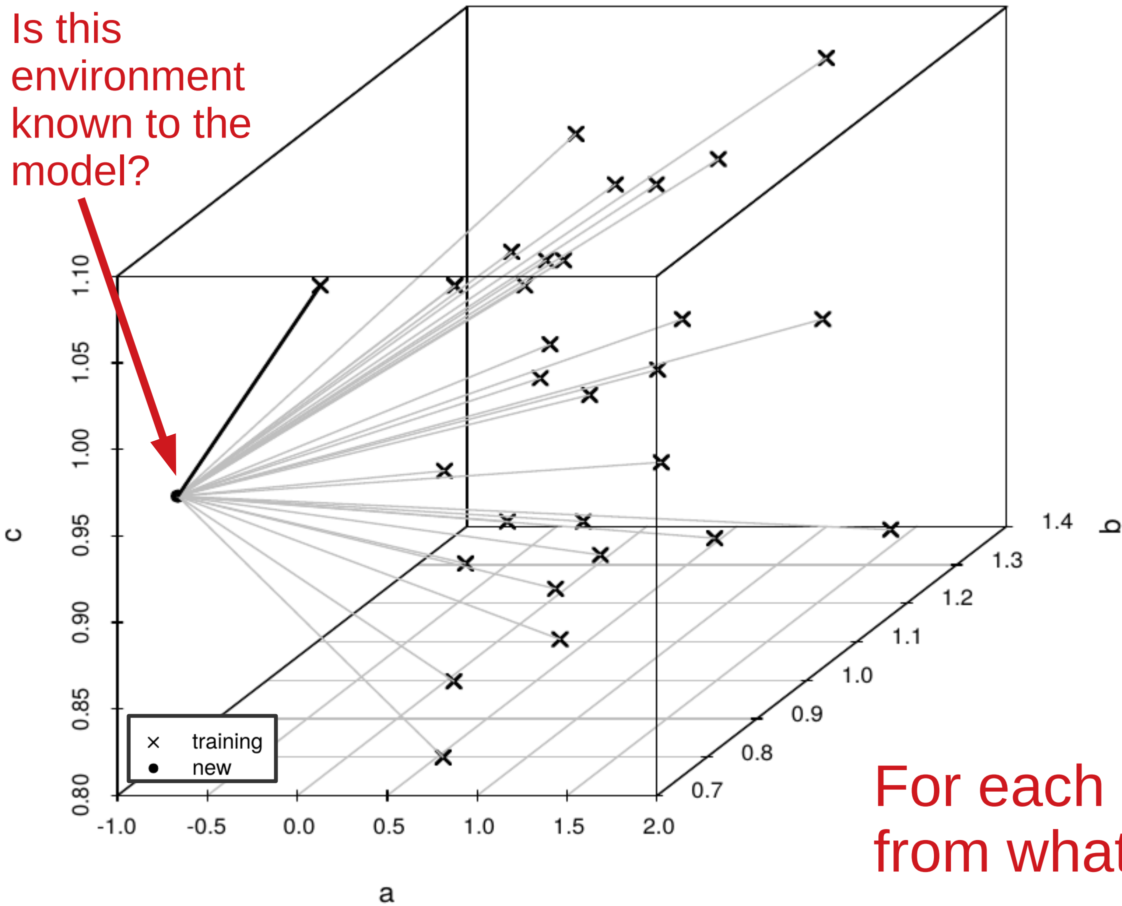


- Unknown space: Environmental conditions that are very different from the training locations
- Suggestion: Dissimilarity Index based on distances in the (weighted) predictor space*

*More details: https://arxiv.org/abs/2005.07939

Distances in feature space as a measure for “unknown space”

Is this environment known to the model?



