



HAL
open science

Predicting intercrops yield, from data gathering to modeling

Rémi Mahmoud, Noémie Gaudio, Pierre Casadebaig, Nadine Hilgert

► To cite this version:

Rémi Mahmoud, Noémie Gaudio, Pierre Casadebaig, Nadine Hilgert. Predicting intercrops yield, from data gathering to modeling. 17. Congress of the European Society for Agronomy, Aug 2022, Potsdam, Germany. 38p. <hal-03798396>

HAL Id: hal-03798396

<https://hal.inrae.fr/hal-03798396v1>

Submitted on 5 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



Predicting intercrops yield, from data gathering to modeling

ESA - 01/09/2022

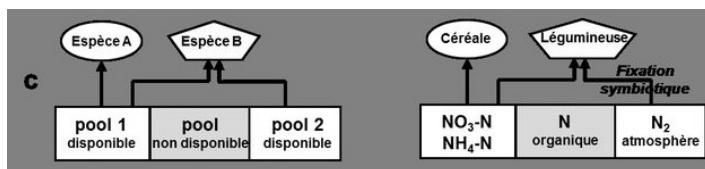
Rémi Mahmoud, Noémie Gaudio, Pierre Casadebaig, Nadine Hilgert

General context

Mixing cereals and legumes in the field decreases the need for exogenous fertilizers

System of interest

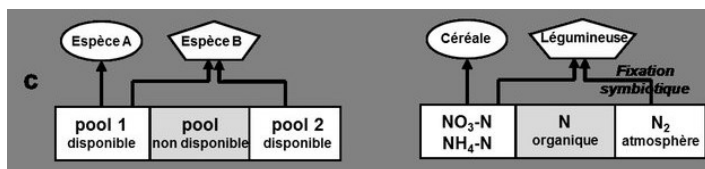
- Cereal-legume intercrops are a promising type of intercrop, in low-input systems:
 - greater yield (Bedoussac, et al. (2015))
 - more stable yield (Raseduzzaman, et al. (2017))
- Legumes rely on symbiotic fixation to use atmospheric N₂ leaving more mineral N for cereal (complementarity)



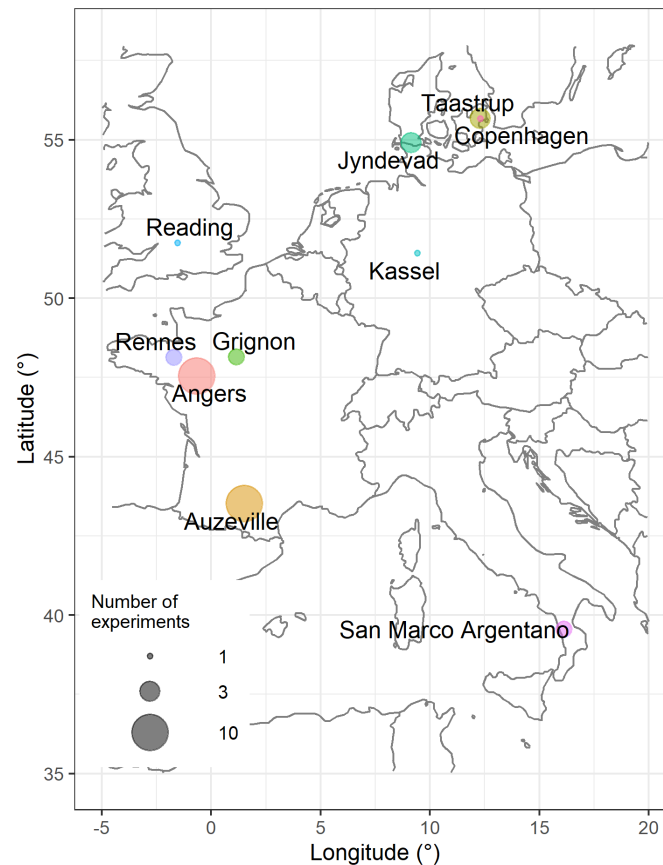
- Data sharing allow us to explore modeling options and validate experimental findings
- **35 trials in 5 European countries**

System of interest

- Cereal-legume intercrops are a promising type of intercrop, in low-input systems:
 - greater yield (Bedoussac, et al. (2015))
 - more stable yield (Raseduzzaman, et al. (2017))
- Legumes rely on symbiotic fixation to use atmospheric N₂ leaving more mineral N for cereal (complementarity)



- Data sharing allow us to explore modeling options and validate experimental findings
- **35 trials in 5 European countries**



Modeling

- Modeling can be a tool to generalize results in various environments
- It can also help to understand processes occurring in complex environments \Rightarrow new experimental ideas

Modeling

- Modeling can be a tool to generalize results in various environments
- It can also help to understand processes occurring in complex environments \Rightarrow new experimental ideas

Models in biology: ‘accurate descriptions of our pathetic thinking’
(Gunawardena, 2014)

- 2 types of modeling
 - *Reverse modeling starts from experimental data and seeks potential causalities suggested by the correlations in the data, captured in the structure of a mathematical model. (phenomenological models)*
 - *Forward modeling starts from known, or suspected, causalities, expressed in the form of a model, from which predictions are made about what to expect. (so-called mechanistic models)*

Modeling

- Modeling can be a tool to generalize results in various environments
- It can also help to understand processes occurring in complex environments \Rightarrow new experimental ideas

Models in biology: ‘accurate descriptions of our pathetic thinking’
(Gunawardena, 2014)

- 2 types of modeling
 - *Reverse modeling starts from experimental data and seeks potential causalities suggested by the correlations in the data, captured in the structure of a mathematical model. (phenomenological models)*
 - *Forward modeling starts from known, or suspected, causalities, expressed in the form of a model, from which predictions are made about what to expect. (so-called mechanistic models)*

Prediction in ecology: promises, obstacles and clarifications (Maris et al., 2018)

- 2 types of predictions
 - **corroboratory prediction is linked to the validation of theories**
 - anticipatory-prediction is linked to the description of possible futures

Modeling of intercrop (IC) systems:

Review: Gaudio, et al. (2019)

Modeling of intercrop (IC) systems:

Review: Gaudio, et al. (2019)

- | Intercrop modeling is just beginning.

