TDMA 2023 - Traitement des Données Massives et Apprentissage - applications en géophysique, écologie et SHS

L'ÉCOLOGIE PRÉDICTIVE À L'HEURE DES DONNÉES MASSIVES

Wilfried THUILLER & Sara Si-MOUSSI

Laboratoire d'Ecologie Alpine
Grenoble, France
wilfried.thuiller@univ-grenoble-alpes.fr

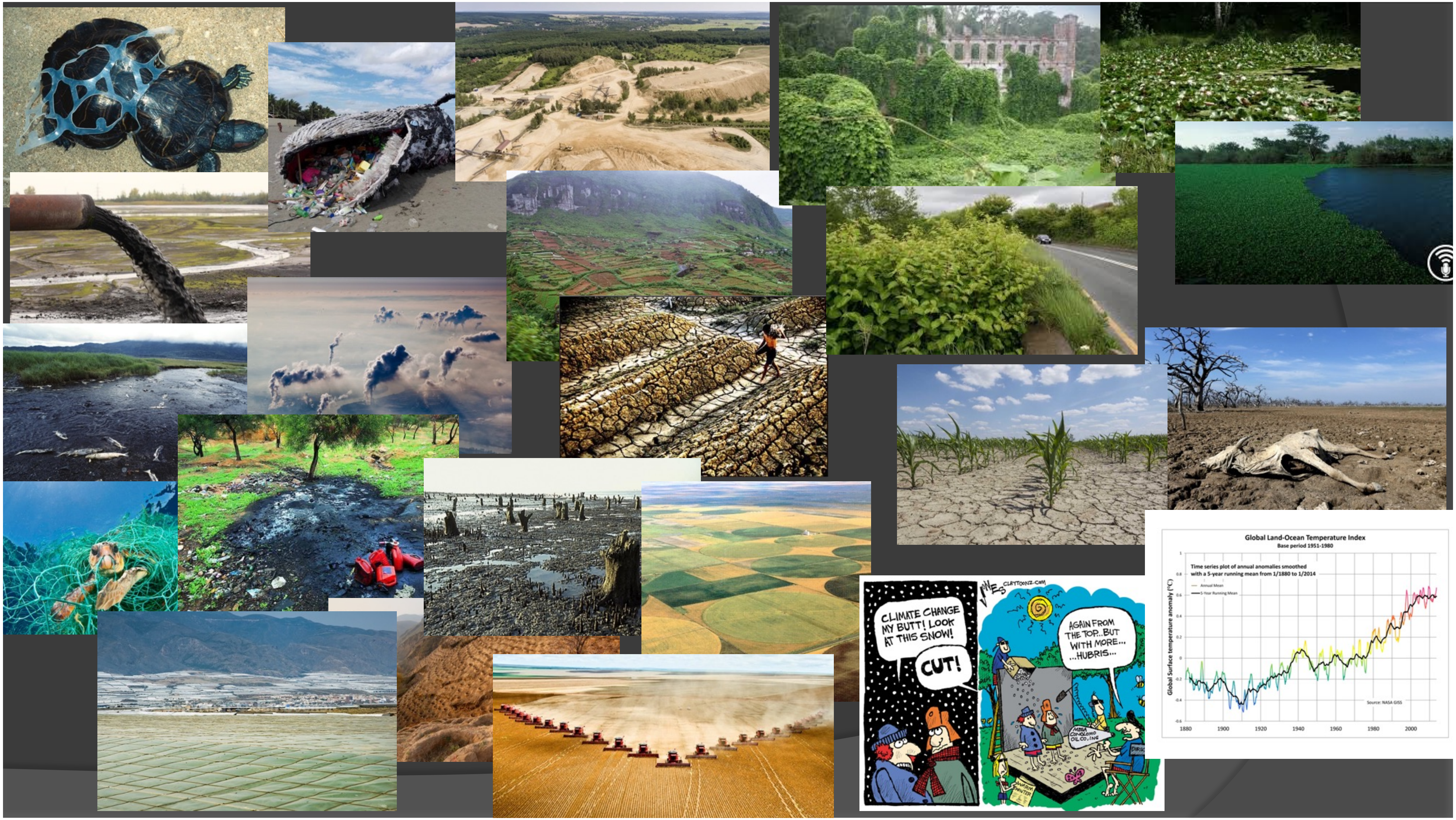
NATURA CONNECT

MIAI Grenoble Alpes

cnrs

LECA Laboratoire d'Ecologie Alpine

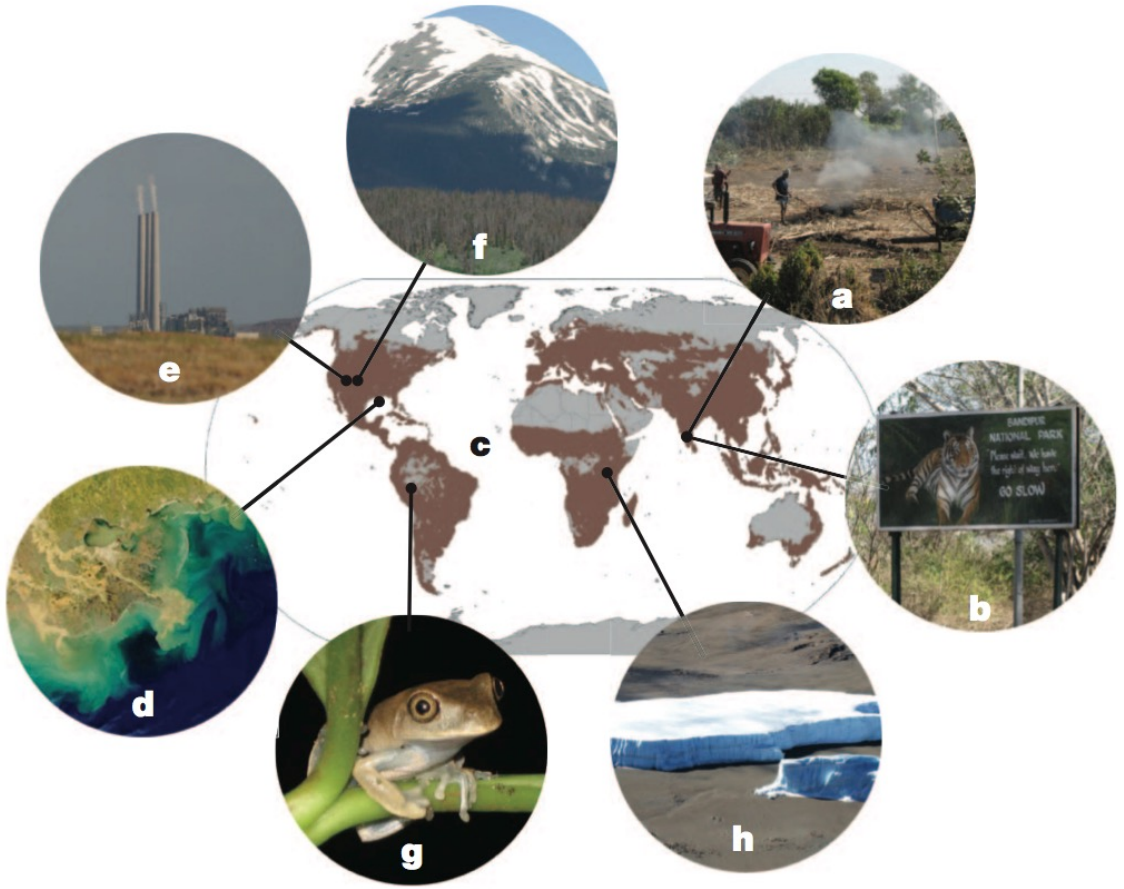
OSUG Observatoire des Sciences de l'Univers de Grenoble

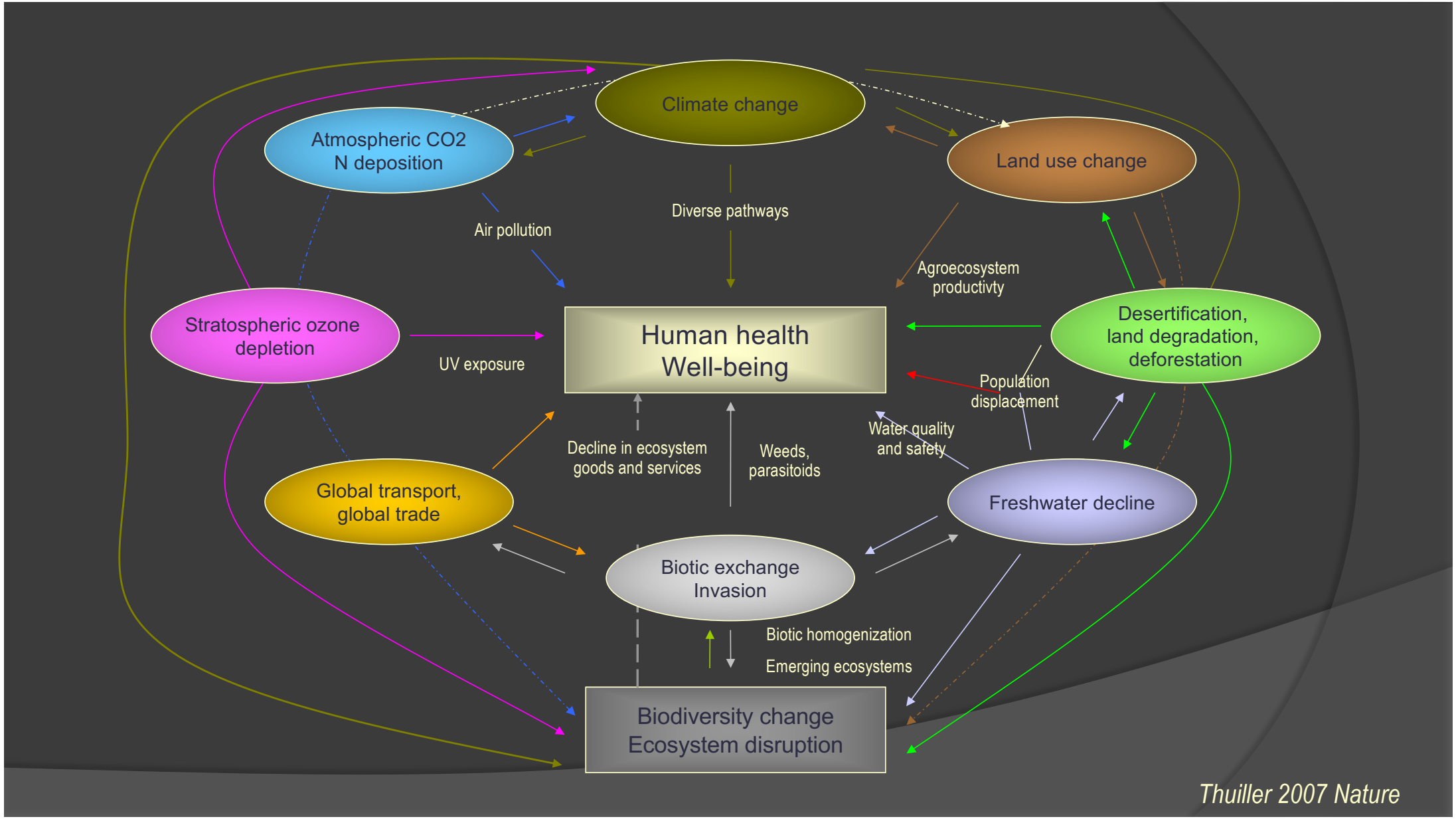


CLIMATE CHANGE MY BUTT! LOOK AT THIS SNOW!
CUT!
JONES
CLAYTOONZ.COM
AGAIN FROM THE TOP...BUT WITH MORE... ...HUBRIS...
MEGA CONGLOMO OIL CO, INC
WARM WINTER
Global Land-Ocean Temperature Index
Base period 1951-1980
Time series plot of annual anomalies smoothed with a 5-year running mean from 1/1880 to 1/2014
Annual Mean
5-Year Running Mean
Global Surface temperature anomaly (°C)
Source: NASA GISS
1880
1900
1920
1940
1960
1980
2000

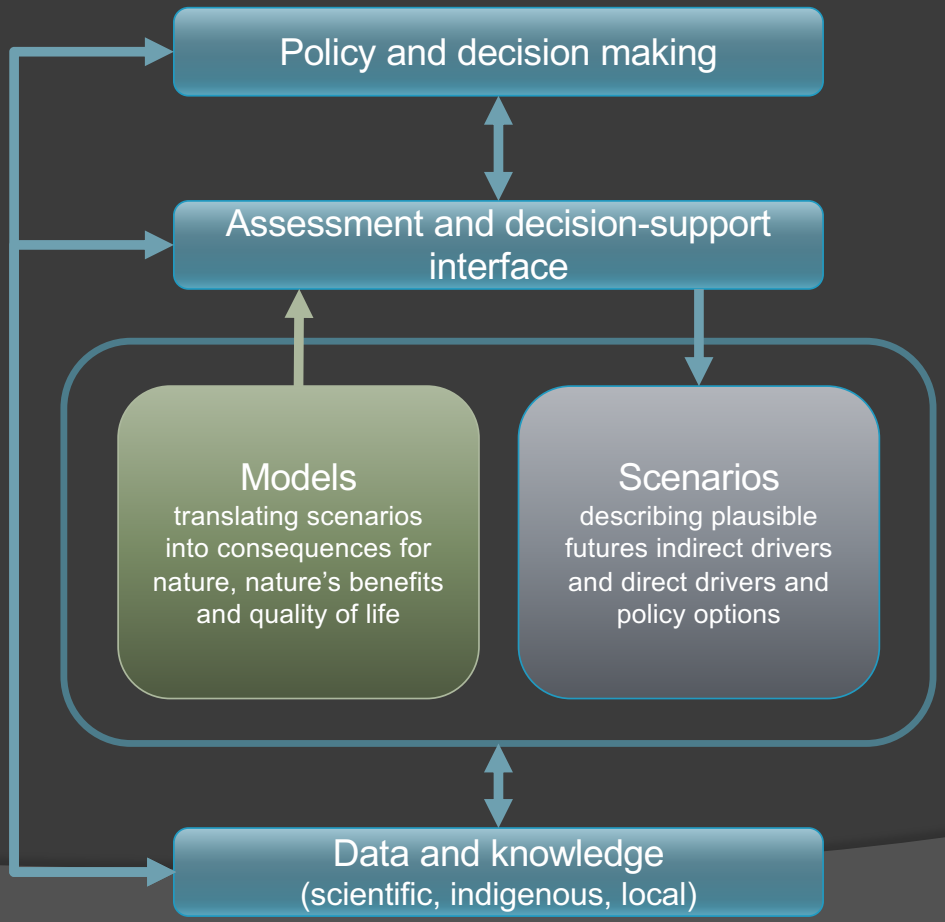
Has the Earth's sixth mass extinction already arrived?

