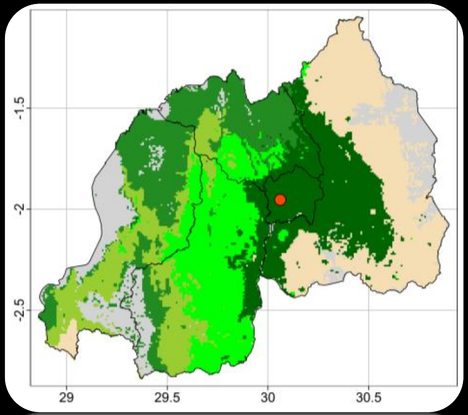
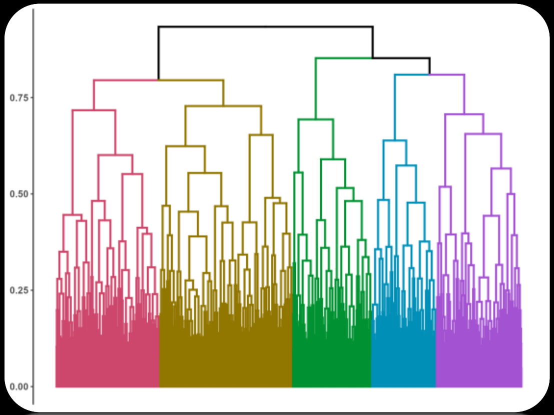
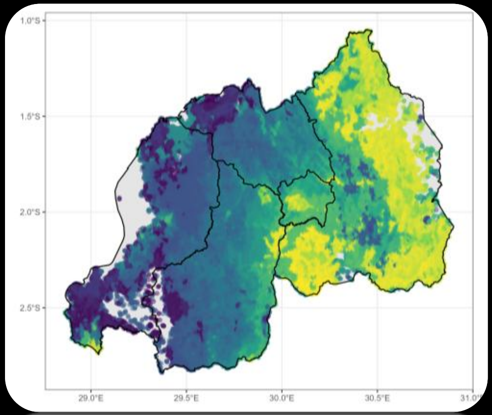
Practical Training on Data-Driven Systems Agronomy





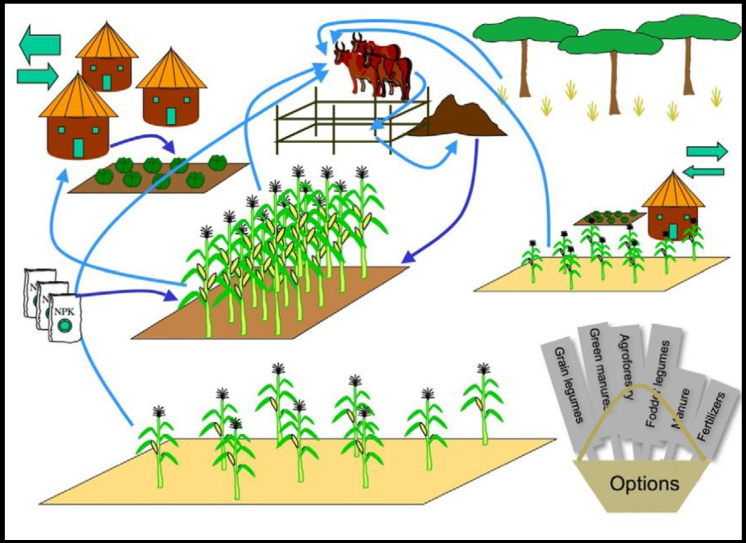


Frédéric Baudron

UM6P, 26 May 2025

Systems agronomy

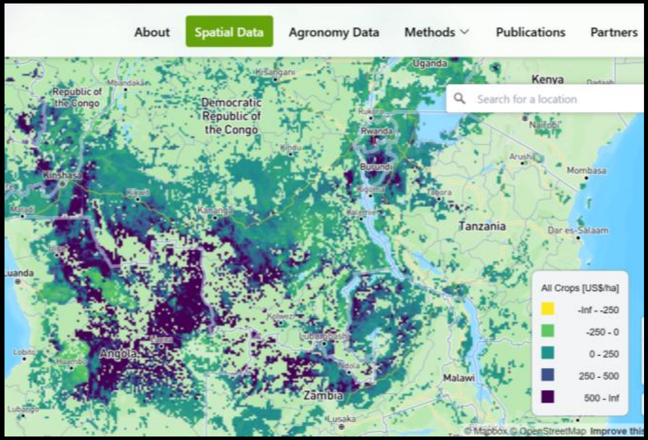
- A methodological approach which seeks to understand cropping, farm household, and farming systems and their interactions, and design a broad basket of options for diverse farming conditions and diverse contexts *(Giller et al., 2011)*
- Which options for what context?
  - Where? For whom? What?



(Giller et al. 2011)

Agronomy-at-scale

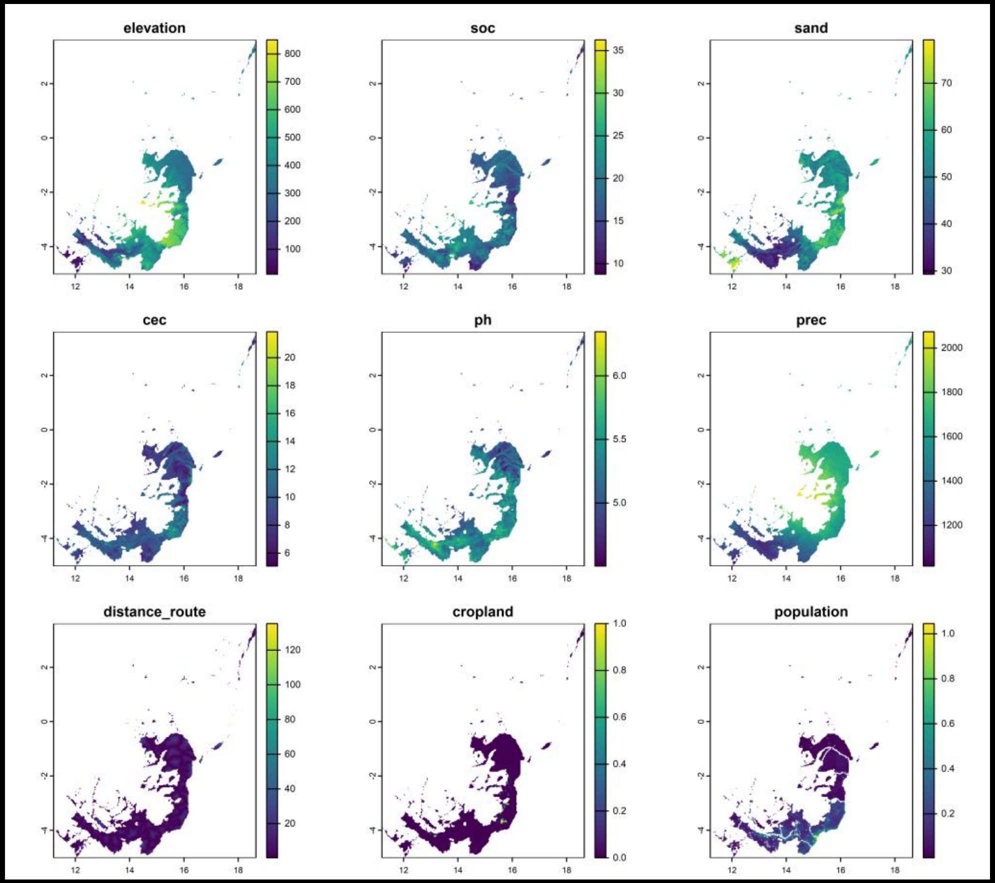
- The use of digital data, geo-spatial tools, analytics, workflows, and interfaces to develop, validate, and disseminate agronomy solutions at scale *(Vanlauwe et al. 2020)*
  - Answering the questions of where? for whom? and what? in a data-driven manner and at scale

