

Modelling urban bird breeding ranges

with remotely sensed heterogeneity in plant traits using a random forest

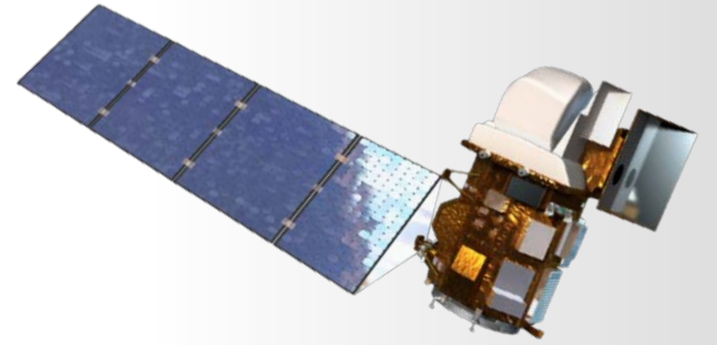
Thilo Wellmann

Thilo.Wellmann@hu-berlin.de
UFZ Leipzig & HU-Berlin
Computational Landscape Ecology

et al.

Dagmar Haase, UFZ & HU-Berlin
Sebastian Scheuer, HU-Berlin
Angela Lausch, UFZ Leipzig

Goal: Transferable and repeatable bird breeding range evaluation



© NASA



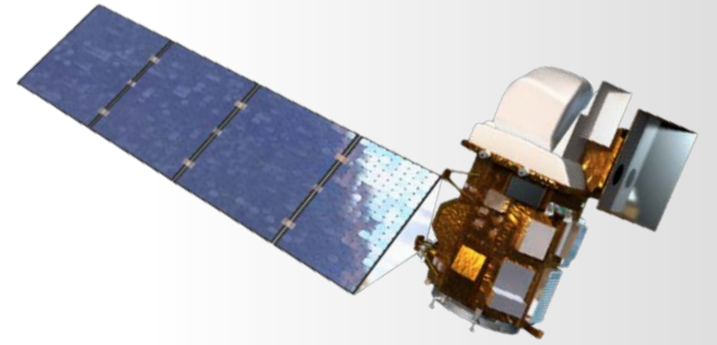
SubGoal1: Identifying potential data sources and indicators

SubGoal2: Find best suited modelling technique.

© Günther Gailberger



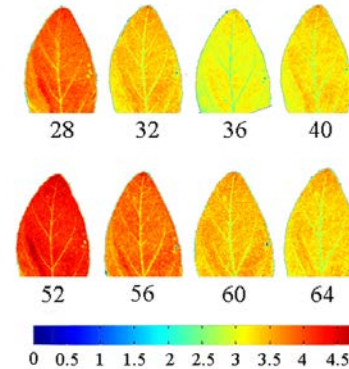
Hypothesis: Variations in **plant traits** are indicative for bird habitat



© NASA

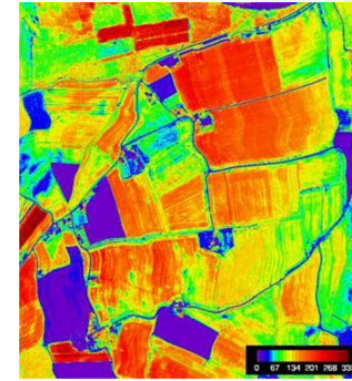
- **Traits** are anatomical, morphological, biochemical, physiological or phenological features and characteristics of individuals, species & populations. (Kattge et al. 2011)
- **Spectral traits** can be recorded using **remote sensing** based techniques.
 - e.g. Chlorophyll content