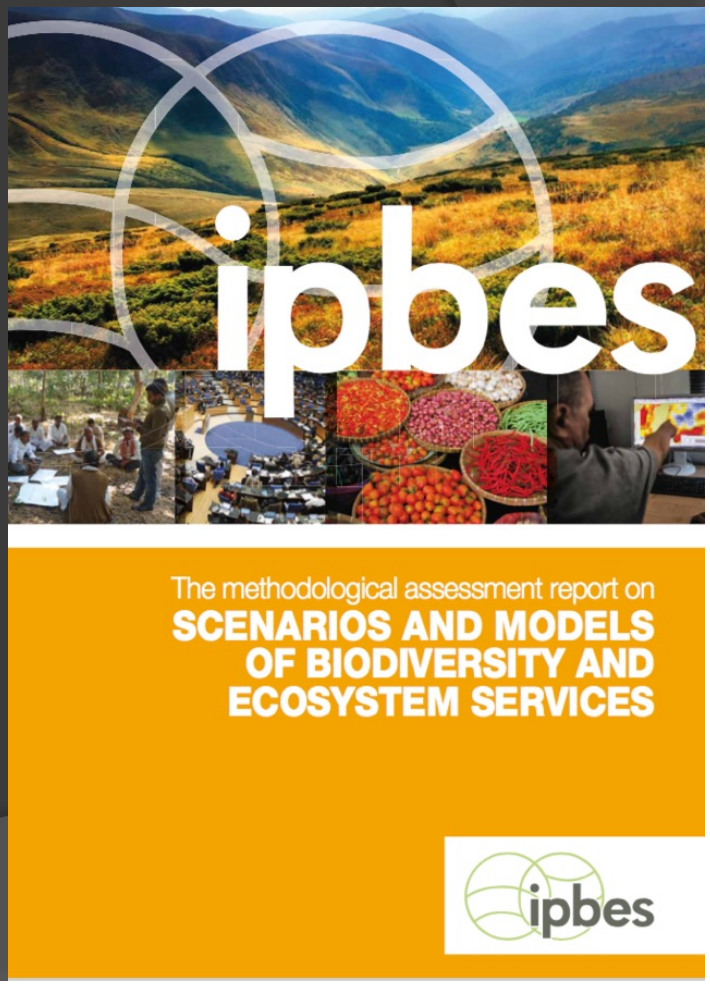
Anthony D. Barnosky[1,2,3], Nicholas Matzke[1], Susumu Tomiya[1,2,3], Guinevere O. U. Wogan[1,3], Brian Swartz[1,2], Tiago B. Quental[1,2]†, Charles Marshall[1,2], Jenny L. McGuire[1,2,3]†, Emily L. Lindsey[1,2], Kaitlin C. Maguire[1,2], Ben Mersey[1,4] & Elizabeth A. Ferrer[1,2]



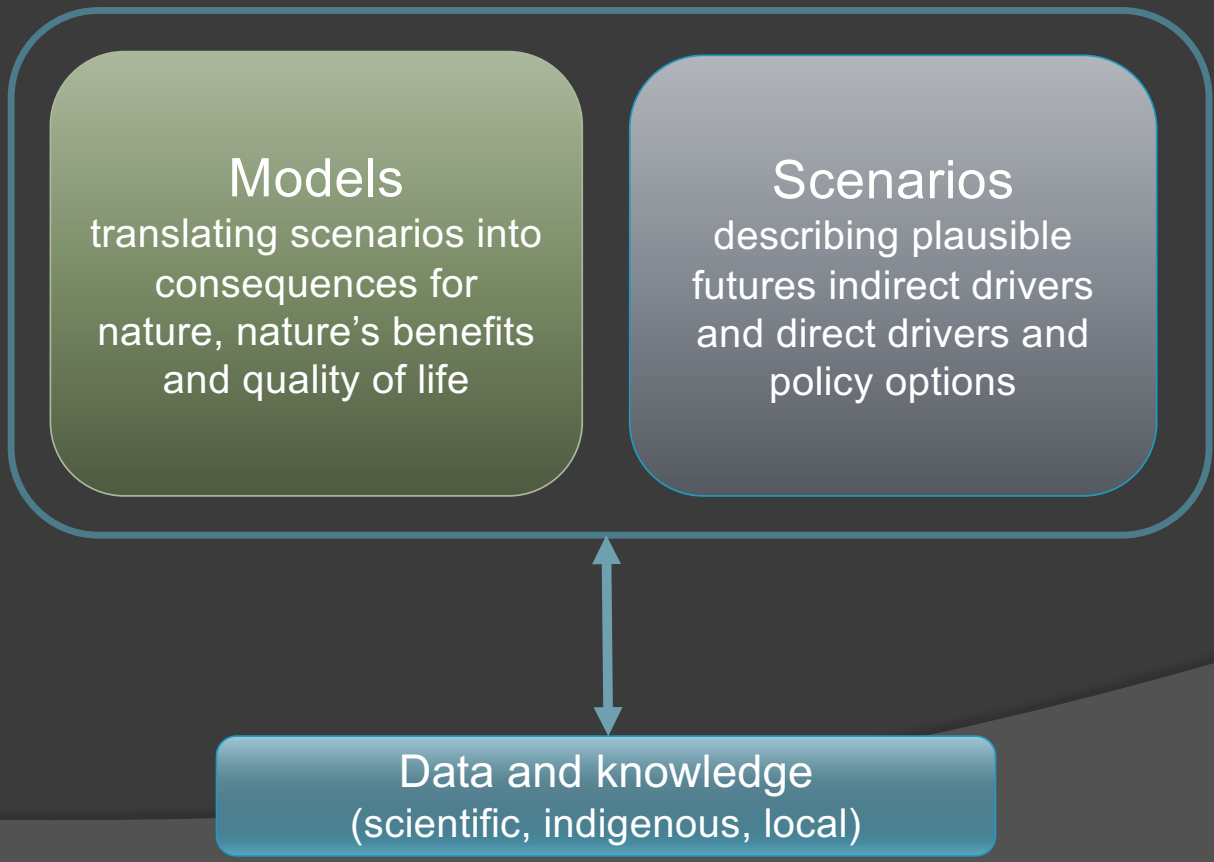

Climate change
Atmospheric CO2
N deposition
Land use change
Diverse pathways
Air pollution
Agroecosystem
productivty
Stratospheric ozone
depletion
Human health
Well-being
Desertification,
land degradation,
deforestation
UV exposure
Population
displacement
Water quality
and safety
Decline in ecosystem
goods and services
Weeds,
parasitoids
Global transport,
global trade
Freshwater decline
Biotic exchange
Invasion
Biotic homogenization
Emerging ecosystems
Biodiversity change
Ecosystem disruption
Thuiller 2007 Nature

The need for biodiversity models and scenarios

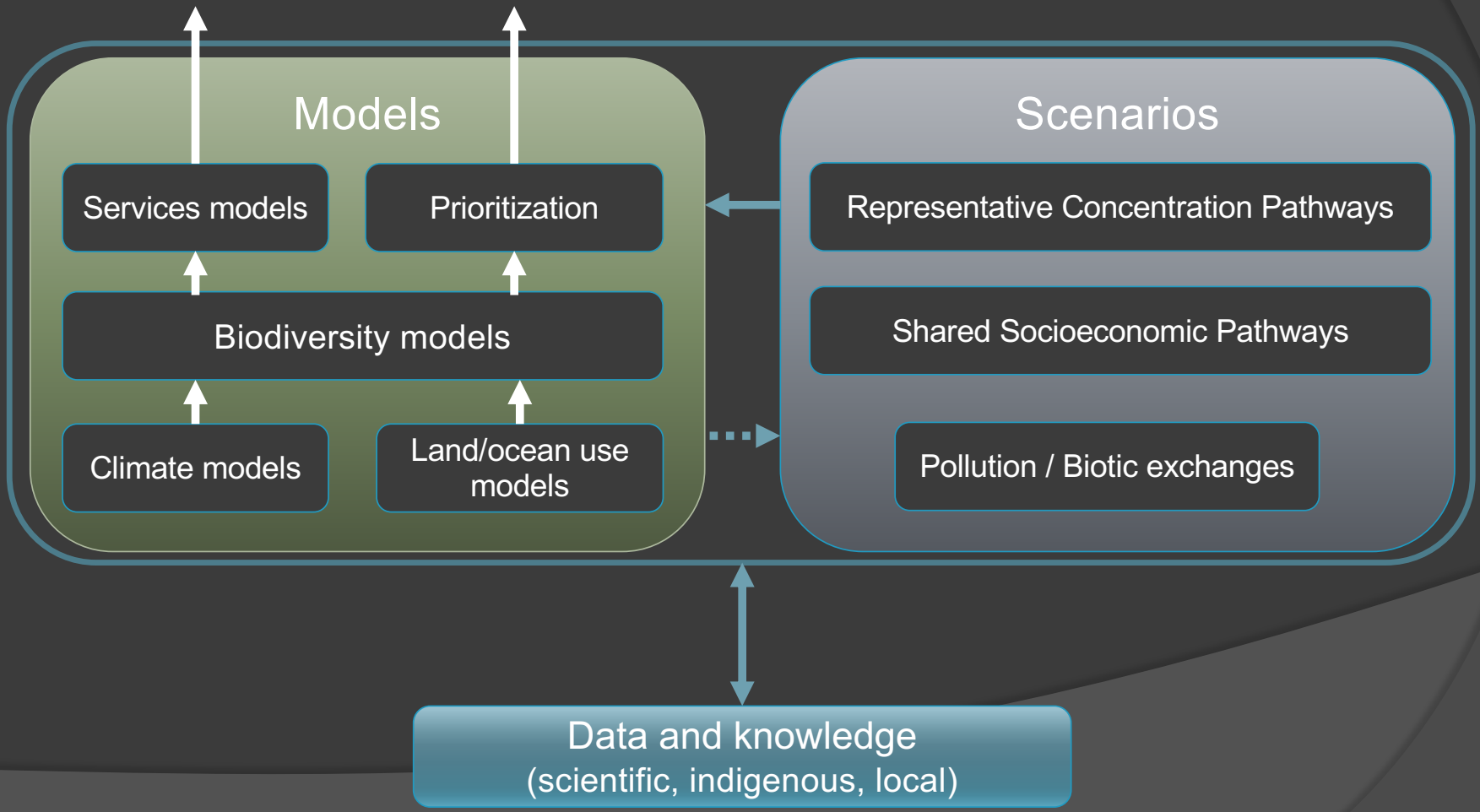




The need for biodiversity models and scenarios



The need for biodiversity models and scenarios



Biodiversity models

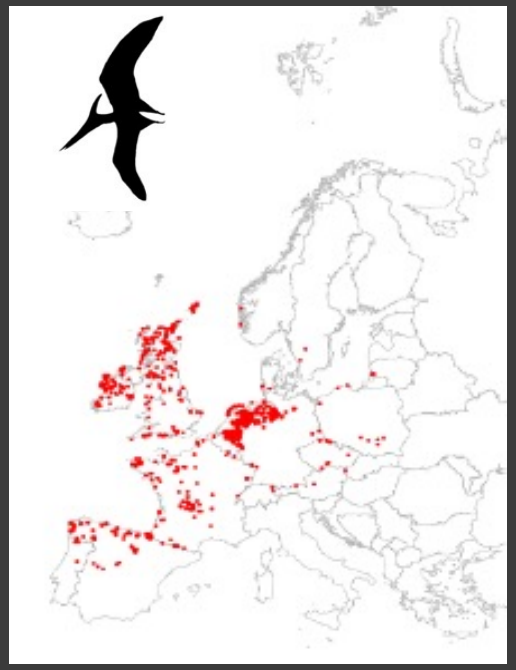
$$Biodiversity = f(X_i, \sum)$$

Biodiversity models

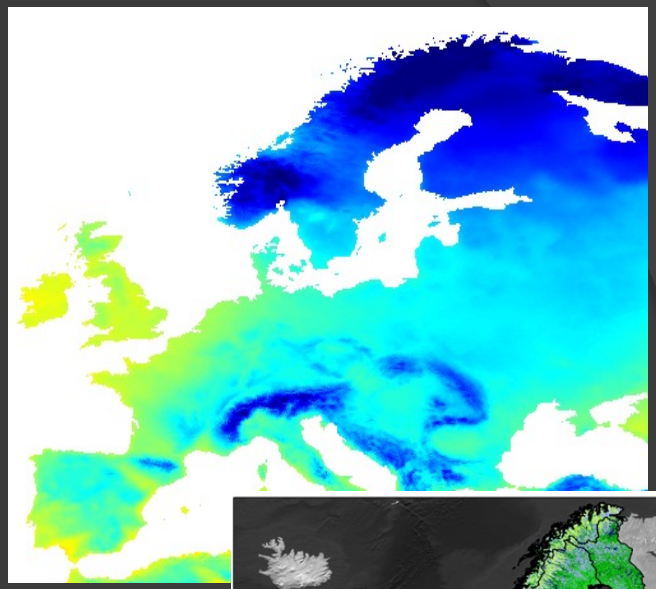
Species, groups, function *Environnement – climat, etc…*

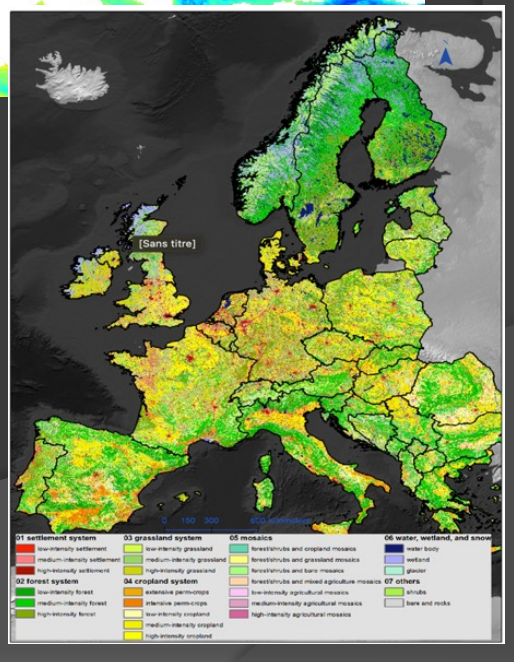
$$\textit{sites}\begin{pmatrix} \vdots & \cdots & \vdots \\ \cdots & \ddots & \cdots \\ \vdots & \cdots & \vdots \end{pmatrix} = f\left(\begin{pmatrix} \vdots & \cdots & \vdots \\ \cdots & \ddots & \cdots \\ \vdots & \cdots & \vdots \end{pmatrix}, \Sigma\right)$$

Biodiversity models



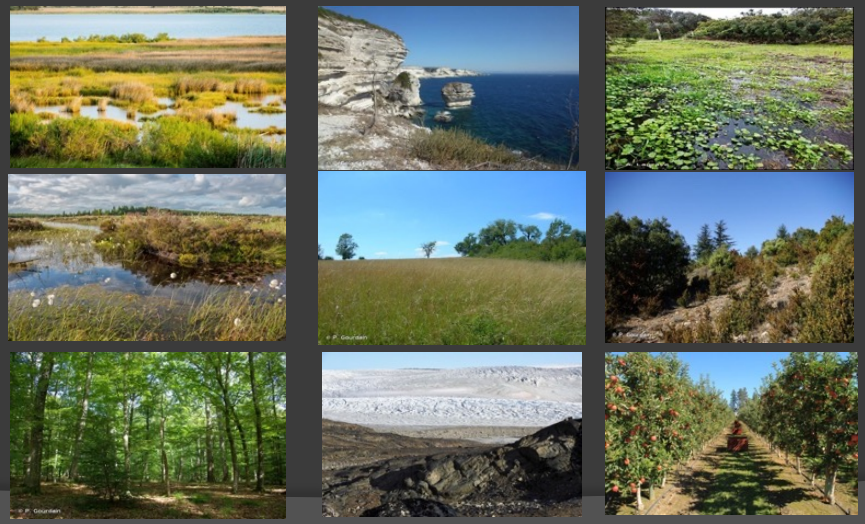
~





Case study 1

Predicting soil diversity in space across the French Alps



Case study 2

Predicting habitat in space across Europe

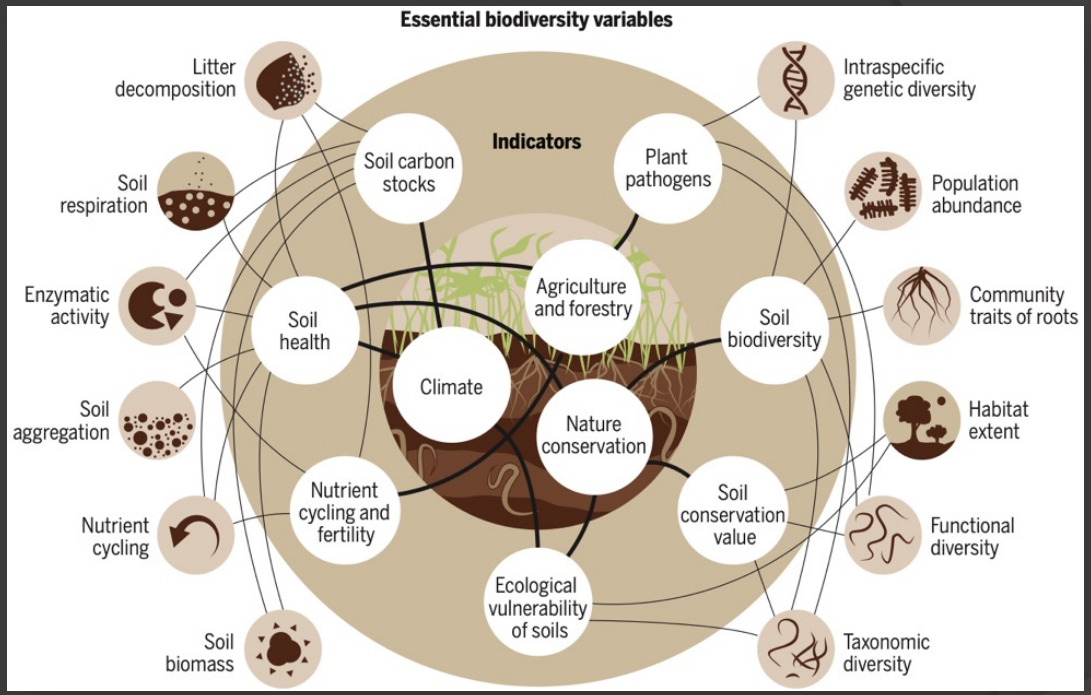
Predicting soil diversity in space across the French Alps

BIODIVERSITY

Tracking, targeting, and conserving soil biodiversity

A monitoring and indicator system can inform policy

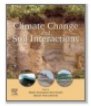
By **Carlos A. Guerra, Richard D. Bardgett, Lucrezia Caon, Thomas W. Crowther, Manuel Delgado-Baquerizo, Luca Montanarella, Laetitia M. Navarro, Alberto Orgiazzi, Brajesh K. Singh, Leho Tedersoo, Ronald Vargas-Rojas, Maria J. I. Briones, François Buscot, Erin K. Cameron, Simone Cesarz, Antonis Chatzinotas, Don A. Cowan, Ika Djukic, Johan van den Hoogen, Anika Lehmann, Fernando T. Maestre, César Marín, Thomas Reitz, Matthias C. Rillig, Linnea C. Smith, Franciska T. de Vries, Alexandra Weigelt, Diana H. Wall, Nico Eisenhauer**



ELSEVIER

Climate Change and Soil Interactions

2020, Pages 1-19



Chapter 1 - Soil biodiversity conservation for mitigating climate change

Jan Frouz [1, 2]