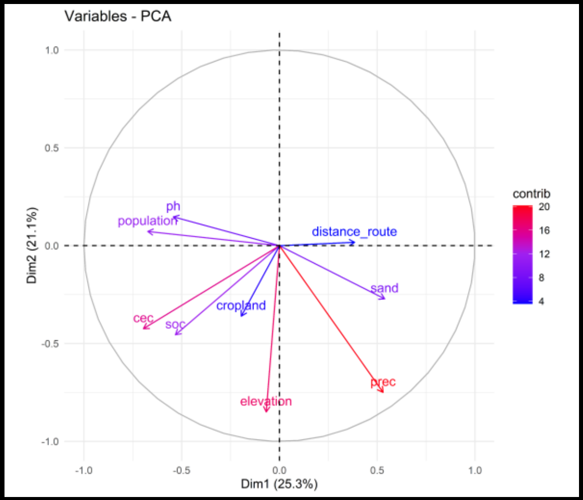


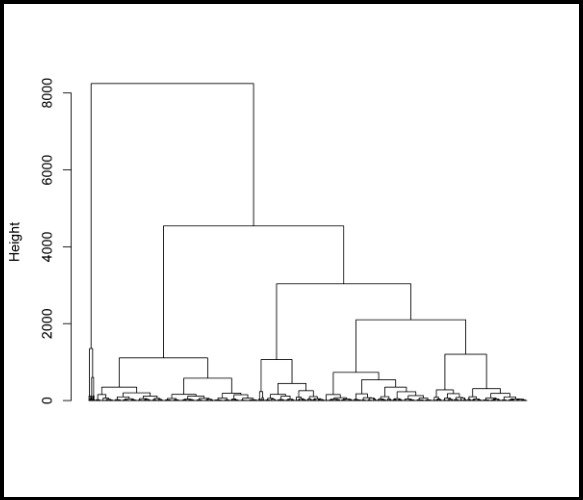
WHERE? Recommendation domains

- Recommendation domains: sections of landscapes sharing similar biophysical and socioeconomic characteristics, where specific interventions *(e.g., technologies, policies, investments)* are likely to be effective *(Muthoni et al., 2017; Notenbaert et al., 2013)*
- Top-down approach: deductive method of clustering using (increasingly available) gridded geospatial layers
- Bottom-up approach: using geospatial models trained with adoption or performance georeferenced data

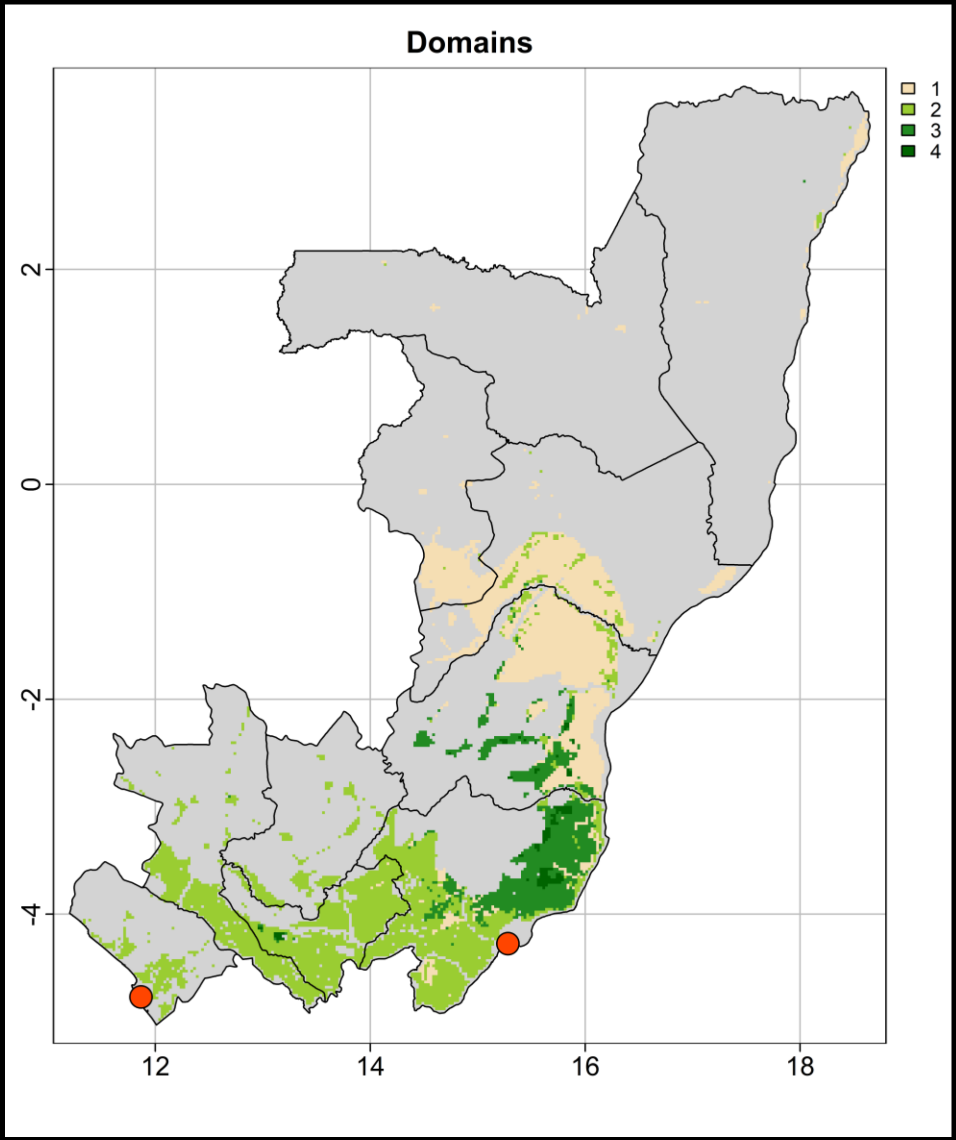
Clustering (PCA + HCA)