Modeling of intercrop (IC) systems:

Review: Gaudio, et al. (2019)

Intercrop modeling is just beginning.

Few crop models and individual-based models exist to simulate crop mixtures (mechanistic models)

Modeling of intercrop (IC) systems:

Review: Gaudio, et al. (2019)

Intercrop modeling is just beginning.

Few crop models and individual-based models exist to simulate crop mixtures (mechanistic models)

Phenomenological approaches seem promising to understand and define assembly rules for intercrops ((Forst, et al., 2019), (Montazeaud, et al., 2020))

Predictions generated by these approaches may be used to understand IC functioning (covariates ranking etc.) ; corroboratory predictions

Modeling of intercrop (IC) systems:

Review: Gaudio, et al. (2019)

Intercrop modeling is just beginning.

Few crop models and individual-based models exist to simulate crop mixtures (mechanistic models)

Phenomenological approaches seem promising to understand and define assembly rules for intercrops ((Forst, et al., 2019), (Montazeaud, et al., 2020))

Predictions generated by these approaches may be used to understand IC functioning (covariates ranking etc.) ; corroboratory predictions

- Modelling approaches and scales are complementary (Gaudio, et al., 2022)

Data gathering and processing workflow

Problems and methods linked to global datasets

Problem

- Multiple experiments
 - Different purposes
 - Different environments
 - Different species/cultivars
 - Different agronomic practices

Problem

- Multiple experiments
 - Different purposes
 - Different environments
 - Different species/cultivars
 - Different agronomic practices
- Multiple data formats
 - Need for data standardization

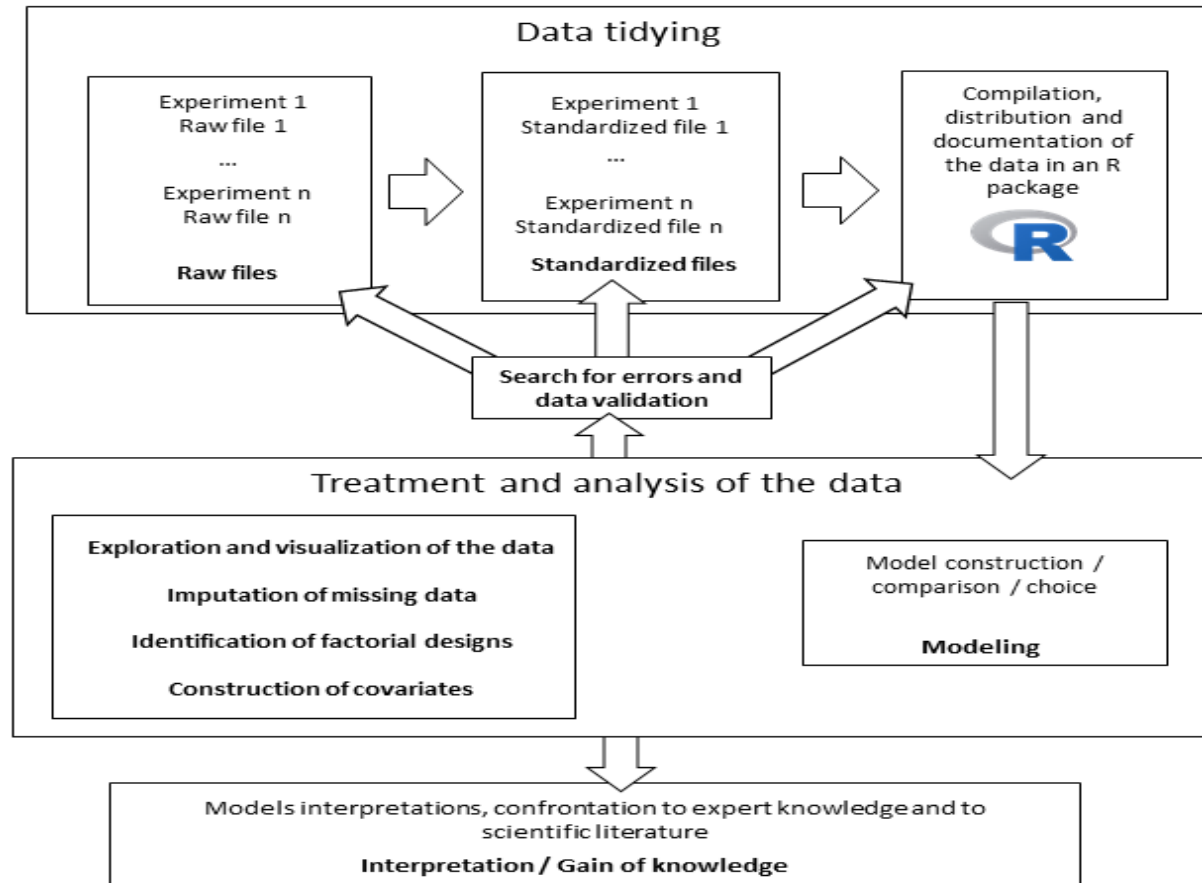
Problem

- Multiple experiments
 - Different purposes
 - Different environments
 - Different species/cultivars
 - Different agronomic practices
- Multiple data formats
 - Need for data standardization
- Multiple research questions
 - Aggregation of several experimental designs \Rightarrow Need to identify subsets inducing complete factorial designs to answer different scientific questions
 - Maximal k-clique enumeration

