Forecasting the risk of locust outbreak using machine learning & agent-based modeling

Lucile Marescot

26 of March 2024

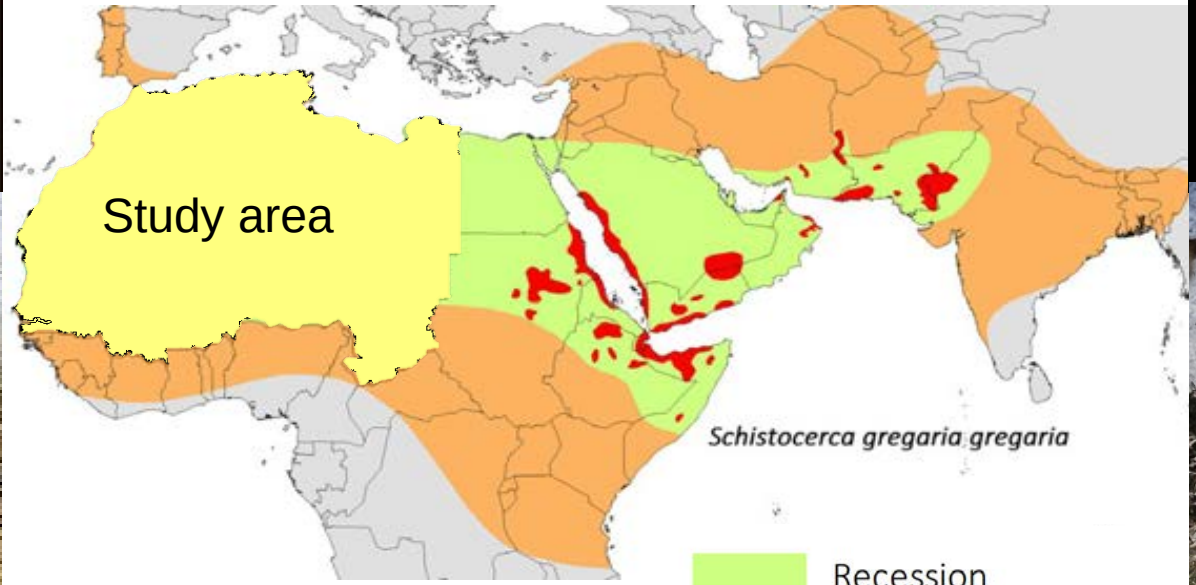
Centre for Biology and Management of Populations
CBGP
Centre de Biologie pour la Gestion des Populations

Cyril Piou, Esther Diouf, Elodie Fernandez, Mamour Touré, Arianne Cease, Lemine Hamouny

cirad
LA RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT

Locust : the eighth plague

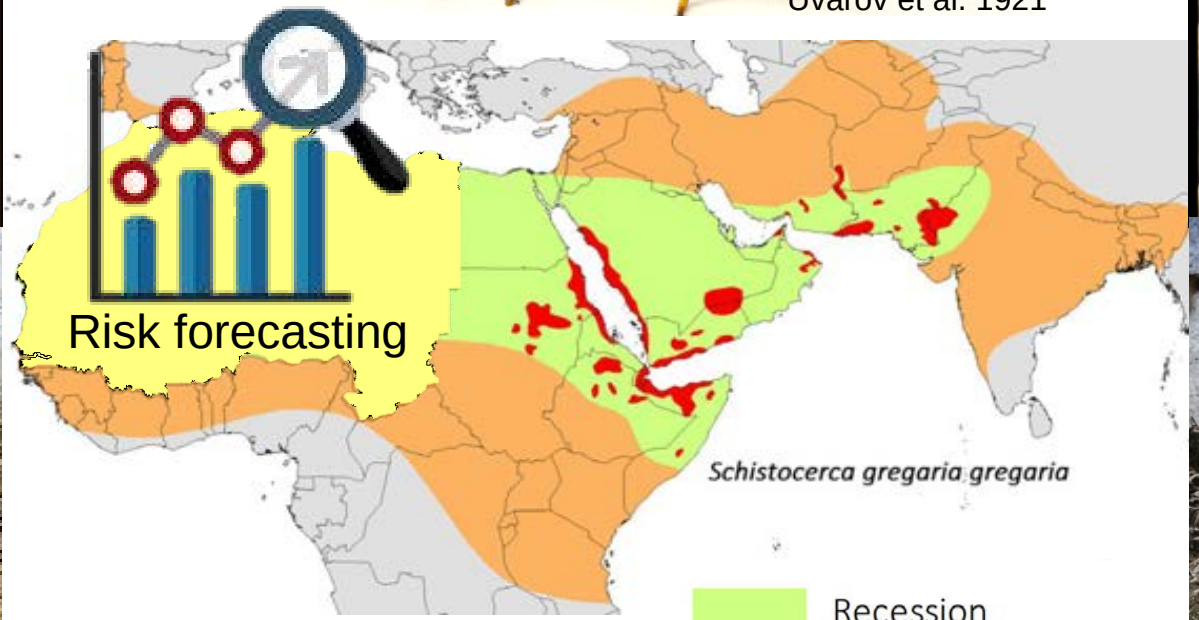
Uvarov et al. 1921



Gay et al. 2021

Locust : the eighth plague

Uvarov et al. 1921



Gay et al. 2021

Objectives

Uncertainties

Forecasted risks

(Piou & Marescot 2023)

Fundamental

Safe agricultural production

- Institutional/governance

Threats for human health
Threats for biodiversity
Economical costs

Who does it ?

Means

Protect crops/pastures from swarms

- Climate change
- Demographic stochasticity

Impacts on production

Management

Global

Proactive

Local

Preventive

Control uncertainty

Outbreak

Gregarization

Breeding

Monitoring

Sampling uncertainty

Presence/absence

solitarious/transiens

bands

swarms

Structural uncertainty

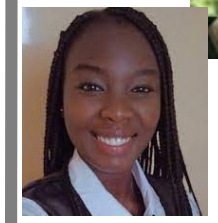
Model Lab

Mental model
Machine learning / Agent Based Model
Experiment

Behavior/Physiology

How to do it ?

Why to do it ?



Objectives

Uncertainties

Forecasted risks

(Piou & Marescot 2023)

Fundamental

Safe agricultural production

- Institutional/governance

Threats for food security
Threats for biodiversity
Economical costs

Means

Protect crops/pastures from swarms

- Climate change
- Demographic stochasticity

Impacts on production

Management

Global

Proactive

Control uncertainty

Outbreak

Local

Preventive

Gregarization

Breeding

Monitoring

Presence/absence

solitarious/transiens

bands

swarms

Sampling uncertainty

Model

AI

Machine learning

Structural uncertainty

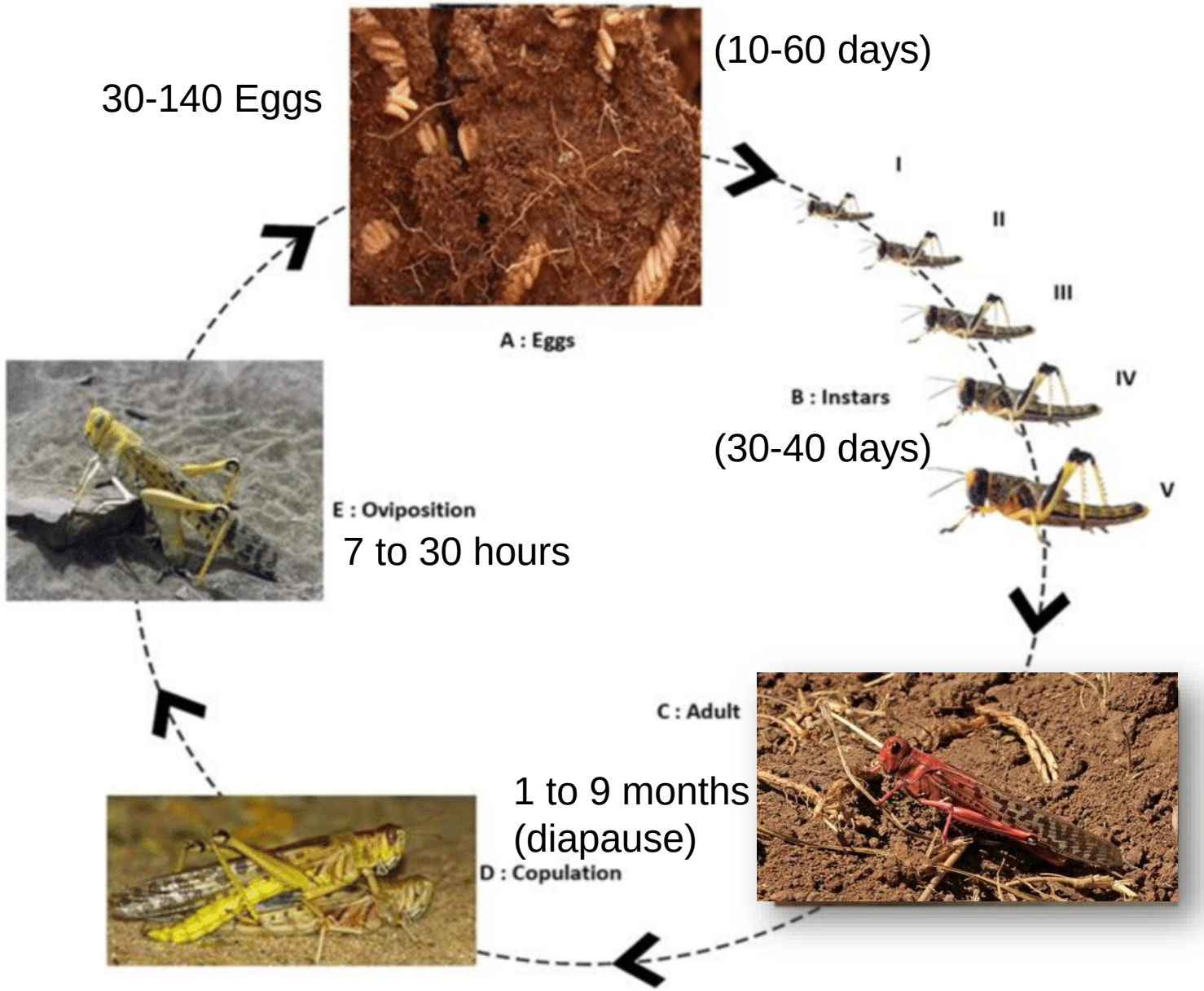
Behavior/Physiology

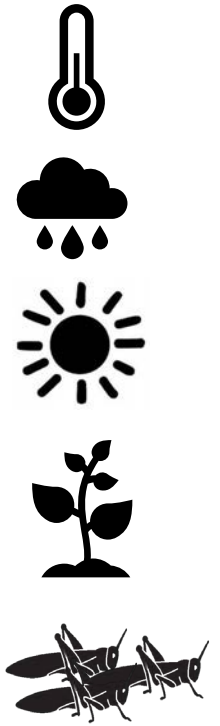
How to do it ?

Why to do it ?

Desert locust life cycle

Schistocerca gregaria

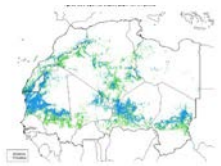




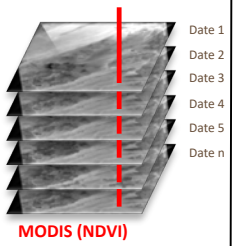
(Shresta et al. 2021)

Step 1: Data processing

Observations



Environmental variables



Train data 80%

Test data 20%

Step 2: Training

Train data

Machine Learning

Optimisation

Trained models

Steps 3 : Evaluation

Test data

Reality

Trained models

Performance Index (Accuray, AUC)

Importance of variables (Gini,permutation)