Individuum



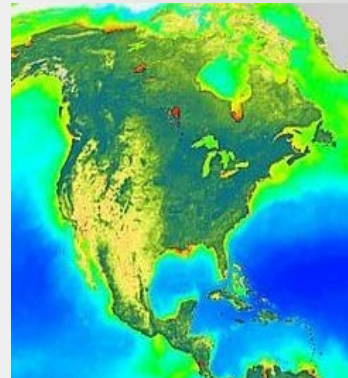
Wen-Juan Pan et al. 2015

Landscape



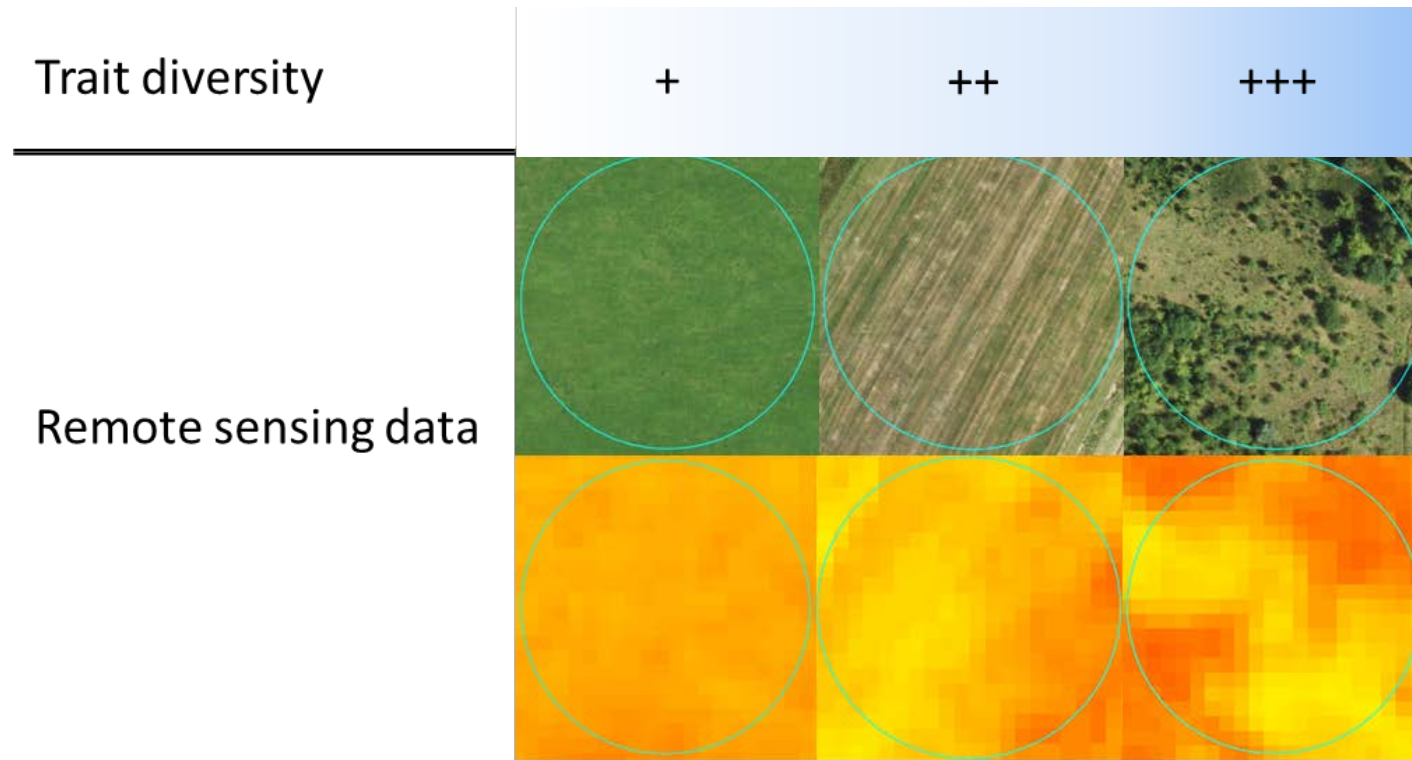
Delegido et al. 2011

Continental



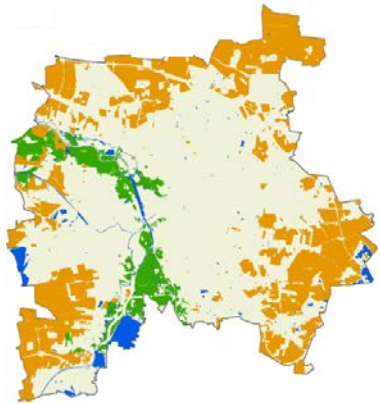
NASA

Research hypothesis



Variations in Spectral traits are indicative for human induced processes and their intensity

Input data



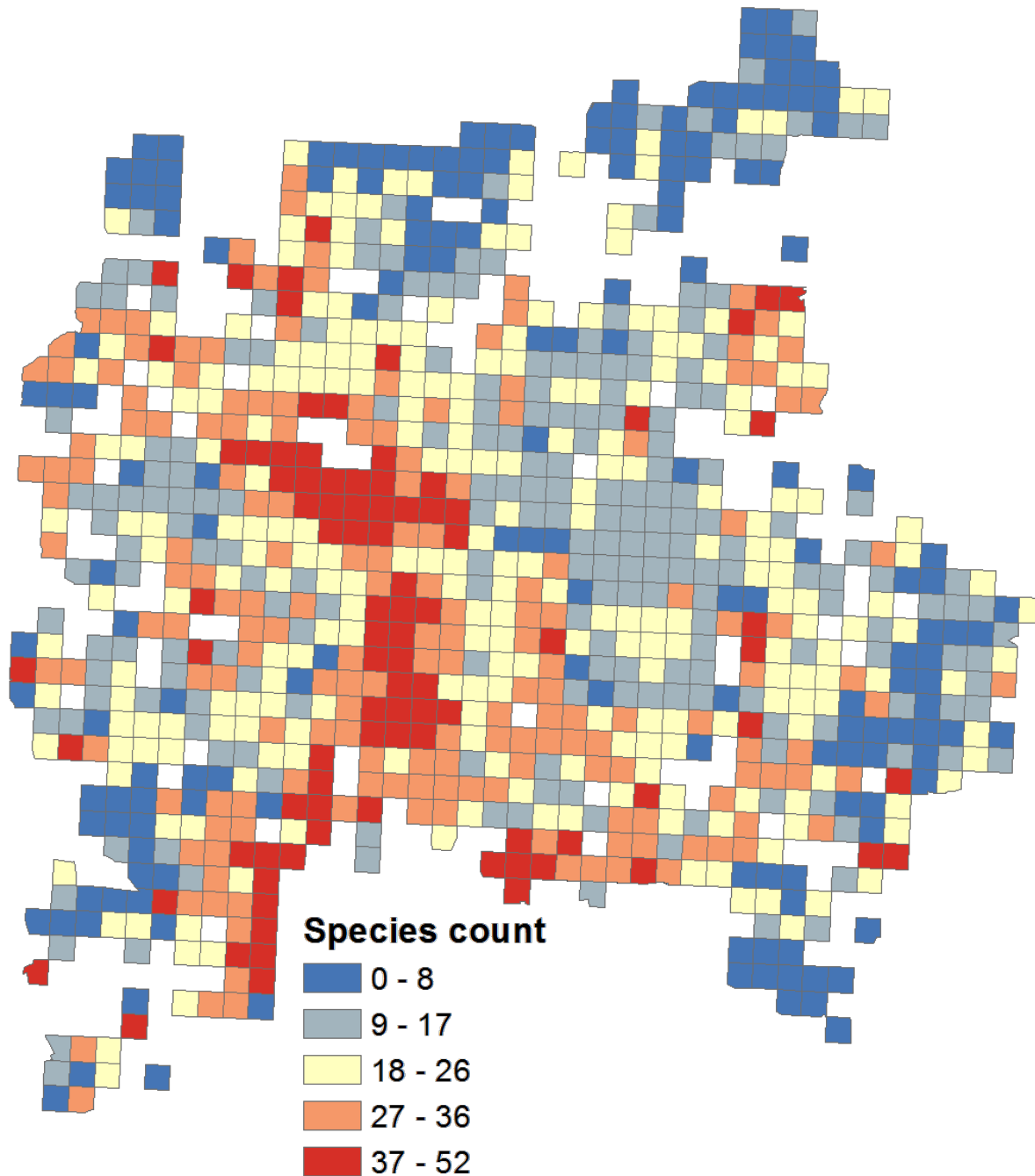
- Fields
- Alluvial Forest
- Waterbodies