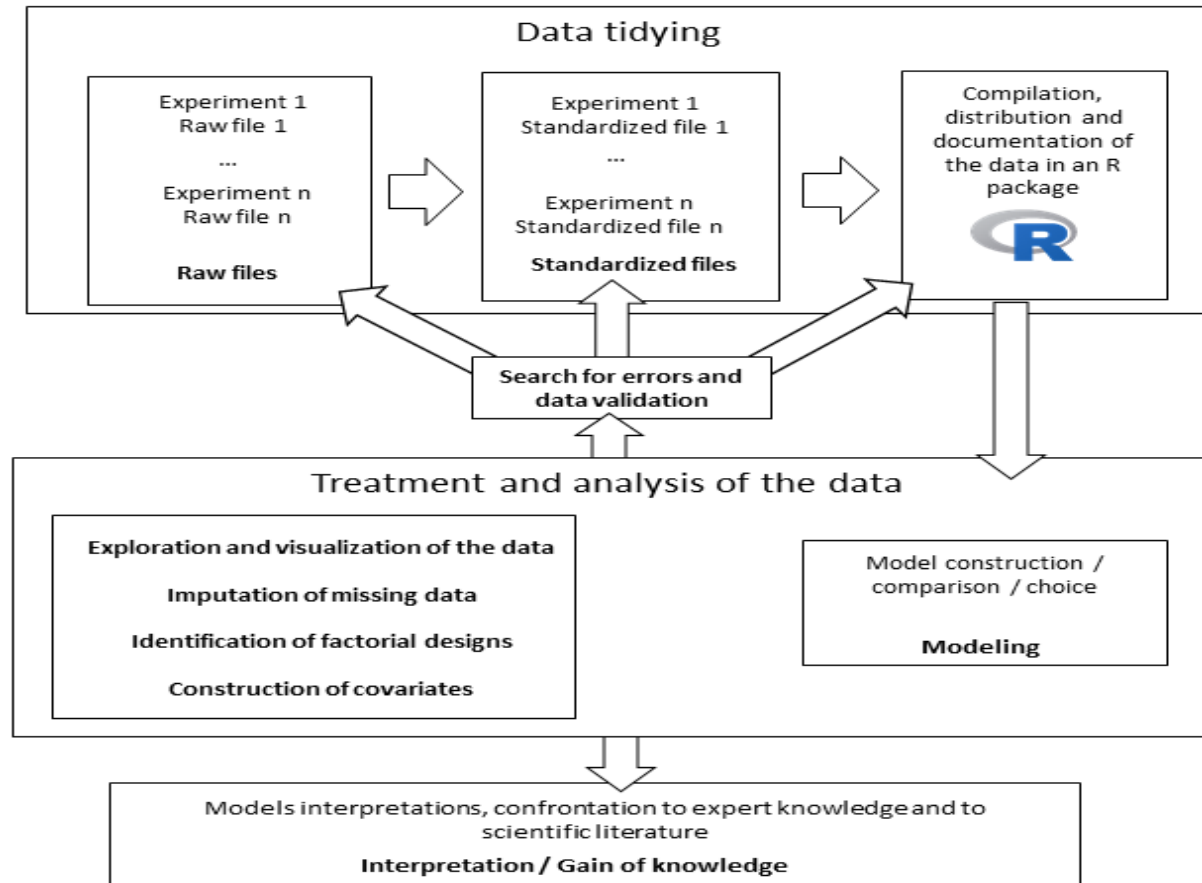
Problem

- Multiple experiments
 - Different purposes
 - Different environments
 - Different species/cultivars
 - Different agronomic practices
- Multiple data formats
 - Need for data standardization
- Multiple research questions
 - Aggregation of several experimental designs \Rightarrow Need to identify subsets inducing complete factorial designs to answer different scientific questions
 - Maximal k-clique enumeration
- Sharing and processing data is key to perform a joint analysis of diverse experiments

Approach



Approach



Preprint

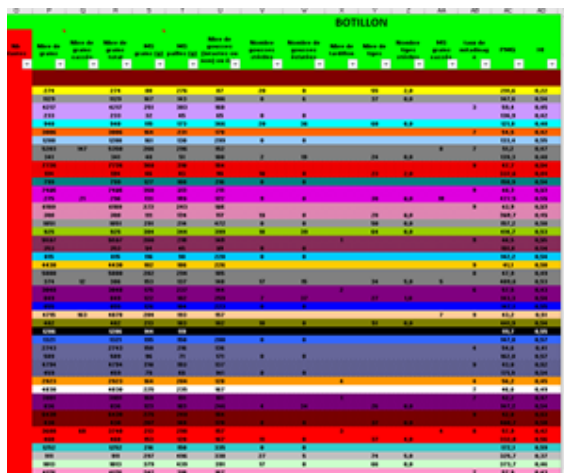
Data standardization

- Raw data to **standardized** data files
- Same variable names, same data structure (Pappagallo, et al., 2021)
- Not possible to automatize

The image shows a screenshot of a data table with a green header bar containing the title 'BOTTILLON'. The table has approximately 15 columns and many rows. The data is presented in a grid format with alternating colors for rows (e.g., red, blue, green, yellow). The columns contain numerical values, and some cells are highlighted in red, possibly indicating missing or specific data points. The overall layout is dense and typical of a raw data export.

