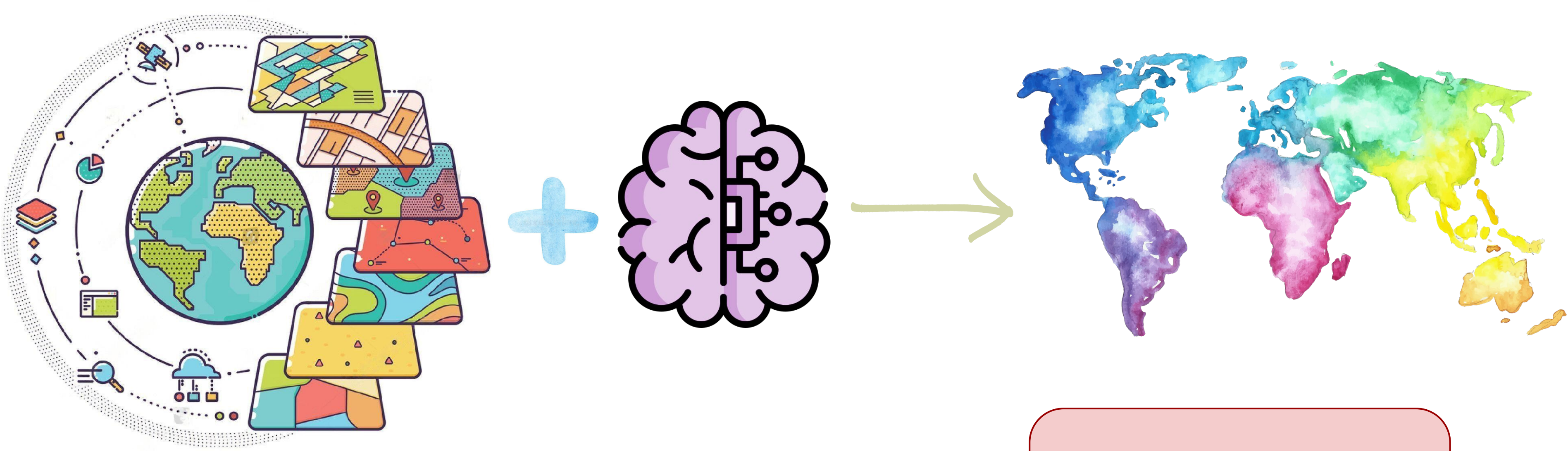
Area of applicability

Laura Martínez-Ferrer
(UV)

Vniversitat de València
IPL Laboratorio de Procesado de Imágenes
Image Signal Processing - ISP

Intro/Motivation



Spatial distribution of measurements remains sparse.

Area Of Applicability

Received: 16 May 2020 | Accepted: 6 May 2021
DOI: 10.1111/2041-210X.13650
RESEARCH ARTICLE
Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

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

Hanna Meyer[1] | Edzer Pebesma[2]