RapidEye image

- Atmospherically corrected
- 6m ground resolution
- 5 spectral bands

Training data

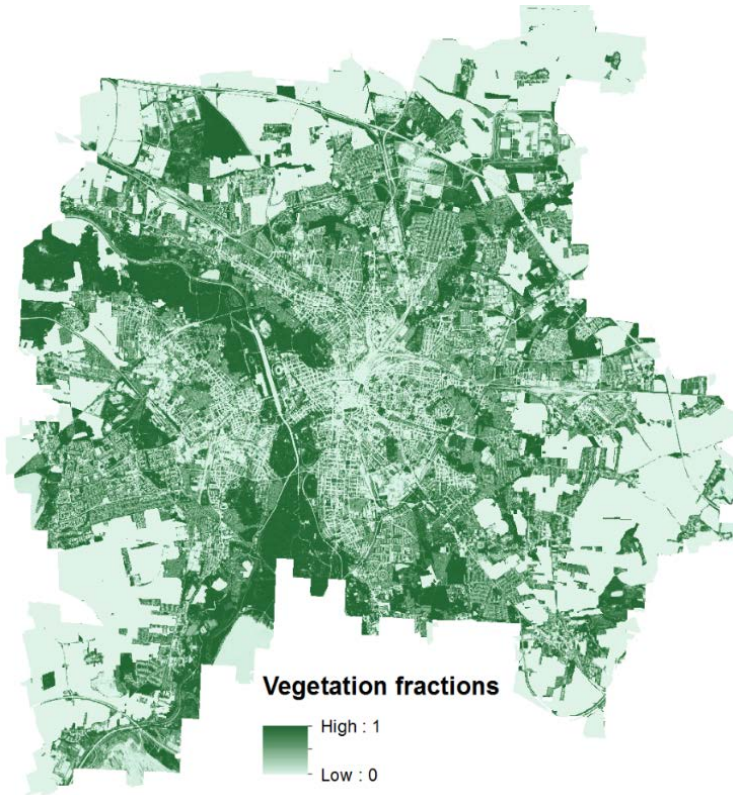


Wellmann et al. in preparation

Bird breeding dataset

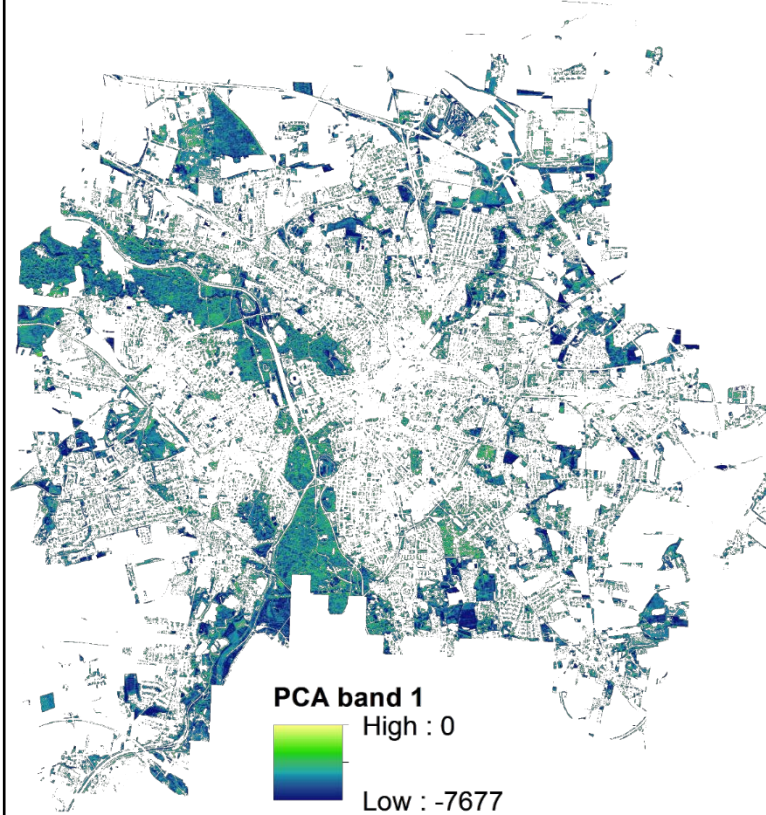
- Three breeding periods between 1992 and 1995
- In total 120 species
- 1300 polygons a 500 by 500 meters

Fractional vegetation cover



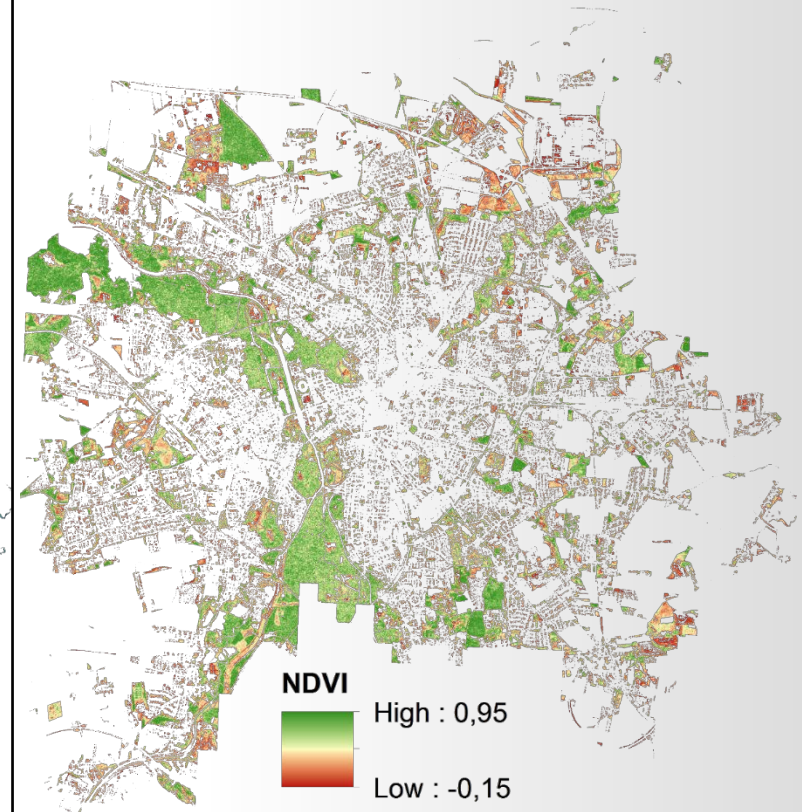
- Used for masking
- Threshold: 75 % cover