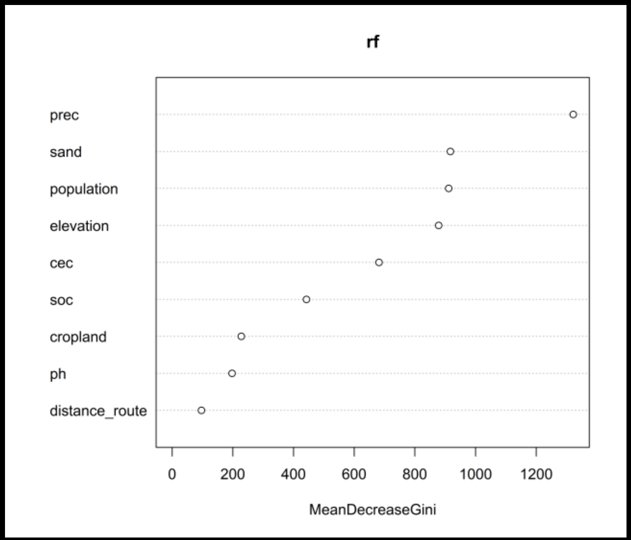


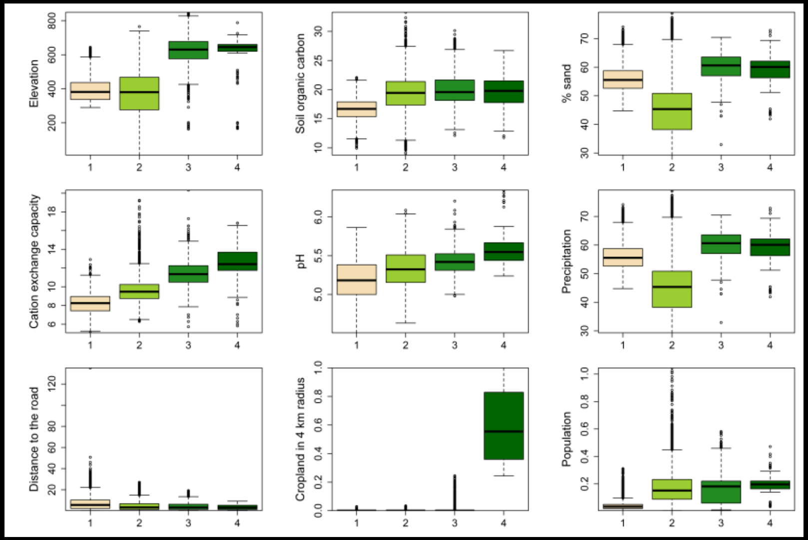




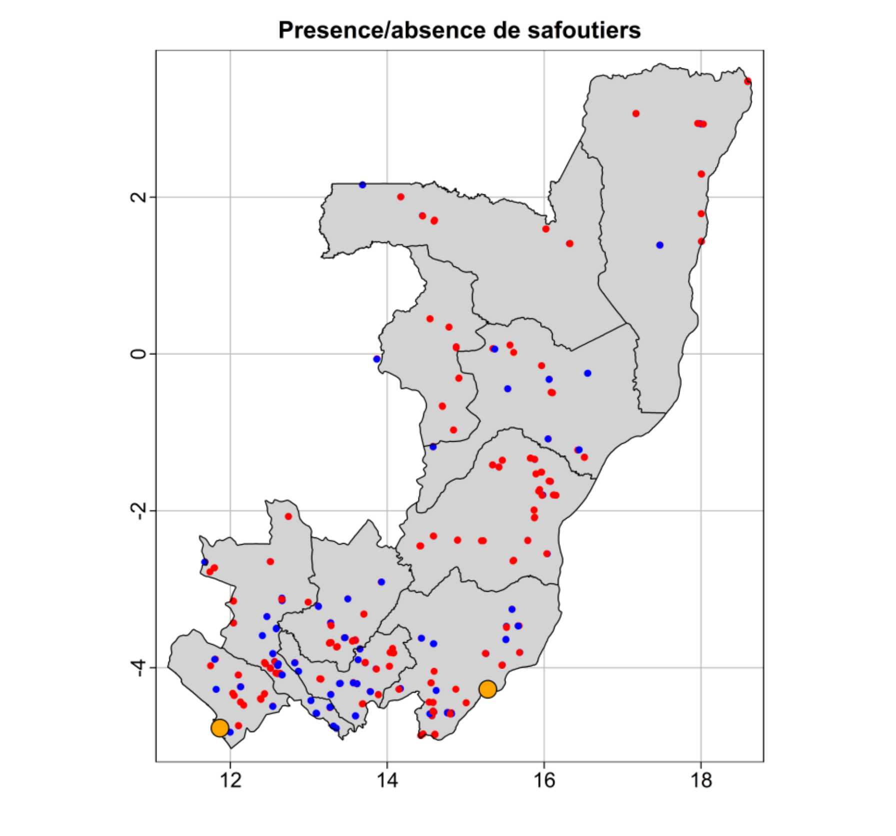
Clustering (PCA + HCA)