Step 4: Forecasting

New data

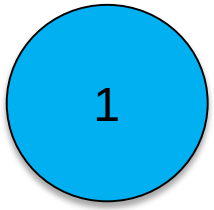
Selected model

Prediction maps

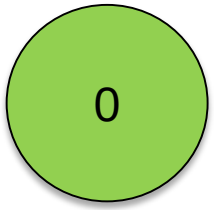
Step 1: Data processing

Response variable

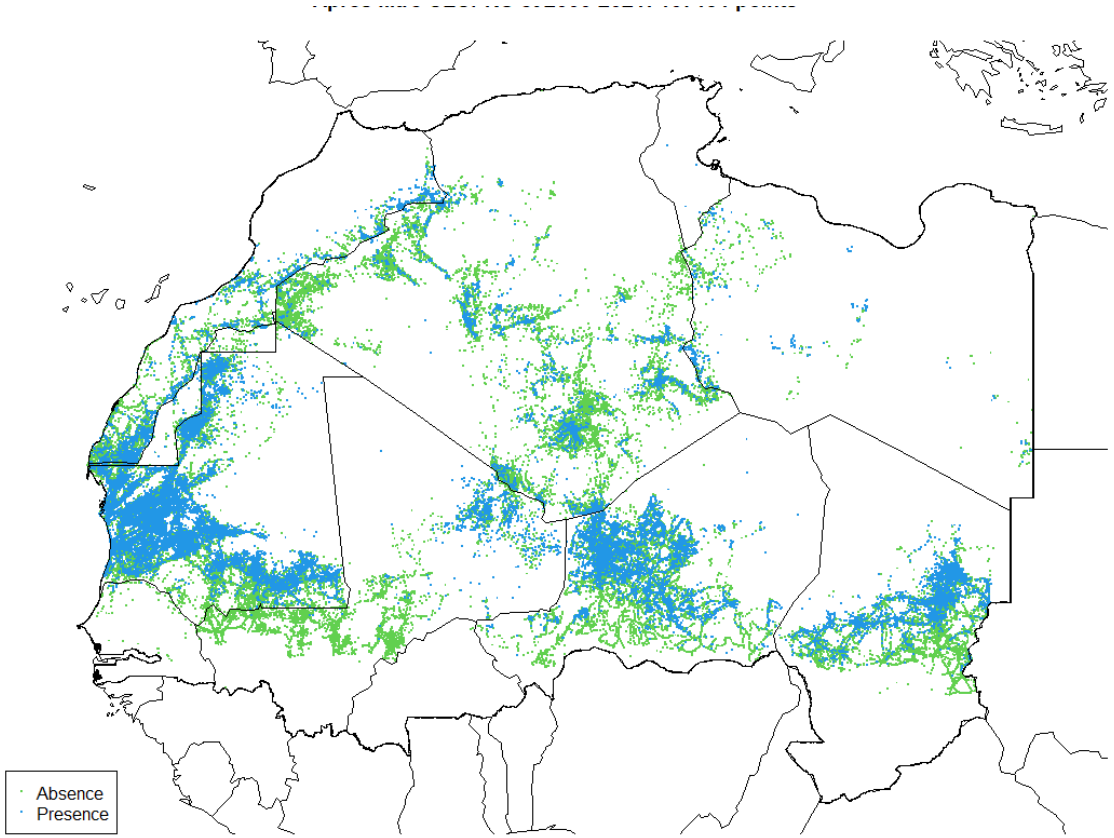
Presence



Absence



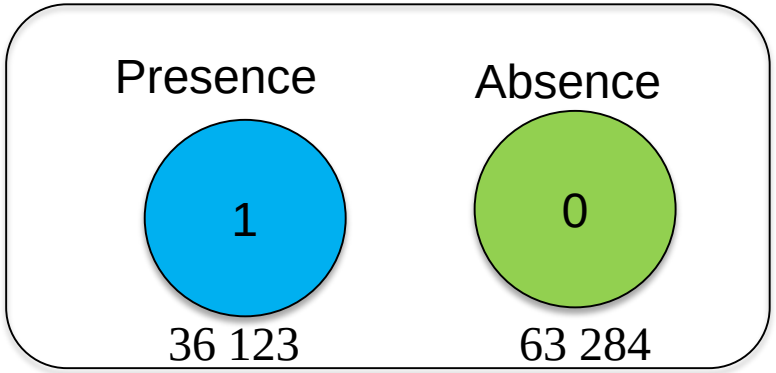
- 10 countries : Mauritania, Morocco, Algeria, Tunisia, Libia, Mali, Niger, Chad, Senegal, Burkina Fasso
- 99 407 prospection points from 2003 to 2022
- 12 861 240 sites



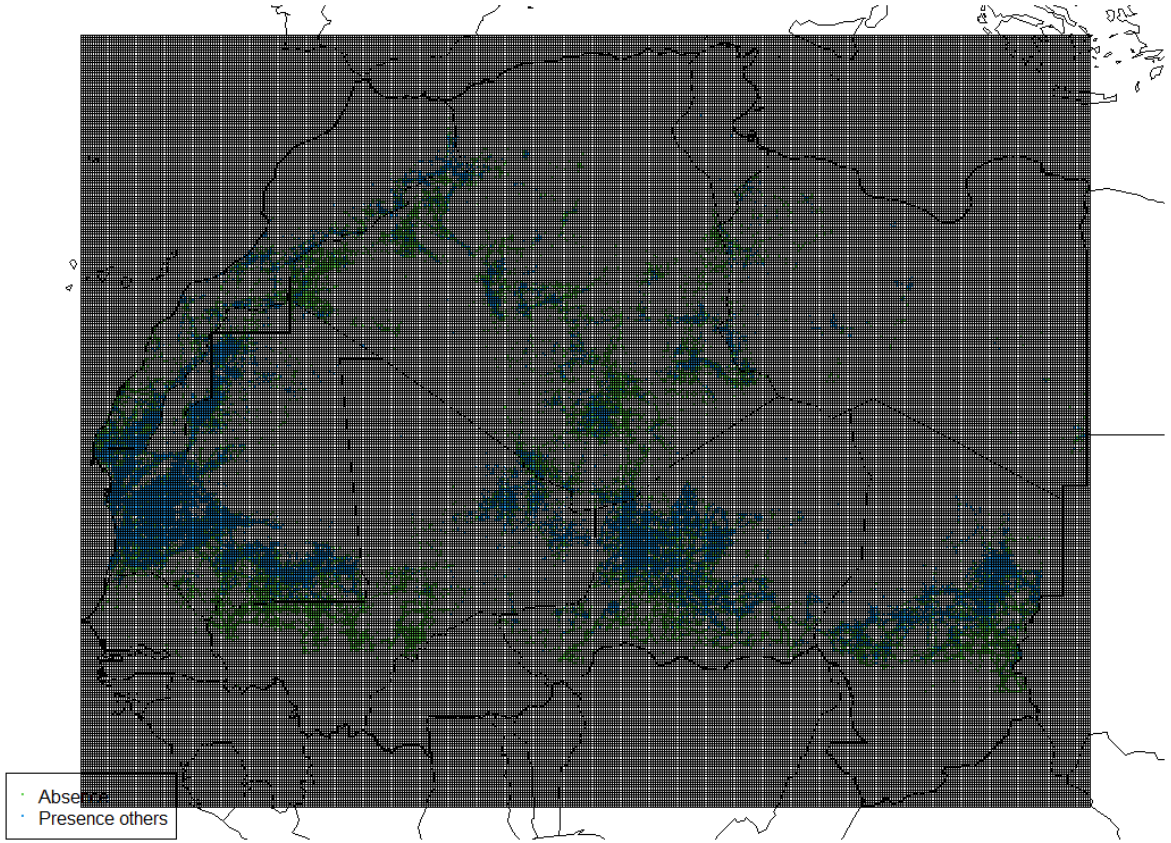
Step 1: Data processing

Response variable

Site of 1 km2 every 10 days



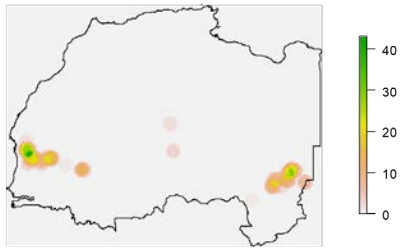
- 10 countries : Mauritania, Morocco, Algeria, Tunisia, Libia, Mali, Niger, Chad, Senegal, Burkina Fasso
- 99 407 prospection points from 2003 to 2022
- 12 861 240 sites



Step 1: Data processing

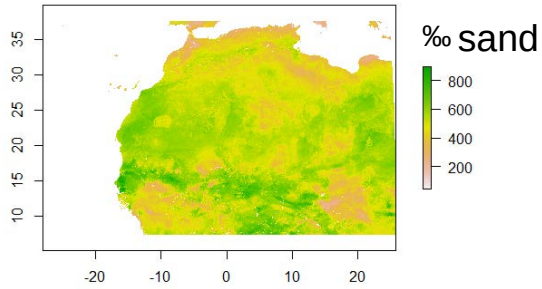
Explanatory variables from remote sensing

Recent observations on 100km buffer

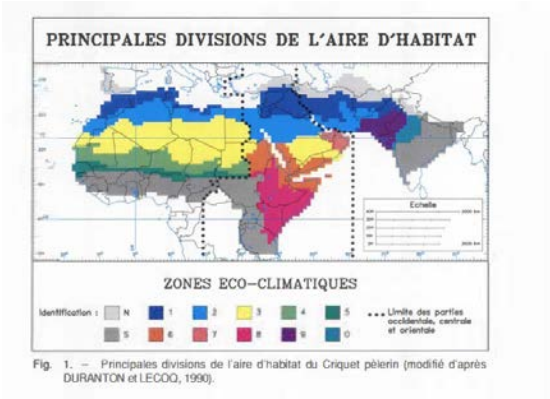


+

Static Variables
- Ecological Units
- Sand cover
- Min and Max NDVI



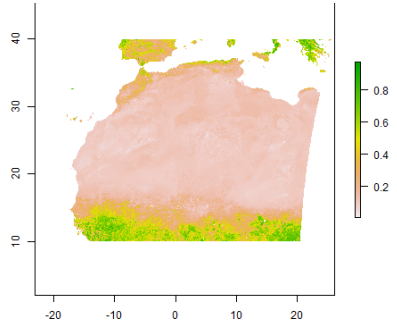
Spatial autocorrélation



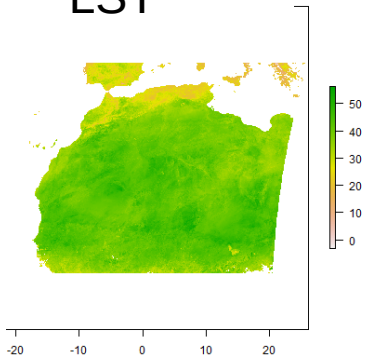
+

Dynamic variables extracted from MODIS and SMOS at time t, t-1, t-2 …t-10 (ten decades = 100 days)

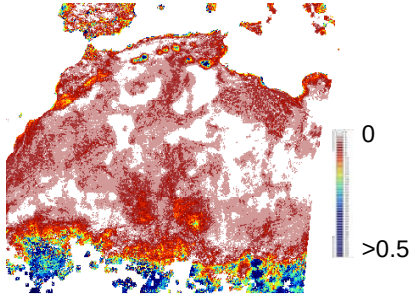
NDVI



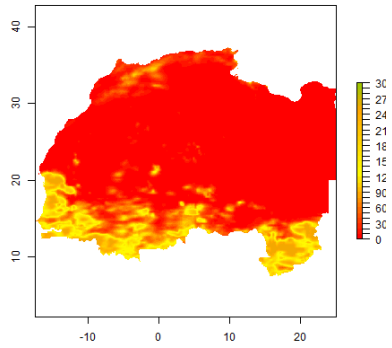
LST



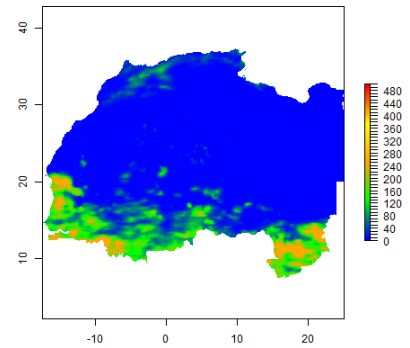
SM



RAIN



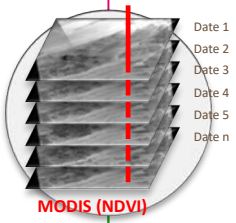
NPP



Step 1 : data processing

Raw data
FAO

Extraction
of variables



Train data
80%

Test data
20%

Step 2 : Training

Train data

Machine learning

Optimisation

Trained models

Step 3 : Evaluation

Test data

Reality

Trained model

Performance Indicator
(Accuracy, AUC)

Variables imortance
(Gini index)

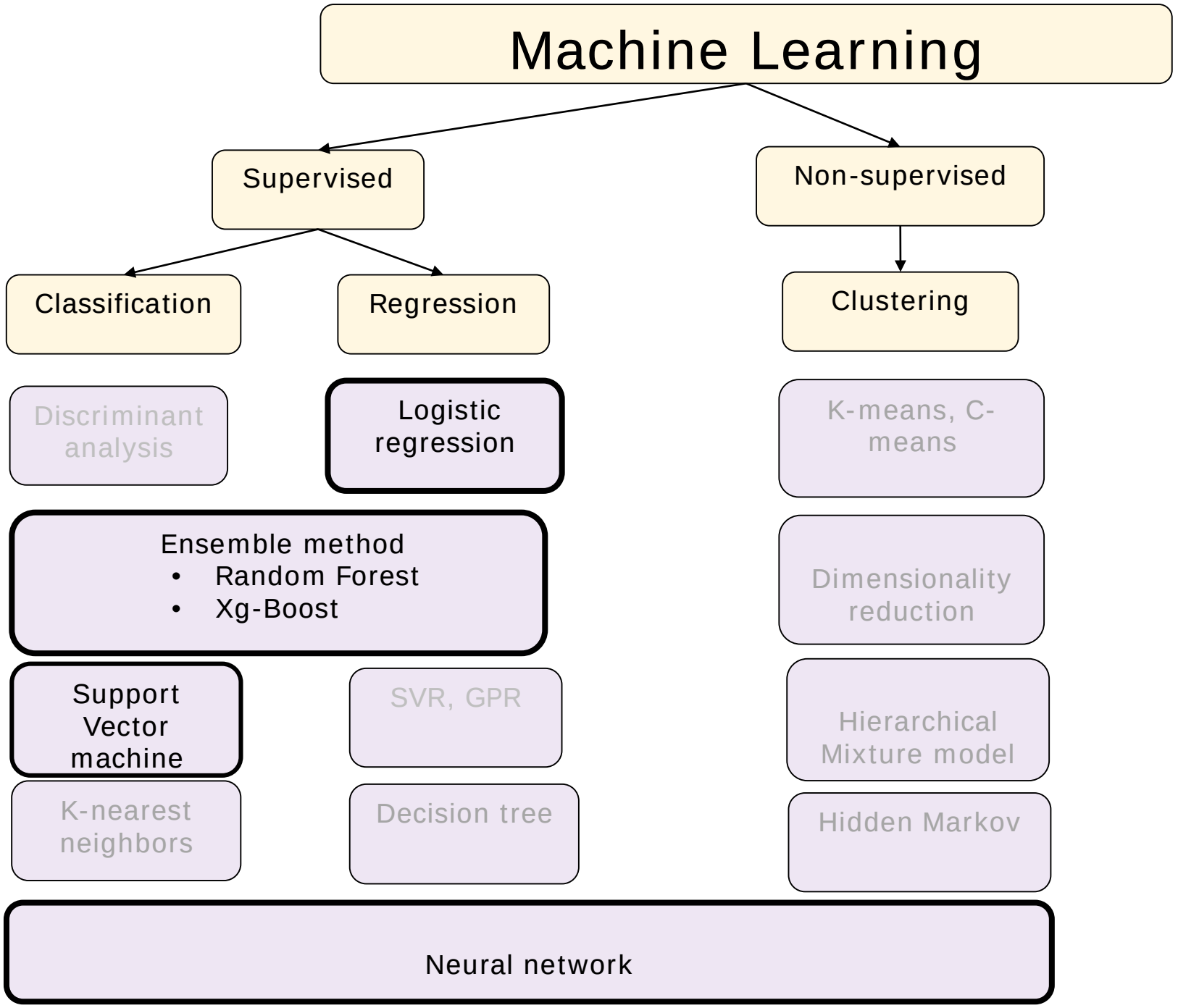
Step 4 : Prediction

New data

selected model

Prediction map

Step 2 : Training



(Breiman 2001)