Principal component analysis (PCA)



- Band 1 & band 2
- Explaining over 99 % of variance

Normalized difference vegetation index (NDVI)



- Photosynthesis capacity
- $$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}$$

Indicator type	Name	Calculation
----------------	------	-------------

Texture metrics

(Haralick et al. 1973)

- GLCM mean

- GLCM variance

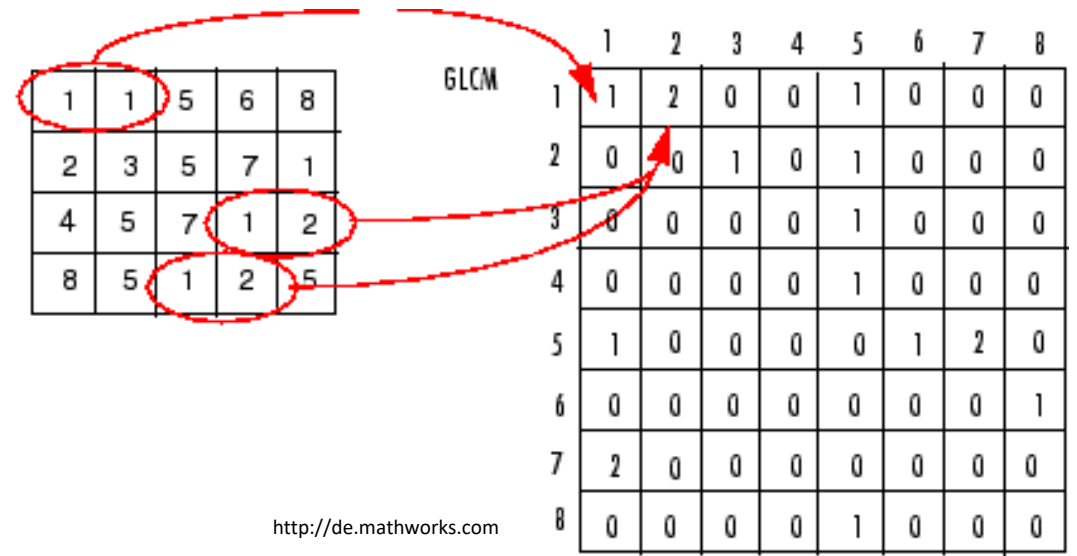
- GLCM homogeneity

- GLCM contrast

- GLCM dissimilarity

- GLCM entropy

- GLCM angular second moment



Grey level co-occurrence matrix