1: Standardization of predictor variables

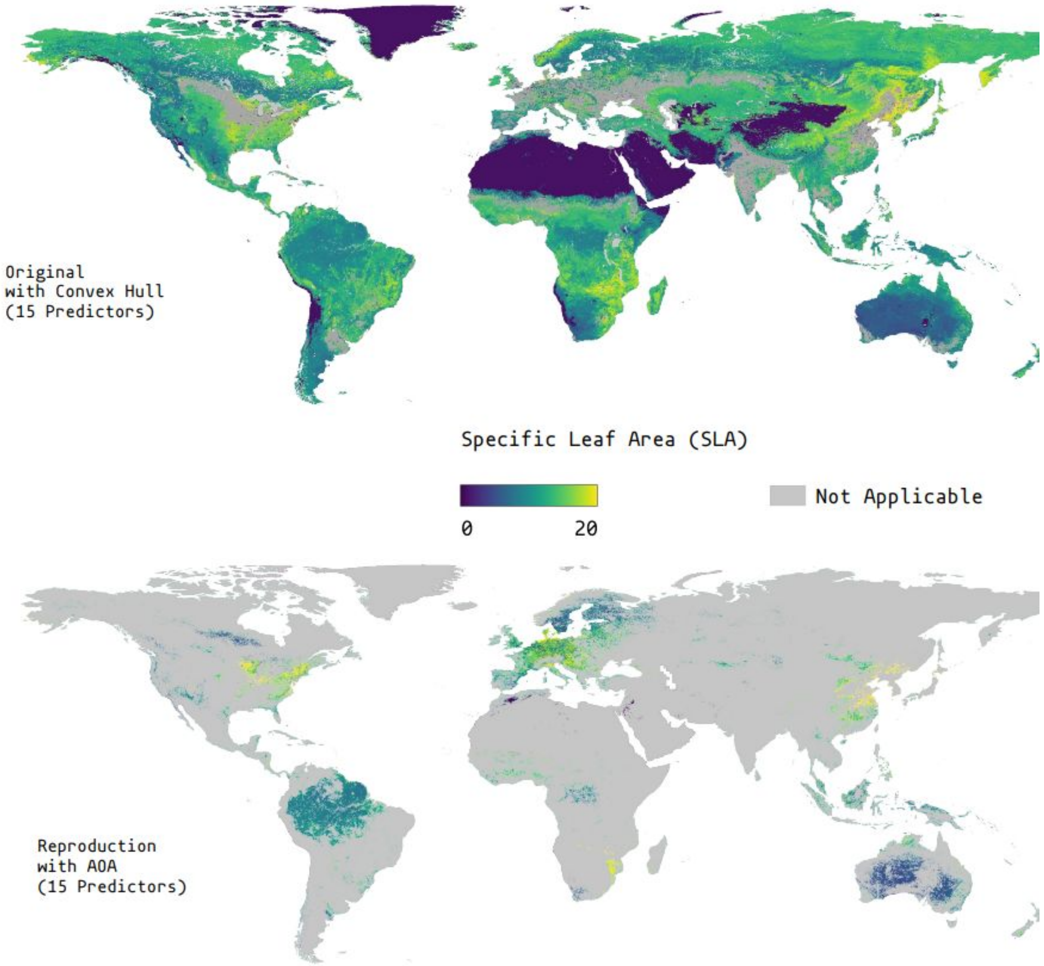
2: Weighting of variables $X_j^{sw} = w_j * X_j^s$

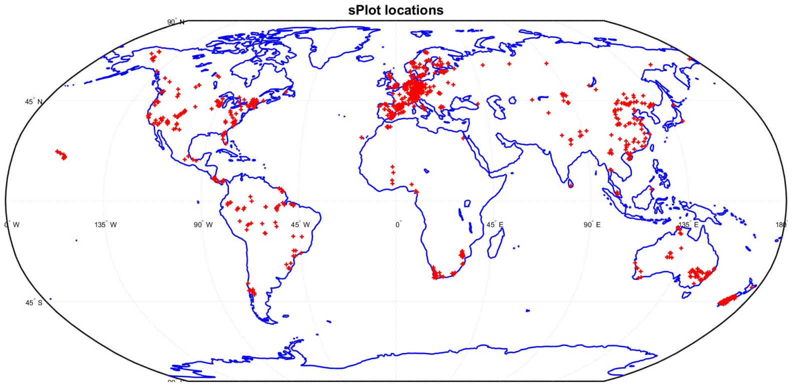
3: Distance from new prediction location k and training dataset $d(k,i) = \sqrt{\sum_{j=1}^{p}(X_{k,j}^{sw} - X_{i,j}^{sw})^2}$

4: Nearest training point $d_k = \arg_i \min d(k,i)$

5: Dissimilarity index $DI_k = d_k/\bar{d}$

AOA


Original
with Convex Hull
(15 Predictors)
Specific Leaf Area (SLA)
0
20
Not Applicable
Reproduction
with AOA
(15 Predictors)


sPlot locations
90° N
45° N
135° W
90° W
45° W
0°
45° E
90° E
135° E
45° S

AOA

1: Standardization of predictor variables

2: Weighting of variables $X_j^{sw} = w_j * X_j^s$

3: Distance from new prediction location k and training dataset $d(k,i) = \sqrt{\sum_{j=1}^{p}(X_{k,j}^{sw} - X_{i,j}^{sw})^2}$

4: Nearest training point $d_k = \arg_i \min d(k,i)$

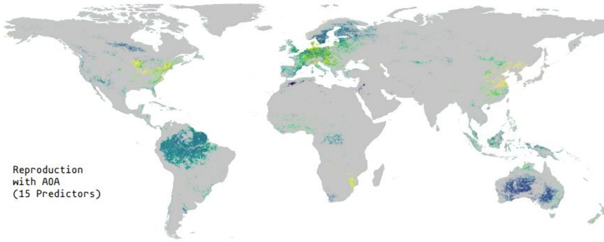
5: Dissimilarity index $DI_k = d_k / \bar{d}$