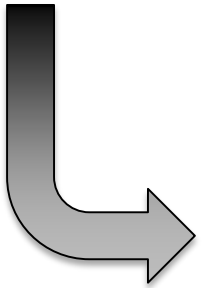
(McCulloch & Pitts 1943)

Step 3 : Evaluation

Choose the best algorithm

- Presence/Absence ~ 53 variables
- 5 machine learning algoritms (*Tidymodel*s package)
- Cross validation and hyperparameter tunning

	NN	RF	Xg-Boost	SVM	LR
Accuracy	0.72	• 0.78	• 0.78	• 0.76	• 0.67



- Recursive feature elimination (caret package)

Step 3 : Evaluation

Select models for presence/absence

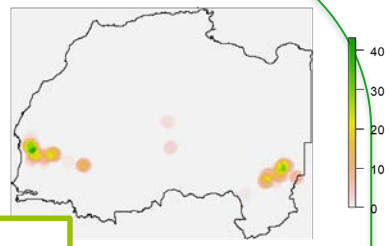
Full model AUC = 0.77

Spatial autocorrélation

+

Seasons

+



Static variables
- Ecological units
- Sand cover
- Min and Max NDVI

+

NDVI

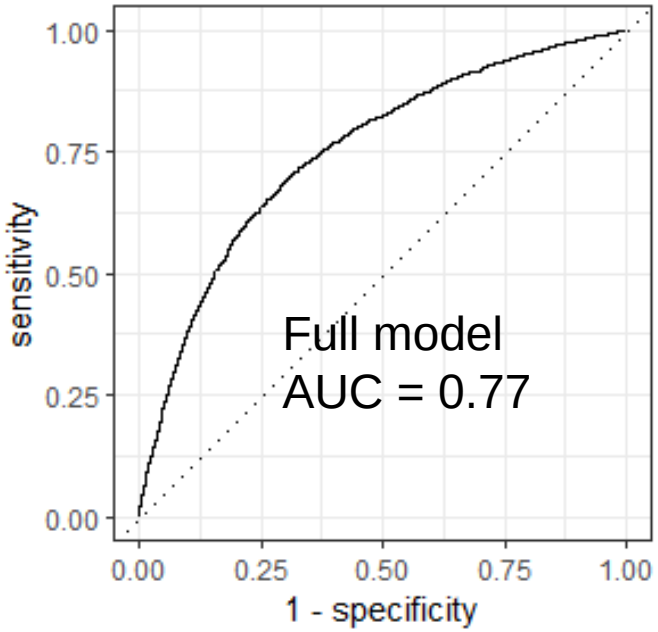
SM

LST

RAIN

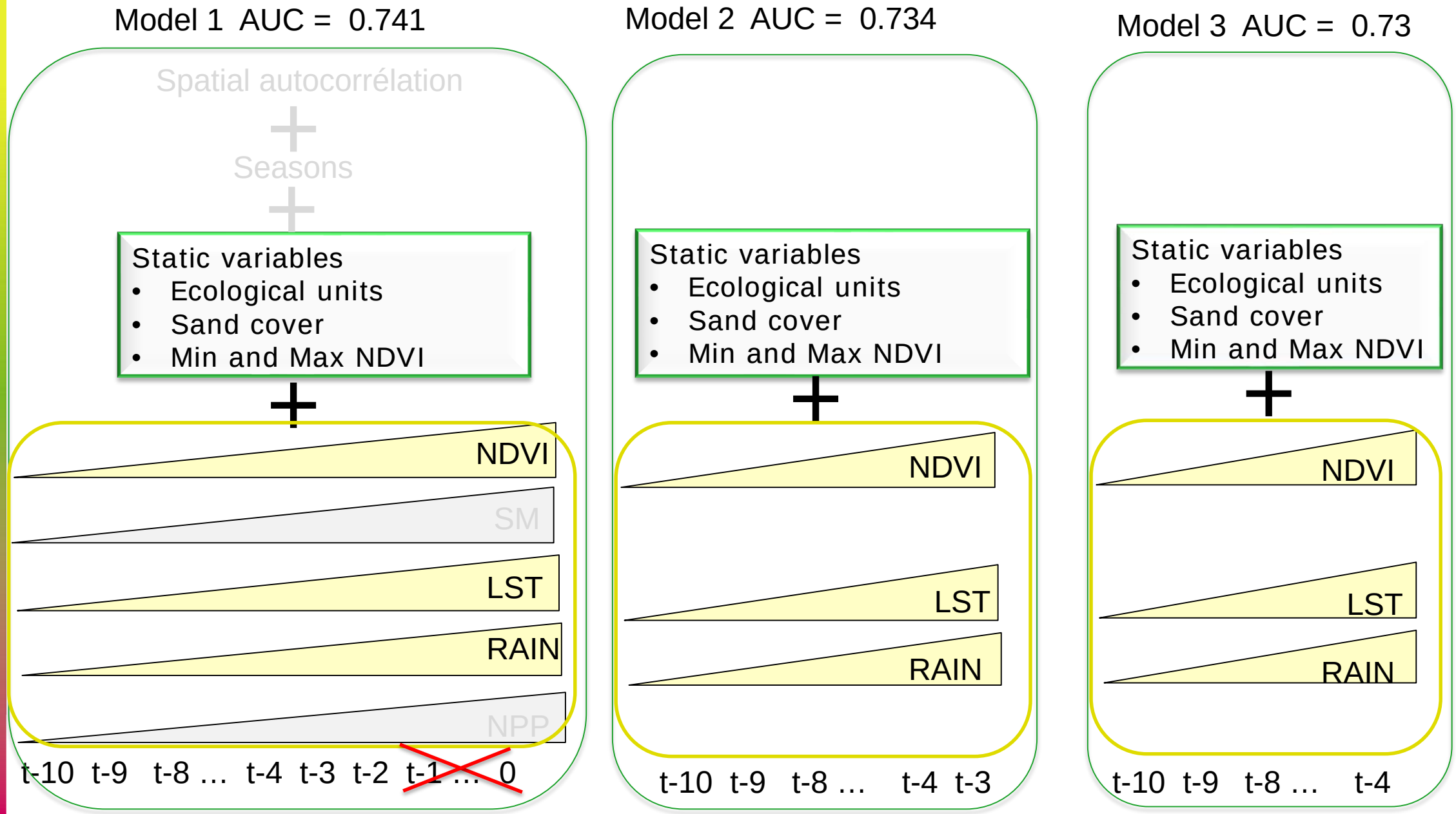
NPP

t-10 t-9 t-8 … t-3 t-2 t-1 … 0



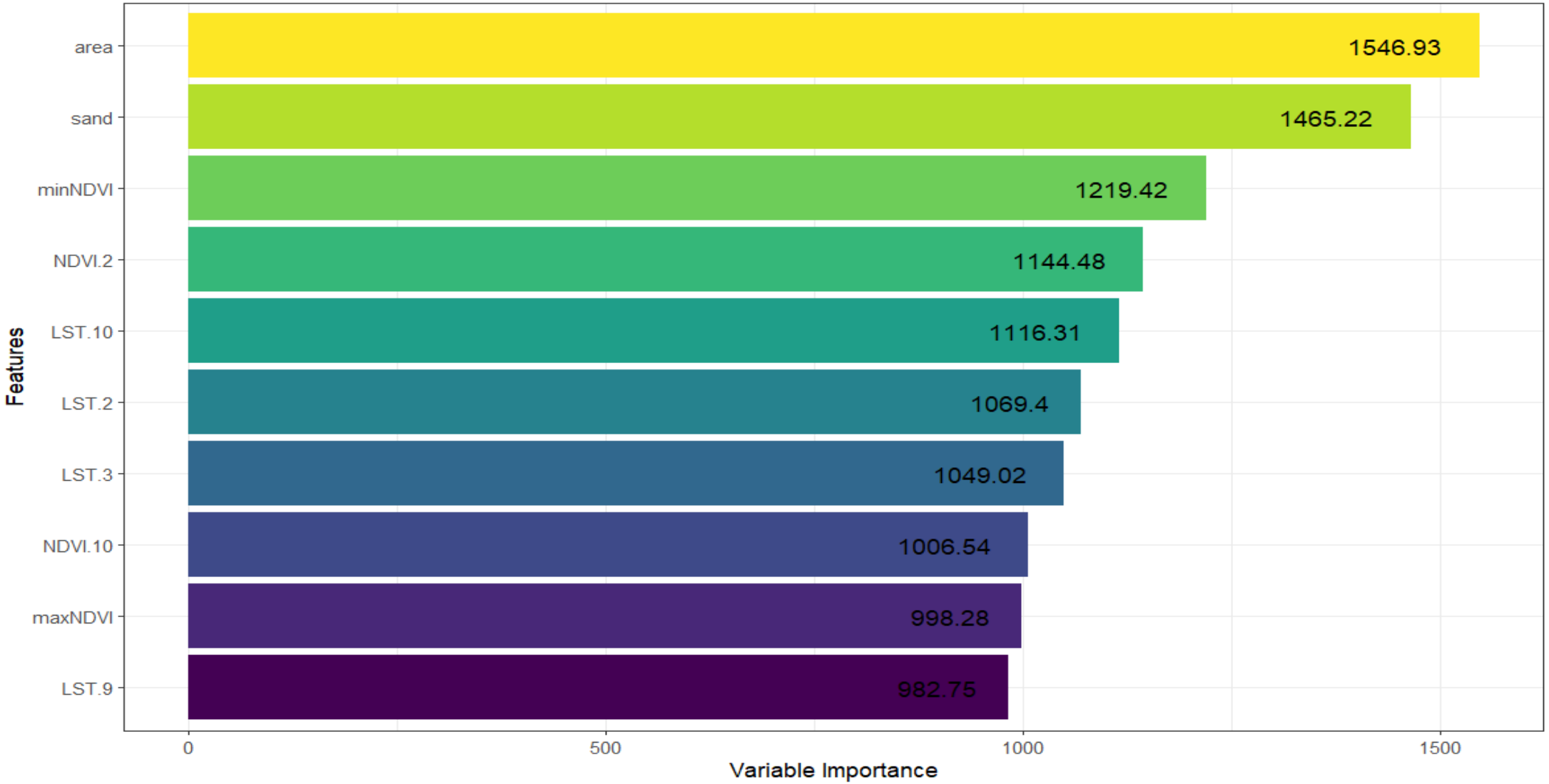
Step 3 : Evaluation

Select operational models

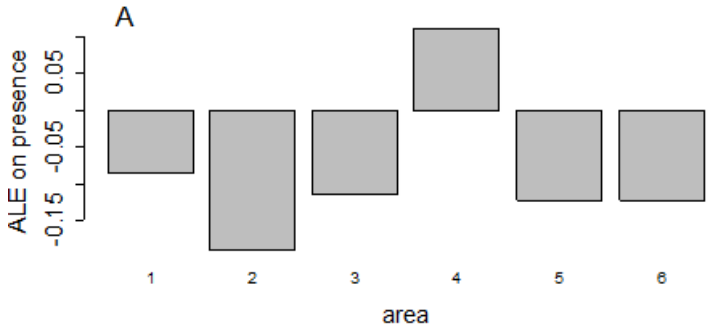