Original Paper | Published: 05 October 2010

Buried treasure: soil biodiversity and conservation

Sophie S. Parker

Biodiversity and Conservation **19**, 3743–3756 (2010) | Cite this article

1772 Accesses | **33** Citations | **10** Altmetric | Metrics

Received: 17 July 2020 | Revised: 24 September 2020 | Accepted: 2 October 2020

DOI: 10.1111/geb.13211

ECOLOGICAL SOUNDINGS

Global Ecology and Biogeography | WILEY

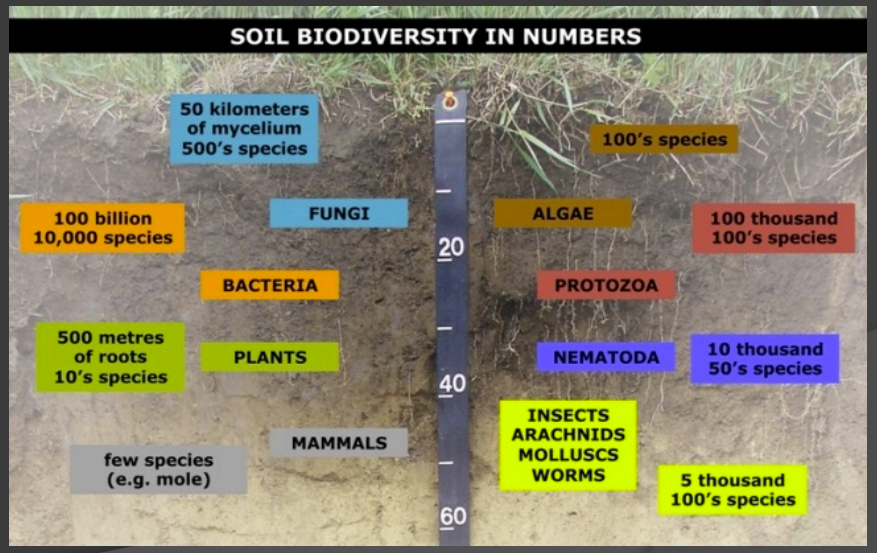
The multidimensionality of soil macroecology

Nico Eisenhauer[1,2] | François Buscot[2,3] | Anna Heintz-Buschart[1,3] | Stephanie D. Jurburg[1] | Kirsten Küsel[1,4] | Johannes Sikorski[5] | Hans-Jörg Vogel[6] | Carlos A. Guerra[1,7]

Predicting soil diversity in space across the French Alps

Major issues in predicting and conserving soil biodiversity:

- Soil biodiversity is a complex system comprising thousand of species in a given sample
- Soil biodiversity is multi-trophic



Predicting soil diversity in space across the French Alps

Major issues in predicting and conserving soil biodiversity:

- Soil biodiversity is a complex system comprising thousand of species in a given sample
- Soil biodiversity is multi-trophic

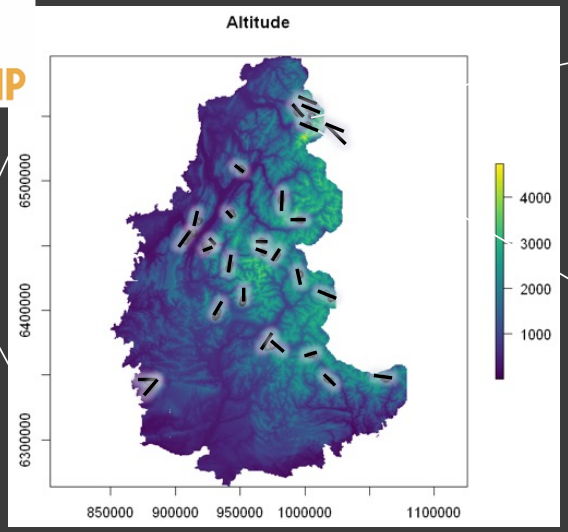
- Strong knowledge gaps:
  - Taxonomy -> Linean shortfall
  - Distribution -> Wallacean shortfall
  - Abundance -> Prestonian shortfall
  - Species traits -> Raunkiæran shortfall
  - Species interactions -> Eltonian shortfall

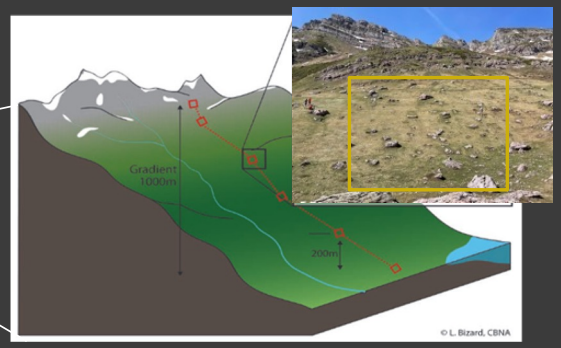
Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity

Joaquín Hortal,[1,2,3,*] Francesco de Bello,[4,5] José Alexandre F. Diniz-Filho,[2] Thomas M. Lewinsohn,[6] Jorge M. Lobo,[1] and Richard J. Ladle[7,8,*]

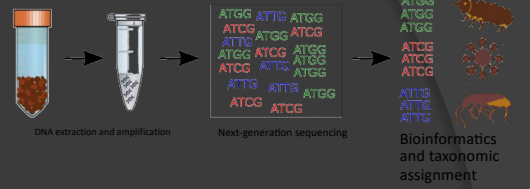
Predicting soil diversity in space across the French Alps

ORCHAMP

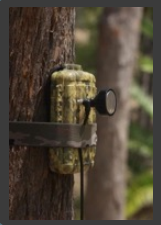




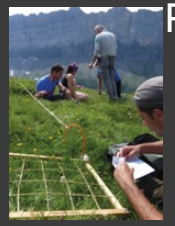
Soil biodiversity with eDNA

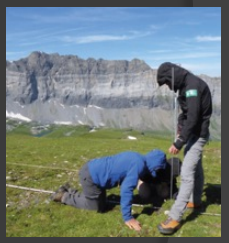


Acoustic landscape

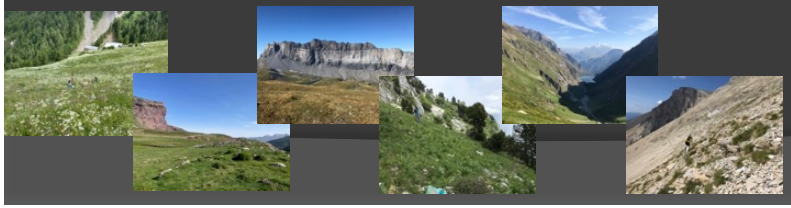


Plant relevés (pin-points)

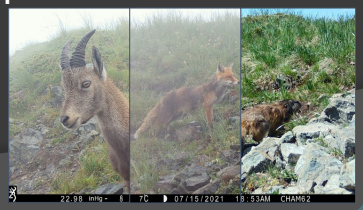




35 gradients
5-8 plots
3 sub-plots
-> 500 soil samples



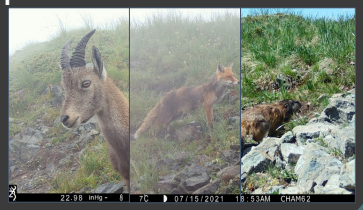
Camera-traps

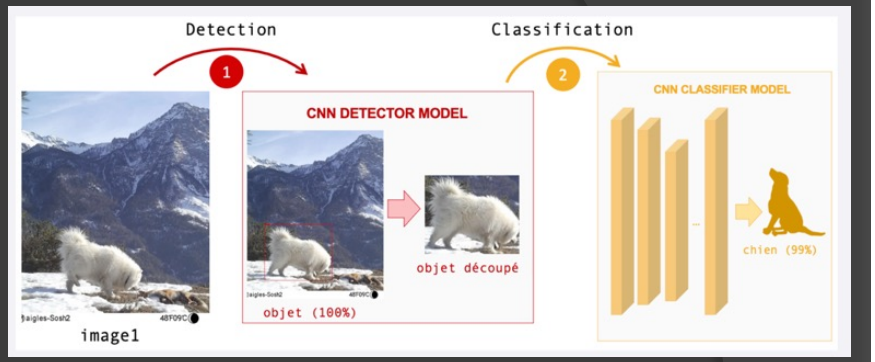


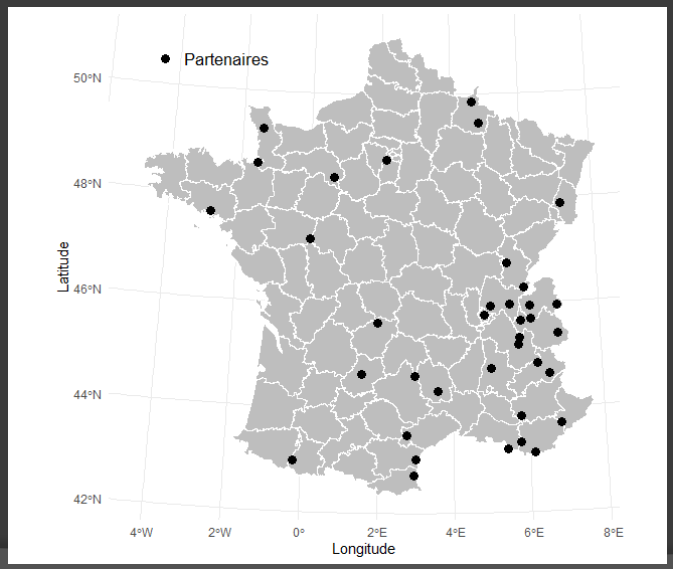
Local conditions: climat, soil physico-chemistry, pedology, usage

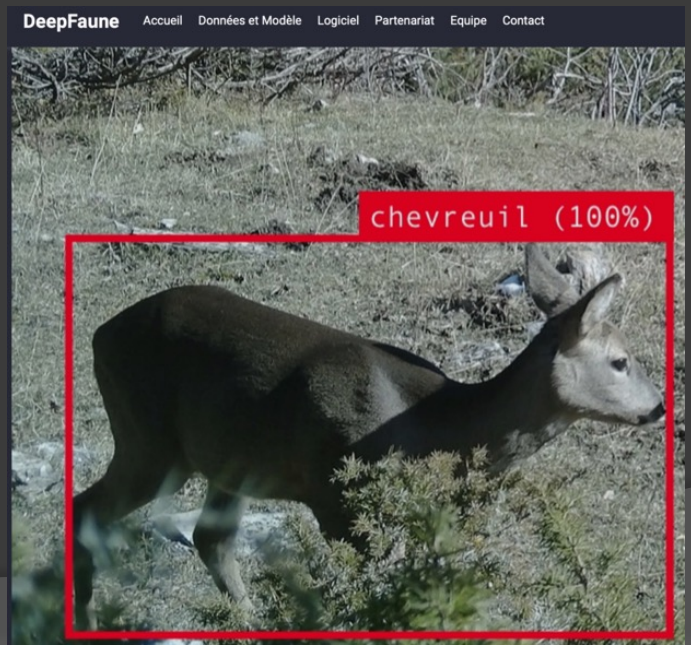
Predicting soil diversity in space across the French Alps

Camera-traps





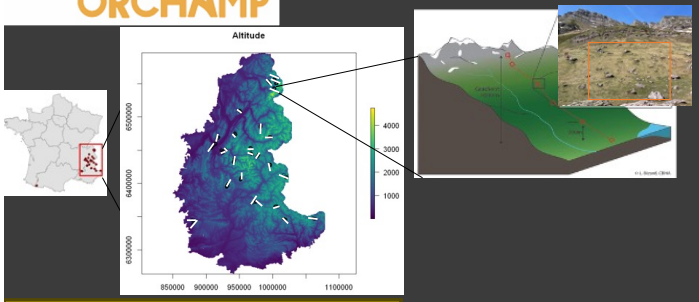




https://www.deepfaune.cnrs.fr

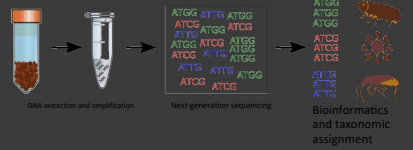
Predicting soil diversity in space across the French Alps

ORCHAMP



35 gradients
5-8 plots
3 sub-plots
-> 500 soil samples

Soil biodiversity with eDNA



DNA markers

- Bacteria — 16S rRNA gene
- Eukaryotes — 18S rRNA gene
- Fungi — ITS1
- Collembola — 16S mitochondrial rDNA
- Oligochaeta — 16S mitochondrial rDNA
- Insecta — 16S mitochondrial rDNA

Received: 23 January 2019 | Revised: 12 July 2019 | Accepted: 15 July 2019
DOI: 10.1111/jbi.13681
RESEARCH PAPER
Journal of Biogeography WILEY

From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices?

Irene Calderón-Sanou[1] | Tamara Münkemüller[1] | Frédéric Boyer[1] | Lucie Zinger[2] | Wilfried Thuiller[1]

Received: 22 January 2021 | Revised: 10 May 2021 | Accepted: 11 May 2021
DOI: 10.1111/1755-0998.13430
RESOURCE ARTICLE
MOLECULAR ECOLOGY RESOURCES WILEY

Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification

Laetitia Mathon[1,2] | Alice Valentini[2] | Pierre-Edouard Guérin[1] | Eric Normandeau[3] | Cyril Noel[4] | Clément Lionnet[5] | Emilie Boulanger[1,6] | Wilfried Thuiller[5] | Louis Bernatchez[3] | David Mouillot[6,7] | Tony Dejean[2] | Stéphanie Manel[1]

Received: 30 April 2020 | Revised: 10 September 2020 | Accepted: 23 September 2020
DOI: 10.1111/mec.15674
SPECIAL ISSUE
MOLECULAR ECOLOGY WI

Effects of soil preservation for biodiversity monitoring using environmental DNA

Alessia Guerrieri[1] | Aurélie Bonin[1] | Tamara Münkemüller[2] | Ludovic Gielly[2] | Wilfried Thuiller[2] | Gentile Francesco Ficetola[1,2]

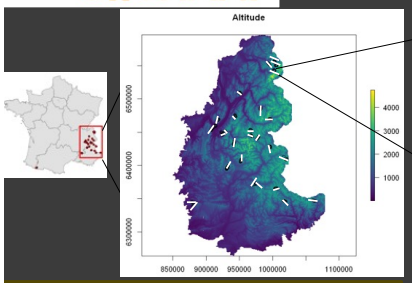
scientific reports
Check for updates
OPEN

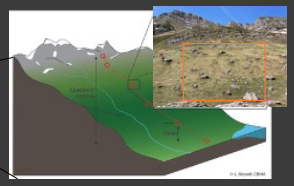
Applying convolutional neural networks to speed up environmental DNA annotation in a highly diverse ecosystem

Benjamin Flück[1,2✉], Laëtitia Mathon[3], Stéphanie Manel[3], Alice Valentini[4], Tony Dejean[4], Camille Albouy[5], David Mouillot[6,7], Wilfried Thuiller[8], Jérôme Murienne[9], Sébastien Brosse[9] & Loïc Pellissier[1,2✉]

Predicting soil diversity in space across the French Alps

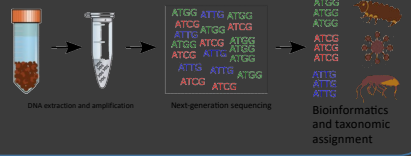
ORCHAMP





35 gradients
5-8 plots
3 sub-plots
-> 500 soil samples

Soil biodiversity with eDNA



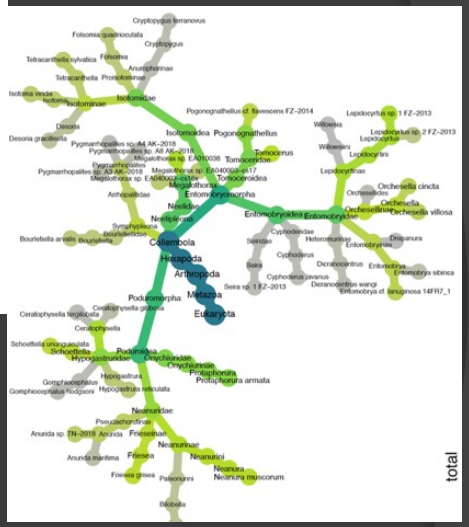
DNA markers

- Bacteria — 16S rRNA gene
- Eukaryotes — 18S rRNA gene
- Fungi — ITS1
- Collembola — 16S mitochondrial rDNA
- Oligochaeta — 16S mitochondrial rDNA
- Insecta — 16S mitochondrial rDNA