Spatial Autocorrelation

- Geary's C

- Moran's I

Descriptive Statistics

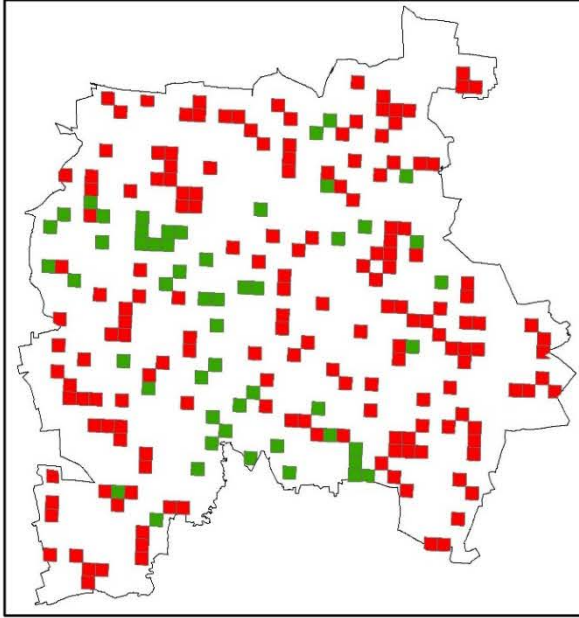
- Standard Deviation

- Coefficient of Variation

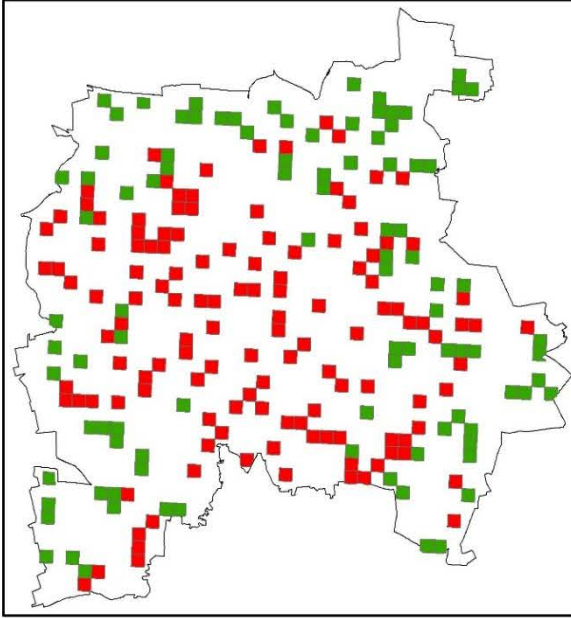
- Sum

Results: Predictions

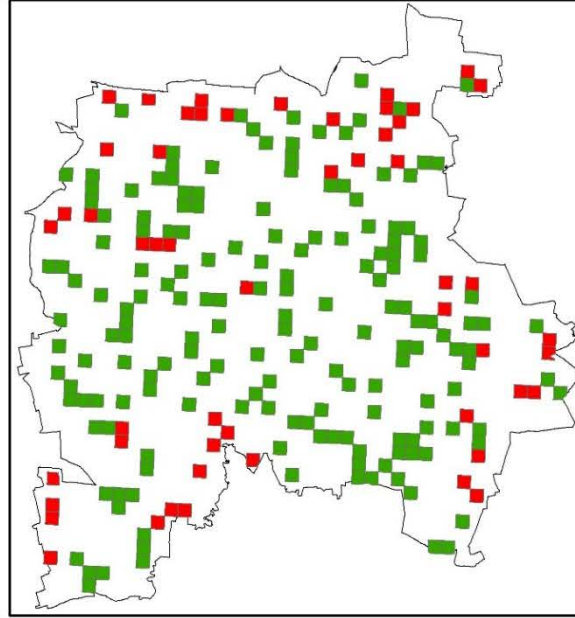
Great spotted woodpecker
(*Dendrocopus major*)



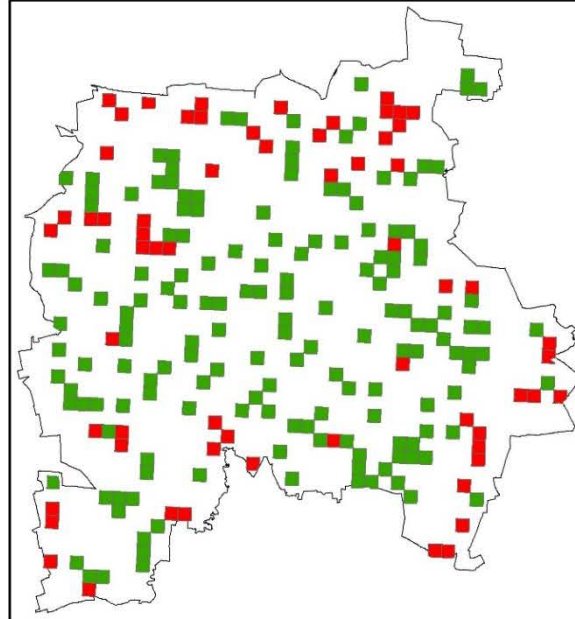
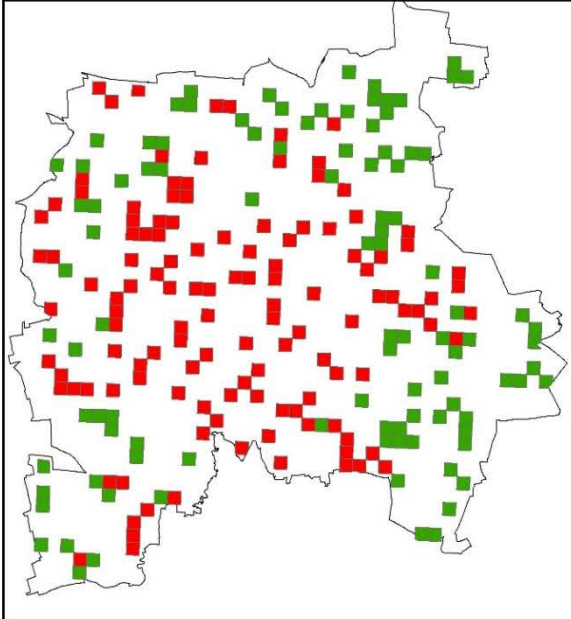
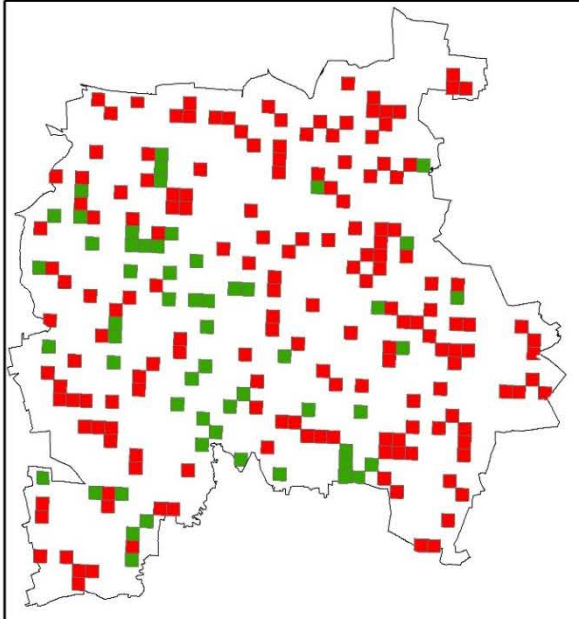
Eurasian skylark
(*Alauda arvensis*)



House sparrow
(*Passer domesticus*)