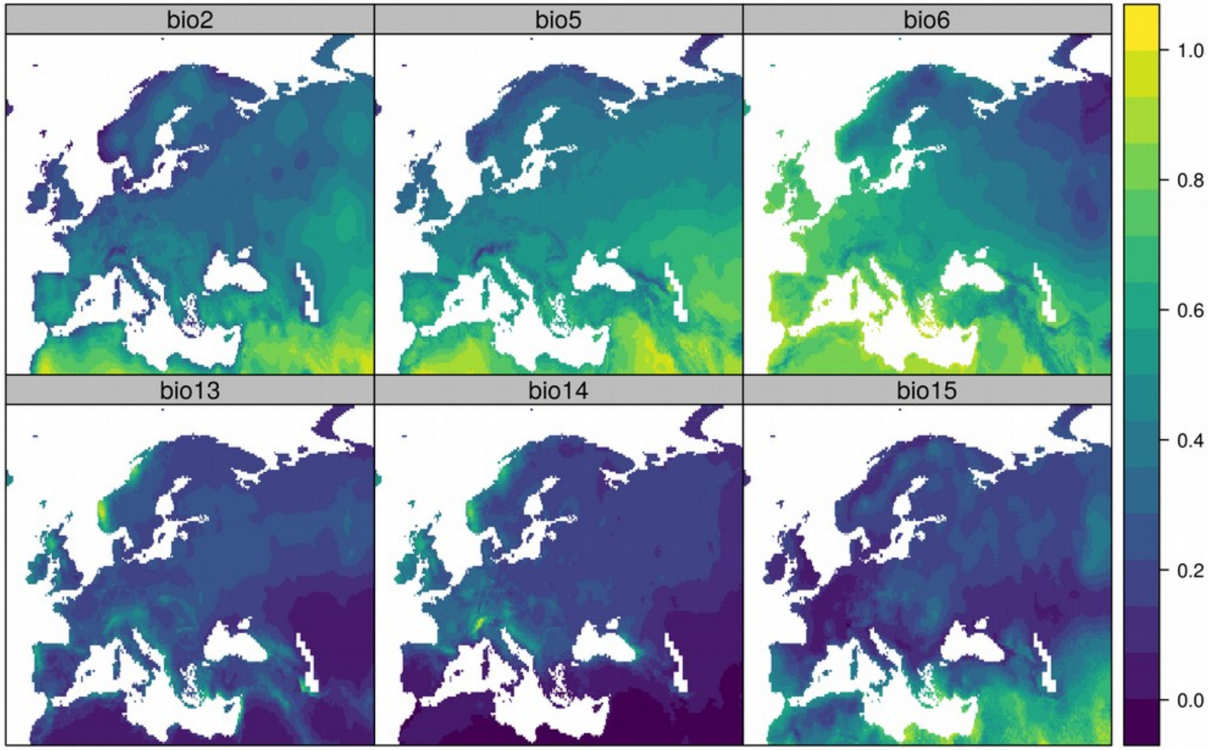
- Unknown space: Environmental conditions that are very different from the training locations
- Suggestion: Dissimilarity Index based on distances in the (weighted) predictor space*

For each new location/pixel: how distant is it from what the algorithm has seen?

*More details: https://arxiv.org/abs/2005.07939

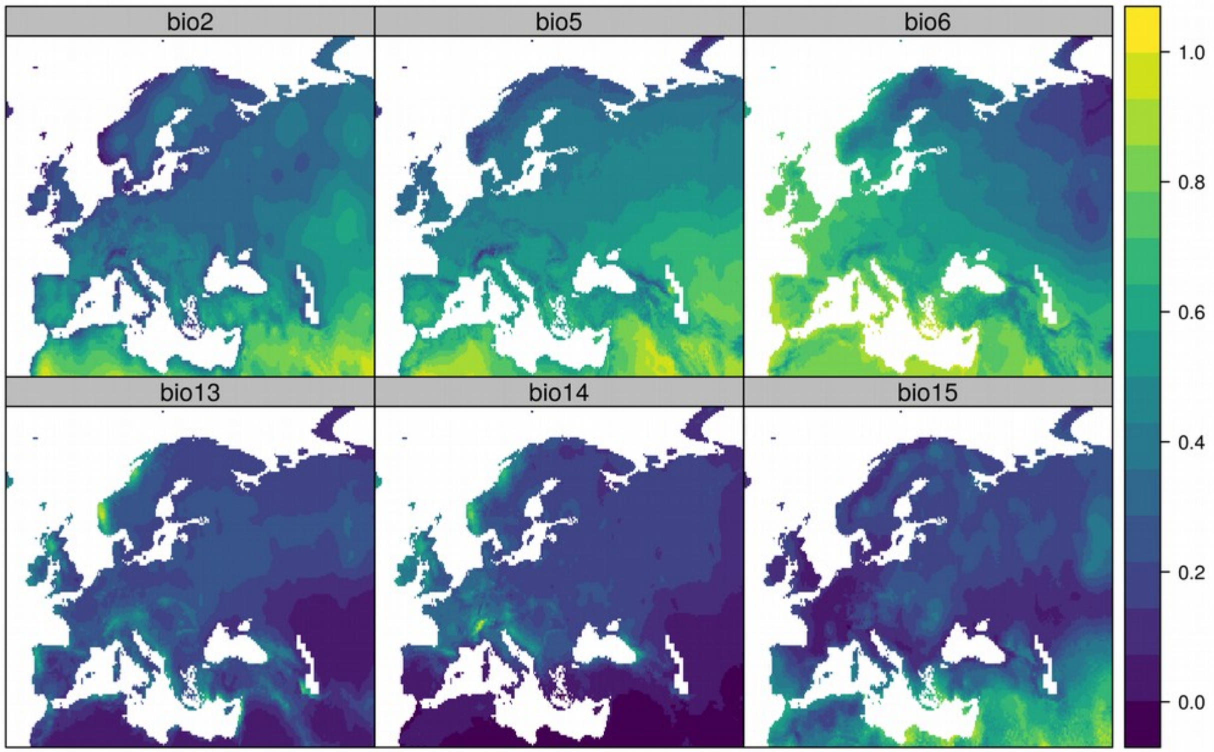
Mapping the area of applicability - Example