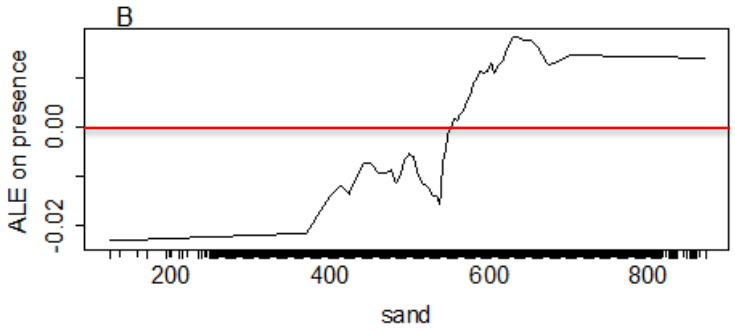
Step 3 : Evaluation

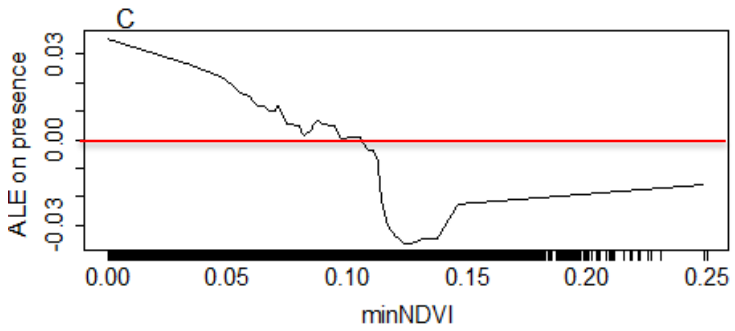
Variable importance

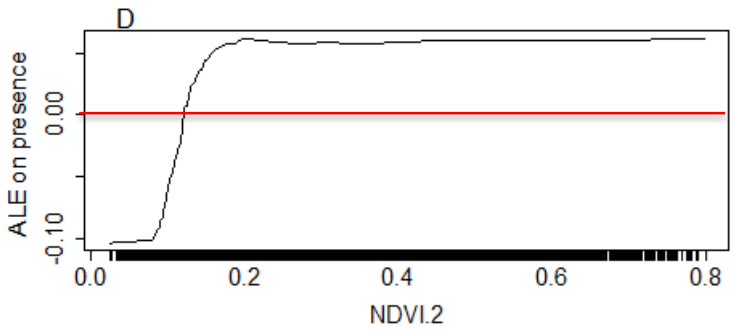


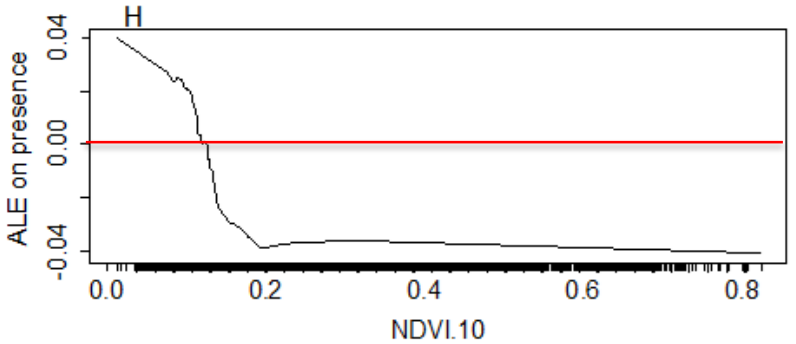
Response curve : accumulated local effect



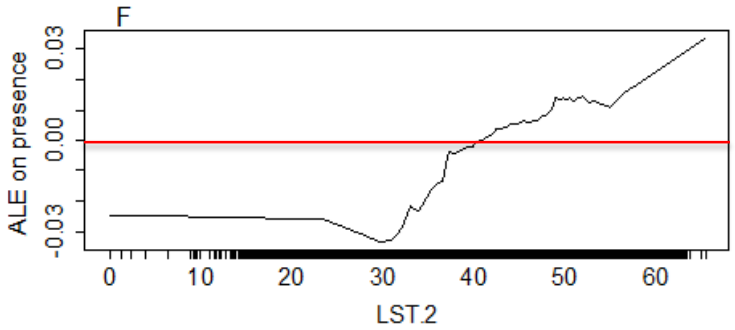


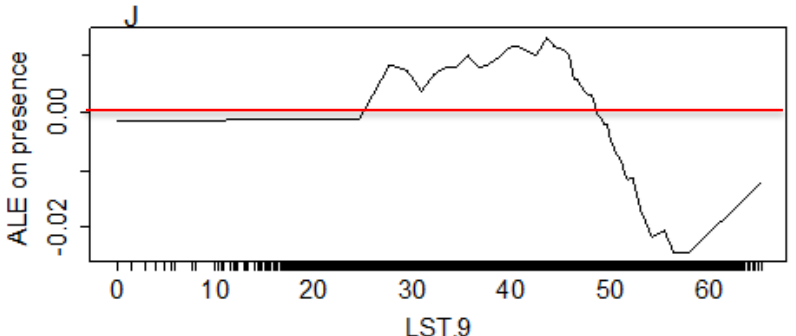






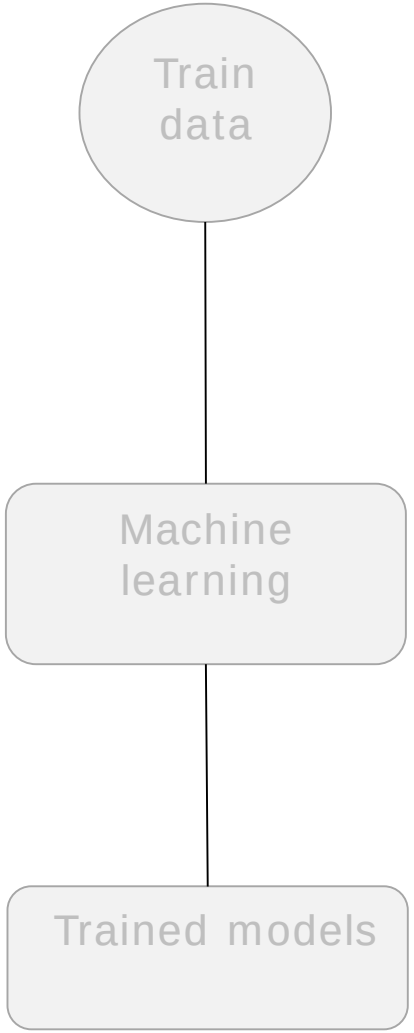
Adaptation to dry habitat
with fast and patchy vegetation
growth



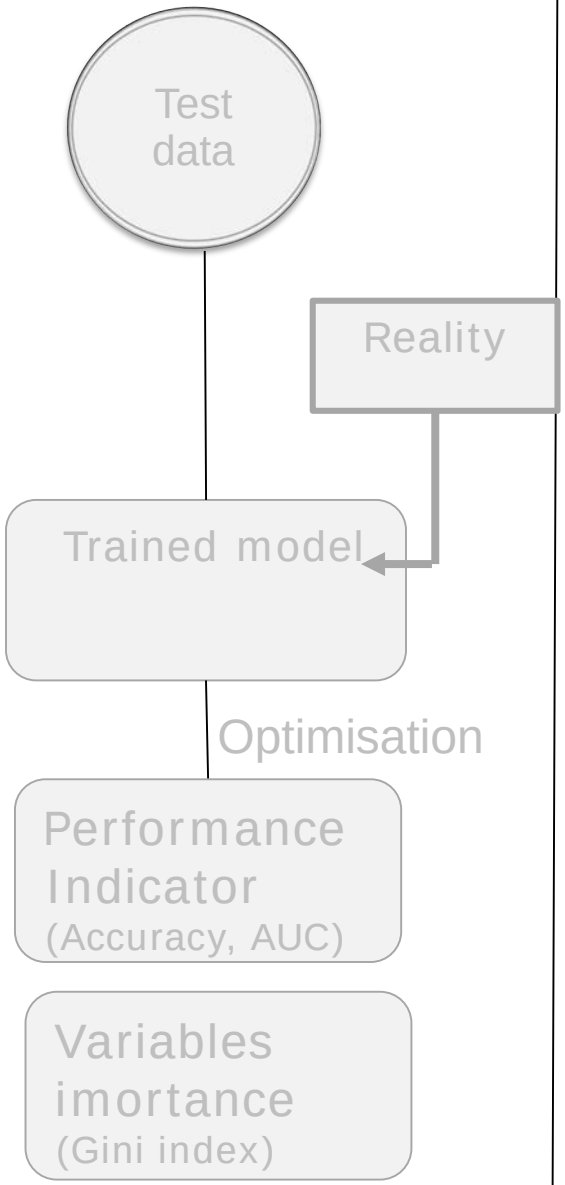


Etape 1 : traitement des données

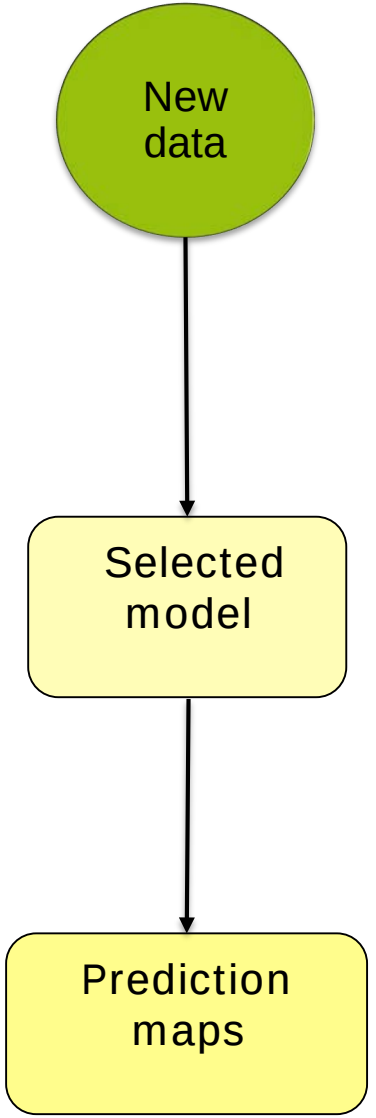
Step 2 : Training



Step 3 : Evaluation

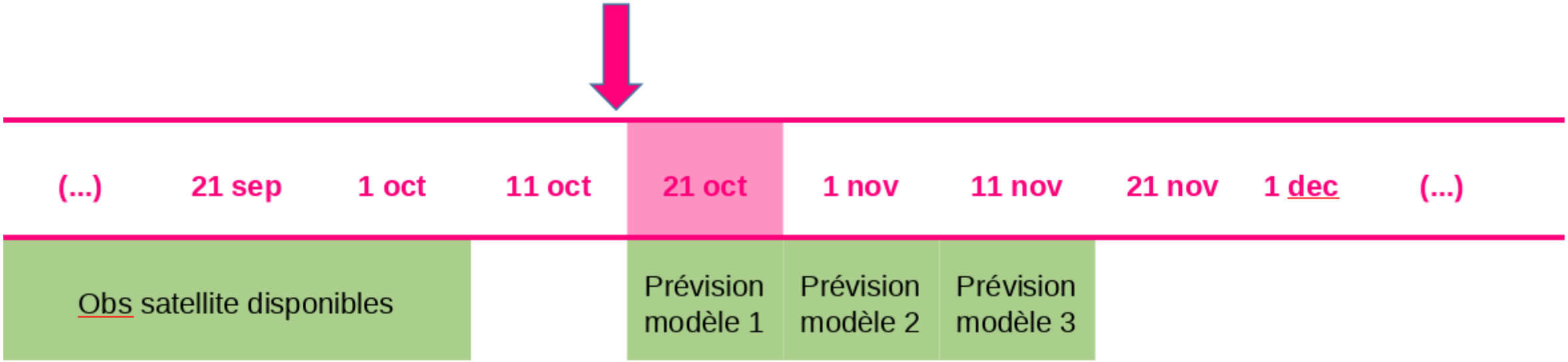


Step 4: Forecasting



@Fernandez

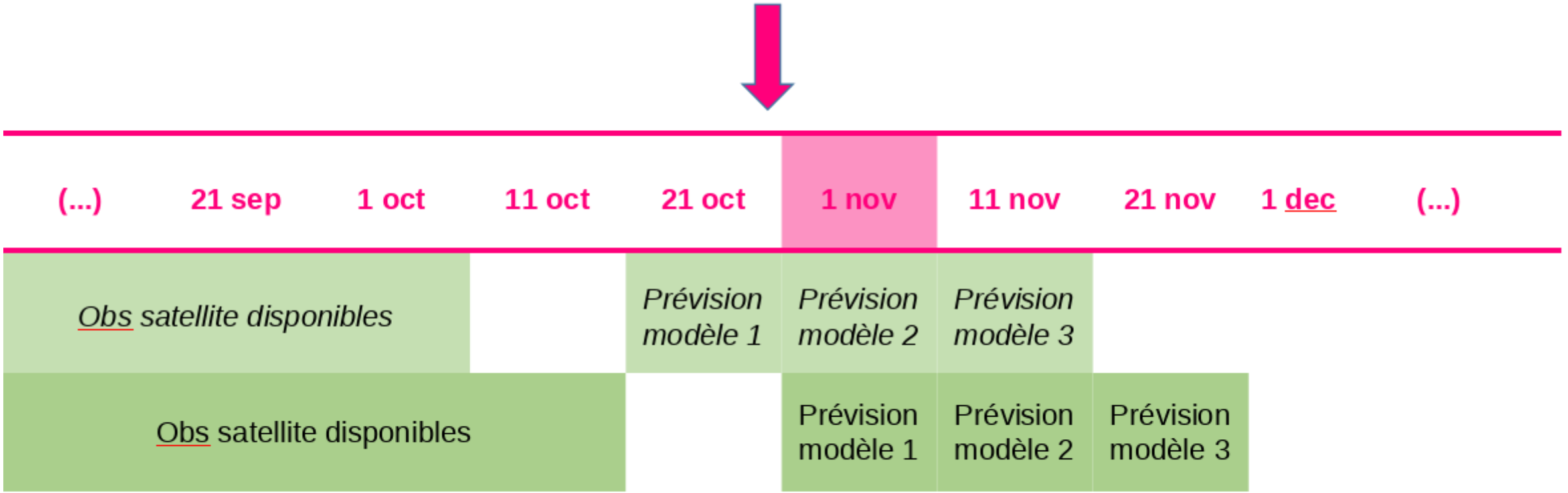
Operational forecast



Model 1: No SAT data for the last decade
Model 2: No SAT data for the last 2 decades
Model 3: No SAT data for the last 3 decades