Data standardization

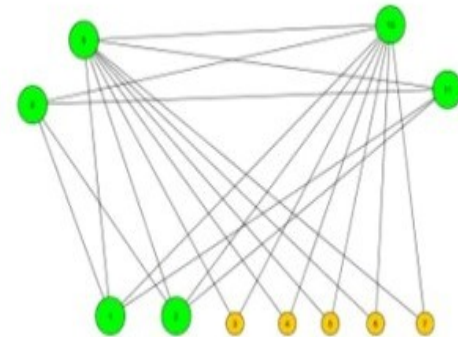
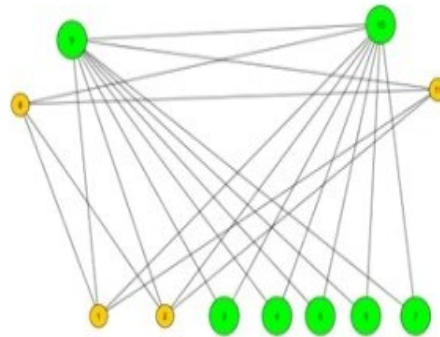
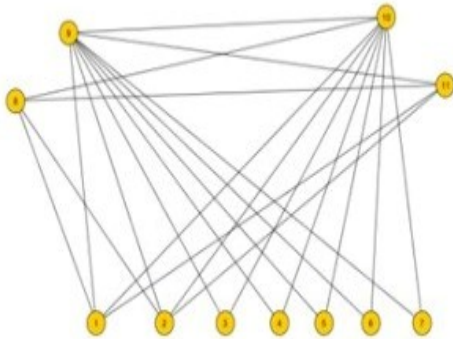
- Raw data to **standardized** data files
- Same variable names, same data structure (Pappagallo, et al., 2021)
- Not possible to automatize



	A	B	C	D	E	F	G	H
code_crop		cultivar	block	x	y	final_density,2003-03-	height,2003-05-	height,2003-06-
SC_B_scarlett_1_NO_PO		scarlett	1	1	1	250	0.30	0.60
SC_B_scarlett_1_NO_PO		scarlett	2	2	1	266	0.36	0.62
SC_B_scarlett_1_NO_PO		scarlett	3	3	1	268	0.31	0.61
SC_B_scarlett_1_NO_PO		scarlett	4	4	1	262	0.36	0.63
SC_P_baccara_1_NO_PO		baccara	1	1	2	110	0.37	0.60
SC_P_baccara_1_NO_PO		baccara	2	2	2	78	0.39	0.54
SC_P_baccara_1_NO_PO		baccara	3	3	2	88	0.35	0.51
SC_P_baccara_1_NO_PO		baccara	4	4	2	76	0.31	0.51
IC_B_scarlett_baccara_0.5_0.5_NO_PO		scarlett	1	1	3	150	0.33	0.64
IC_B_scarlett_baccara_0.5_0.5_NO_PO		scarlett	2	2	3	188	0.35	0.62
IC_B_scarlett_baccara_0.5_0.5_NO_PO		scarlett	3	3	3	156	0.32	0.66
IC_B_scarlett_baccara_0.5_0.5_NO_PO		scarlett	4	4	3	178	0.35	0.66
IC_B_scarlett_baccara_0.5_1_NO_PO		scarlett	1	1	4	130	0.37	0.64
IC_B_scarlett_baccara_0.5_1_NO_PO		scarlett	2	2	4	130	0.32	0.71
IC_B_scarlett_baccara_0.5_1_NO_PO		scarlett	3	3	4	160	0.34	0.67
IC_B_scarlett_baccara_0.5_1_NO_PO		scarlett	4	4	4	144	0.38	0.68
IC_P_scarlett_baccara_0.5_0.5_NO_PO		baccara	1	1	3	34	0.35	0.50
IC_P_scarlett_baccara_0.5_0.5_NO_PO		baccara	2	2	3	50	0.31	0.49
IC_P_scarlett_baccara_0.5_0.5_NO_PO		baccara	3	3	3	50	0.30	0.47
IC_P_scarlett_baccara_0.5_0.5_NO_PO		baccara	4	4	3	52	0.35	0.50
IC_P_scarlett_baccara_0.5_1_NO_PO		baccara	1	1	4	84	0.38	0.48

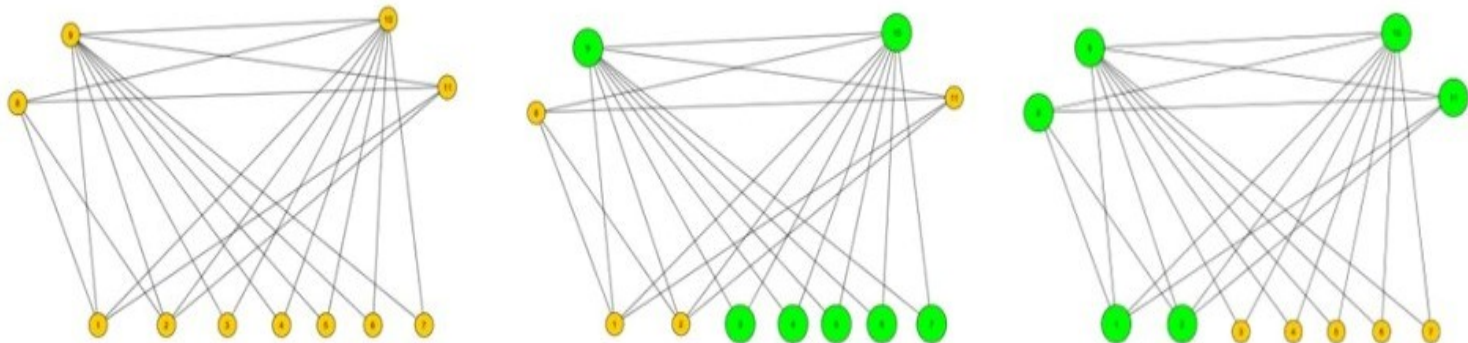
Enumerate complete factorial designs

- First, some elements of graph theory:
 - A **k-partite** graph and associated **maximal k-cliques**