Massive data

```
>gi|224967179|gb|AC234315.2| Homo sapiens FOSMID clone ABC14-5019070OJ6 from
chromosome x, complete sequence
TTCCCAATAGGCTGGACTGCTTACCACCCCATGTGGCCTCAAAGAGCTCCAGTCACTCCTTTACGAACCC
AATCACTCCAGAACTTTAGAACAAAGTTTCTGAGTTACTCCTTGTAATAGGCTAAATAATGGCTCCCAAA
GATATTAGGATTTGATTCCCAGAACCTATAAATATTACCTTATTTGGAAAACGGTTCTTAGCAGATGTGA
TTGAGTTAAGGATATTGAGATGCAGAGATTATTTTAGATTATCTAGACTATCTGGGTGGATGTATTGGTC
AGGGTTCTTCAGAGGACAGAGCCAATAGGATATATGTATATAAAAAGGGAGTTAATTAGGGAGAATTGGC
TCACATGATTACAAGGTGAAGTCCCACGATAGGCCGTCTGCAAACTGGGGAGAGAAGCTAGTTGTGTGGC
TCAGTCCAAATCCAAAAGCCTCAAAACTGGAGAAGCTGACAGTACAAGCCCTAGTCTGAGGCCAAAGGTC
CAAGAGCCCCTGAGAGGCTGCTGGTGCAAGTTCCAGAGTCCAAAGGTTAACAAACCTGAAGTCTGGTGTC
CAAAGGCAGGAGGAGAGGAAGCAGACAGGAAGAGAGAAAGCAAACAGACTCAGCAAGAAAGCTGCTGTTC
TTCCACCTGCTTTGTTCTAGCCACGCTGGCAGTCAATTGCATGGTGCCCATCCACACTGAGGGTGGATCT
TCCTCTAACAGTCAAACACTGACTCAAATGTCATCTTCTCTGGCAACACCCTCACAGACACACCCAGAAA
CAATGCTTCACCAGCCATCTATGCAGCCCTCAATCCAGTCAAGGTGACACCTAATGGTTAATGGTTATTA
ACCACGGTTAATAACCATGACAGTGGGTTCTAAATGTAATCACGTGTATCCTTATAAAAAAAGAGGCAG
AGGGAGATTTGAAGAGCTATACAGAGGAGAAGACACGTGAAGATGGAGGAGAGAGAAATTTGGCCATCA
```

Different taxonomic resolutions



samples

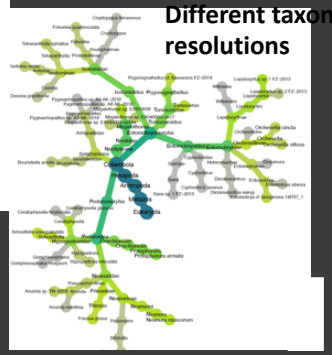
	sp1	sp2	sp3
S1			
S2			
S3			
S4			

Community matrix per marker

Predicting soil diversity in space across the French Alps

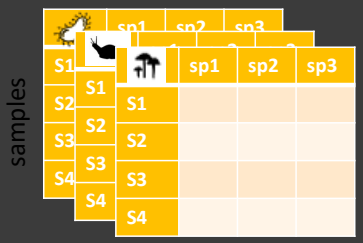
The quest

Massive data



Different taxonomic resolutions

MOTU

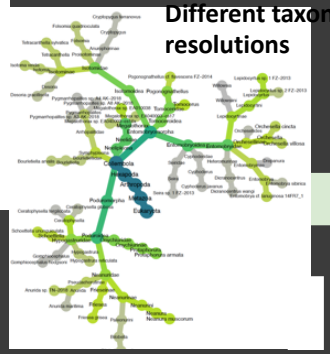


Community matrix per marker

Predicting soil diversity in space across the French Alps

The quest

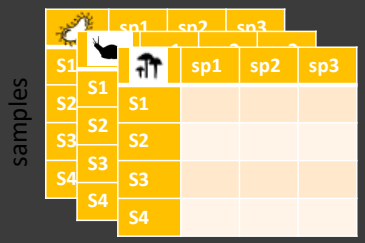
Massive data



Different taxonomic resolutions

Metacommunity across markers / plots

MOTU

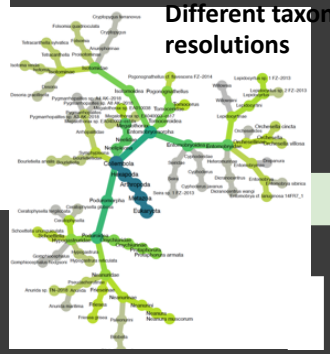


Community matrix per marker

Predicting soil diversity in space across the French Alps

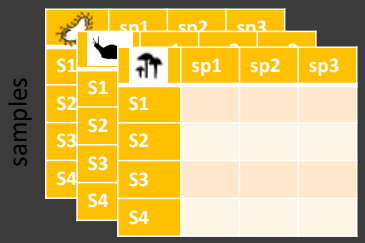
The quest

Massive data



Different taxonomic resolutions

MOTU



Community matrix per marker

Metacommunity across markers / plots

Trophic group assignment

Trophic groups
Trophic classes

Predicting soil diversity in space across the French Alps

Unstructured (textual) data

TaxoNERD

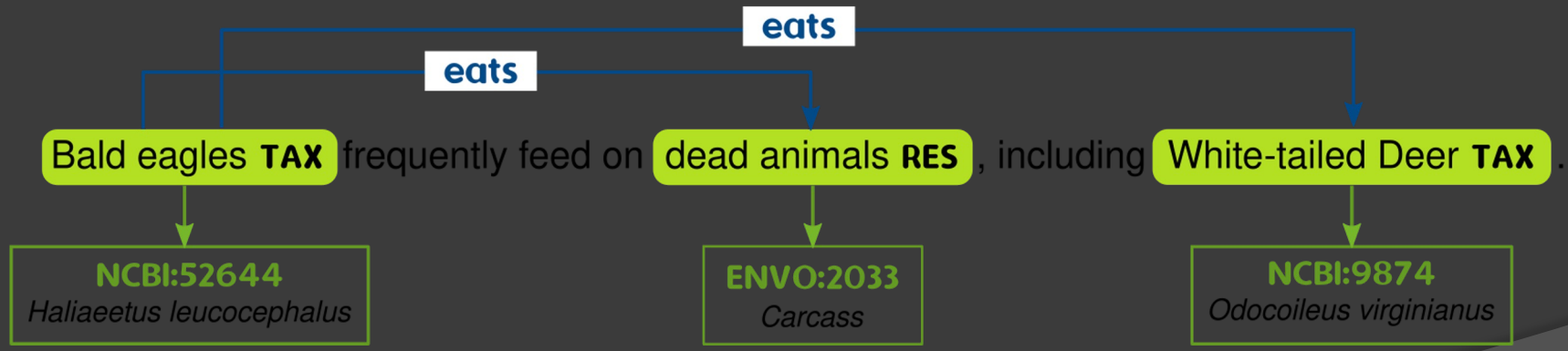
https://github.com/nleguillarme/taxonerd

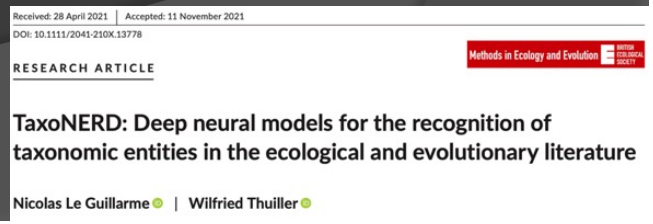
WIP

NAMED ENTITY RECOGNITION → ENTITY NORMALISATION → RELATION EXTRACTION

Structured (tabular) data

CSV





Received: 28 April 2021 | Accepted: 11 November 2021

DOI: 10.1111/2041-210X.13778

RESEARCH ARTICLE

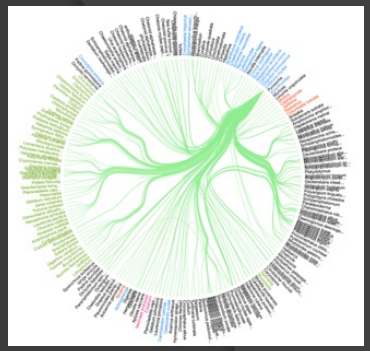
Methods in Ecology and Evolution

TaxoNERD: Deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature

Nicolas Le Guillarme | Wilfried Thuiller

Predicting soil diversity in space across the French Alps

- How to represent and map the knowledge



TaxNERD

GLOBI

Mangal

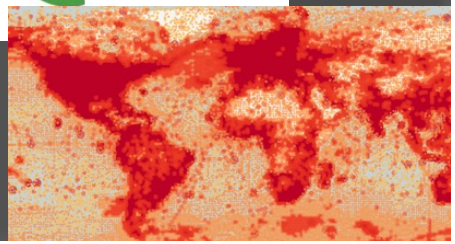
DATA INTEGRATION TOOLS

Trait and trophic web database (FungalTrait, FunGuild, Protax, NemaGuild)

GBIF

CSV

Expert knowledge



Predicting soil diversity in space across the French Alps

Trophic knowledge graph creation


RESEARCH ARTICLE

TaxoNERD: Deep neural models for the recognition of taxonomic entities in the ecological and evolutionary literature

Nicolas Le Guillarme | Wilfried Thuiller

Trophic information from heterogeneous data sources

Different formats, vocabularies, taxonomies, levels of structuredness…

agroportal.lirmm.fr/ontologies/SFWO

SOILFOODWEB ONTOLOGY

inteGraph

A library of components for ontology-based biodiversity data integration

https://github.com/nleguillarme/inteGraph

GRATIN : a GRAph of Trophic INformation

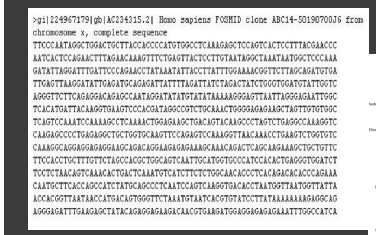
Single format, vocabulary, taxonomy, contains both stated and inferred facts

Le Guillarme et al. (2021). Building a Trophic Knowledge Graph to Support Soil Food Web Reconstruction. In *S4BioDiv 2021: 3rd International Workshop on Semantics for Biodiversity*

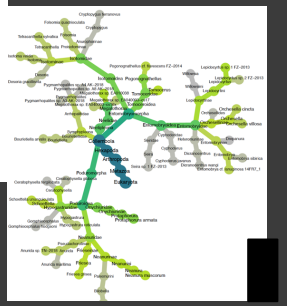
Le Guillarme et al. (2021) STWO: an Ontology for Soil Food Web Reconstruction. In *S4BioDiv 2021: 3rd International Workshop on Semantics for Biodiversity*

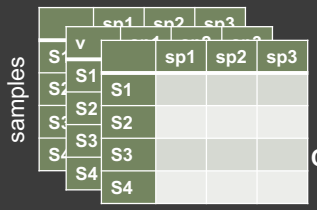
Predicting soil diversity in space across the French Alps

Massive data



Different taxonomic resolutions

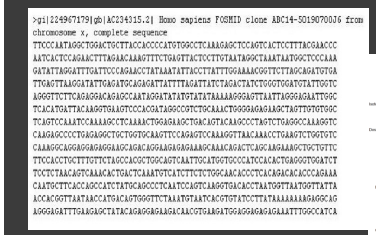




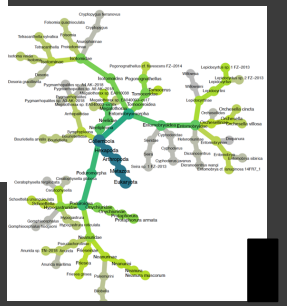
Community matrix per marker

Predicting soil diversity in space across the French Alps

Massive data



Different taxonomic resolutions



samples

	sp1	sp2	sp3
S1			
S2			
S3			
S4			

Community matrix per marker

Trophic information from heterogeneous data sources

Different formats, vocabularies, taxonomies, levels of structuredness...

inteGraph

A library of components for ontology-based biodiversity data integration

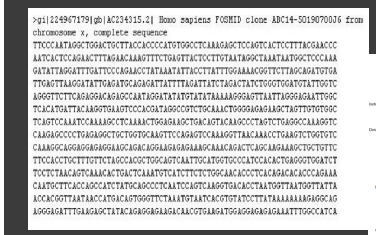
https://github.com/nleguillarme/inteGraph

GRATIN : a GRAph of Trophic INformation

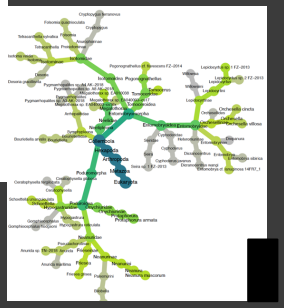
Single format, vocabulary, taxonomy, contains both stated and inferred facts

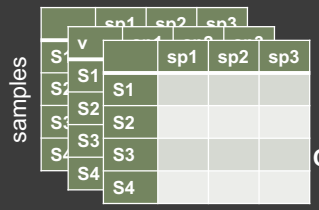
Predicting soil diversity in space across the French Alps

Massive data



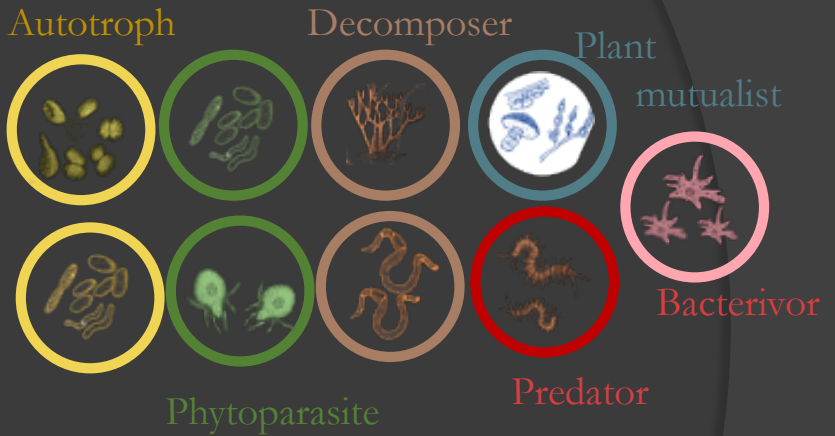
Different taxonomic resolutions





Community matrix per marker

500 x 300,000



55 trophic groups
11 trophic classes

Trophic information from heterogeneous data sources

Different formats, vocabularies, taxonomies, levels of structuredness...

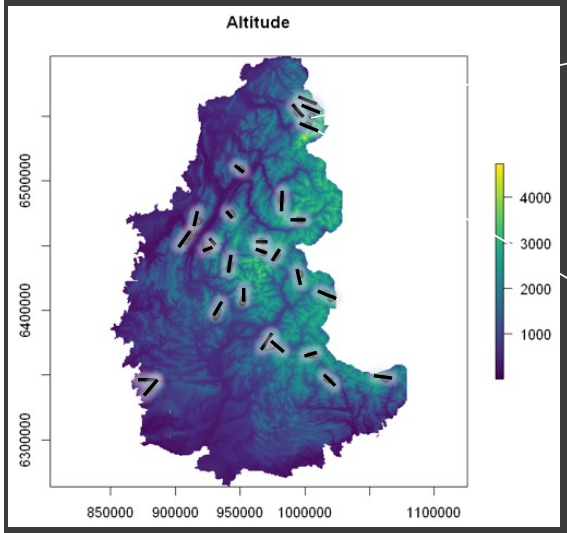
inteGraph
A library of components for ontology-based biodiversity data integration

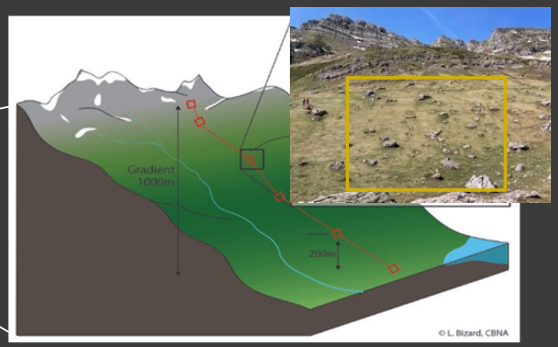
https://github.com/nleguillarme/inteGraph

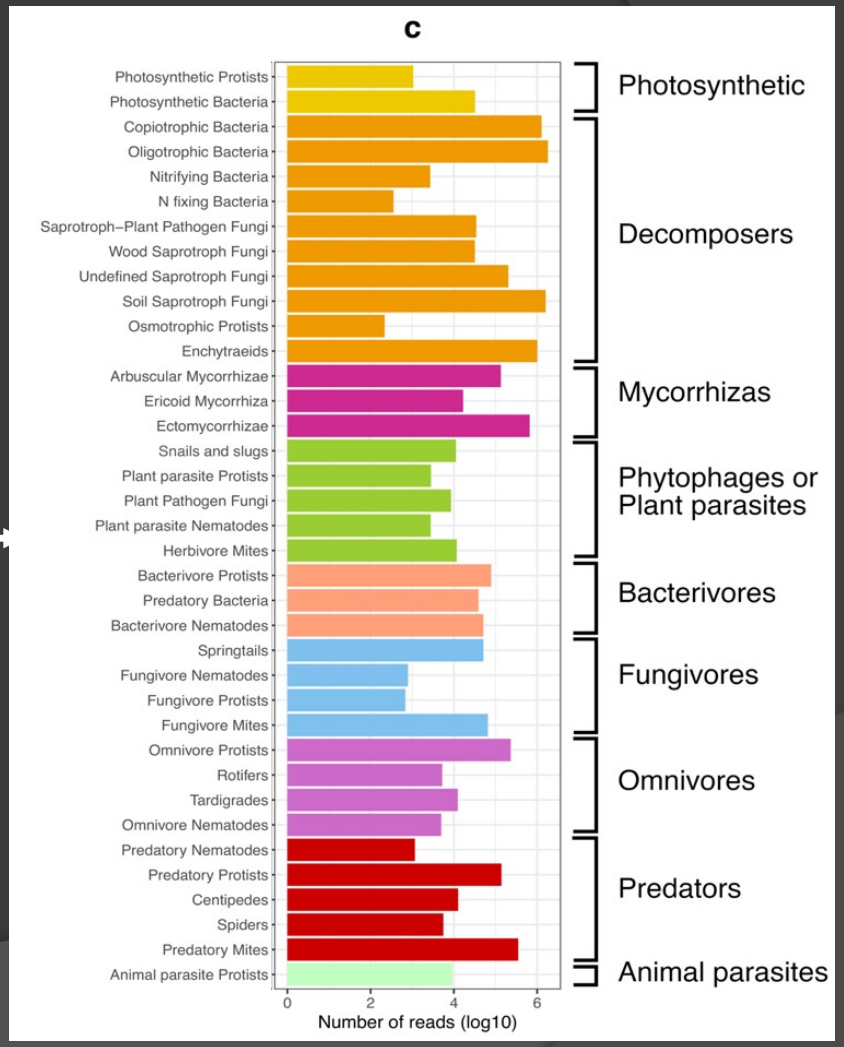
GRATIN : a GRAph of Trophic INformation

Single format, vocabulary, taxonomy, contains both stated and inferred facts

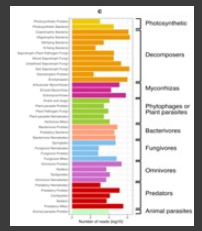
Predicting soil diversity in space across the French Alps