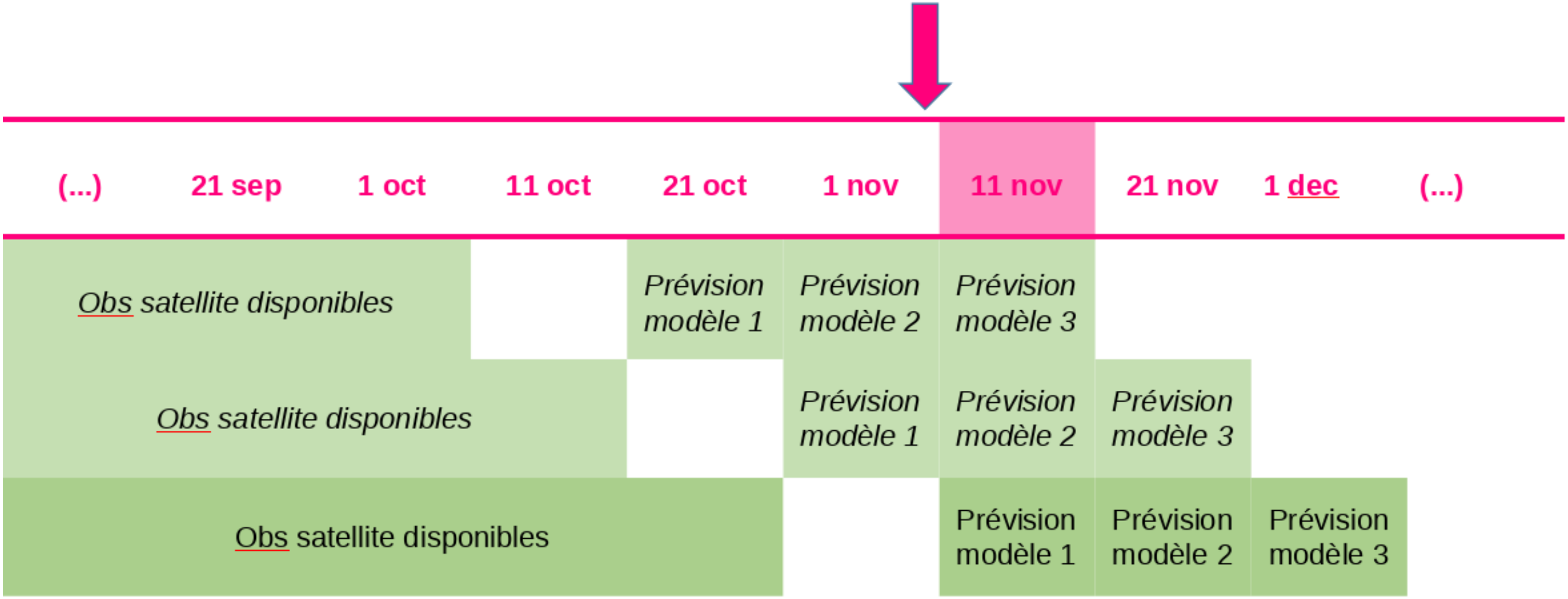
@Fernandez

Operational forecast



@Fernandez

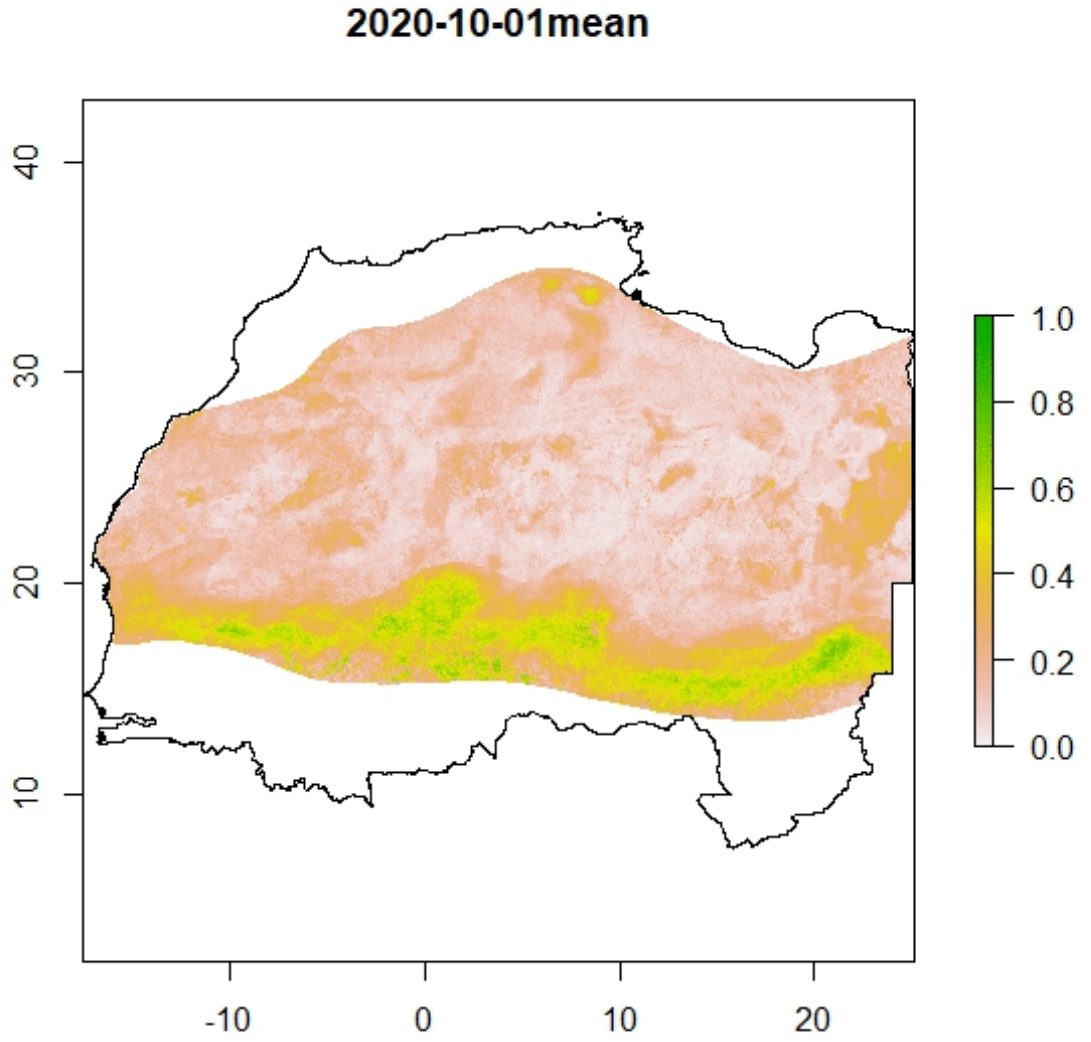
Operational forecast



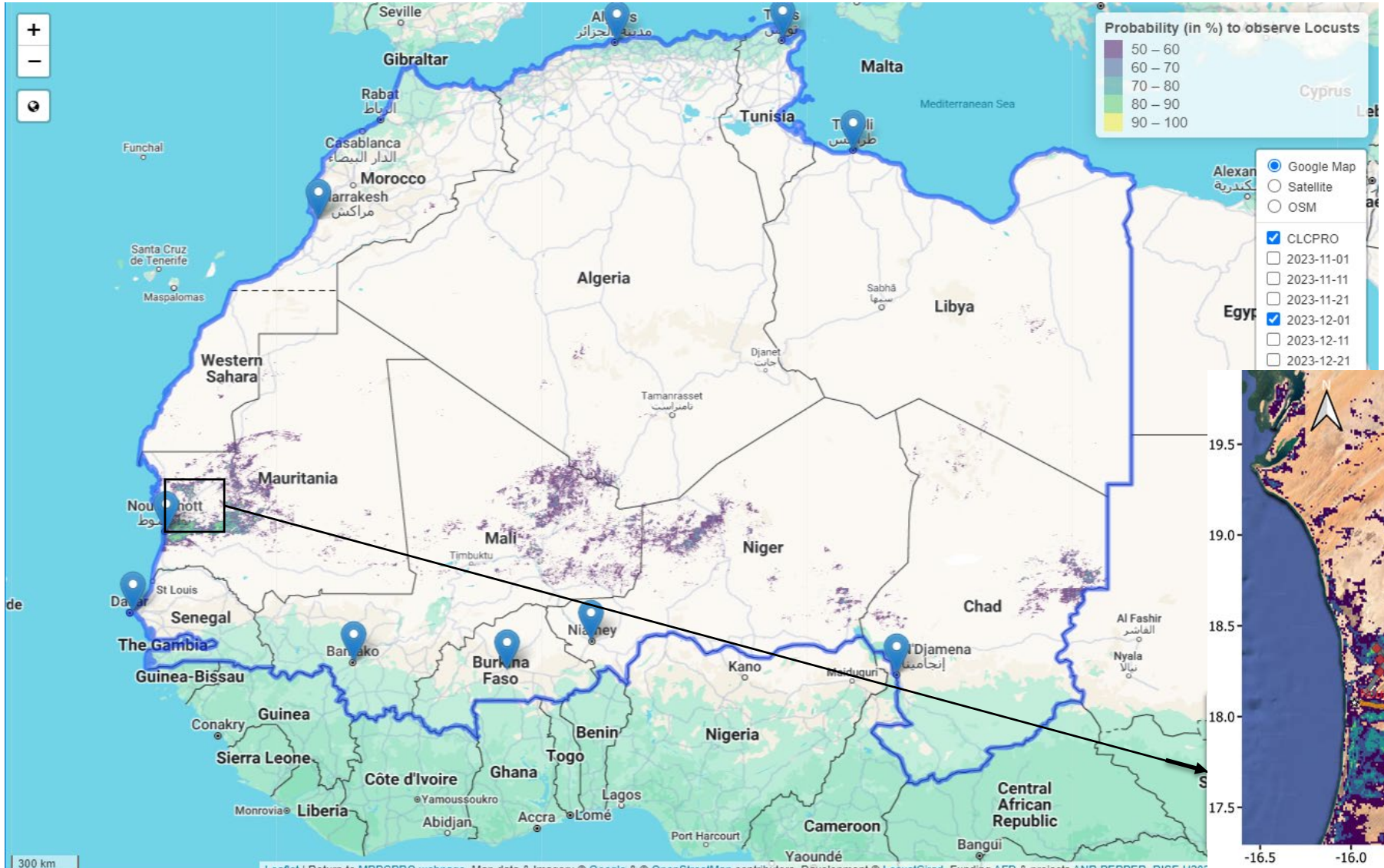
Retrospective validation

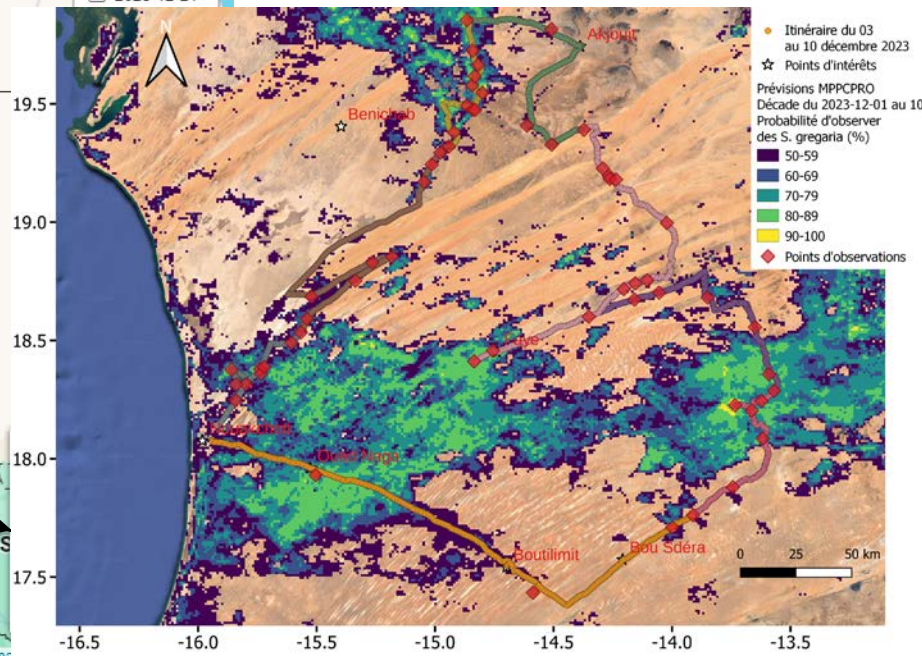
Where are Desert locust most likely to be found ?