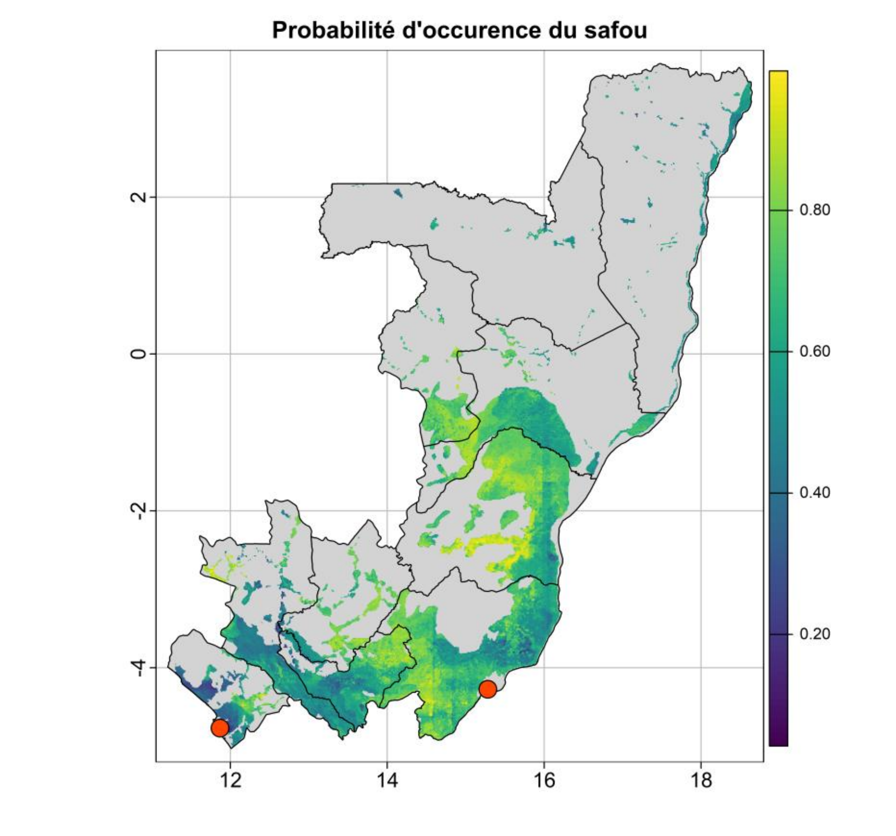






Bottom-up approach

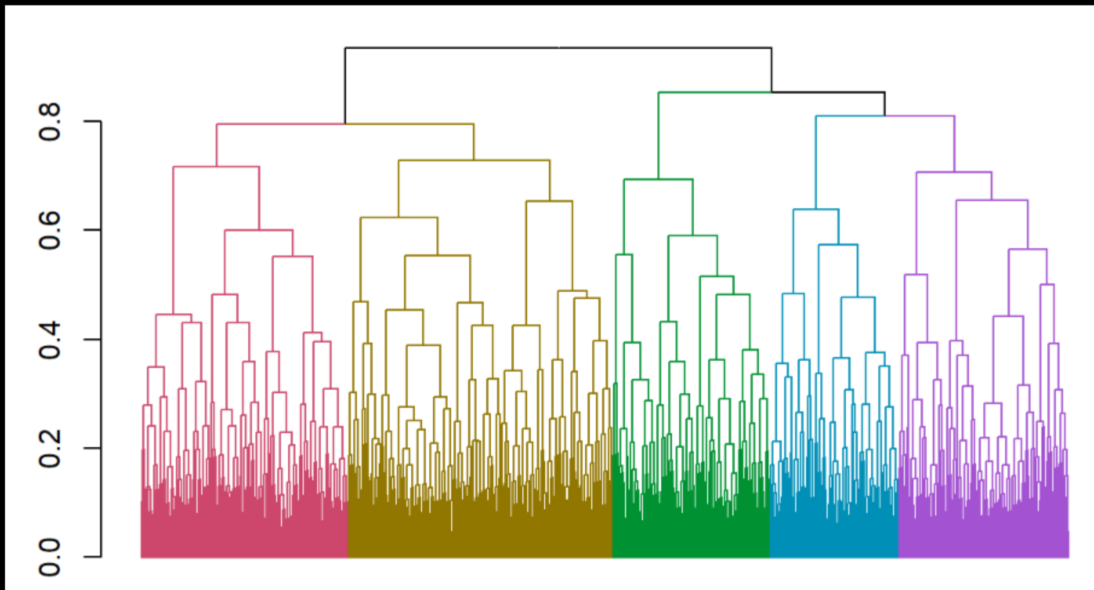


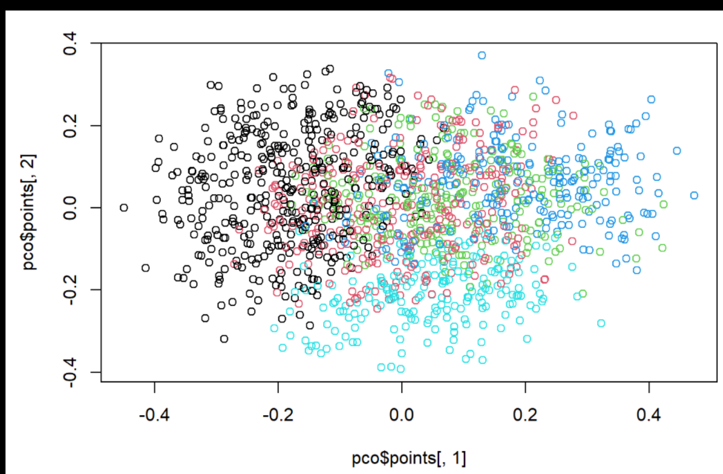


FOR WHOM? Farm typologies

- Farm typologies: characterizing the diversity of farming systems and their distribution in heterogenous communities, as a basis for prioritizing interventions *(Hammond et al., 2020)*
- Quantitative statistical typologies, without a priori selection of discriminant variables *(Tittonell et al., 2010)*
- Qualitative expert-based typologies involving farmers and other knowledgeable stakeholders, based on a specific research hypothesis *(Kebede, 2009)*

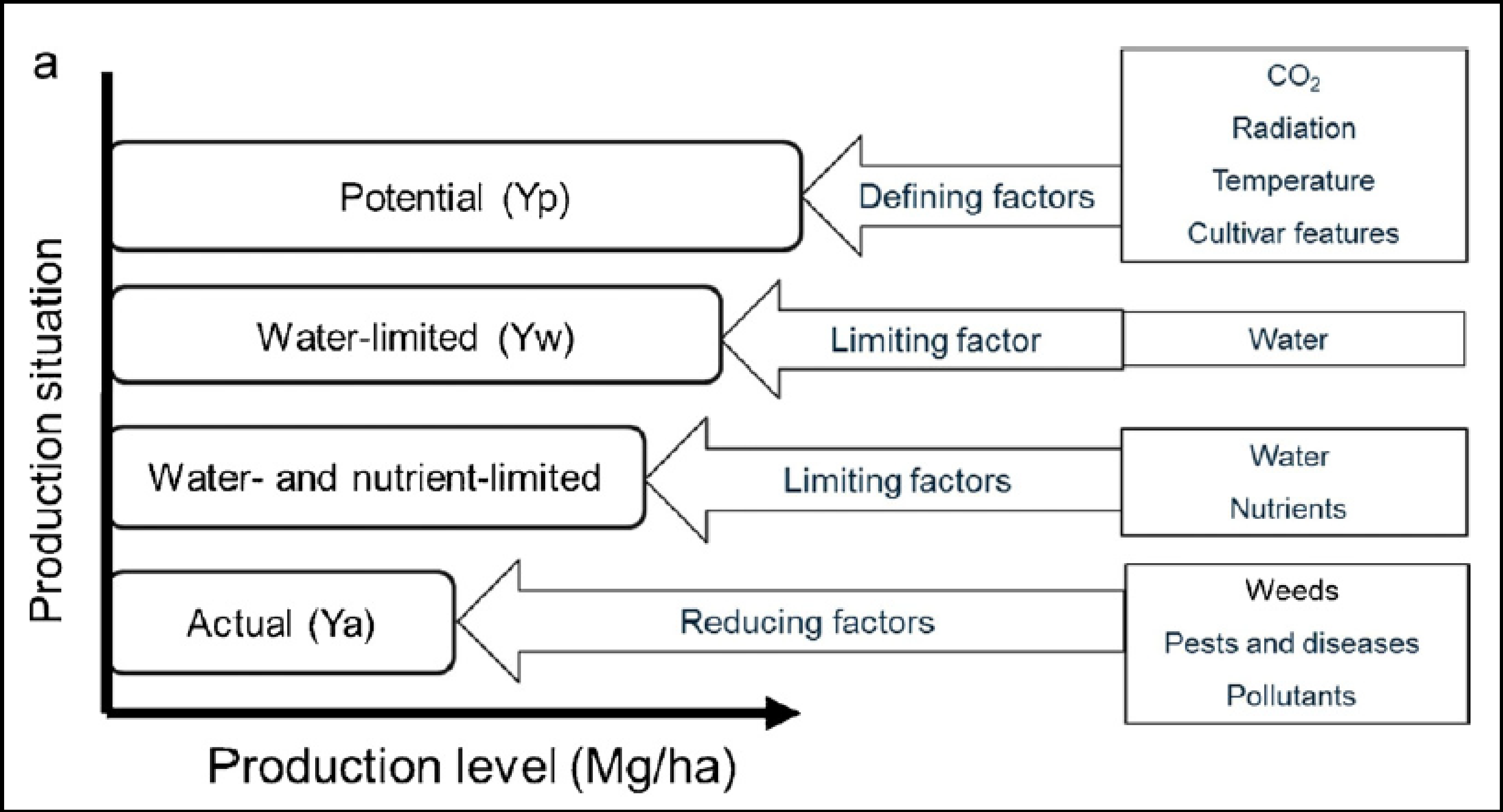
Clustering (nMDS + HCA) and interpretation