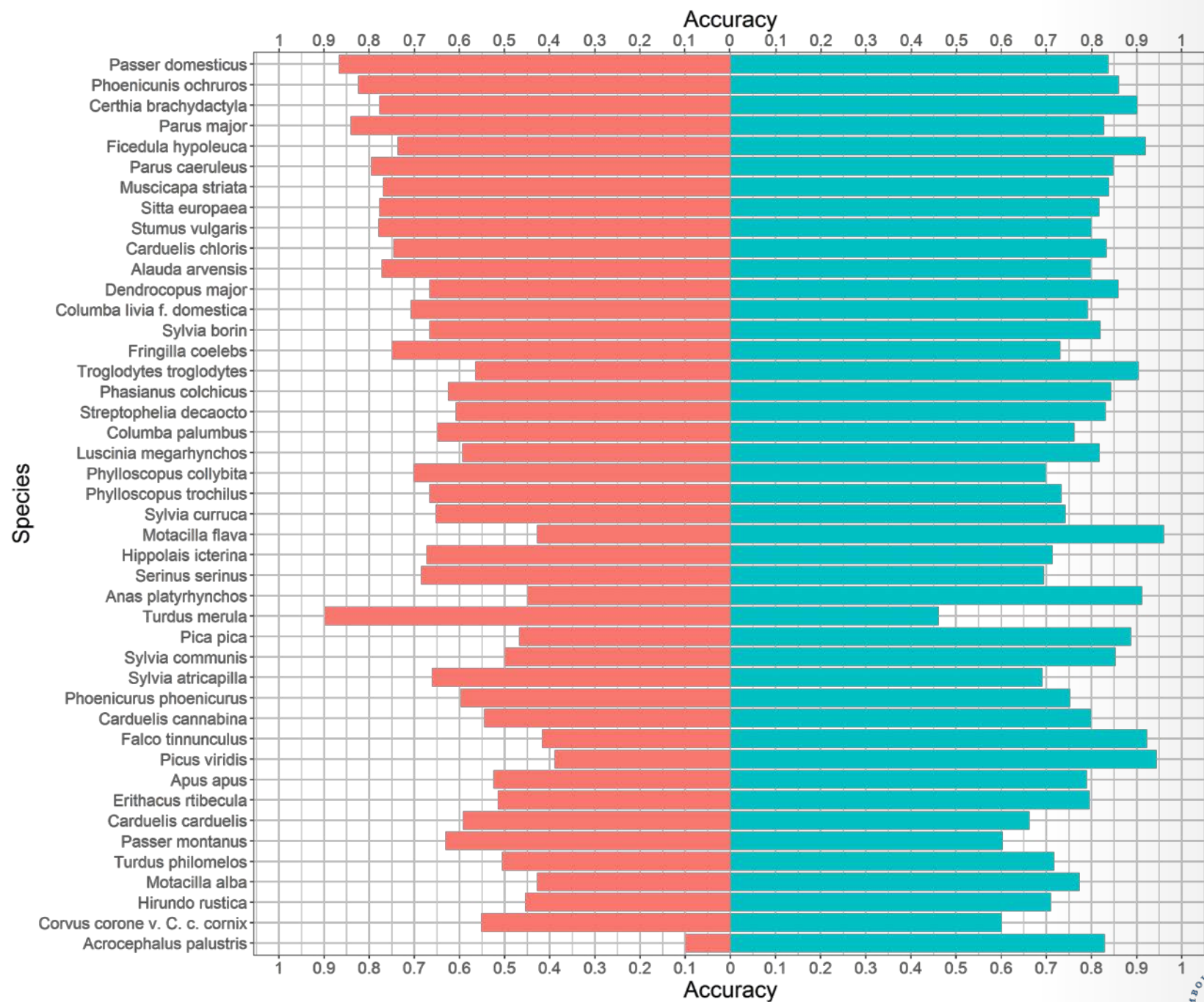


Prediction



Reference

Results: Accuracies



Modelling

Modeling technique	Fine-tuned hyper parameters	computation time per model in seconds	R package name
Random forest	<ul style="list-style-type: none">• Number of trees• Nodesize• Mtry	1.2	randomForest
CART classification tree		7.5	evtree
Support vector machine	<ul style="list-style-type: none">• Kernel type	0.08	e1071
Neural network	<ul style="list-style-type: none">• Hidden layer• Number of repetitions• Stepmax	140	neuralnet
Ensemble model		150	

Results: Comparison

Modeling technique	Mean accuracy	Mean sensitivity	Mean specificity
Random forest	0.78	0.70	0.77
Globally optimal trees	0.76	0.61	0.80
Support vector machine	0.78	0.65	0.78
Neural network	0.62	0.47	0.60
Mixed modelling	0.78	0.77	0.77