Predictors

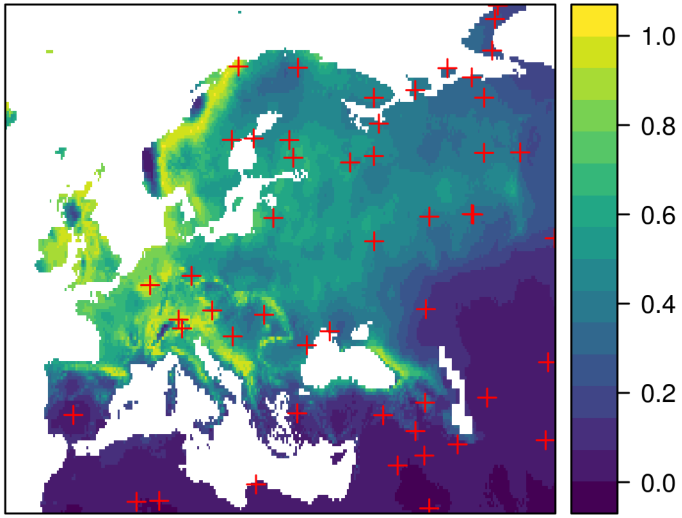


Mapping the area of applicability - Example

Predictors

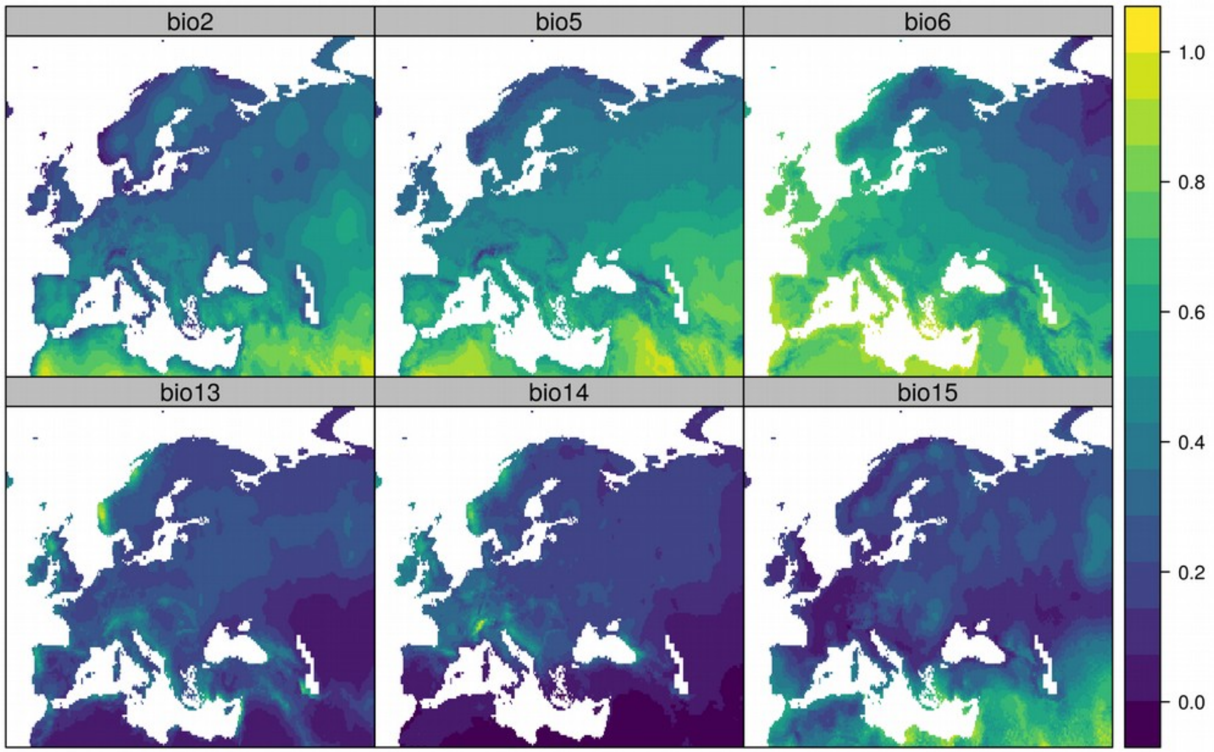


Virtual Response and simulated samples