- Mostly present in the saharo-sahalian zone
- Less than 5 % of false negative



Forecasting map for field prospection

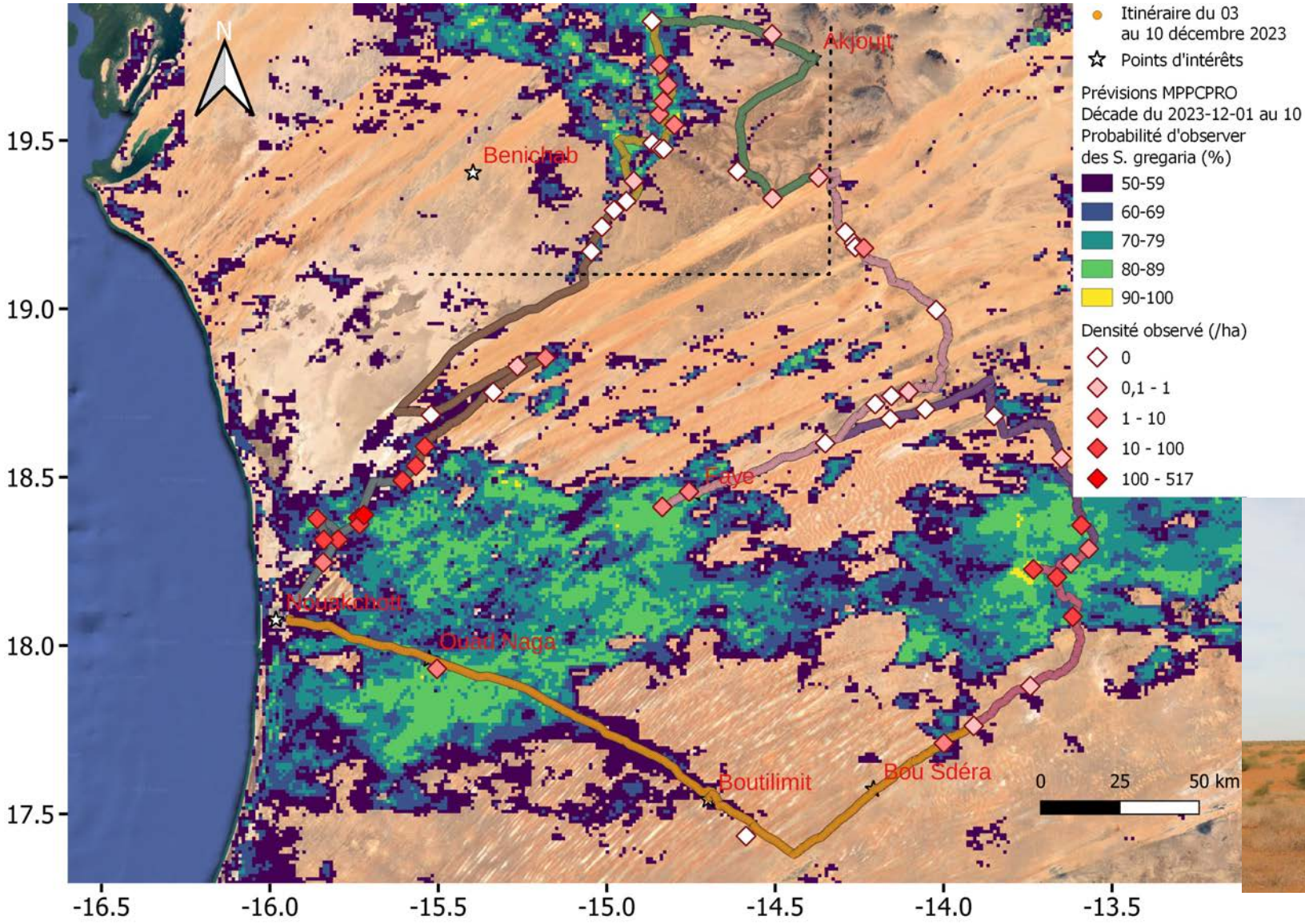




- 1500km drive
- Locust density with 3 prospectors
- 58 sites

@Piou
@Fernandez

Field validation



- 81 % of correct predictions
- Correlation Classes probability of presence - Density (Kendall tau= 0.64)

Objectives | Uncertainties | Forecasted risks | (Piou & Marescot 2023)

Fundamental

Safe agricultural production

- Institutional/governance

Threats for food security
Threats for biodiversity
Economical costs

Means

Protect crops/pastures from swarms

- Environmental variability
- Demographic stochasticity

Impacts on production

Management

Proactive

Preventive

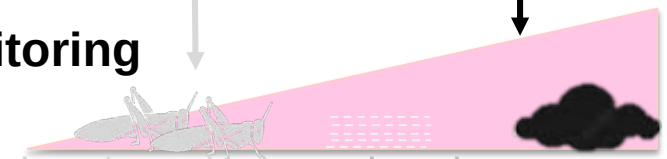
Outbreak

Control uncertainty

Gregarization

Breeding

Monitoring



solitarious/transiens

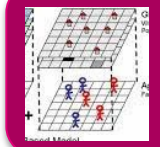
bands

swarms

Presence/absence

Sampling uncertainty

Model



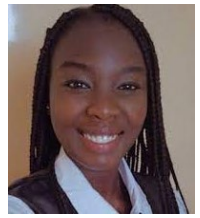
Agent based model

Structural uncertainty

Behavior/Physiology

How to do it ?

Why to do it ?



Oedaleus senagelensis

Objectives

Uncertainties

Forecasted risks

Fundamental

Safe agricultural production

- Institutional/governance

Threats for food security
Threats for biodiversity
Economical costs

Means

Protect crops/pastures from swarms

- Environmental variability
- Demographic stochasticity

Management

Proactive

Preventive

Control uncertainty

Impacts on production

Outbreak

Gregarization

Breeding

Monitoring

solitarious/transiens

bands

swarms

Sampling uncertainty

Presence/absence

Model

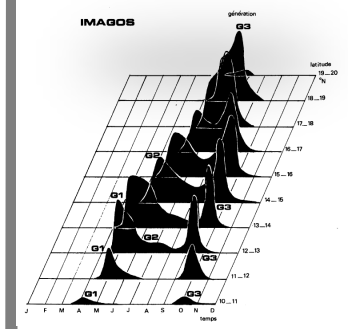
Agent based model

Structural uncertainty

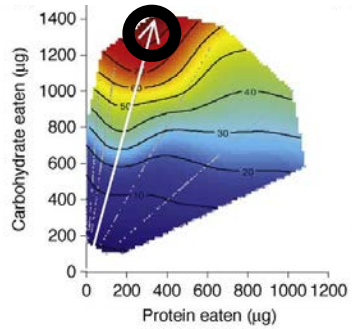
Behavior/Nutrition

How to do it ?

Why to do it ?

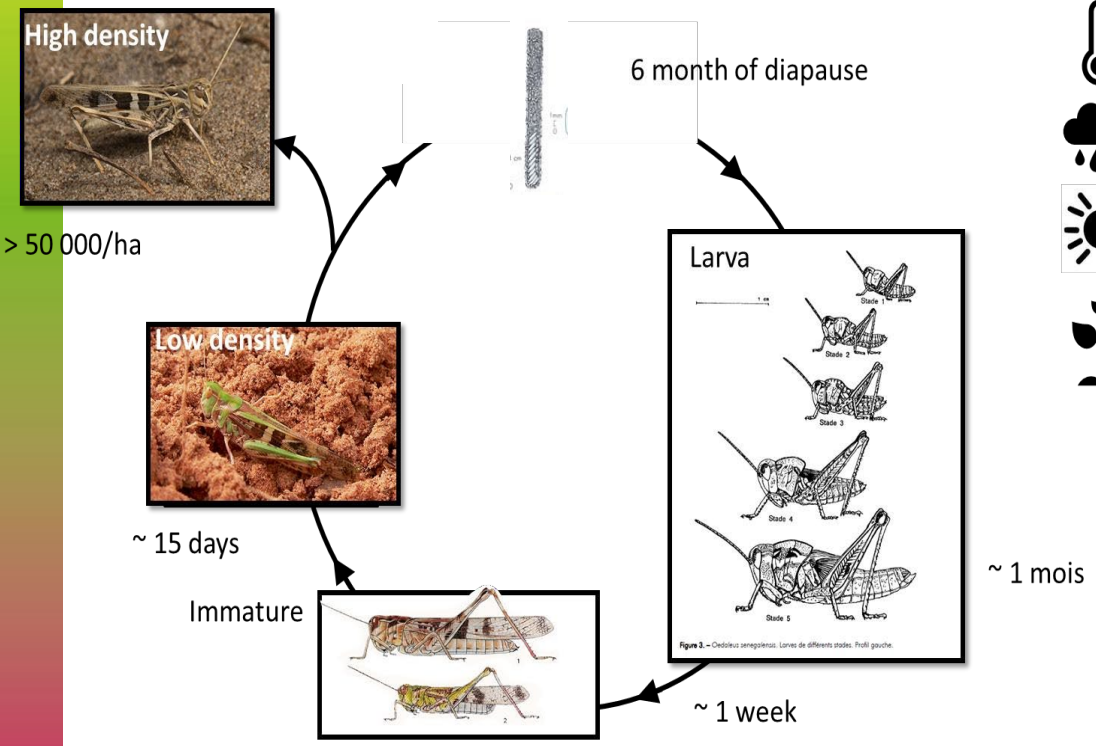


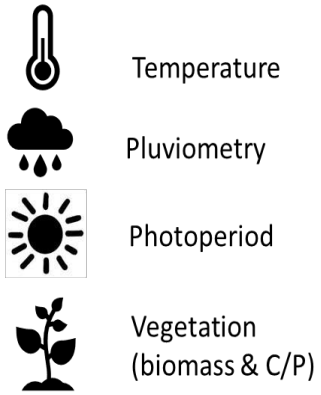
Oedaleus senagelensis



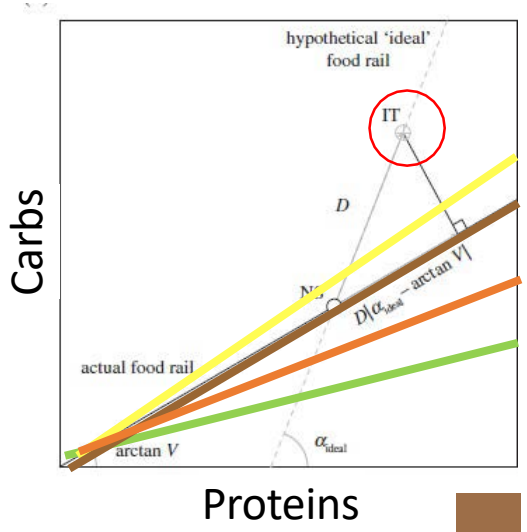
Agent-based model linking nutritional ecology and population dynamics

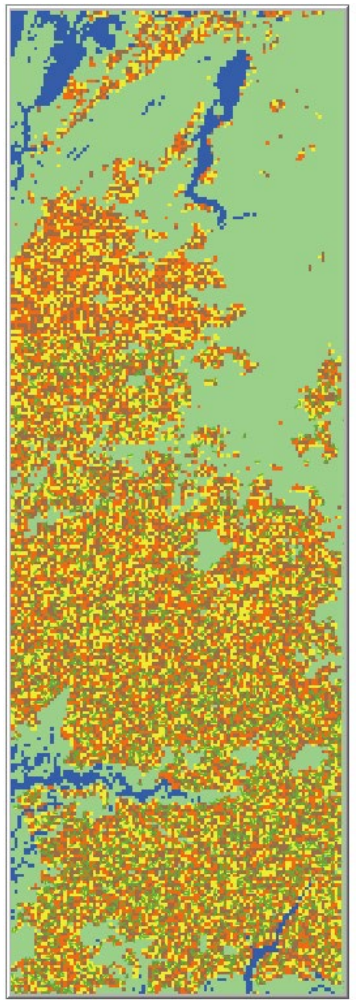
Can soil amendment affect population dynamics and behaviour and reduce crop damage ?





Nutritional geometric framework

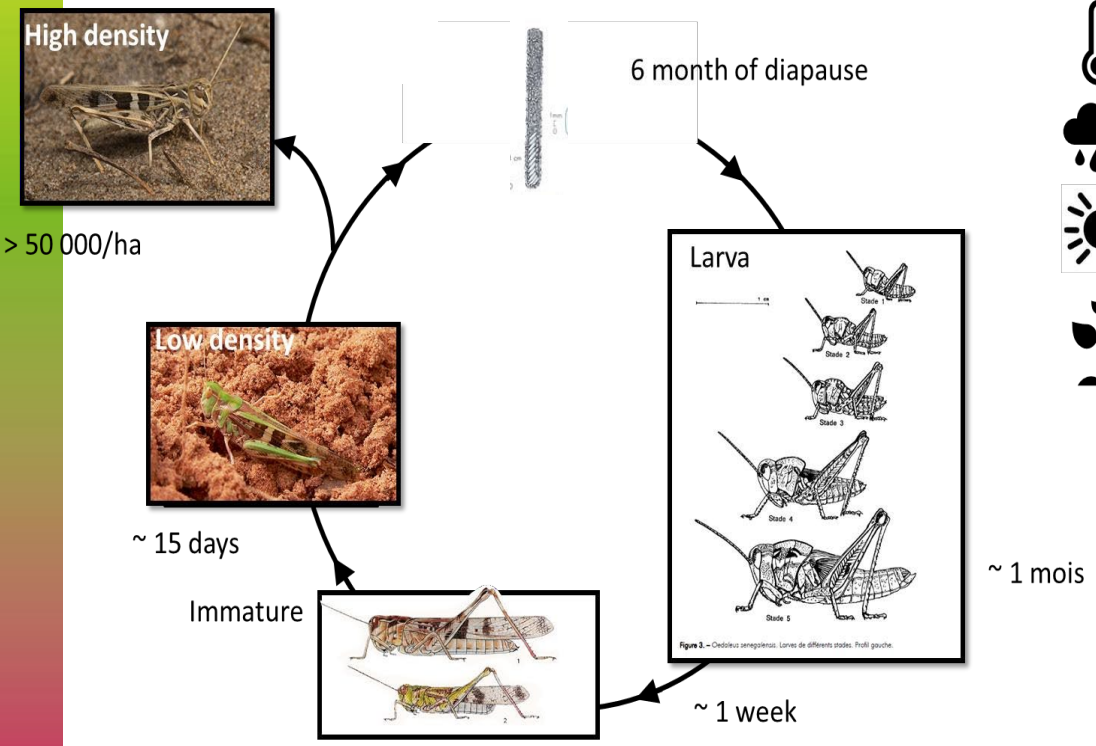


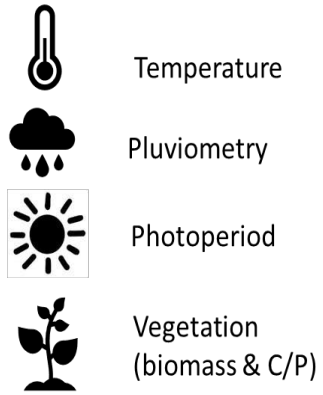


(Launois 1979, Simpson et al. 2010, Legall et al. 2019, 2021)