- **Type 1:** migrants, with low income, renting land, mainly in savanna biome.
- **Type 2:** land owners with low income (vulnerable).
- **Type 3:** Land owners with high income, diverse farms, perenial crops, stable cultivated area.
- **Type 4:** Land owners with high income, diverse farms, perenial crops, expanding.
- **Type 5:** pluriactivity, mainly sbsistence farming, high income from natural resources, mainly in forest and mosaic biomes.

WHAT? Yield gap analysis

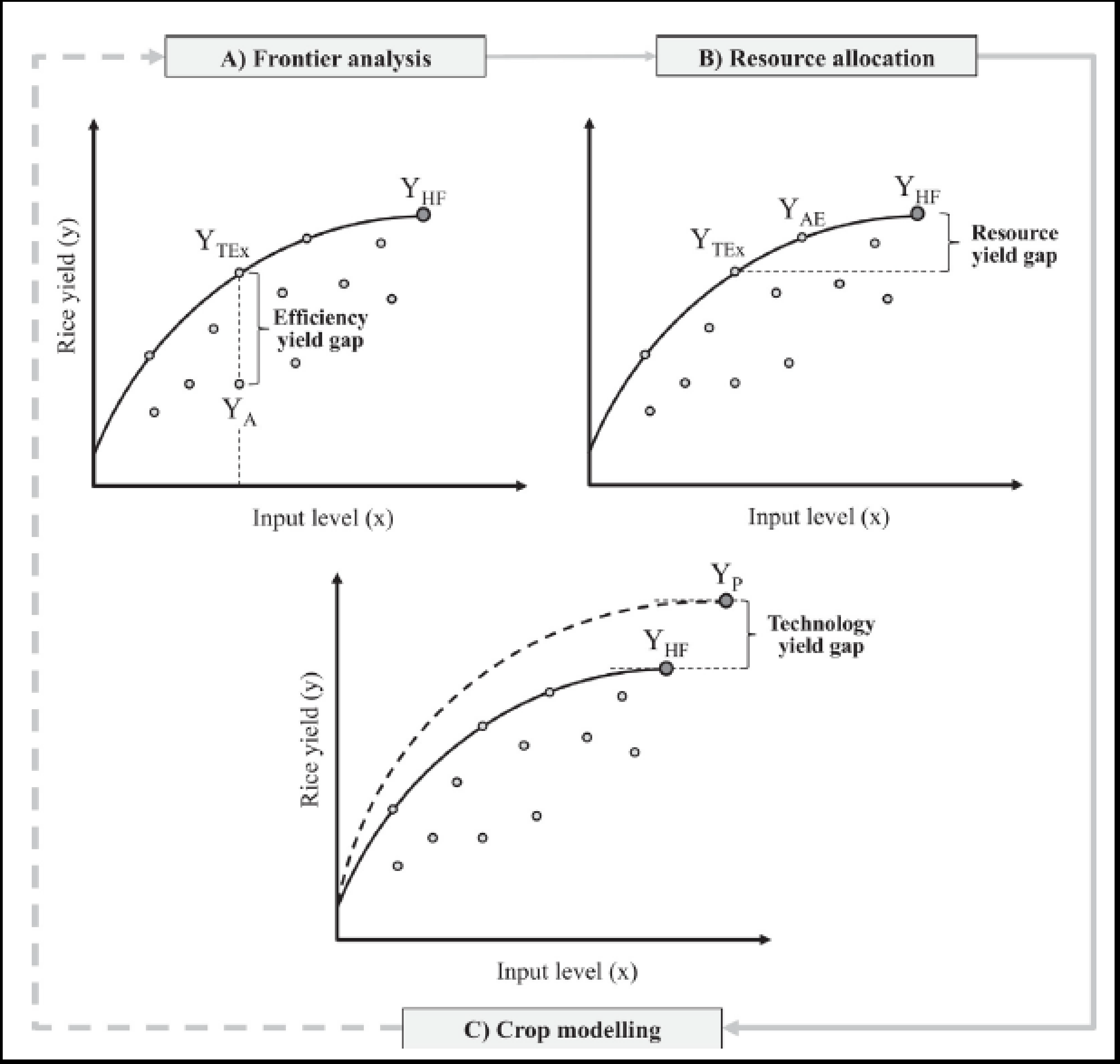


(van Ittersum et al. 2013)

WHAT? Yield gap analysis

- Frontier analysis to decompose yield gaps (efficiency, resource and technology yield gap) and understand the relative importance of biophysical and crop management factors explaining these gaps *(Silva et al., 2017)*
- Machine learning using random forest and post hoc calculation of SHapley Additive exPlanations (SHAP) values, reflecting the contribution of each variable to the yield in each field *(Nayak et al., 2024).* Spatial predictions using different scenarios.