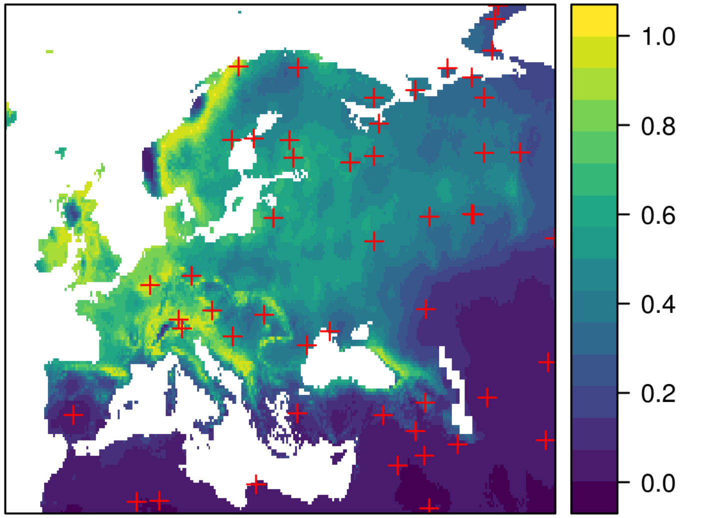


Mapping the area of applicability - Example

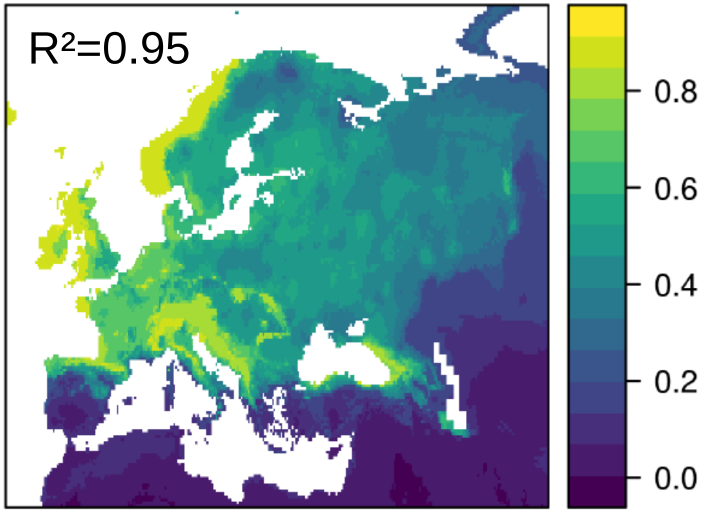
Predictors



Virtual Response and simulated samples