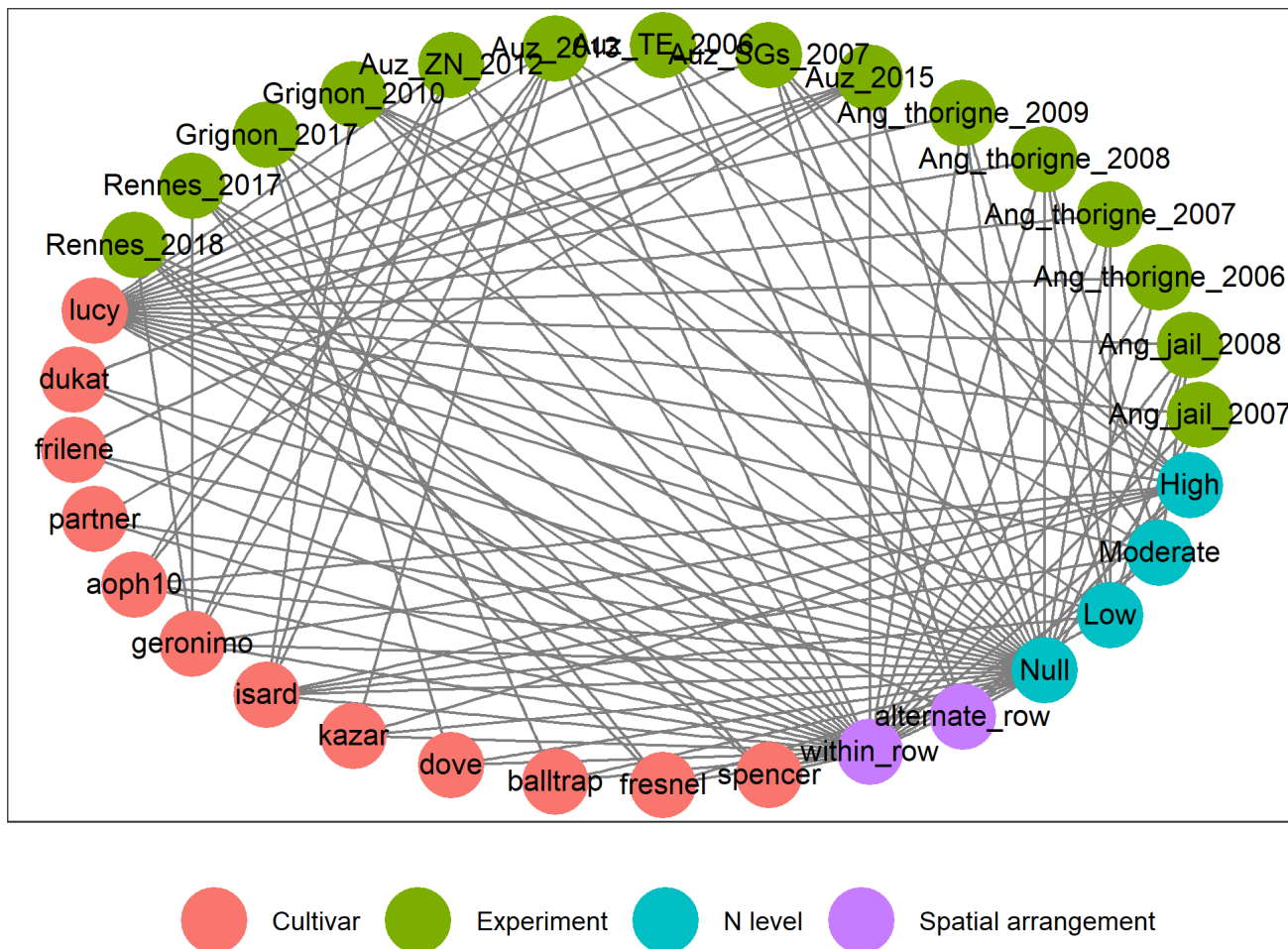
Enumerate complete factorial designs

- First, some elements of graph theory:
 - A **k-partite** graph and associated **maximal k-cliques**



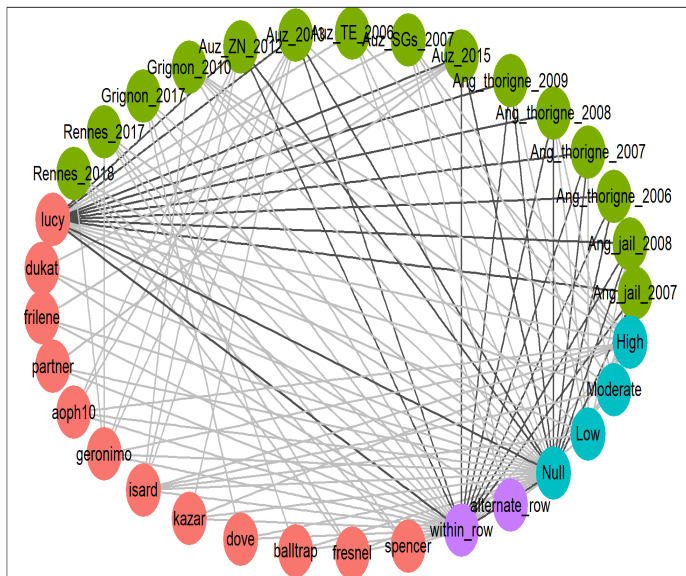
- Enumerate all maximal k-cliques \Rightarrow MMCE algorithm (Phillips, et al., 2019)