Stochastic frontier analysis



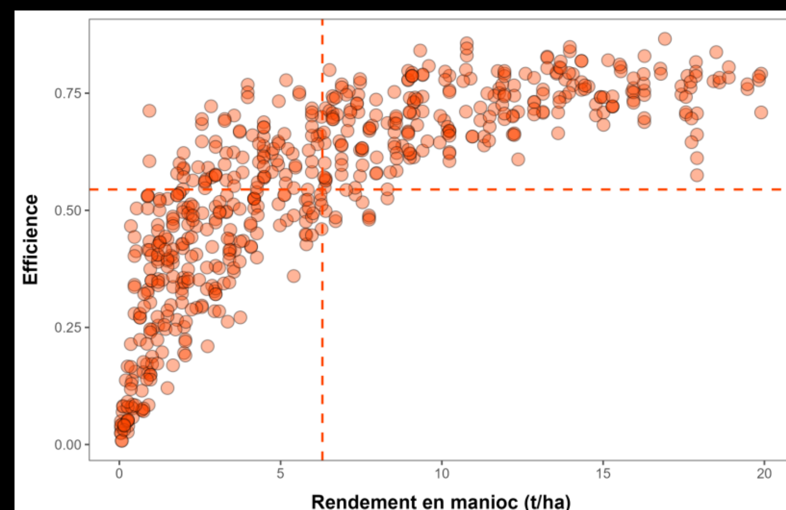
(Silva et al. 2017)

Results

Tableau 1 – effets des différents facteurs biophysiques et de gestion de la culture sur le rendement en manioc. Le model a été construit avec la fonction sfa() to package R frontier. Les facteurs en gras sont ceux ayant un effet significatif à P-value < 0.1.

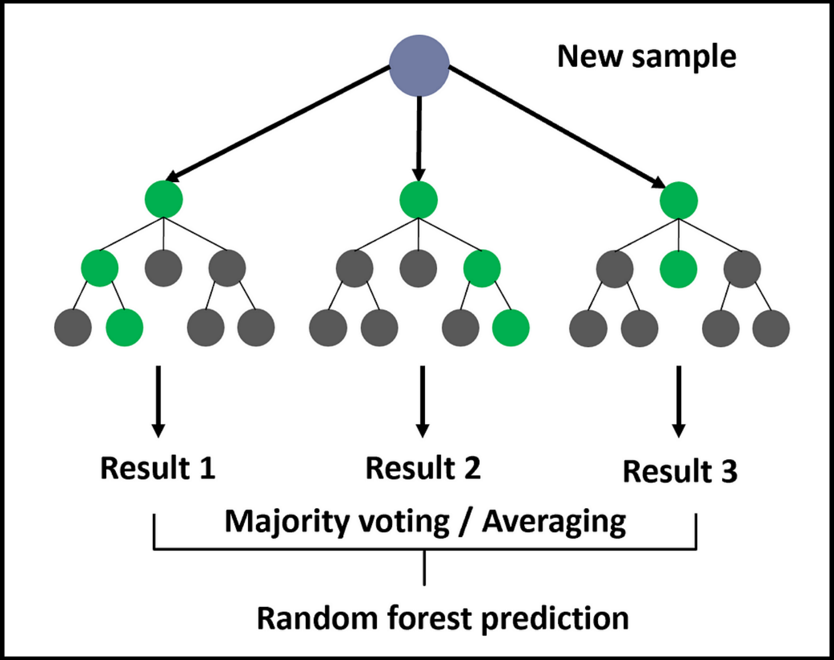
	Coefficient	Erreur std	z value	Pr(>|z|)
Frontière de production				
Intercept	0.310	0.171	1.807	0.07
Sarclage: 2 sarclages	0.478	0.124	3.860	< 0.01
Sarclage: 3 sarclages	0.508	0.139	3.653	< 0.01
Sarclage: 4 et + sarclages	0.251	0.161	1.557	0.12
Sarclage: pas de sarclage	0.213	0.165	1.297	0.19
Végétation précédente: forêt dense	0.029	0.142	0.201	0.84
Végétation précédente: savane arbustive	0.186	0.100	1.856	0.06
Végétation précédente: savane herbacée	-0.268	0.140	-1.907	0.06
Durée de jachère: 1-2 ans	-0.302	0.213	-1.419	0.16
Durée de jachère: 3-5 ans	-0.114	0.115	-0.998	0.32
Durée de jachère: 6-11 ans	0.110	0.114	0.962	0.34
Durée de jachère: Pas de jachère/végétation naturelle	-0.355	0.128	-2.774	0.01
Utilisation du tracteur	0.579	0.202	2.869	< 0.01
Association tubercules	0.072	0.095	0.757	0.45
Association céréales	-0.074	0.092	-0.804	0.42
Association légumineuses	0.012	0.093	0.128	0.90
Association légumes	-0.012	0.083	-0.145	0.88
Association banane plantain	-0.068	0.091	-0.746	0.46
Association arbres (non-fruitiers)	0.233	0.090	2.604	< 0.01
Domaine: mosaïque	-0.348	0.116	-2.994	< 0.01
Domaine: savane	-0.477	0.170	-2.807	< 0.01
Altitude (log)	0.889	0.142	6.279	< 0.01
Carbone organique du sol (log)	0.655	0.257	2.551	0.01
Sable (log)	0.141	0.368	0.384	0.70
pH (log)	-2.738	1.394	-1.965	0.05
Température moyenne (log)	8.249	1.707	4.832	< 0.01
Précipitation	-3.657	0.567	-6.448	< 0.01
Effets d'inefficience				
Intercept	-97.578	303.843	-0.321	0.75
Surface cultivée (log)	3.406	9.819	0.347	0.73
Propriété	15.710	47.735	0.329	0.74
Evaluation du modèle				
$\sigma^2 = \sigma^2_v + \sigma^2_u$	73.021	219.220	0.333	0.74
$\gamma = \sigma^2_u / \sigma^2$	0.995	0.014	72.585	< 0.01

- Mean YLD = 6295 kg/ha
- HFY = 16864 kg/ha
- Efficiency = 54.4 %
- Mean tech. eff. YLD = 10439 kg/ha
- Mean eff. YG = 3970 kg/ha
- Mean res. YG = 7083 kg/ha