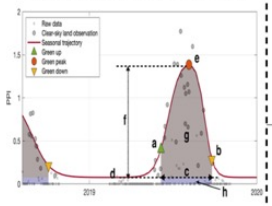


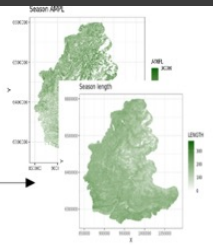
Predicting soil diversity in space across the French Alps



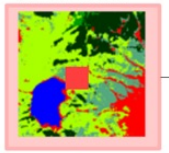
Sentinel2 - MSI

Vegetation phenology

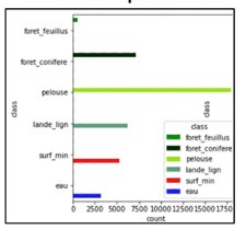


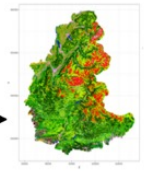


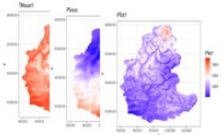
Land cover



Class frequencies

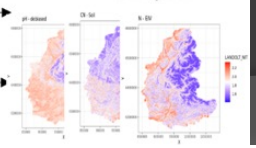






Climate CHELSA

Soil ISRIC World Soil Information



Ensemble forecasting framework

Model interpretability

- Variable importance
- Response / Turnover functions

Multi-species distribution models

Random Forest

XG Boost

ANN

GLM

Ensemble model

Taxonomic groups

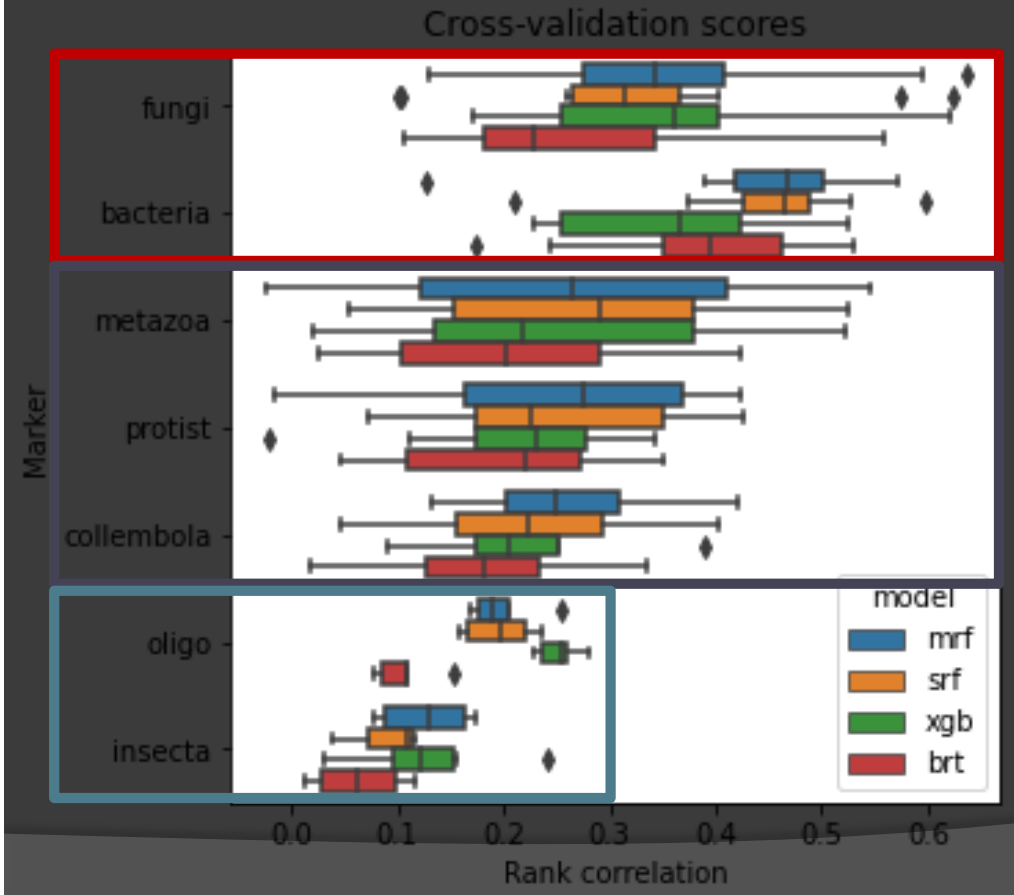
Trophic groups

1 2 ... 9 10

cross-validation stratified per altitude

- Distribution maps
- Uncertainty maps
- Diversity maps
- Model comparison

Predicting soil diversity in space across the French Alps



Very good predictions for microbes

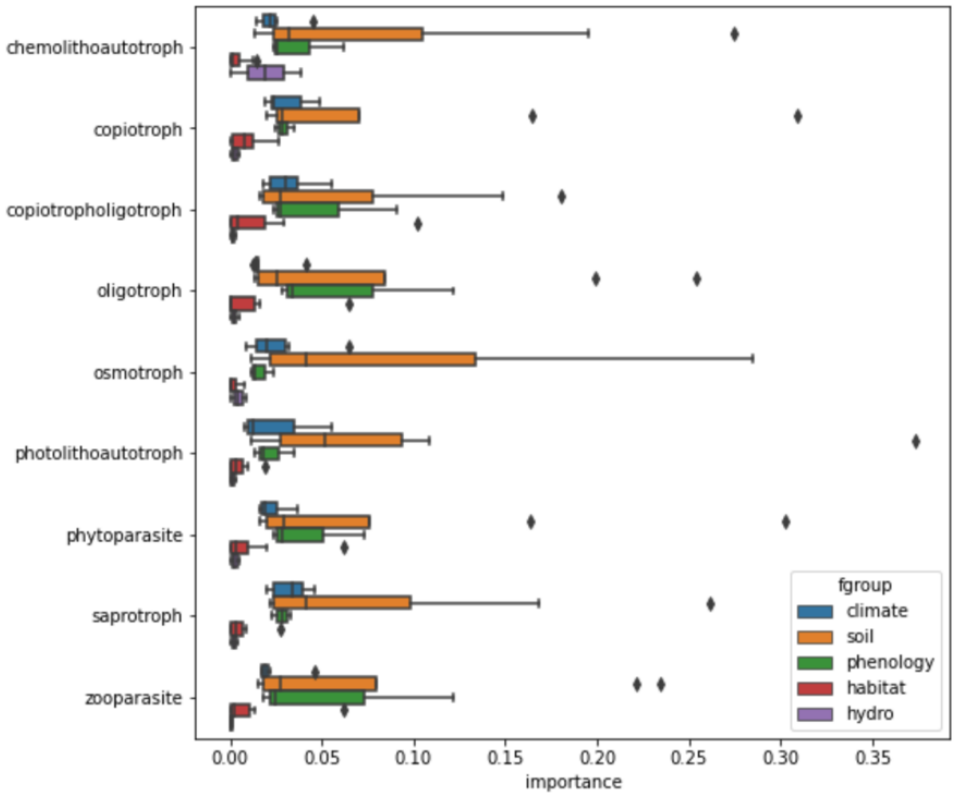
Good predictions for micro-to macro organisms

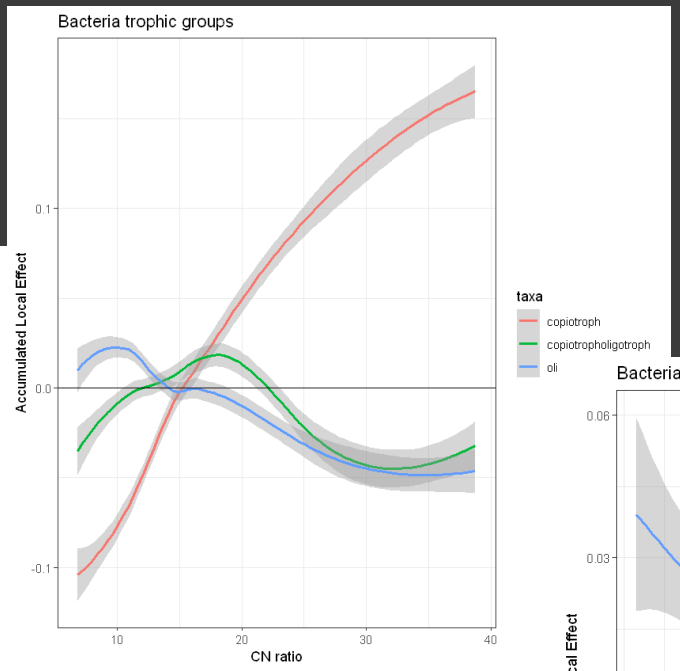
Moderate predictions for macro-organisms

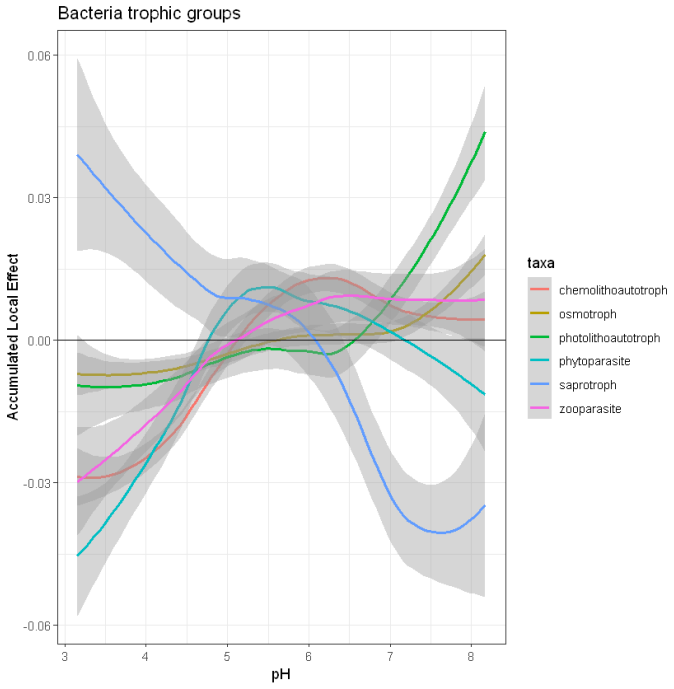
Predicting soil diversity in space across the French Alps

Variable importance and ALE

Bacteria

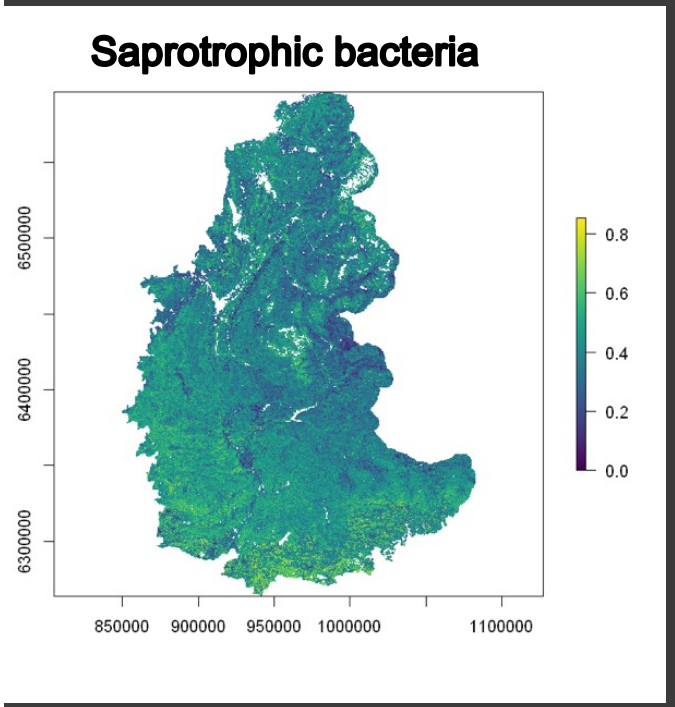


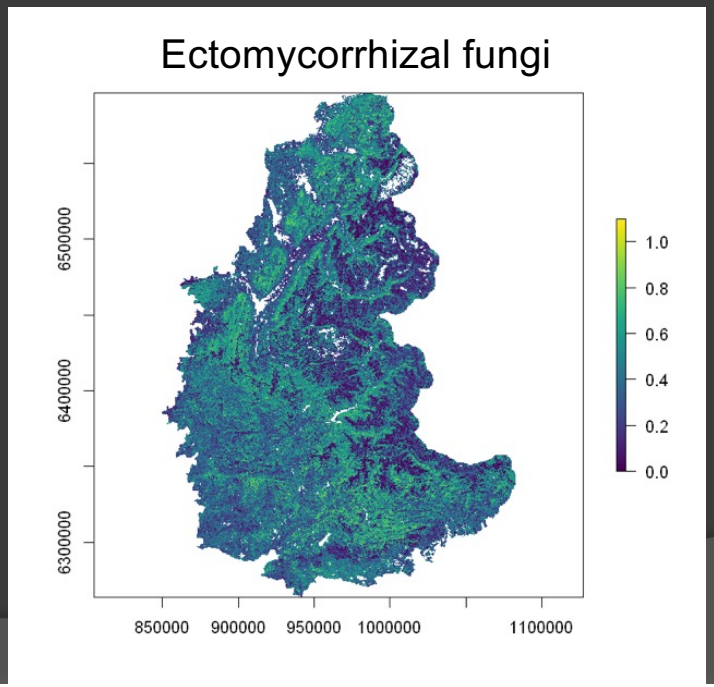


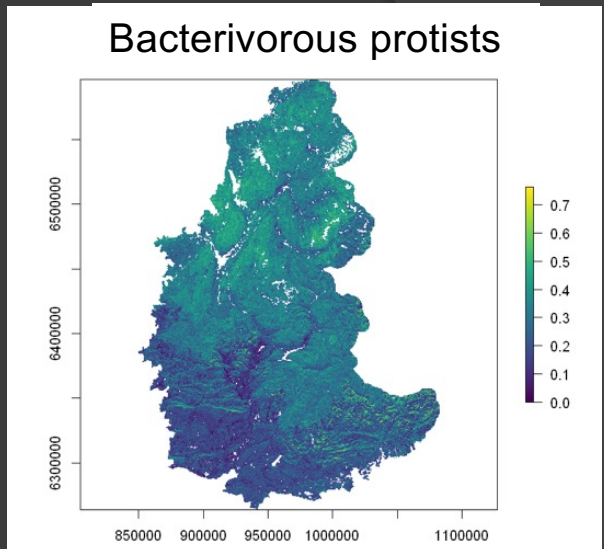


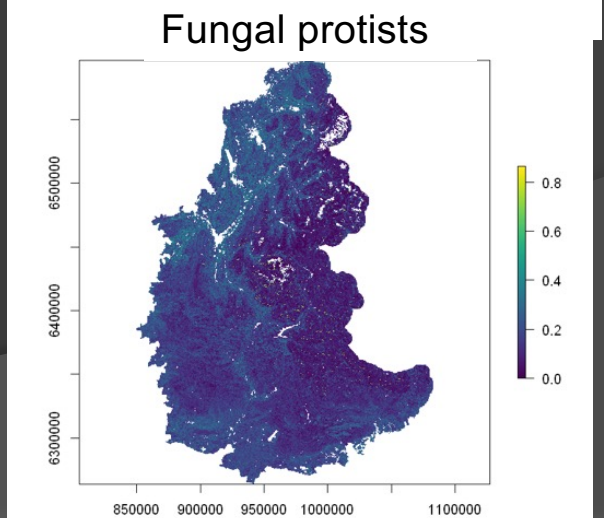
Predicting soil diversity in space across the French Alps