In my case, each factor is a partite set, each factor-level is a node

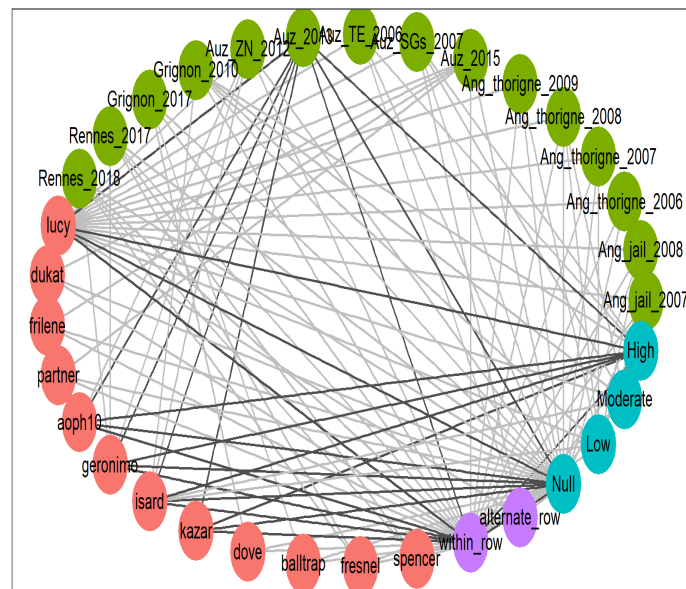


In my case, each factor is a partite set, each factor-level is a node

Enumerating all complete factorial designs \iff Enumerate all maximal k-cliques



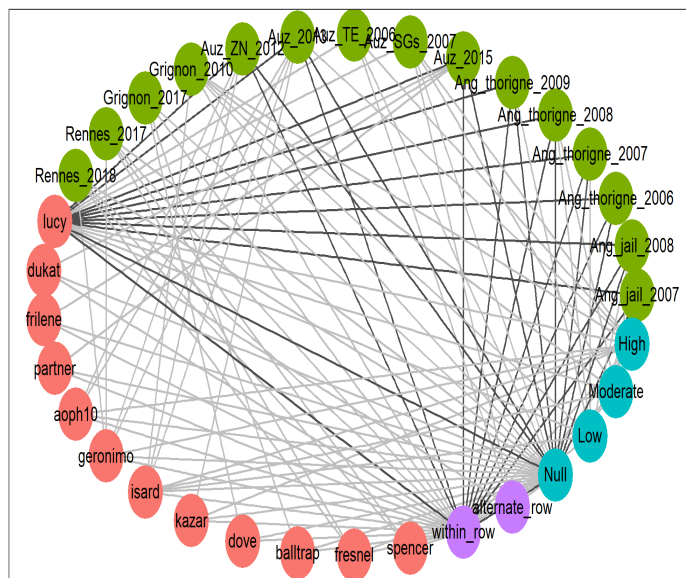
1 cultivar ; 9 experiments



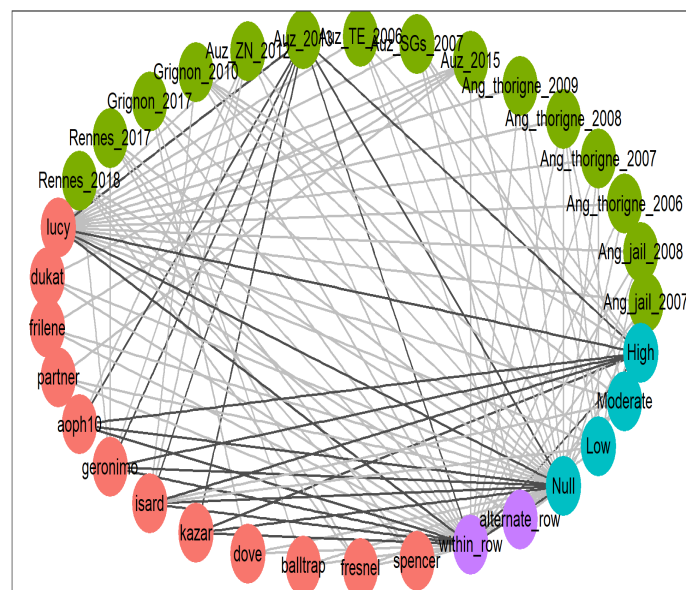
5 cultivars ; 1 experiment ; 2 N levels

In my case, each factor is a partite set, each factor-level is a node

Enumerating all complete factorial designs \iff Enumerate all maximal k-cliques



1 cultivar ; 9 experiments



5 cultivars ; 1 experiment ; 2 N levels

Code available and R package installable at <https://github.com/RemiMahmoud/kclique>

Modeling

How to use available data and ecological knowledge to predict the yield of intercrops

Global modeling approach

Purpose: Predict the yield of various intercrops, in different environments

Global modeling approach

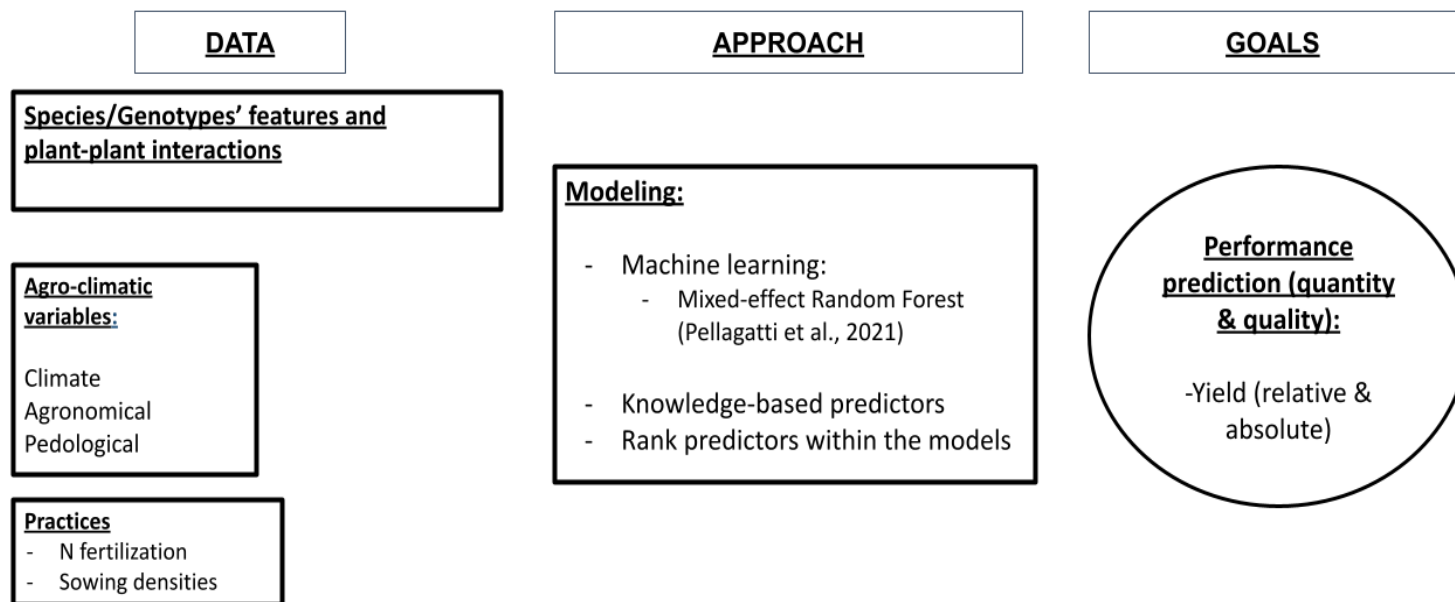
Purpose: Predict the yield of various intercrops, in different environments