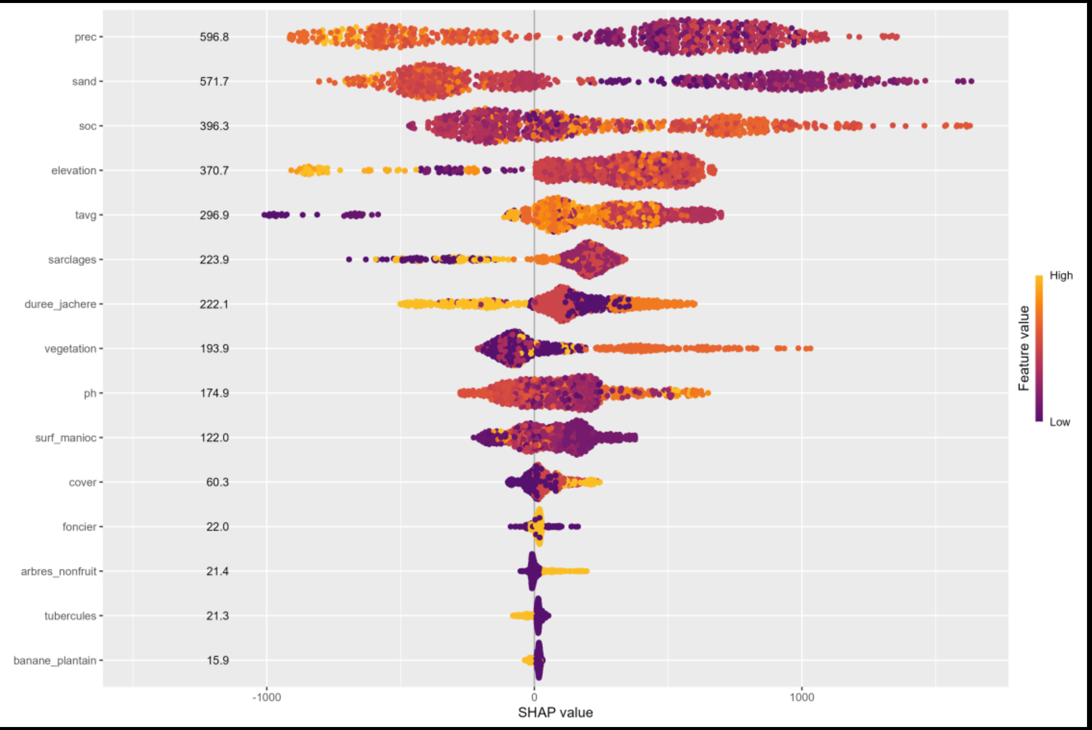


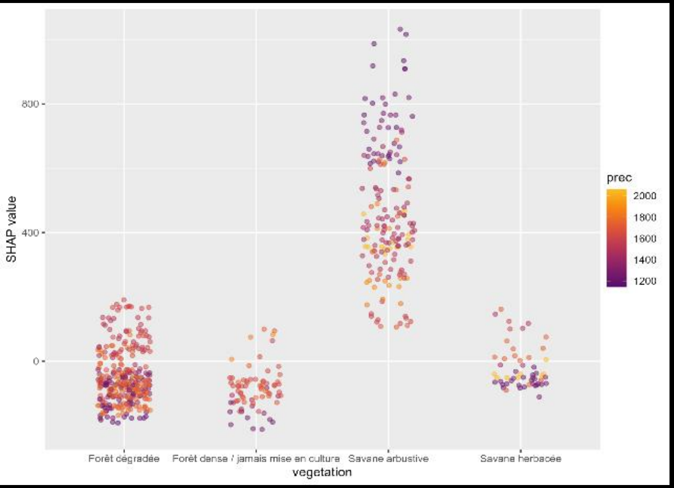
Random forests

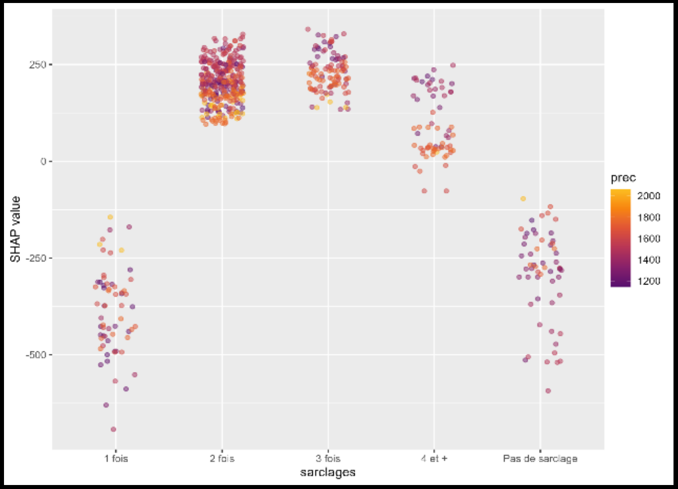
1. Creation of ntree bootstrap samples
2. Construction of one tree per sample using mtry explanatory variables selected randomly
3. Calculation of OOB erreur *(mean square error or % of observation misclassified)*
4. Aggregation of predictions *(means or majority voting)*
5. Calculation of mean OOB