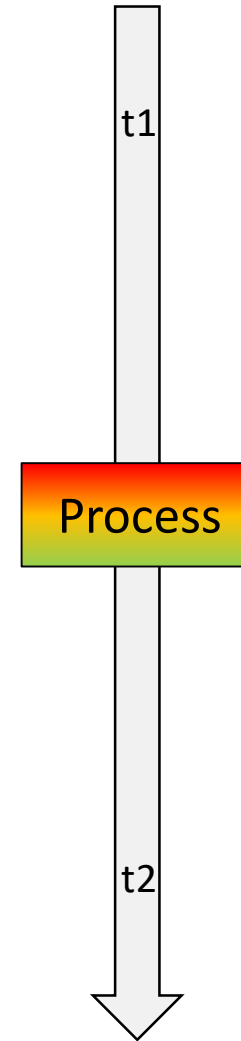
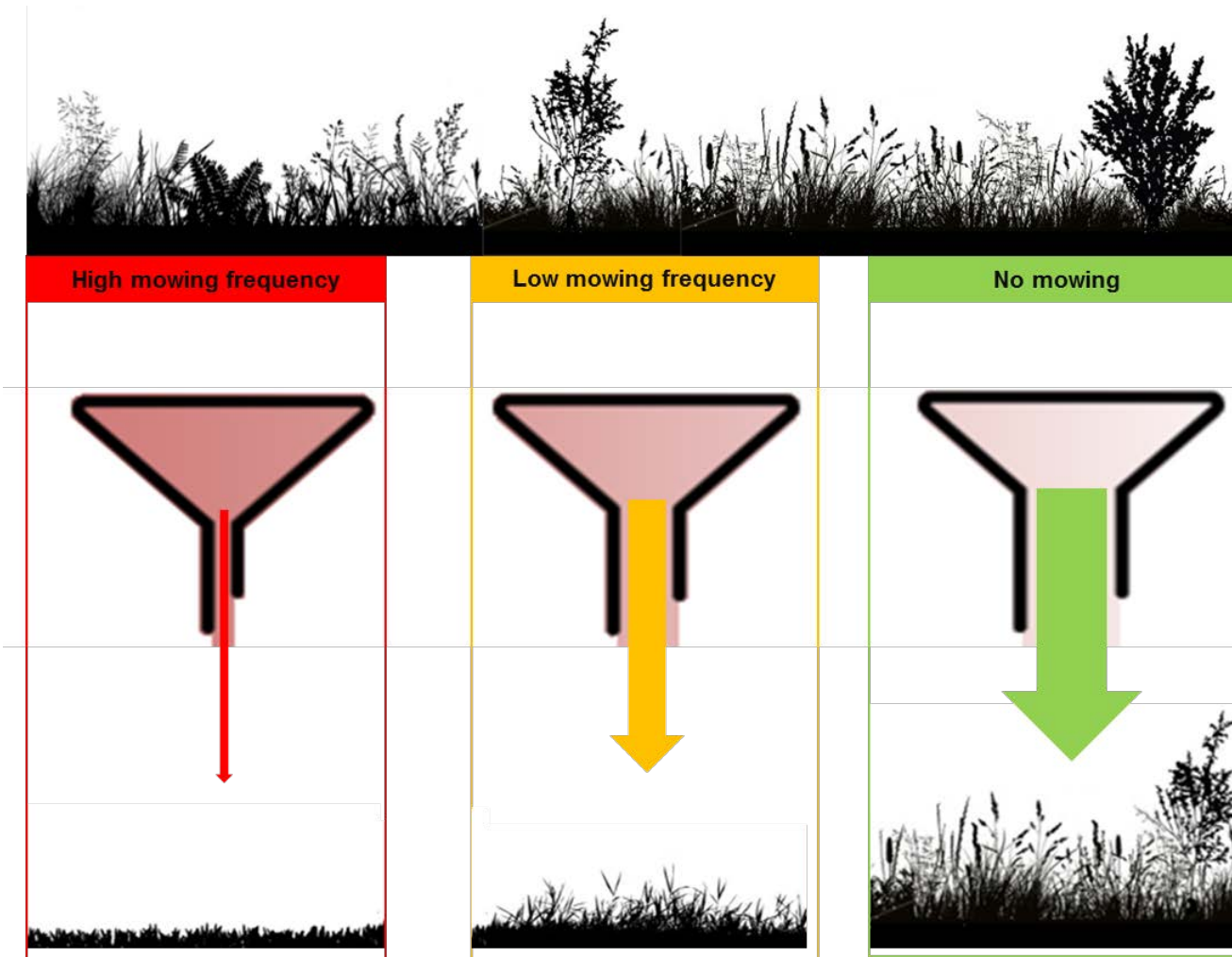
- For various species reliable habitat models can be derived at surprising accuracies while independent of other data.
- Suggested indicators advance previous studies based on preclassified LULC data
- Because of non linear interactions and high collinearity machine learning approach favorable
- Up next: Transferability studies to Halle and Berlin

Additional slides



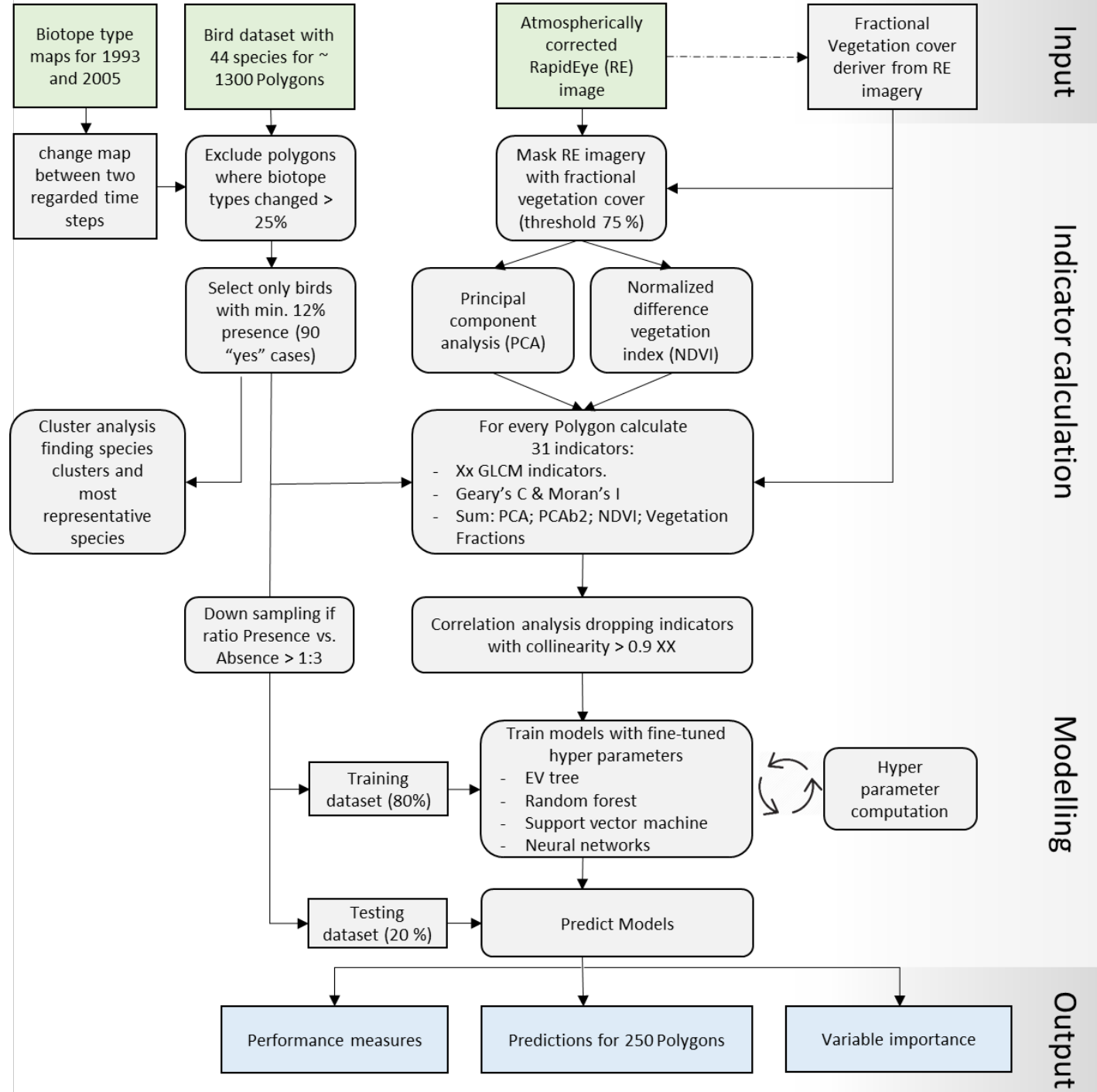
Was sagt der Indikator im Raum???