Available data: Environment x Plant variables x Agronomic practices

Global modeling approach

Purpose: Predict the yield of various intercrops, in different environments

Available data: Environment x Plant variables x Agronomic practices



Features computations, using ecological knowledge

Problematic: Reduce dimension of variables measured several times across the cropping season

Features computations, using ecological knowledge

Problematic: Reduce dimension of variables measured several times across the cropping season

Steps:

1. Relevant fit for each curve
2. Extraction of key features (lag phase λ , slope μ at inflection point) of the curves
3. Compute plant-plant interactions' predictors from these curves

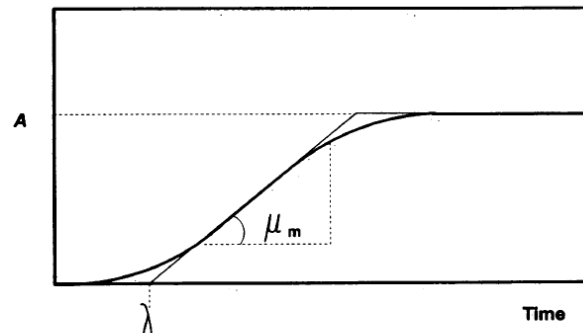
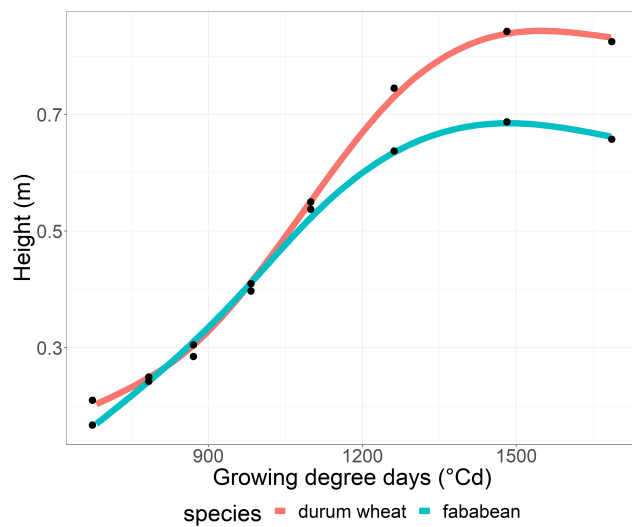


FIG. 1. A growth curve.

Feature computations: plant-plant interactions

Within intercrops differences between curve parameters \Rightarrow complementarity/competition indicators (ex.

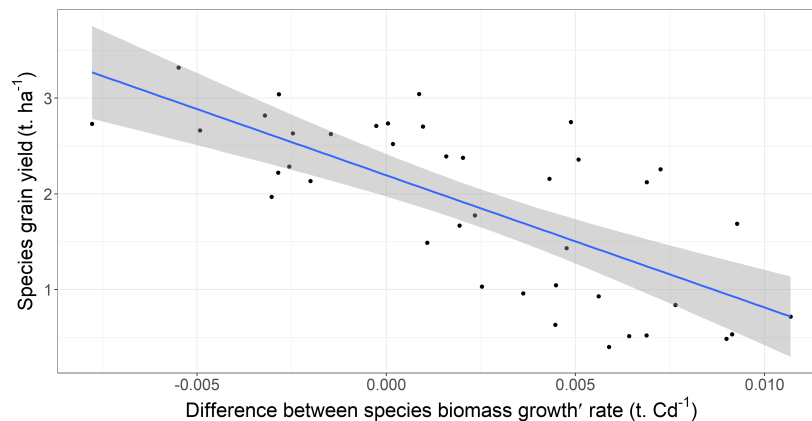
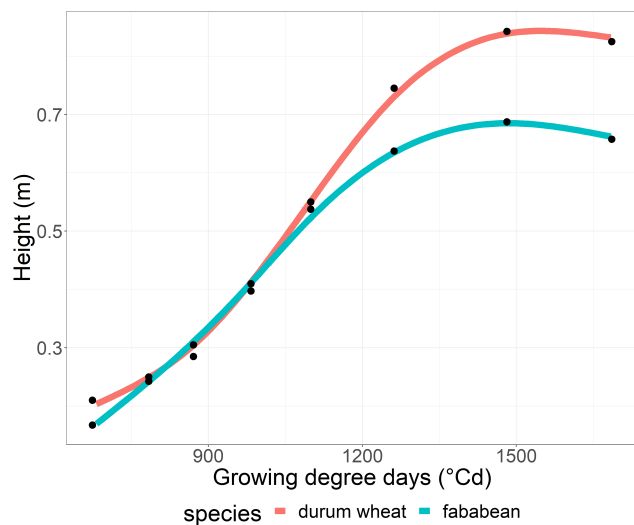
$$\Delta\mu = \mu_C - \mu_L)$$