Trophic group distributions

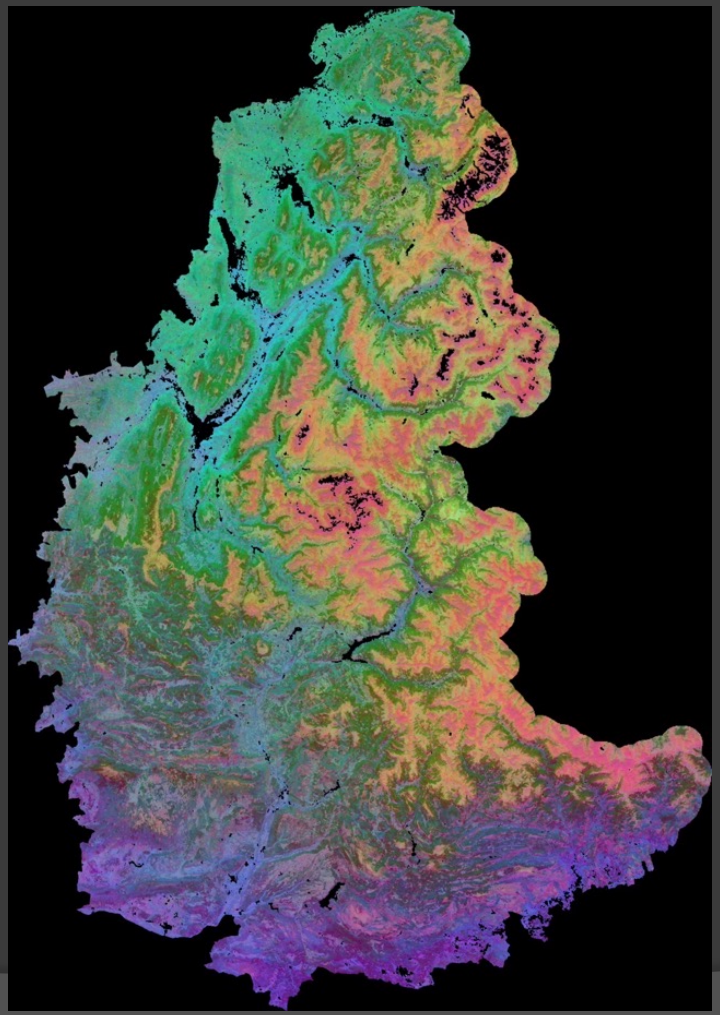


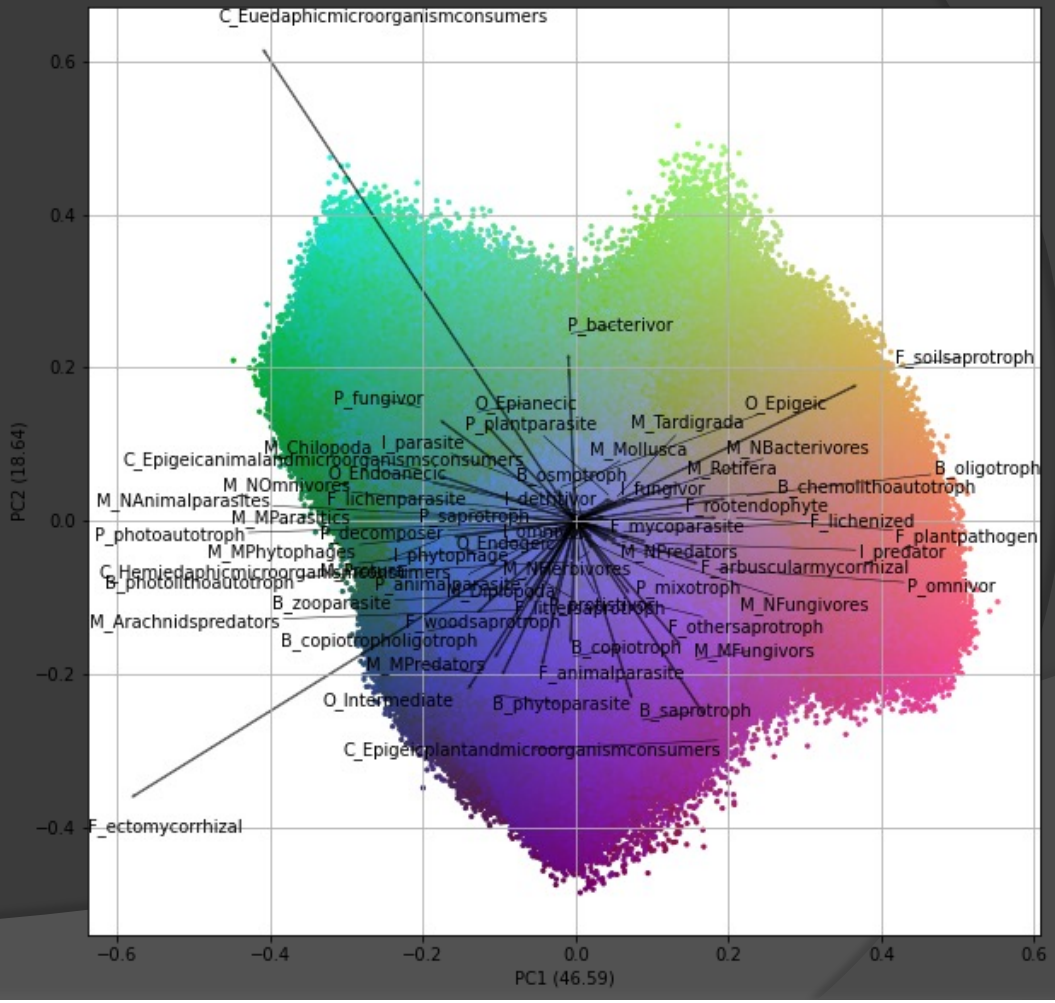






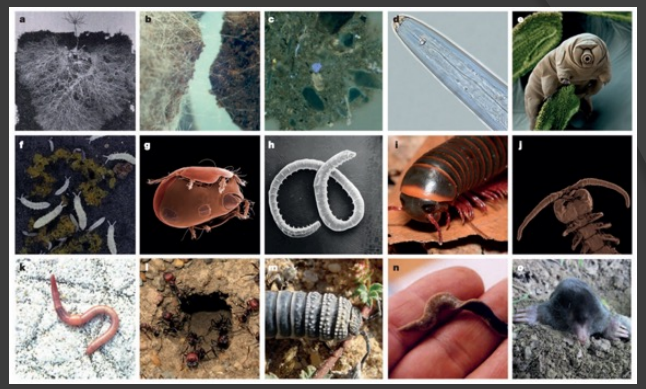
Predicting soil diversity in space across the French Alps

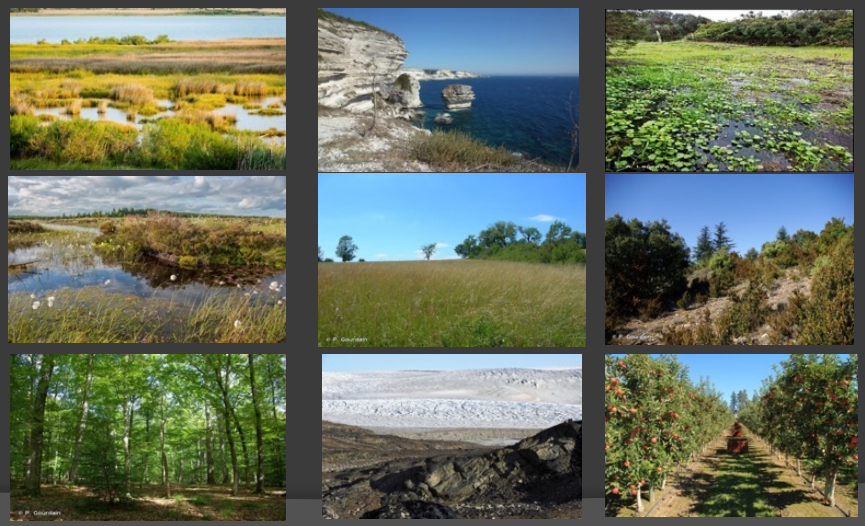




Case study 1

Predicting soil diversity in space across the French Alps

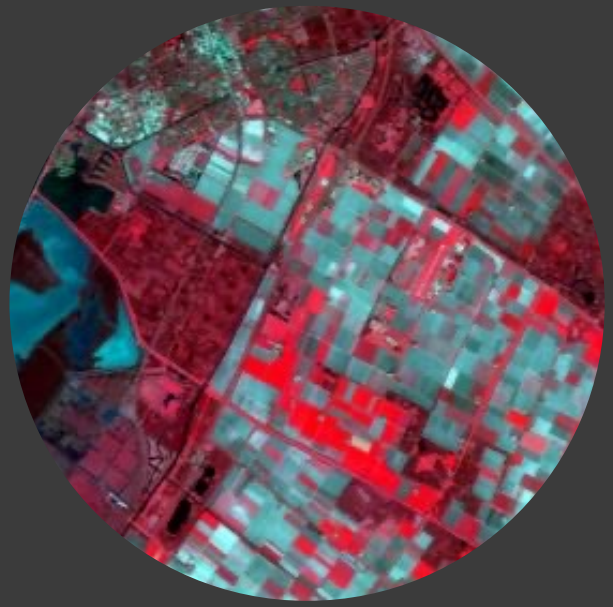


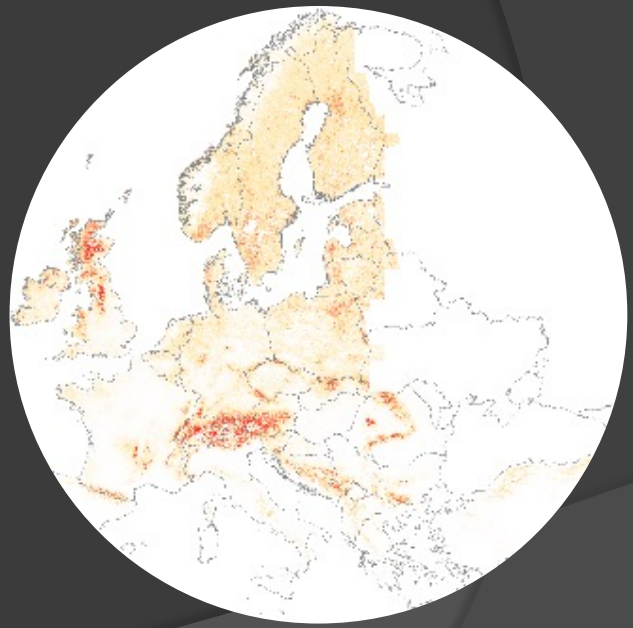


Case study 2

Predicting habitat in space across Europe

European & regional habitat modelling





esa UNIVERSITY OF TWENTE. ITC vito cnrs WAGENINGEN UNIVERSITY & RESEARCH LUND UNIVERSITY UNEP WCMC SYKE University of Zurich RSL

European & regional habitat modelling

Context:

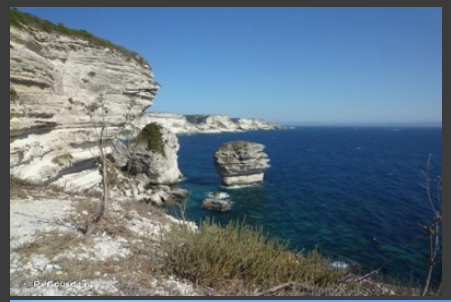
- Strong decline in biodiversity partly due to the degradation of the ecosystems and habitats that support it (27% with unfavorable conservation status).
- EU's biodiversity strategy for 2030 and the new Nature Restoration law aims to protect nature and reverse the degradation of ecosystems (restore > 15%).

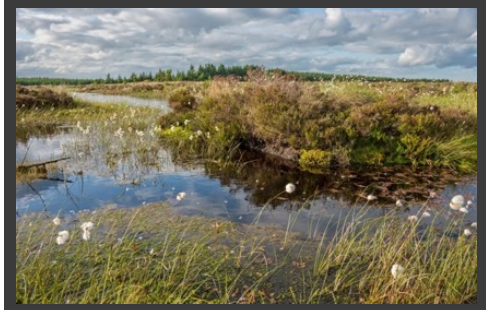
-> WANTED! Information about distribution and extent of habitats and ecosystems to inform policy makers.

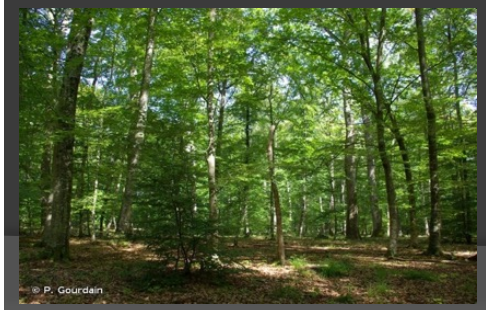
-> Remote-sensing offers a unique opportunity to monitor the extend and state of habitats.

European & regional habitat modelling

EUNIS habitat classification system




Code 2021	Name 2021
MA	Marine habitats (#11)
N	Coastal habitats (#25)
P	Freshwater habitats(#10)
Q	Wetlands (#20)
R	Grasslands and lands dominated by forbs, mosses or lichens (#54)
S	Heathland, scrub and tundra (#43)
T	Forest and other wooded land (#46)
U	Inland habitats with no or little soil and mostly with sparse vegetation(#26)
V	Vegetated man-made habitats (#12)

VEGETATION SURVEY | Open Access

EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats

Milan Chytrý, Lubomír Tichý, Stephan M. Hennekens, Ilona Knollová, John A. M. Janssen,

European & regional habitat modelling

Hierarchy

EUNIS level 1

- MA *Marine habitats*
- N *Coastal habitats*
- P *Inland waters*
- Q *Wetlands*
- R *Grasslands and lands dominated by forbs, mosses or lichens*
- S *Heathlands, scrub and tundra*
- T *Forests and other wooded land*
- U *Inland habitats with no or little soil and mostly with sparse vegetation*
- V *Vegetated man-made habitats*

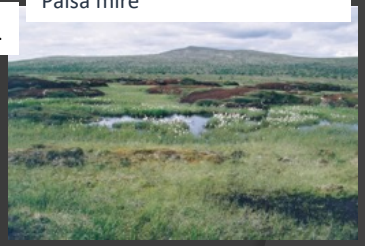
EUNIS level 2

- N1 Coastal dunes and sandy shores
- N2 Coastal shingle
- N3 Rock cliffs, ledges and shores, including the supralittoral
- Q1 Raised and blanket bogs
- Q2 Valley mires, poor fens and transition mires
- Q3 Palsa and polygon mires
- Q4 Base-rich fens and calcareous spring mires
- Q5 Helophyte beds
- Q6 Periodically exposed shores
- T1 Broadleaved deciduous forests
- T2 Broadleaved evergreen forests
- T3 Coniferous forests

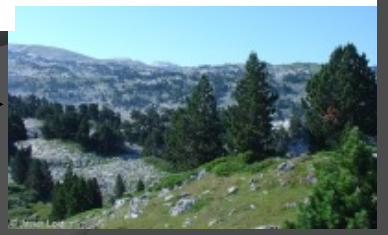
EUNIS level 3

N32 Mediterranean and Black Sea rocky sea cliff and shore

Q31 Palsa mire



T34 Temperate subalpine Larix, Pinus cembra and Pinus uncinata forest



European & regional habitat modelling

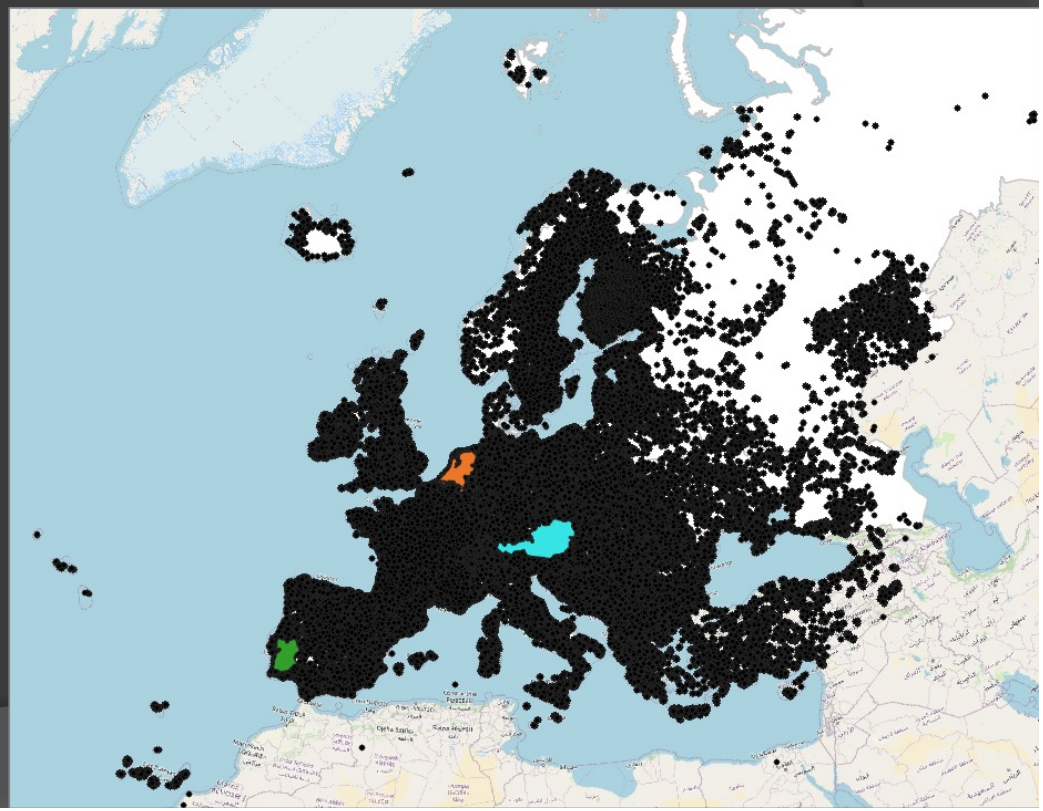
DATA

Training:

Eur. Vegetation Archive (EVA)
2M+ plots, only part used
2000-todate
Location uncertainty<100m

Evaluation with local EUNIS maps:

- Netherlands
- Austria
- South Portugal

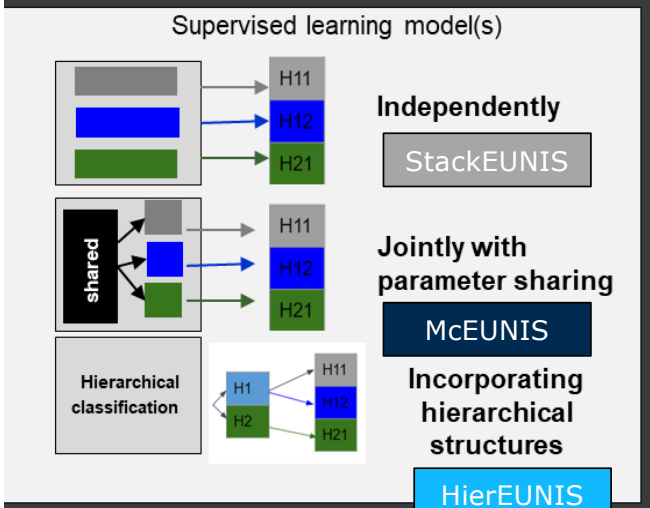


European & regional habitat modelling

- Large-scale assessment of habitat modelling approaches
- Use of advanced RS products such as HR-VPP
- Evaluation of a broad range of algorithmic approaches and classification strategies
- Production of high resolution maps of EUNIS habitats

European & regional habitat modelling

Question: What is the best way to model all EUNIS habitats at level 3 ?