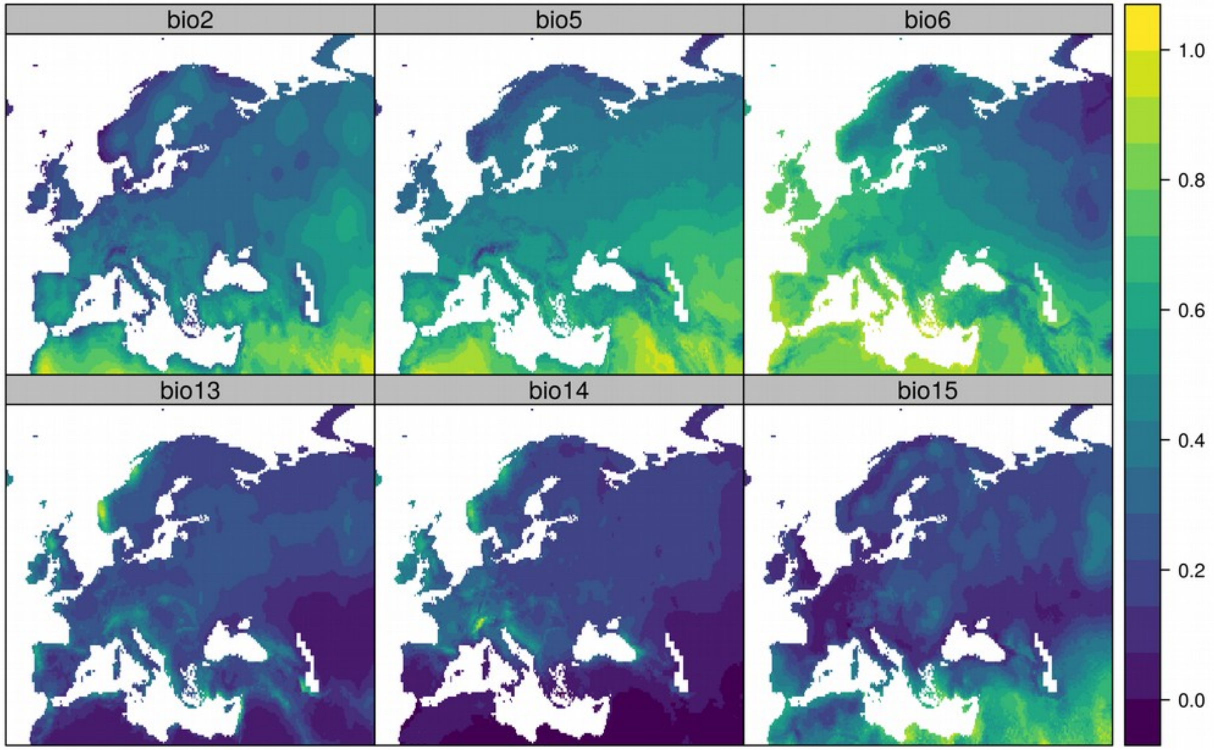


Prediction

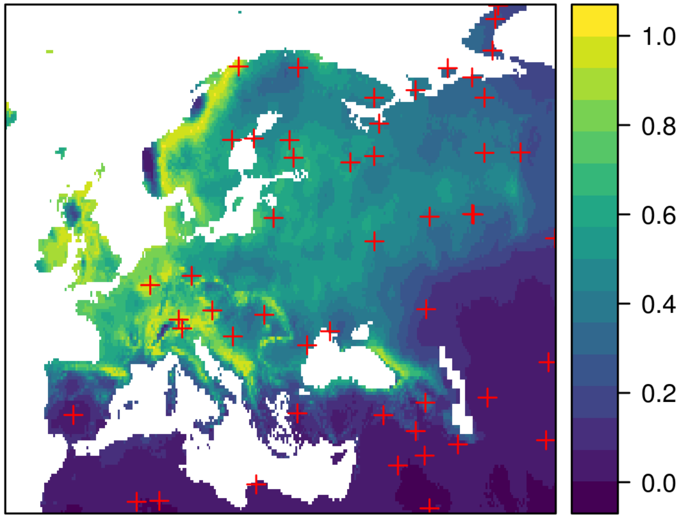


Mapping the area of applicability - Example

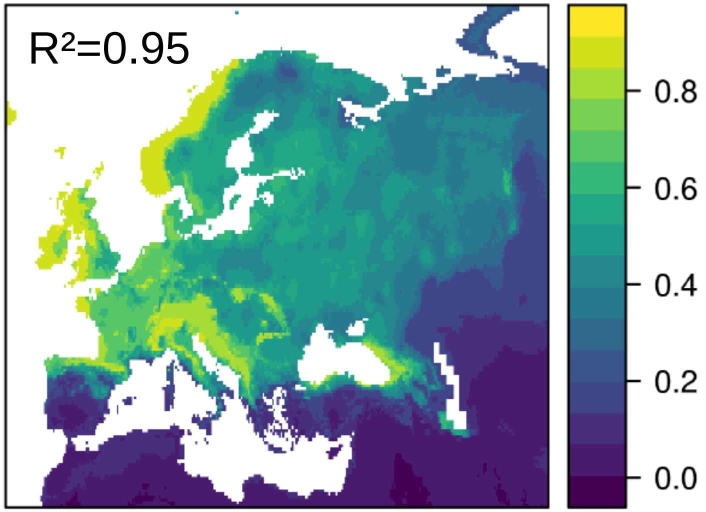
Predictors