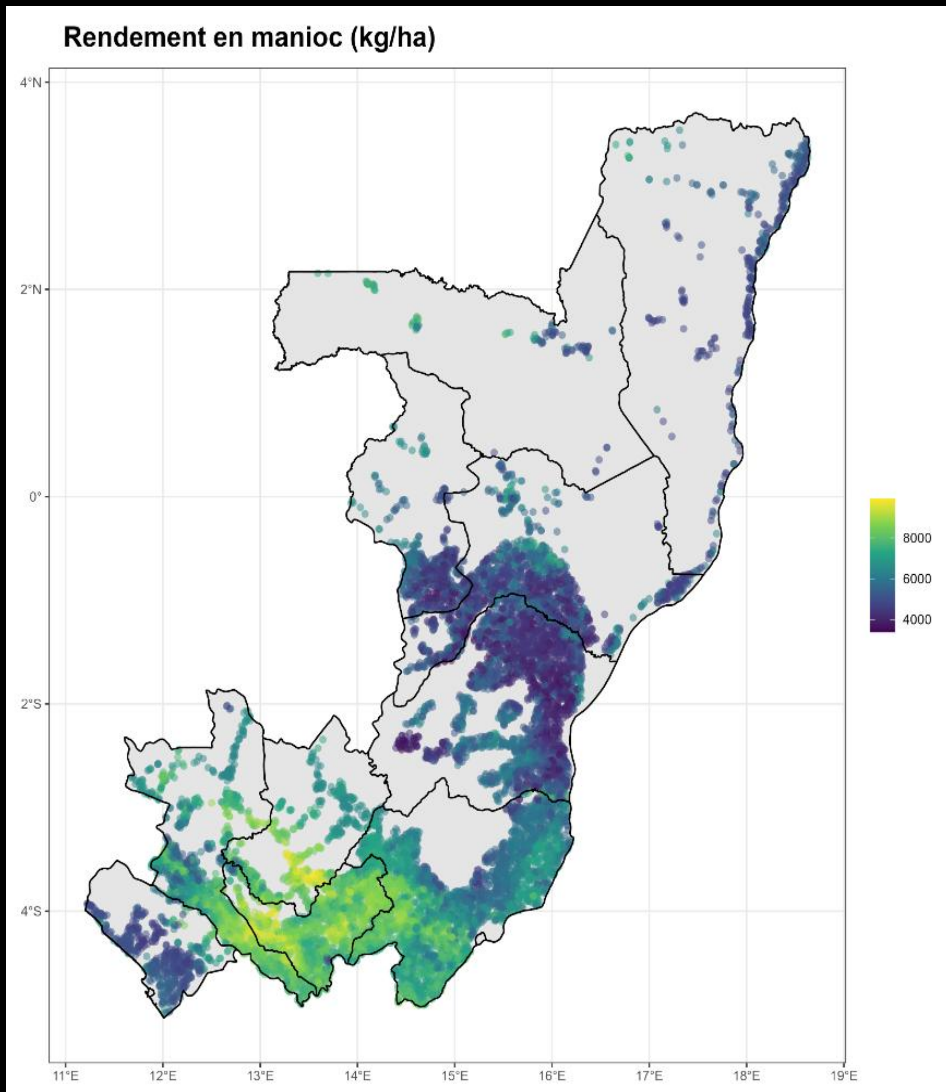
SHAP values from RF

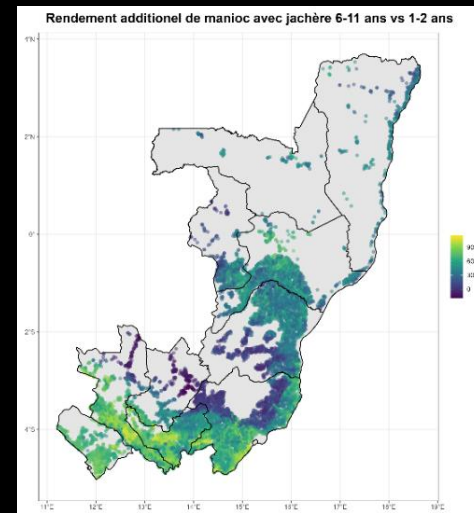


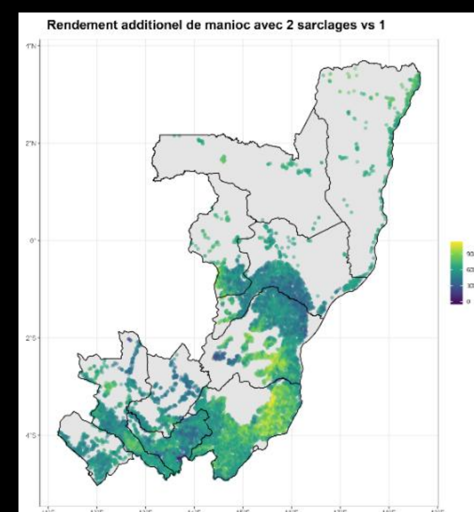




Spatial predictions from RF

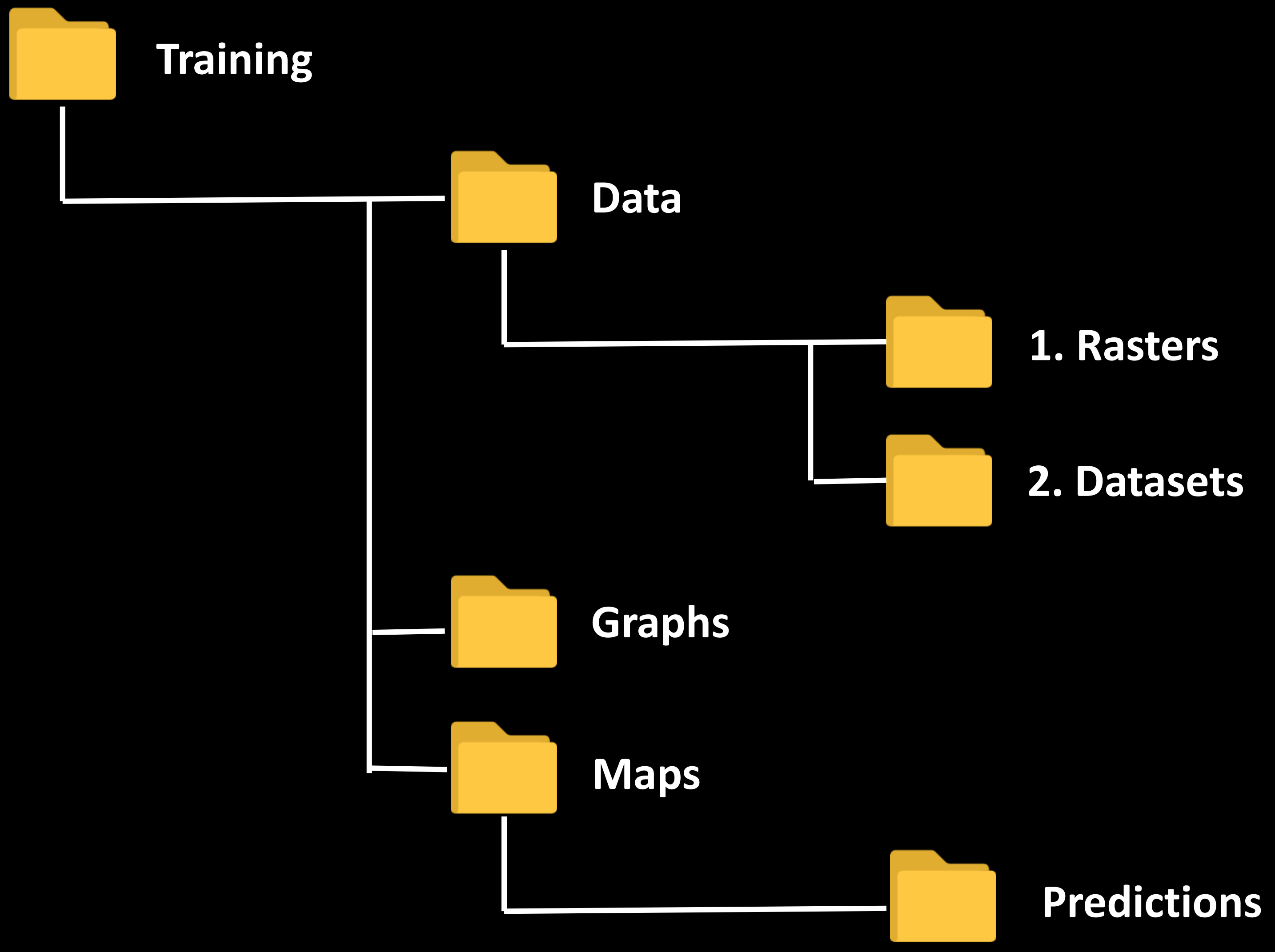






Thank you for your interest!

frederic.baudron@cirad.fr


Training
Data
1. Rasters
2. Datasets
Graphs
Maps
Predictions