Stacked EUNIS

200+ models

HierEUNIS

9 models

- **By dividing the problem into less ambiguous sub-problems ?**

How to resolve the ambiguities of the multi-class model?

- **By incorporating landscape (landcover composition and spatial layout) information ?**

Multi-class EUNIS

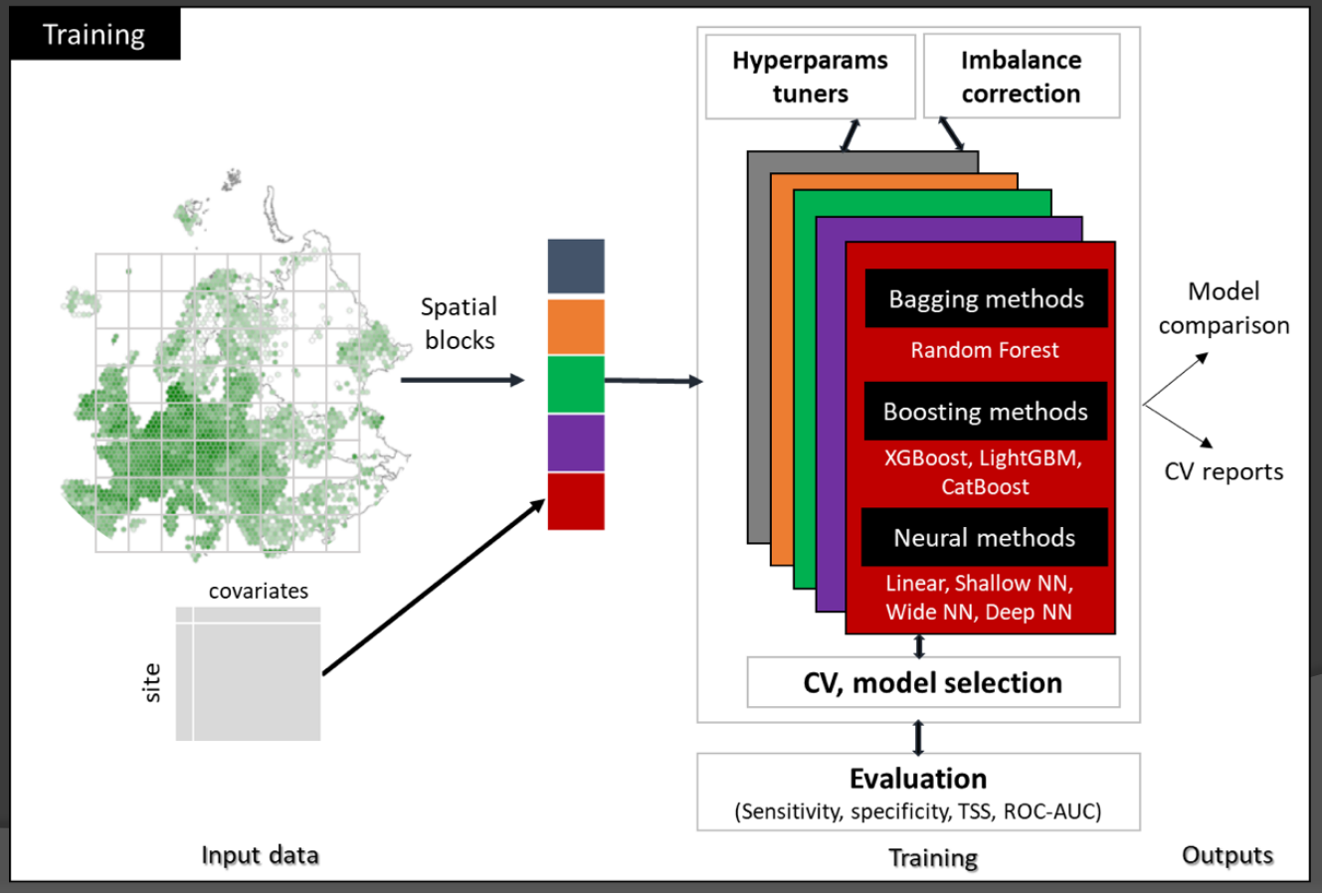
1 global model

Landscape composition → McEUNIS

Landscape structure → DeepEUNIS

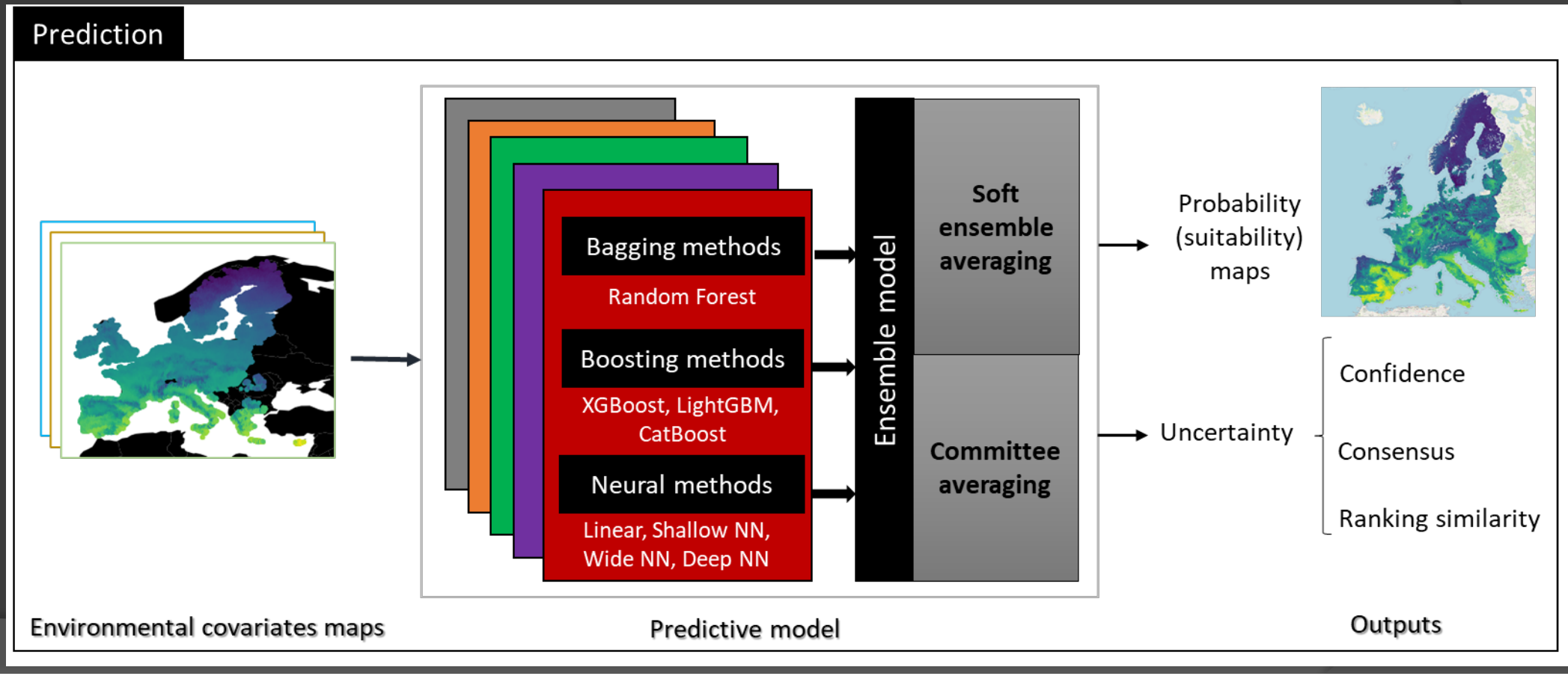
European & regional habitat modelling

General modelling framework – training phase on point data



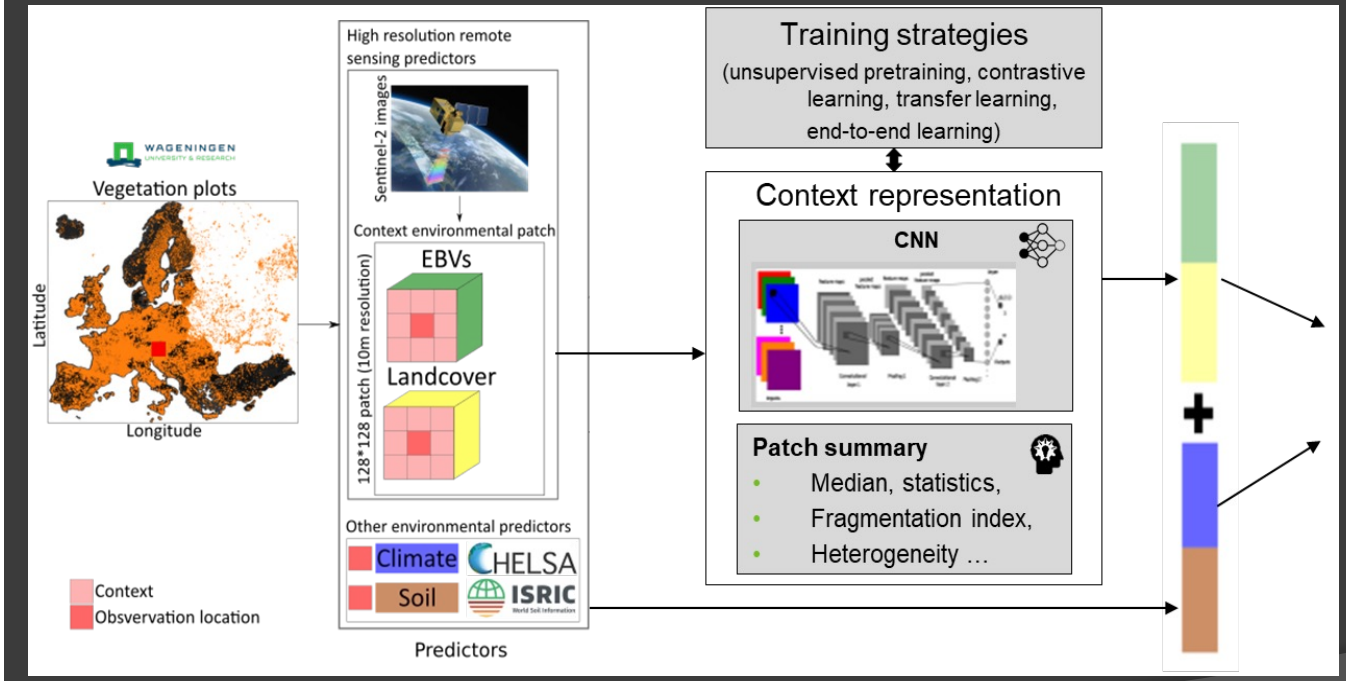
European & regional habitat modelling

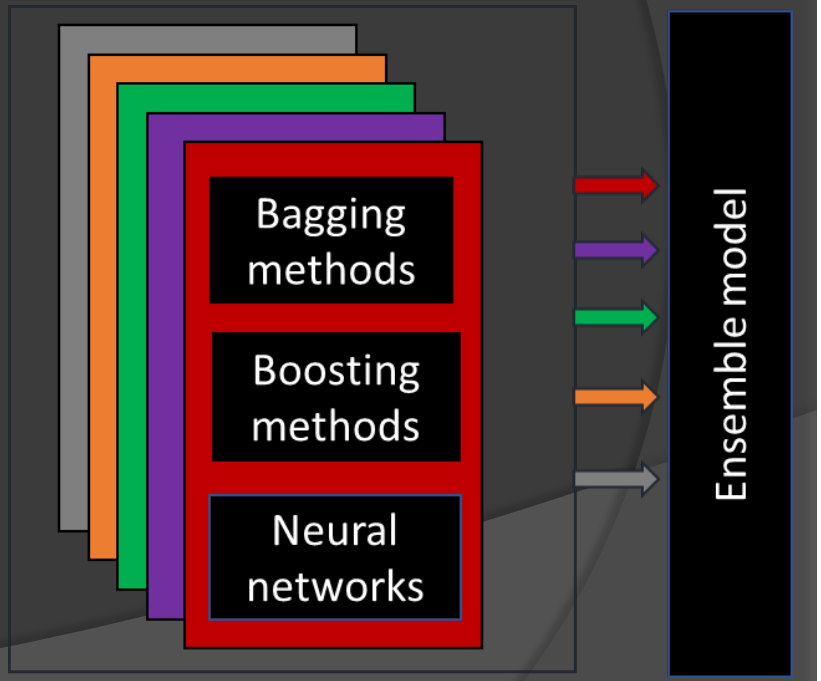
General modelling framework – predicting phase on raster data