Virtual Response and simulated samples

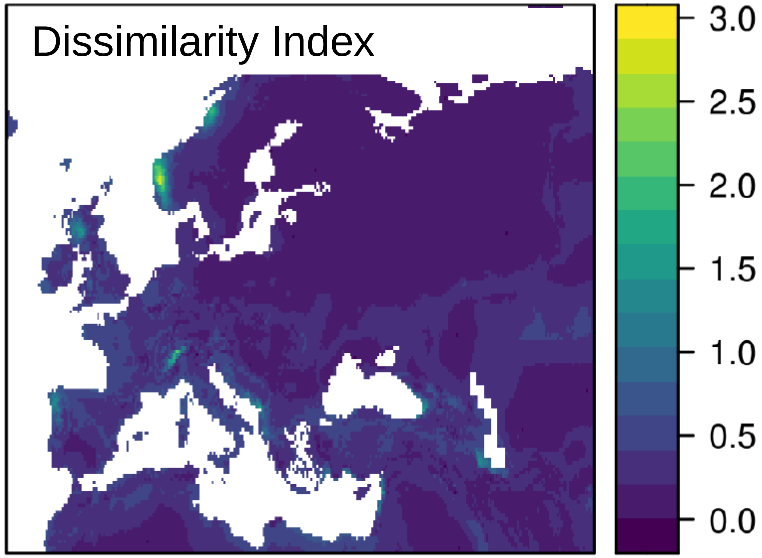


Prediction

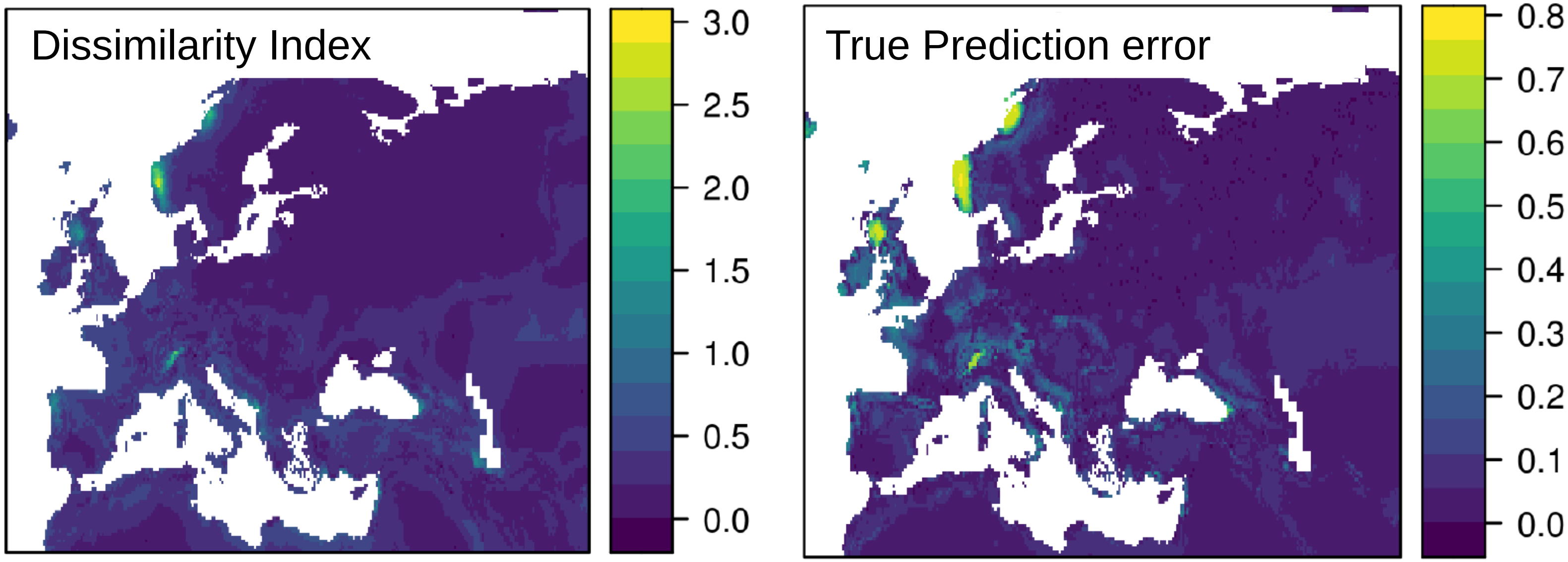


Where can we trust the predictions
and where should we better not?