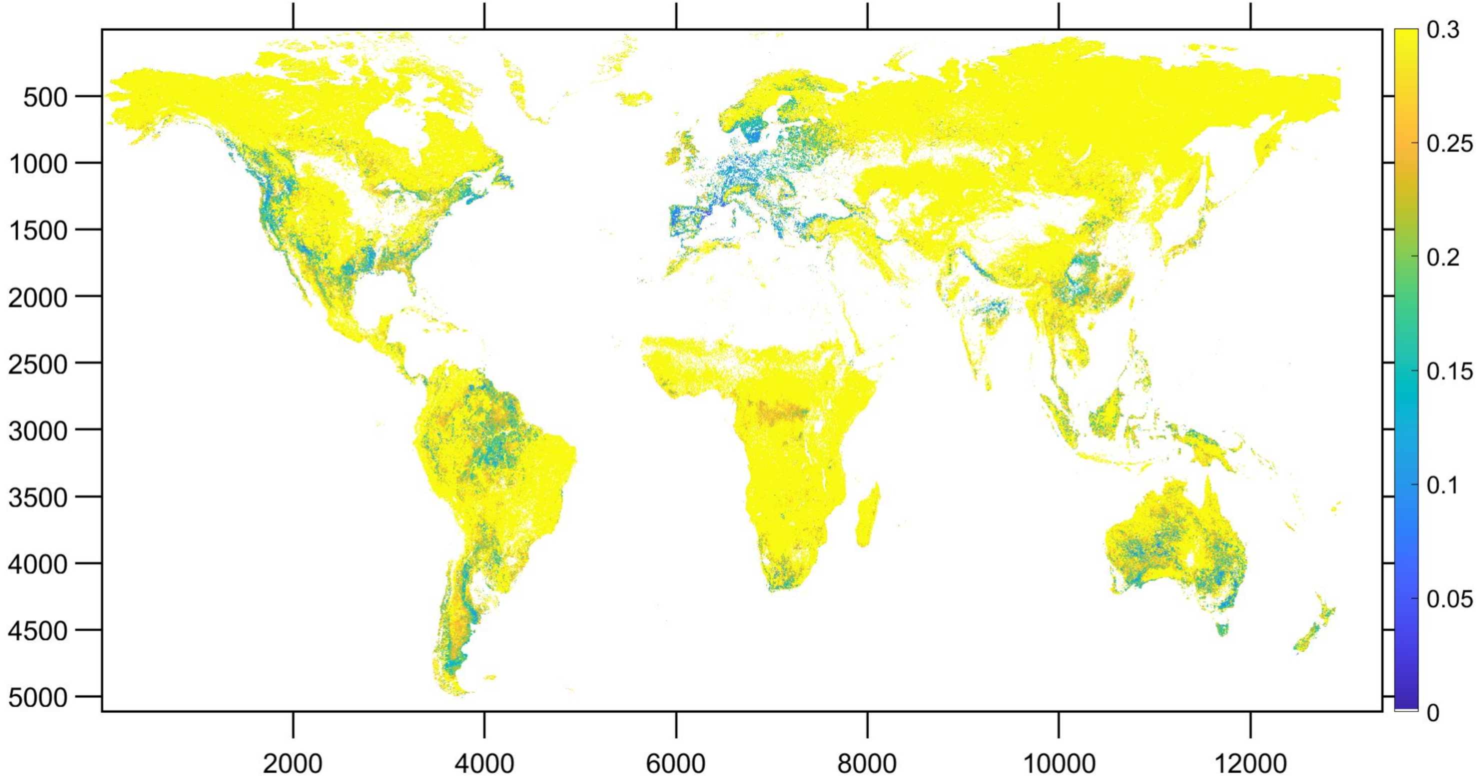
Kernels

Jacobians

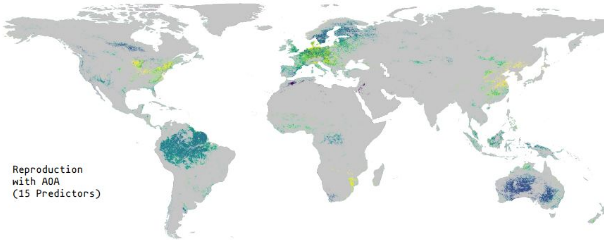
Latent space

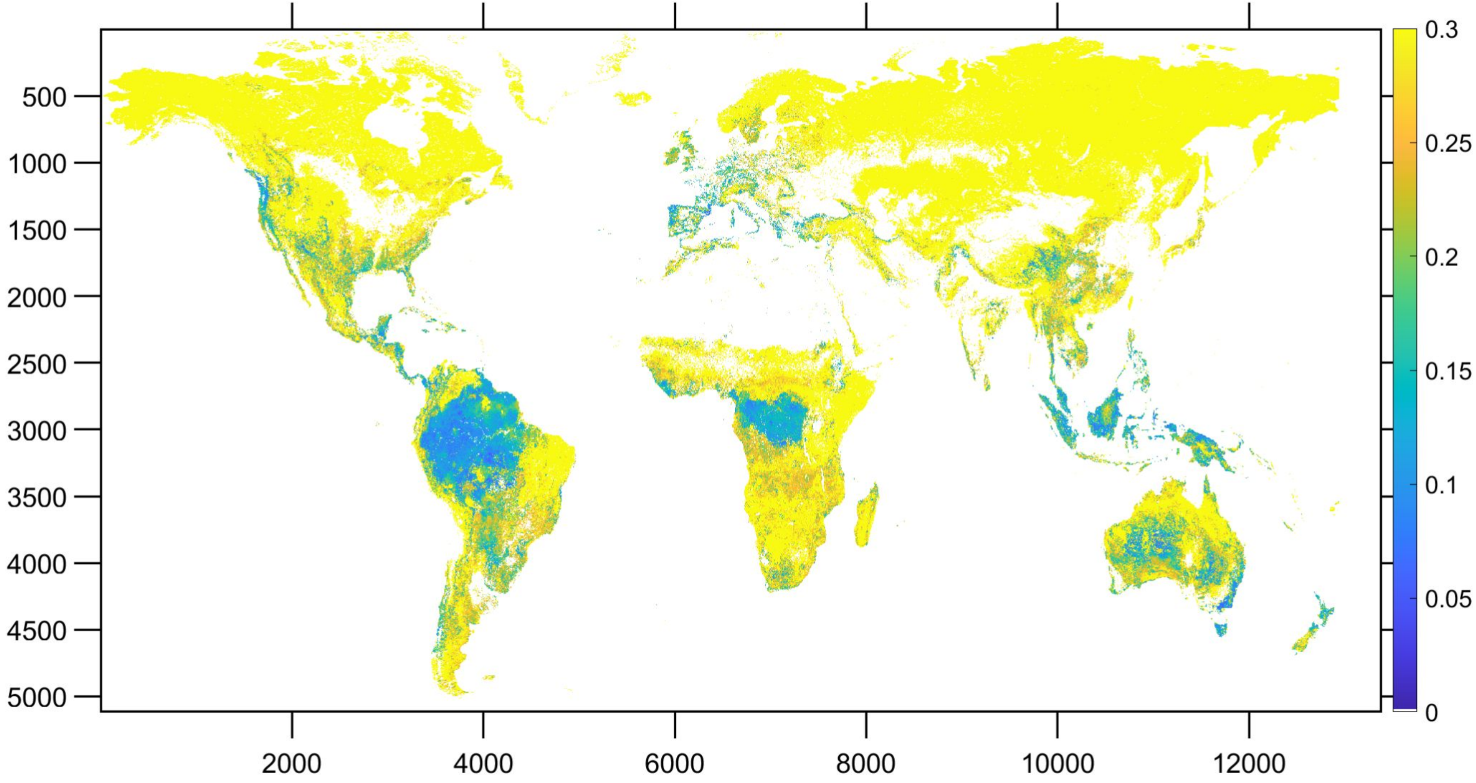
NN Jacobian


Reproduction
with AOA
(15 Predictors)


500
1000
1500
2000
2500
3000
3500
4000
4500
5000
2000
4000
6000
8000
10000
12000
0.3
0.25
0.2
0.15
0.1
0.05
0