European & regional habitat modelling

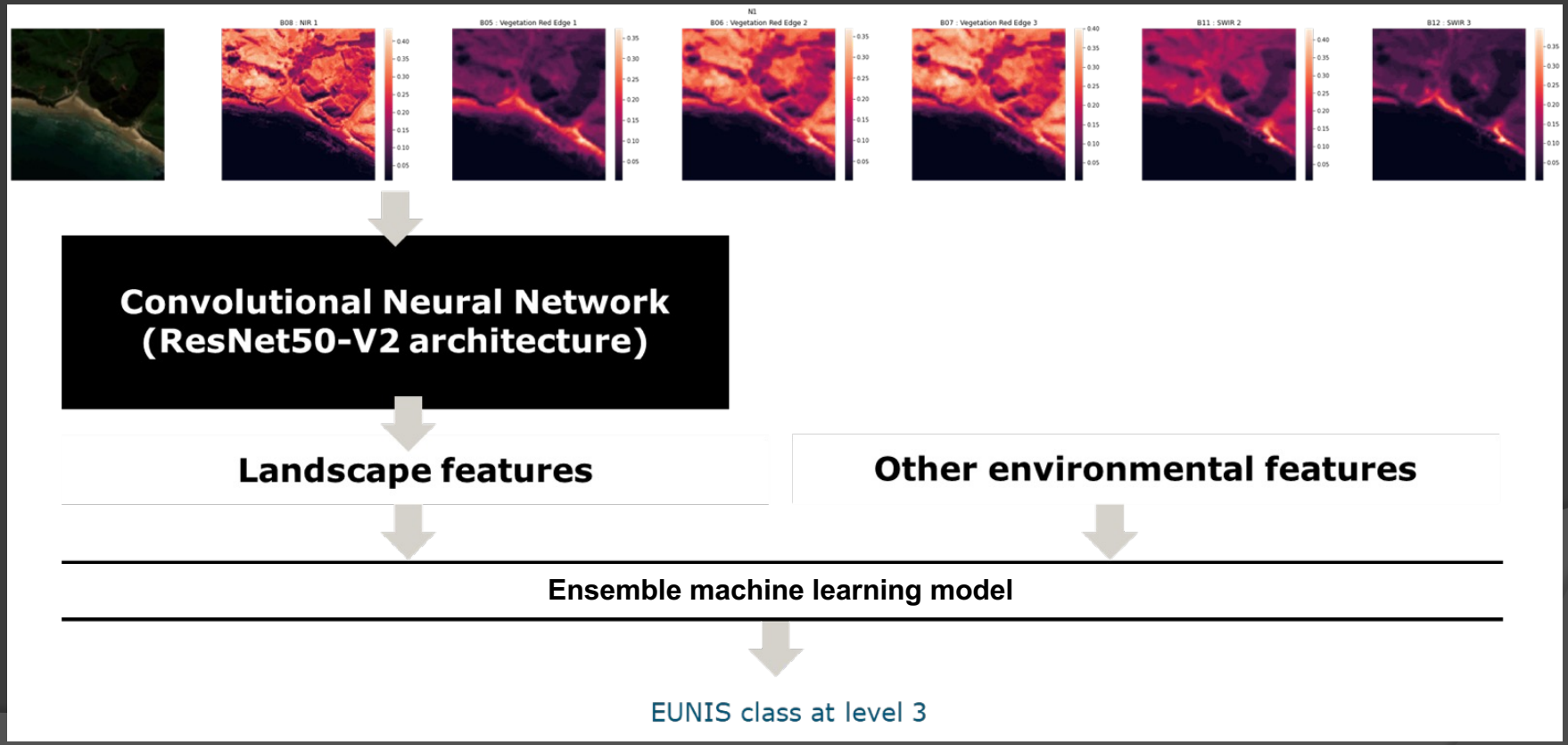
Landscape representation





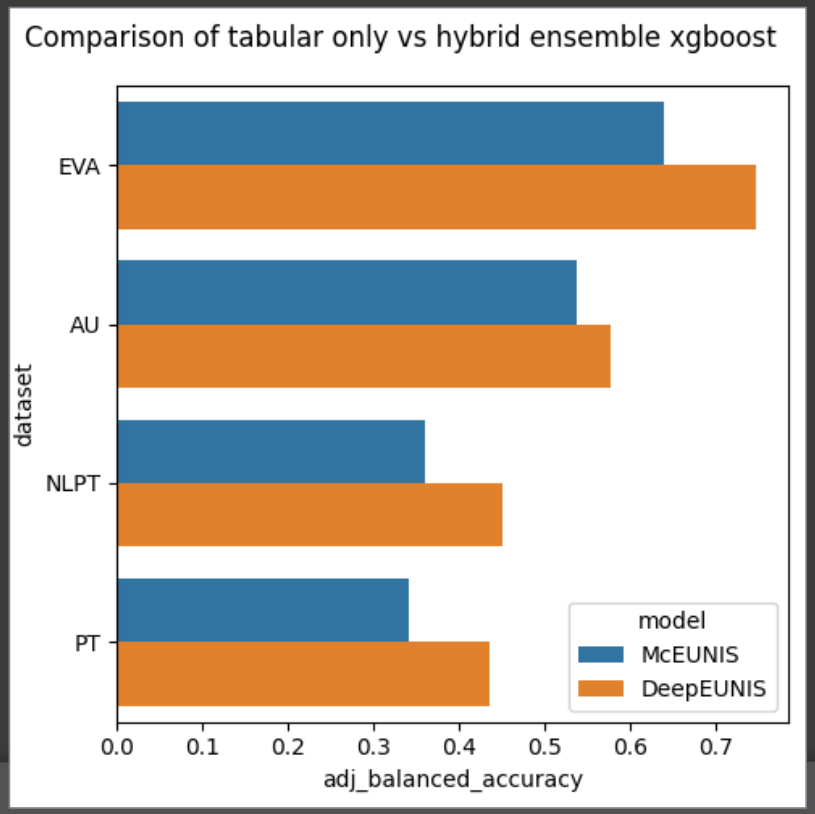
European & regional habitat modelling

Learning landscape features using a deep learning model trained on MSI patches



European & regional habitat modelling

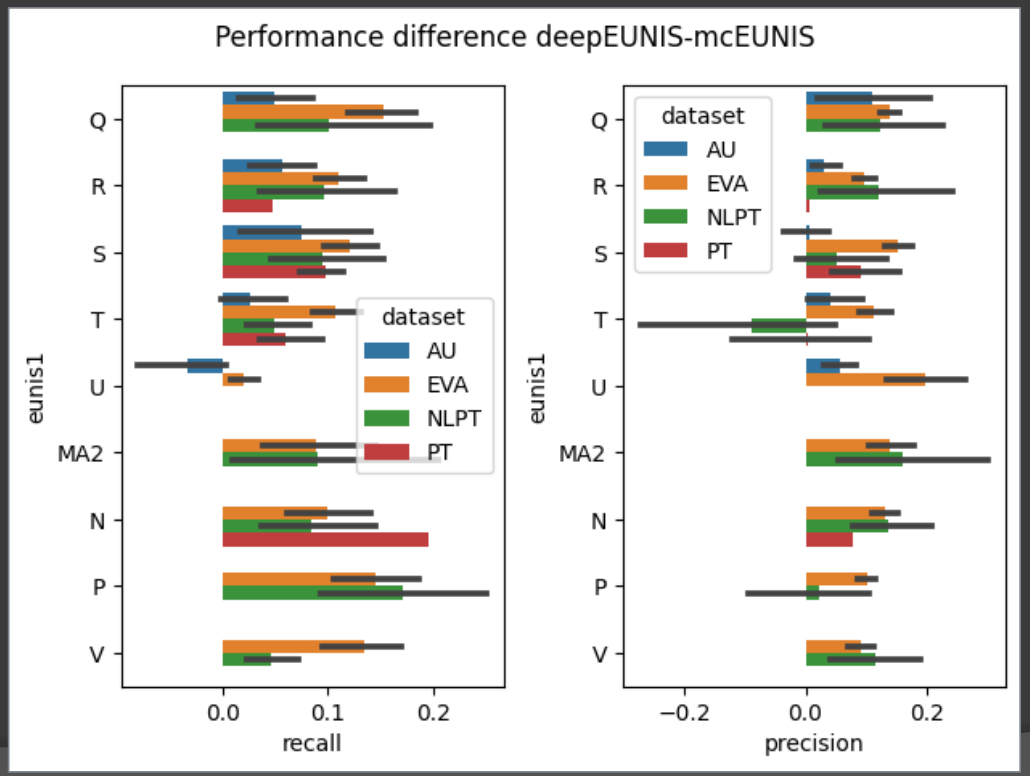
McEUNIS vs DeepEUNIS: does incorporating multi-spectral information and its spatial layout improve habitat discrimination ?



Yes! In general, accounting for habitat structure through MSI information and DL model work better

European & regional habitat modelling

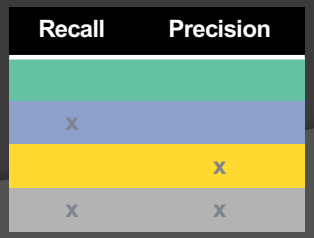
McEUNIS vs DeepEUNIS: does incorporating multi-spectral information and its spatial layout improve habitat discrimination ?

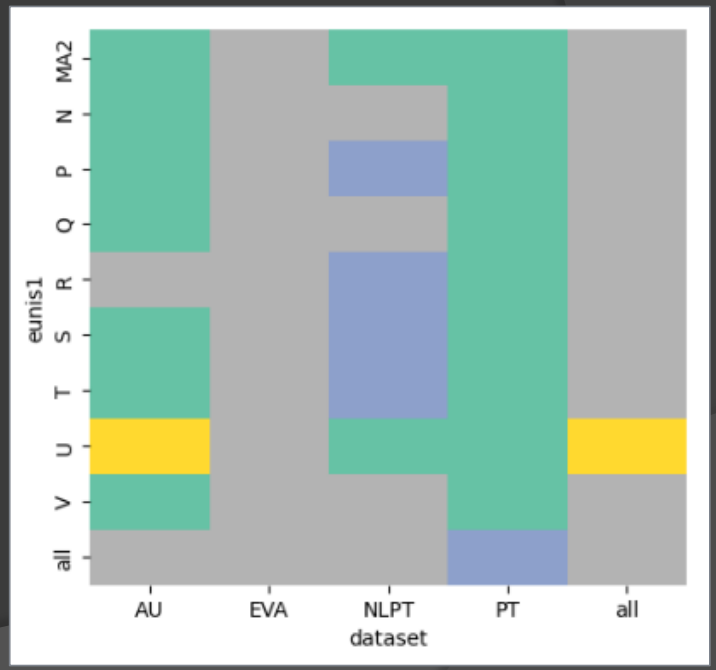


Wilcoxon ranking test:

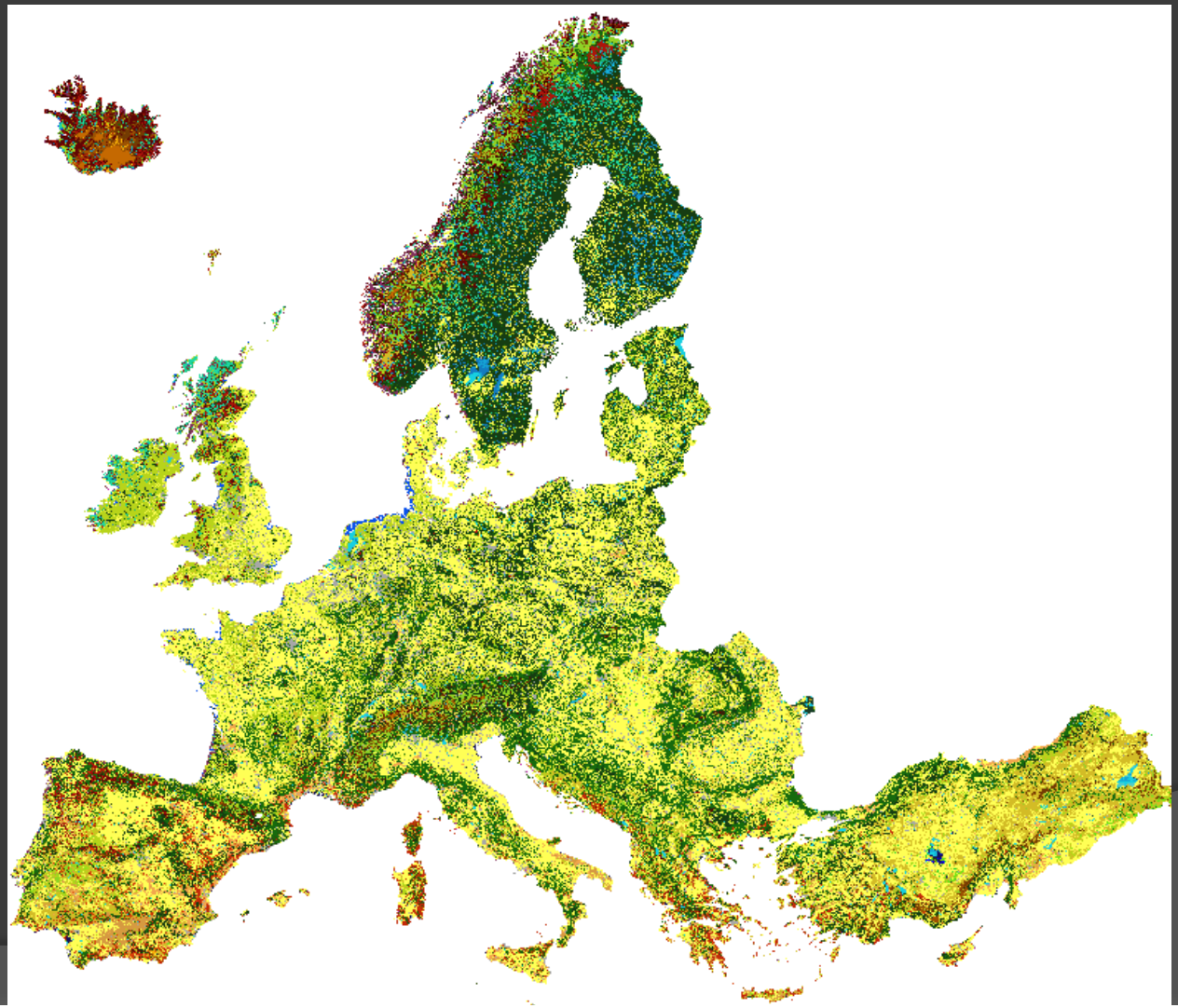
« deepEUNIS > mcEUNIS »

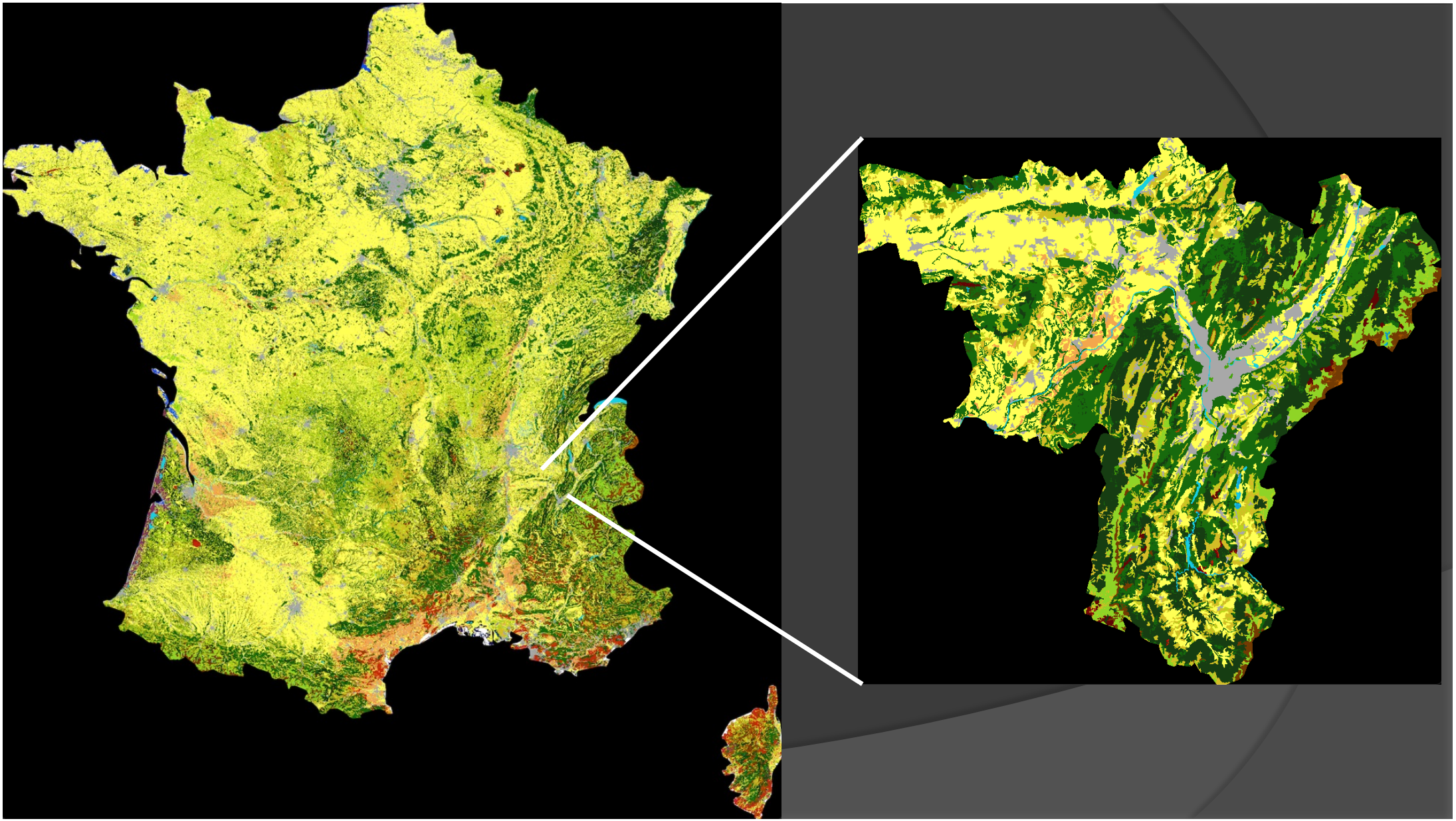
α = 0.01



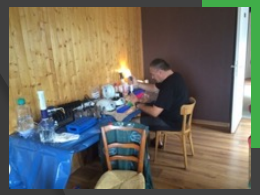


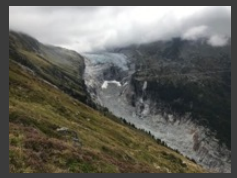
European & regional habitat modelling

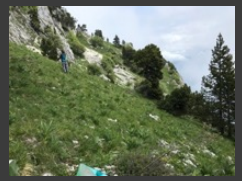


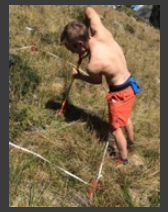


Remerciements









cnrs

OSUG
Observatoire des Sciences de l'Univers de Grenoble

erc

AGENCE NATIONALE DE LA RECHERCHE
ANR

MIAI
Grenoble Alpes
Multidisciplinary Institute in Artificial intelligence

OFB
OFFICE FRANÇAIS DE LA BIODIVERSITÉ

Horizon Europe