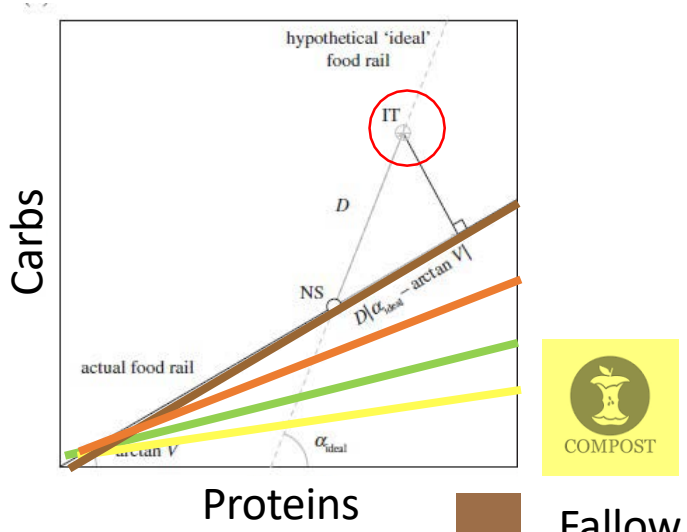
Agent-based model linking nutritional ecology and population dynamics

Can soil amendment affect population dynamics and behaviour and reduce crop damage ?



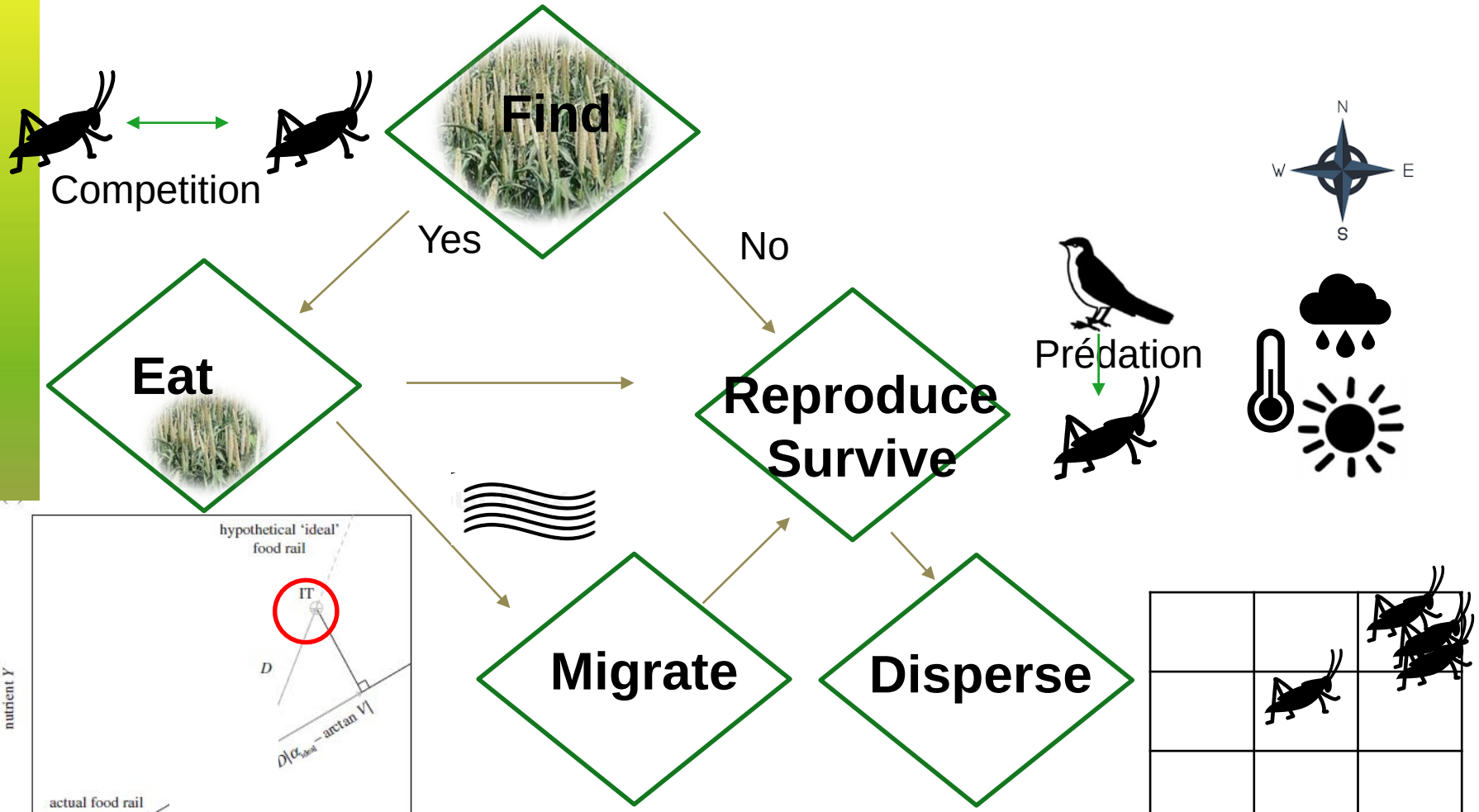


Nutritional geometric framework

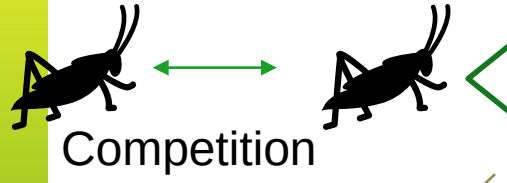


(Launois 1979, Simpson et al. 2010, Legall et al. 2019, 2021)

Model structure



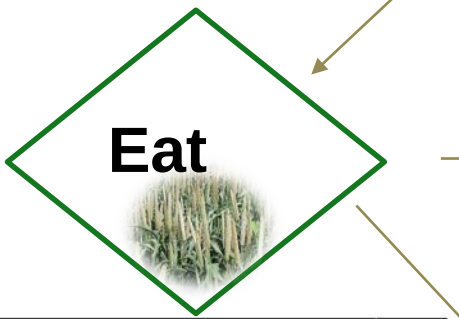
Model structure

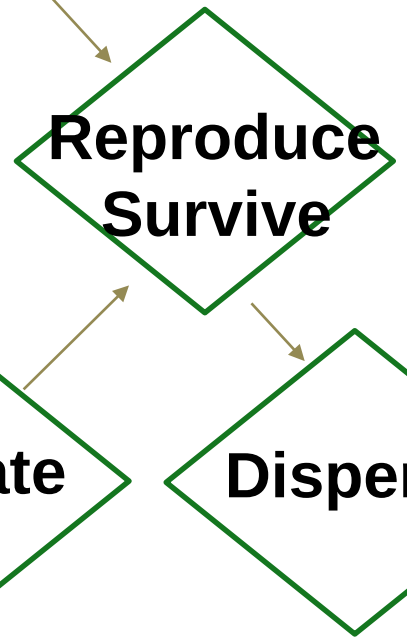


Competition

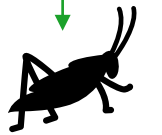
Yes

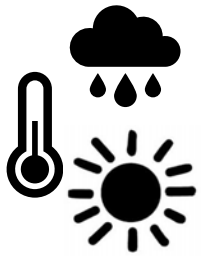
No

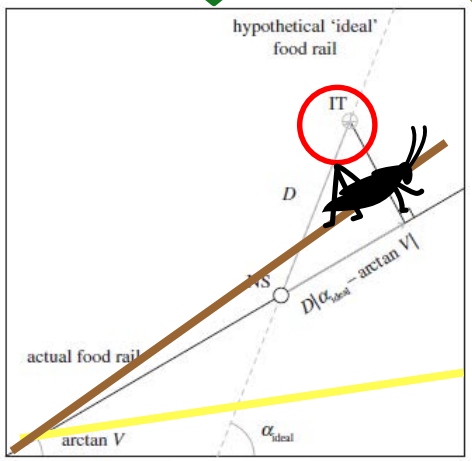


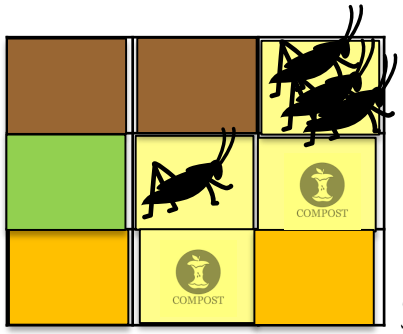


Prédation

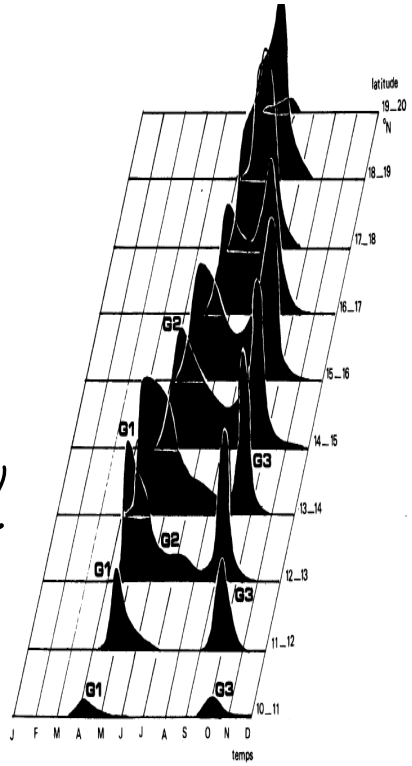








Launois et al. 1979



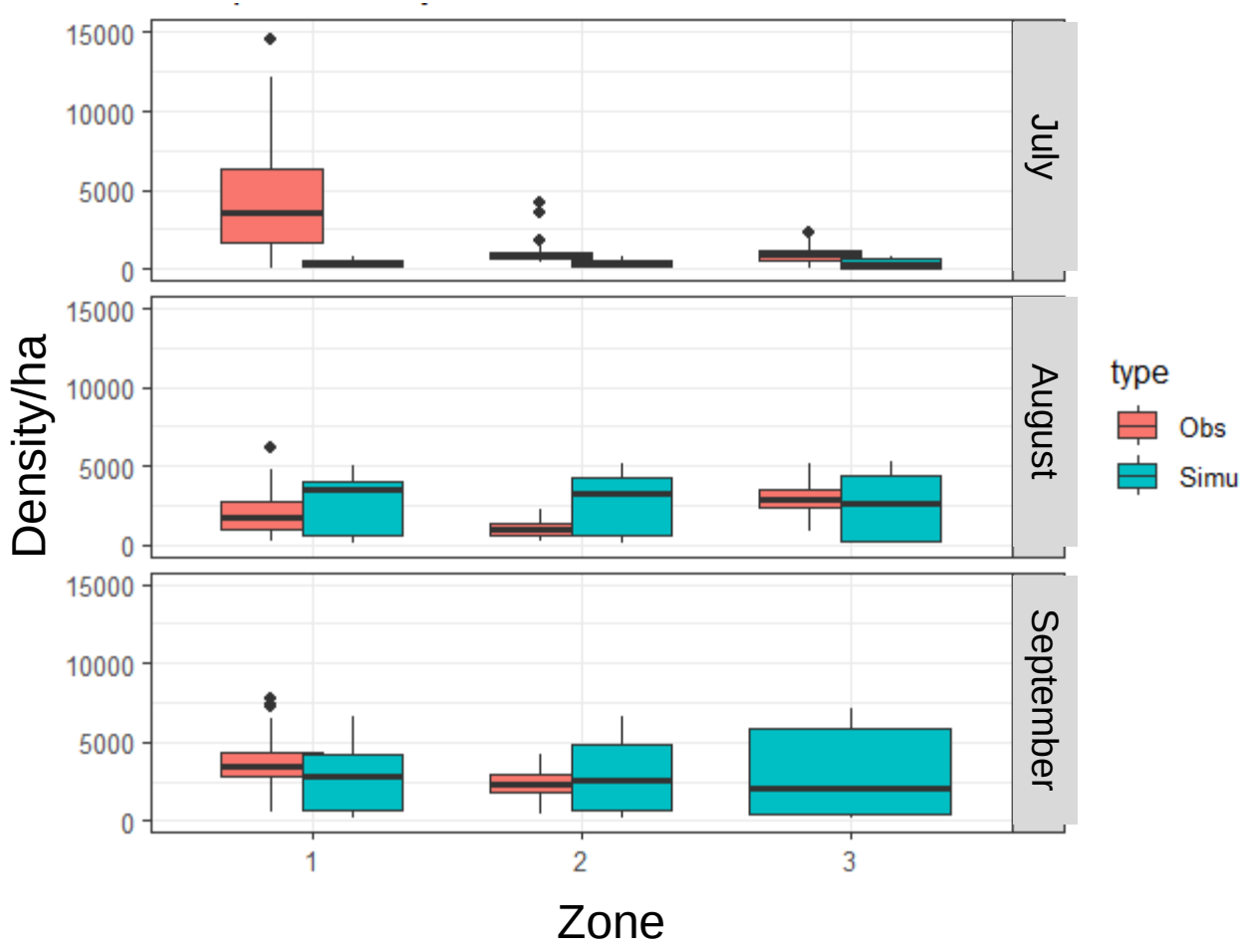
Above-ground biomass and yield for millet (Lavarenne J., model SARRAH)