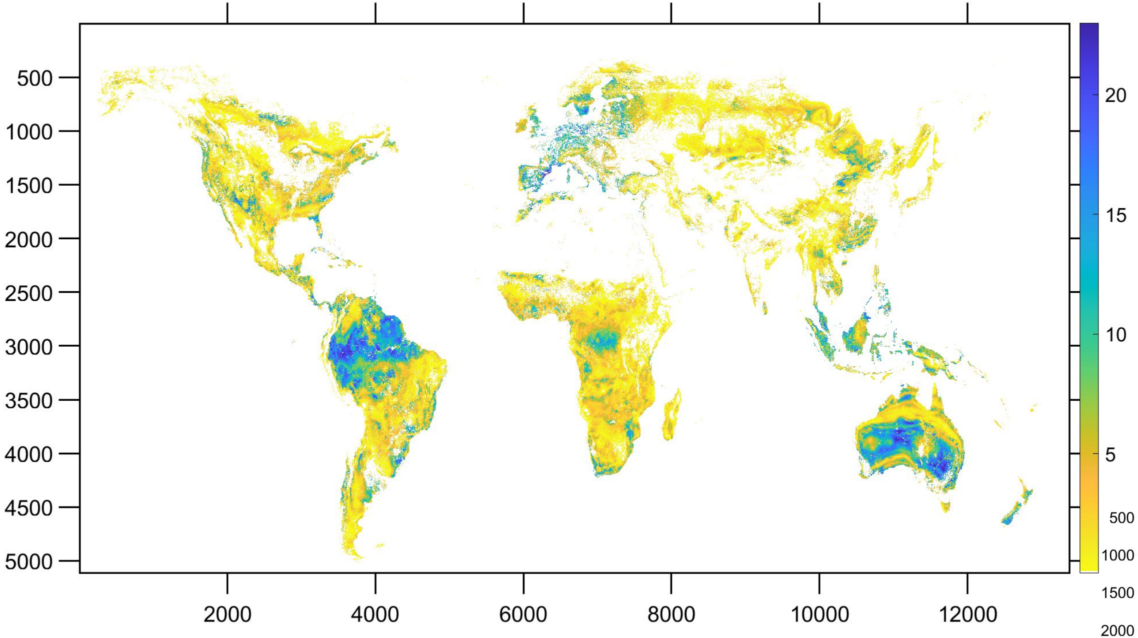
NN Latent Space

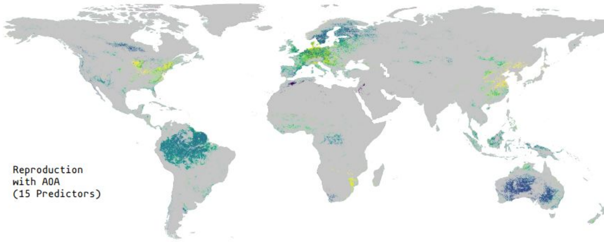




GPs

GP-Dissimilarity Index

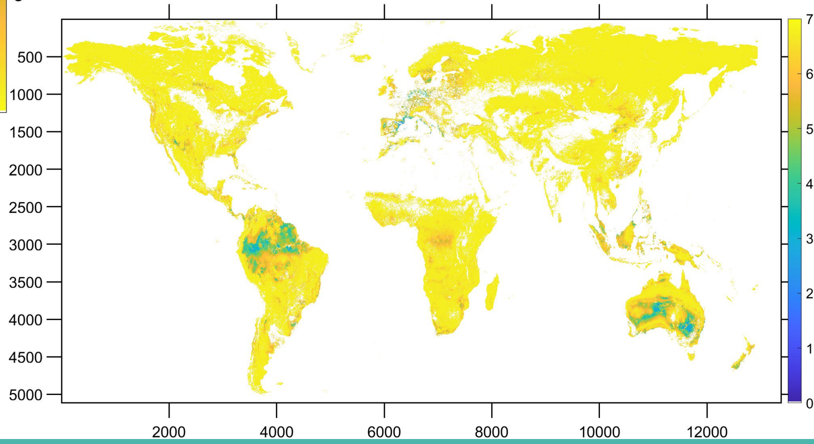




ARD Kernel

$$k(x_i, x_j|\theta) = \sigma_f^2 \exp\left[-\frac{1}{2}\sum_{m=1}^{d}\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}\right]$$

GP variance



Conclusions/open questions

★ Where is the advantage if the AOA coincides with the locations of the training data?
★ Meyer proposed to find the predictors combination that maximizes the AOA
★ Different ML methods/approaches lead to similar results