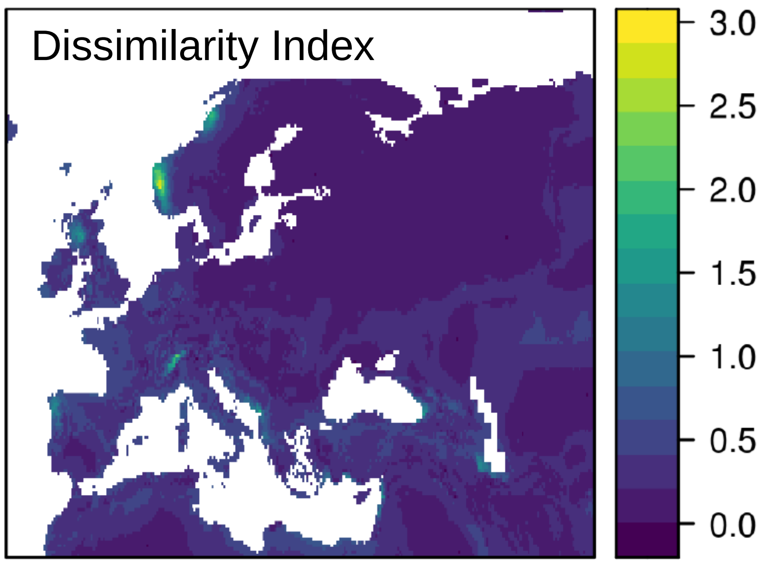
Mapping the area of applicability - Example

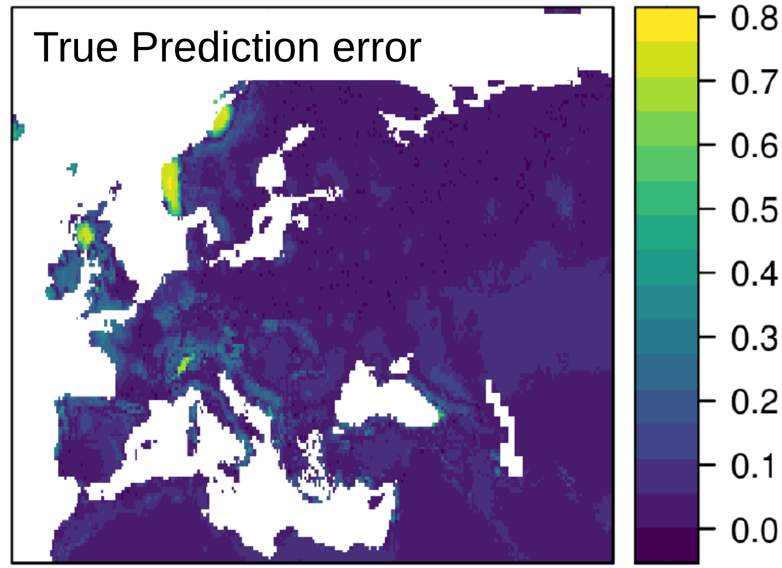


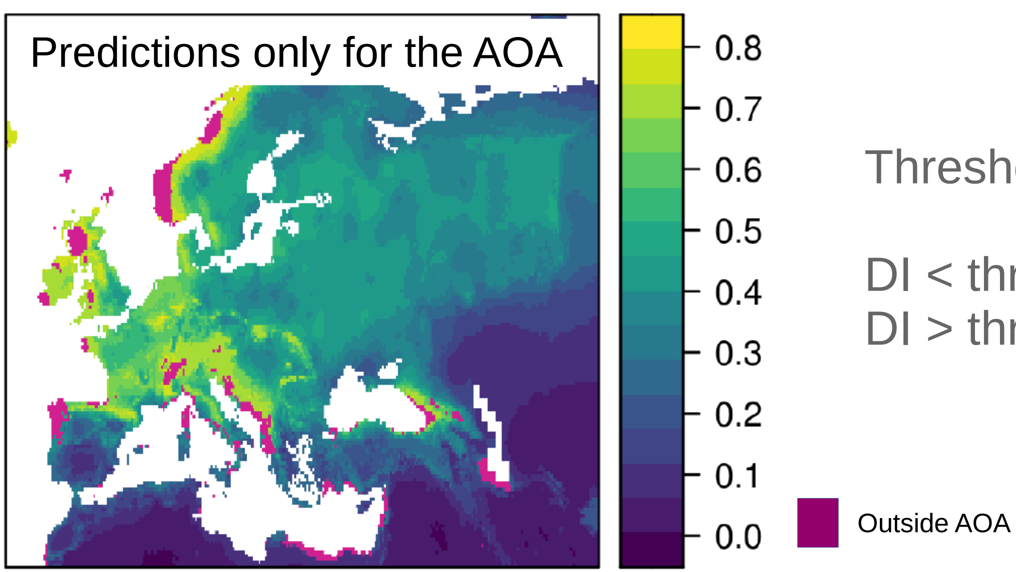
Mapping the area of applicability - Example