(Diouf work in progress)

Model Calibration using pattern-oriented modelling

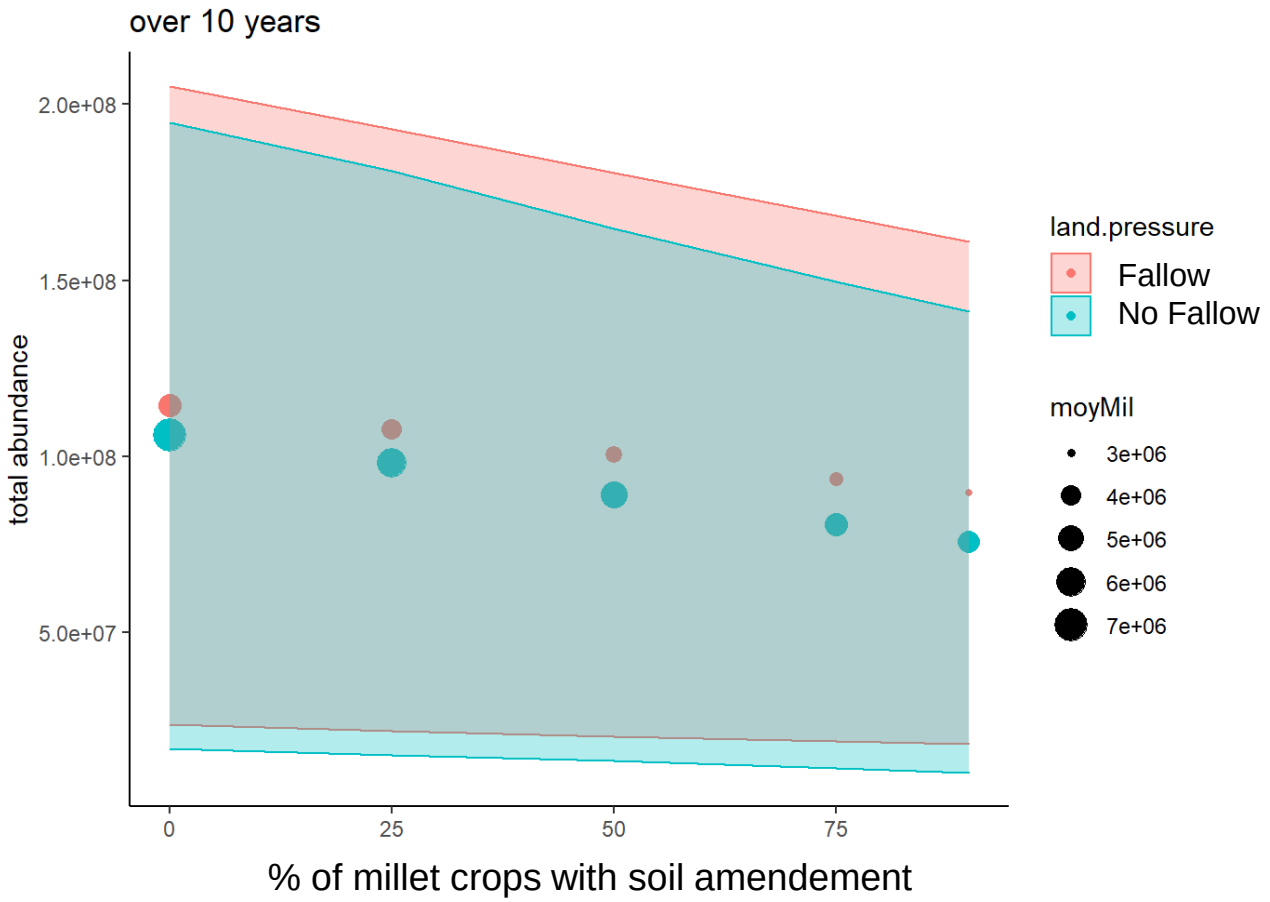
- Sensitive to initial population size
- Late migration date
- Small migration capacity
- Narrow orientation

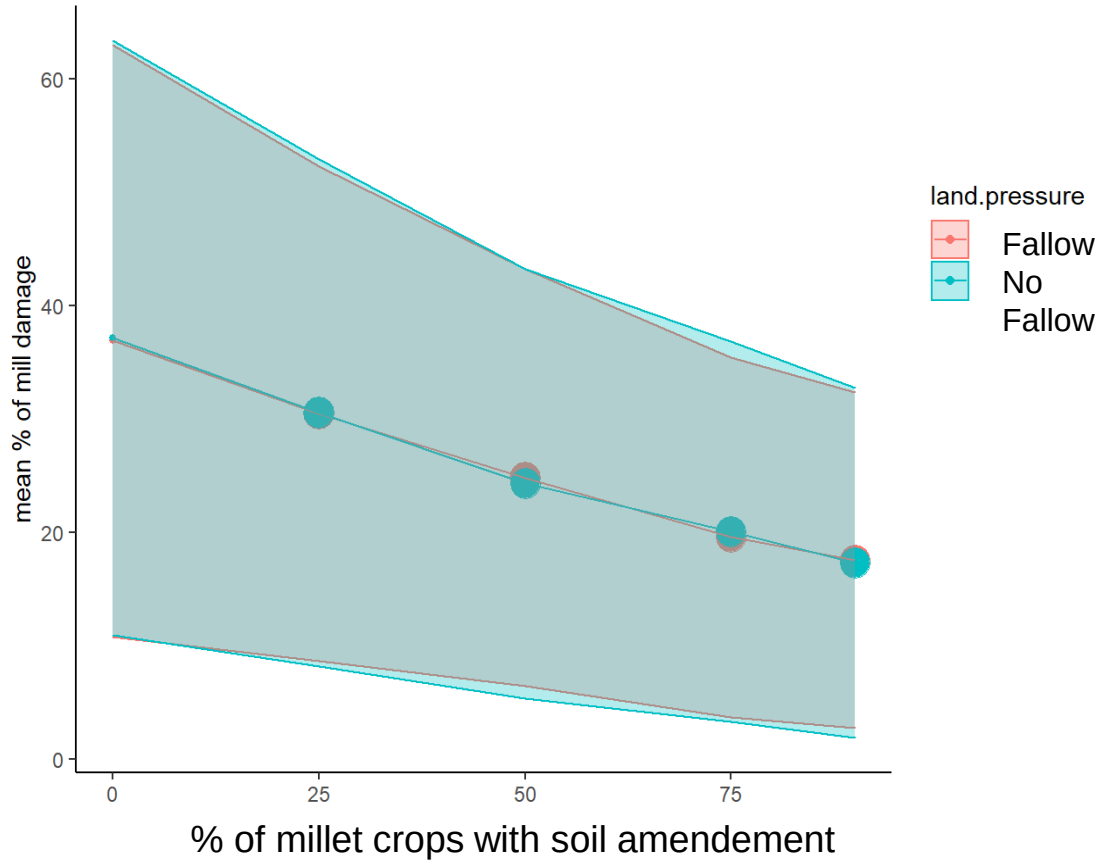
- Robust to cohesion factor and competition



(Touré et al in prep)

Preliminary results





- Soil amendements : no thresholds to prevent outbreaks
- But it can reduce impact on crops by changing grasshopper diet

Objectives

Uncertainties

Forecasted risks

Fundamental

Safe agricultural production

- Institutional/governance

Threats for food security
Threats for biodiversity
Economical costs

How to do it ?

Means

Protect crops/pastures from swarms

Management

Proactive

Preventive

Monitoring

solitarious/transiens bands swarms

Model

Machine learning
Agent based model
Hidden Markov model

- Environmental variability
- Demographic stochasticity

Control uncertainty

Sampling uncertainty

Structural uncertainty

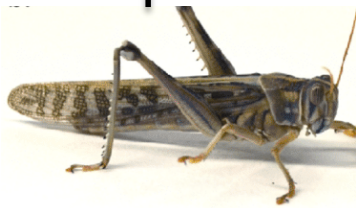
Impacts on production

Outbreak

Gregarization Breeding

Presence/absence

Oedaleus senagelensis



Schistocerca gregaria

Conclusion : which model for which risk level ?

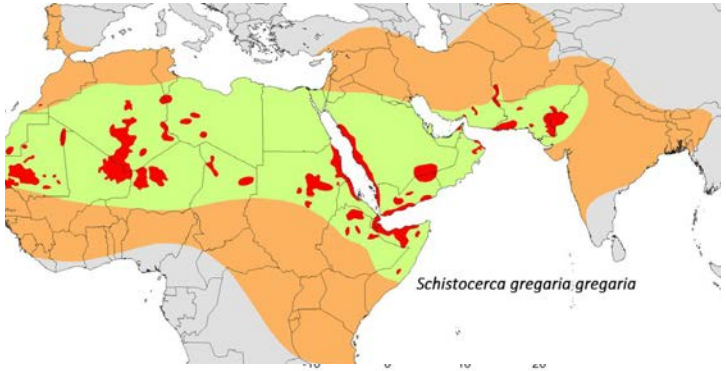
Acridian situation

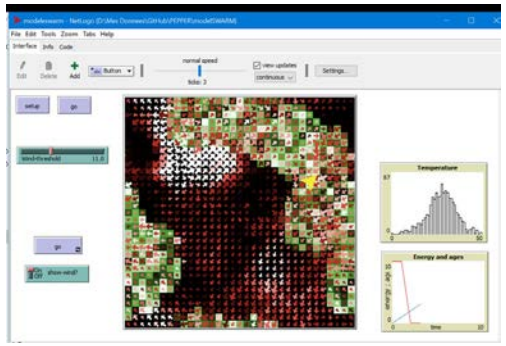
Invasion

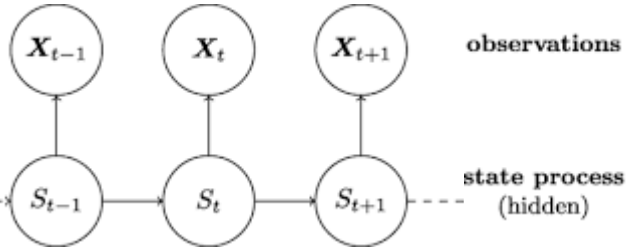
Resurgence: proaction

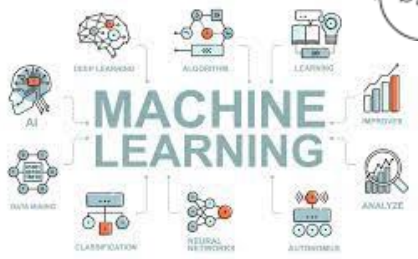
Remission : prevention

Recession : surveillance









Risk levels

Occurence

Group/Phase

Population

Agroecosystems

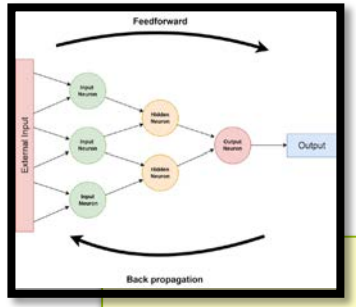
Conlusions and perspectives

- Remote sensing and machine learning are powerful tools to forecast distribution (Piou et al. 2013, 2019, Sun et al. 2021)
- Provide maps with probability of detecting the species
- Used to orientate team in the field and strengthen preventive management
- Less powerful to forecast risk of gregarization (Véran et al. 2015, Lawton et al. 2021)
- Accounting for imperfect and heterogenous detection using occupancy models
- More mechanistic approaches are necessary to predict outbreak (Gay et al. 2021)
- Need for pattern oriented modeling to become operational (Grimm et Railsback 2005)
- Ecology and human forces must stay at the heart of locust management

Thanks for your attention

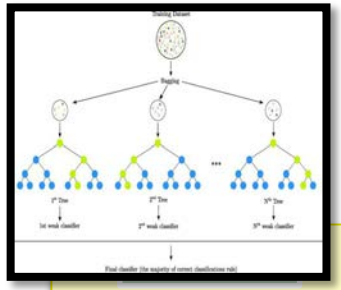
Step 2 & 3

Pros and Cons between algorithms



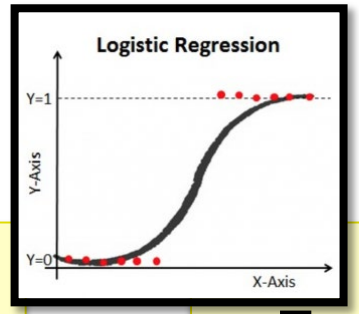
NN

- No hypothesis
- Large data sets
- Any data type
- Adapt to new data/ update prediction
- Long to train
- Difficult interpret
- Overfitting



RF

- No hypothesis
- Large data sets
- Unbalanced data
- Many variables correlated, lagged
- Few preprocessing
- Parallelized
- Robust to outliers
- Overfitting
- Interpretation



LR

- Fast to train
- Deal with NA
- Response Curve
- Effect size
- Linearity