Feature computations: plant-plant interactions

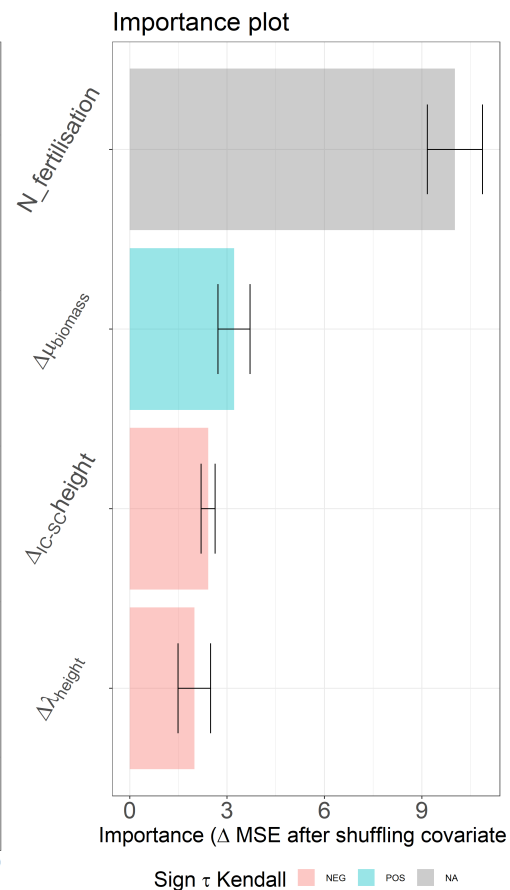
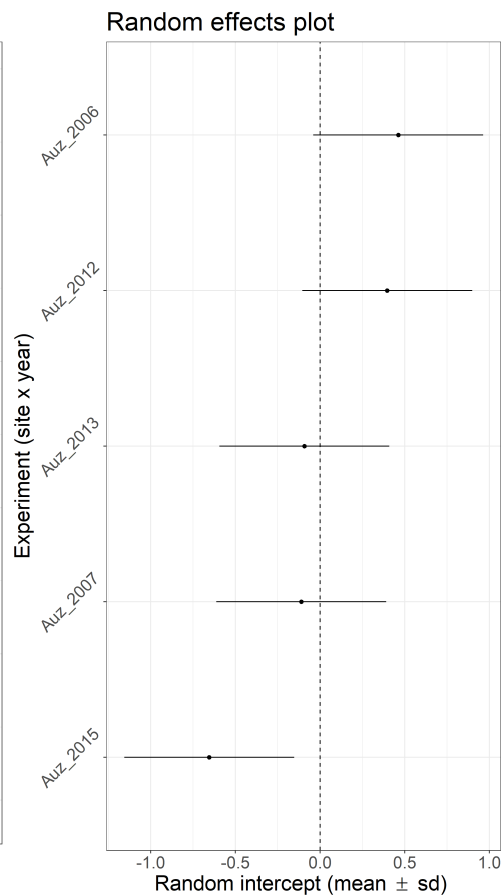
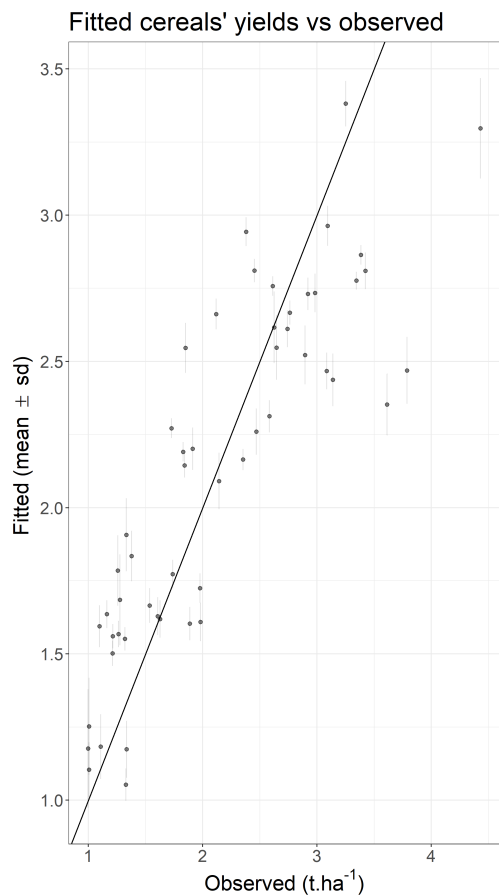
Within intercrops differences between curve parameters \Rightarrow complementarity/competition indicators (ex.

$$\Delta\mu = \mu_C - \mu_L)$$



Results

Results, durum wheat / pea mixtures, Y = wheat' yield ($\overline{RMSE} = 0.46 \text{ t. ha}^{-1}$, $\overline{EF} = 0.7$)



Partial conclusions & Perspectives

- Results highly depend on species mixtures and on the covariates ($RMSE \in [0.1; 0.8]$)
- Environment included as random effect in the model for now
 - Regression approaches tested to reduce environmental data dimension until now, with little success
 - Other approaches are tried (FPCA ; Clustering of curves ; Wavelets decomposition etc.)
- Random forest + variable selection + random effect is promising

Partial conclusions & Perspectives

- Results highly depend on species mixtures and on the covariates ($RMSE \in [0.1; 0.8]$)
- Environment included as random effect in the model for now
 - Regression approaches tested to reduce environmental data dimension until now, with little success
 - Other approaches are tried (FPCA ; Clustering of curves ; Wavelets decomposition etc.)
- Random forest + variable selection + random effect is promising
- Any remark, questions ?
 - remi.mahmoud@inrae.fr