Research hypothesis



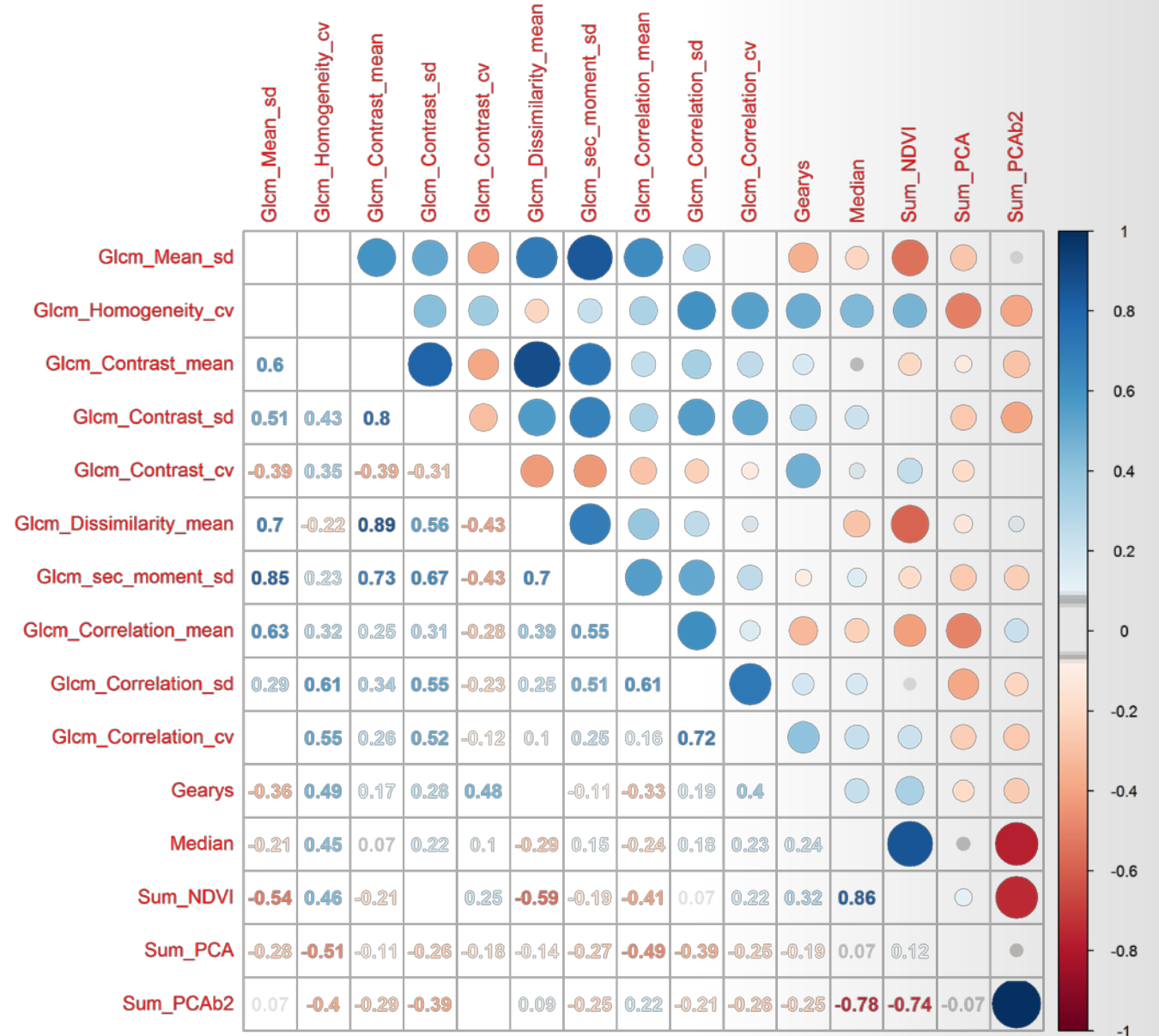
Trait variations are indicative for human induced processes and their intensity

Additional slide: Flowchart



Indicator selection

Indicator type	Name	Reference
Texture metrics	GLCM mean	
	GLCM variance	
	GLCM homogeneity	
	Grey level co-occurrence matrix	Haralick et al., 1973
	GLCM contrast	
	GLCM dissimilarity	
	GLCM entropy	
GLCM angular second moment		
Spatial	Geary's C	Geary, 1954
Autocorrelation	Moran's I	Moran, 1950
Descriptive Statistics	Standard Deviation	
	Coefficient of Variation	Datt, 1998
	Sum	



Wellmann et al. in preparation



Additional slide: Accuracies

Accuracy measure	RF	EV Tree	NN	SVM	3 Models
Overall accuracy	0.78	0.76	0.62	0.78	0.78
Balanced accuracy	0.74	0.70	0.54	0.67	0.77
Precision	0.61	0.59	0.46	0.58	0.59
Sensitivity	0.70	0.61	0.47	0.65	0.77
Specificity	0.77	0.80	0.60	0.78	0.77