Mapping the area of applicability - Example







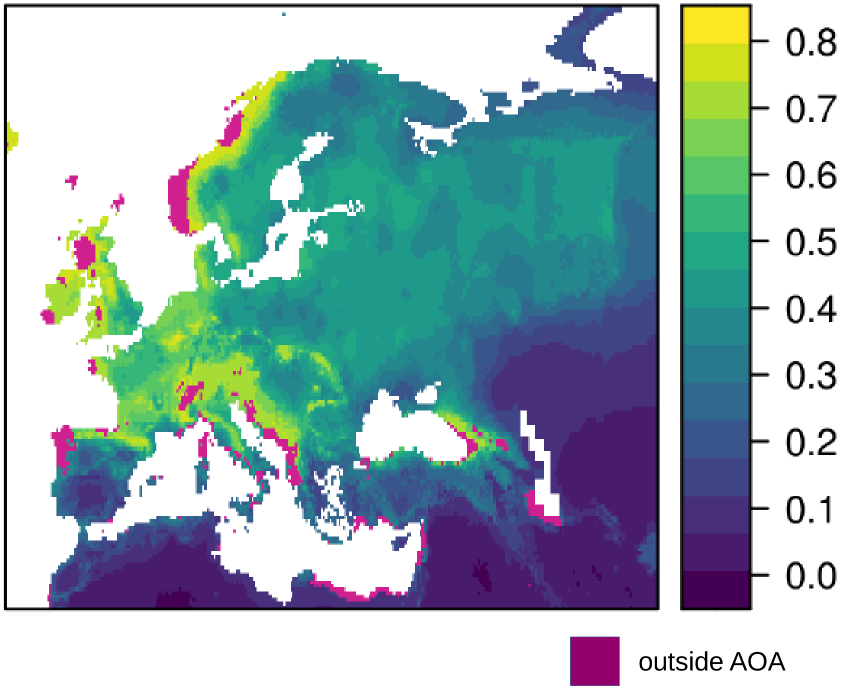
Threshold = DI of cross-validated training data

DI < threshold = inside AOA
DI > threshold = outside AOA

Why is it relevant to map the area of applicability?

Results are not just nice maps but used for…

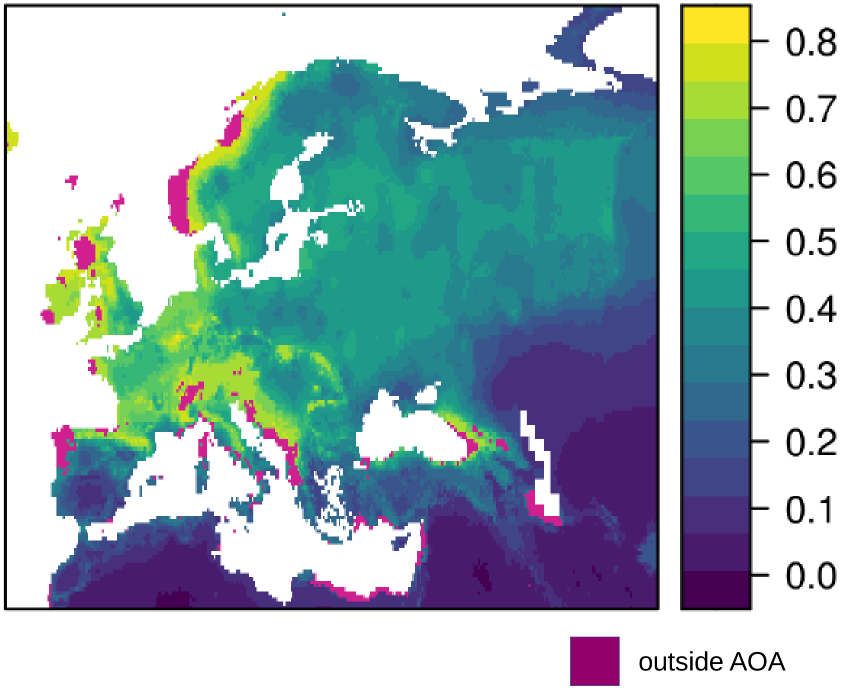
- subsequent modeling
- nature conservation
- risk assessment
- ...



Why is it relevant to map the area of applicability?

Results are not just nice maps but used for…

- subsequent modeling
- nature conservation
- risk assessment
- ...



Predictions should only be presented for the area of applicability to avoid error propagation or misplanning (and to keep trust in the methods)!

More information

- Meyer H, Pebesma E: Predicting into unknown space? Estimating the area of applicability of spatial prediction models. https://arxiv.org/abs/2005.07939
- Method implemented in the R Package “CAST”: https://CRAN.R-project.org/package=CAST
- Tutorial: https://github.com/HannaMeyer/OpenGeoHub_2020