Differences in growth features between species are driving cereal-legume intercrop yield.

Pierre Casadebaig, Noémie Gaudio, Rémi Mahmoud, Nadine Hilgert

INRAE, University of Toulouse, UMR AGIR, France

18th Congress of the European Society for Agronomy in Rennes, France

Increase plant diversity at the crop level ?

Results from diversification experiments in ecology can lead to overly optimistic interpretations for agriculture [Cardinale2007, Loreau2021]

Diversification *per se* is not the cause of productivity increase [McGuire2023, Dee2023]

Cover crops
best mixture ~ best monoculture [Florence2020]

Intercrops
add a legume species, esp. in low N management, or extend the cropped time with relay cropping [Jones2023, McLaren2023]

System description (agronomy)

Concepts

Production relative to :

Area (grain yield, Yi)

Area and management

land equivalent ratio = $\sum_{i=1}^{m} \frac{I(Y_i)}{S(Y_i)}$

Practice: Data curation and tidying

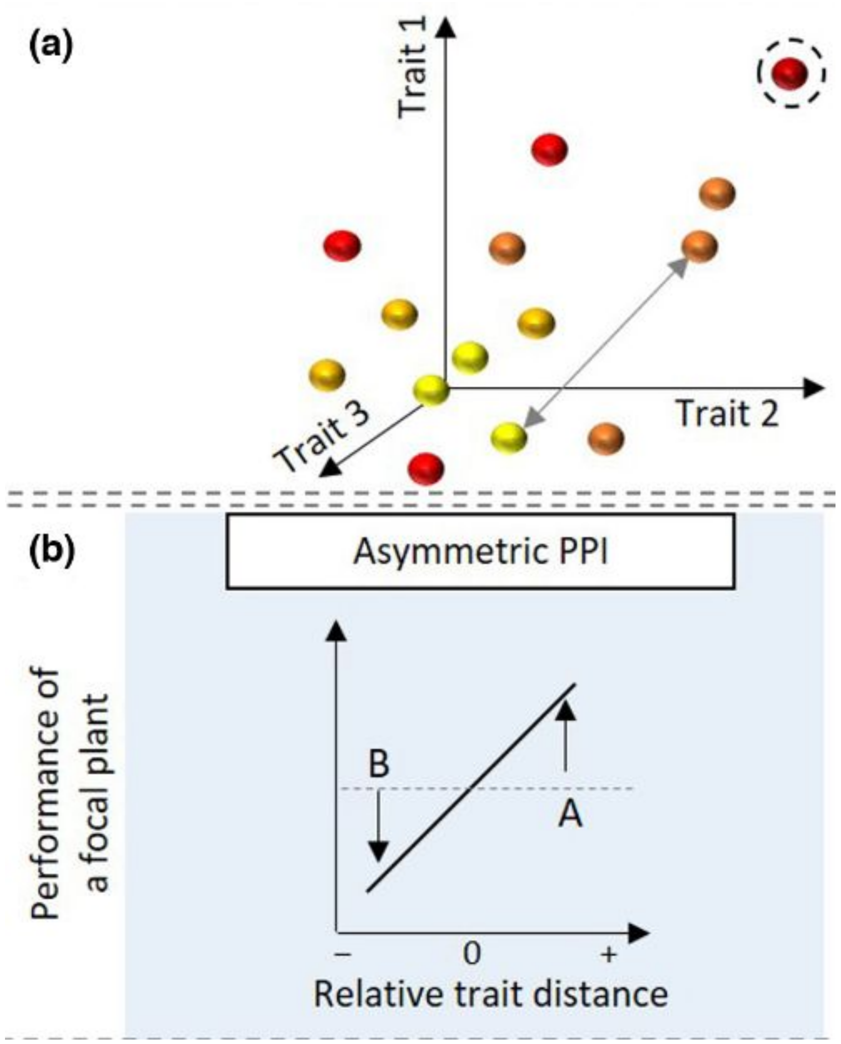

Peer Community Journal
Section: Mathematical & Computational Biology
Research article
Published 2024-02-29
Cite as
Rémi Mahmoud, Pierre Casadebaig, Nadine Hilgert and Noémie Gaudio (2024) A workflow for processing global datasets: application to intercropping, Peer Community Journal, 4: e24.
Correspondence
pierre.casadebaig@inrae.fr
A workflow for
datasets: applic
intercropping
Rémi Mahmoud [1], Pi
Hilgert[2], and Noémie
Volume 4 (2024), article e2
https://doi.org/10.24072/p
zenodo
Search records...
Communities
My dashboard
Published June 26, 2023 | Version 1.0.0
Dataset
Open
A global dataset gathering 37 field experiments involving cereal-legume intercrops and their corresponding sole crops.
Noémie Gaudio[1]; Rémi Mahmoud[1]; Laurent Bedoussac[2]; Eric Justes[3];
Etienne-Pascal Journet[1]; Christophe Naudin[4]; Henrik Hauggaard-Nielsen[5];
Erik Steen Jensen[6]; Elise Pelzer[7]; Guénaëlle Corre-Hellou[4]; Bochra Kammoun[8];
Loïc Viguier[9]; Romain Barillot[10]; Antoine Couëdel[11]; Philippe Hinsinger[12];
Pierre Casadebaig[1]
Show affiliations

Experiments in 5 countries, 2001-2017, ~ 600 units {location, year, management}, open-data [Gaudio2023, Mahmoud2024]

Measurements of performance (seeds, shoots), components (leaf area index) as a function of time (biomass, height, nitrogen content)

System description (ecology)

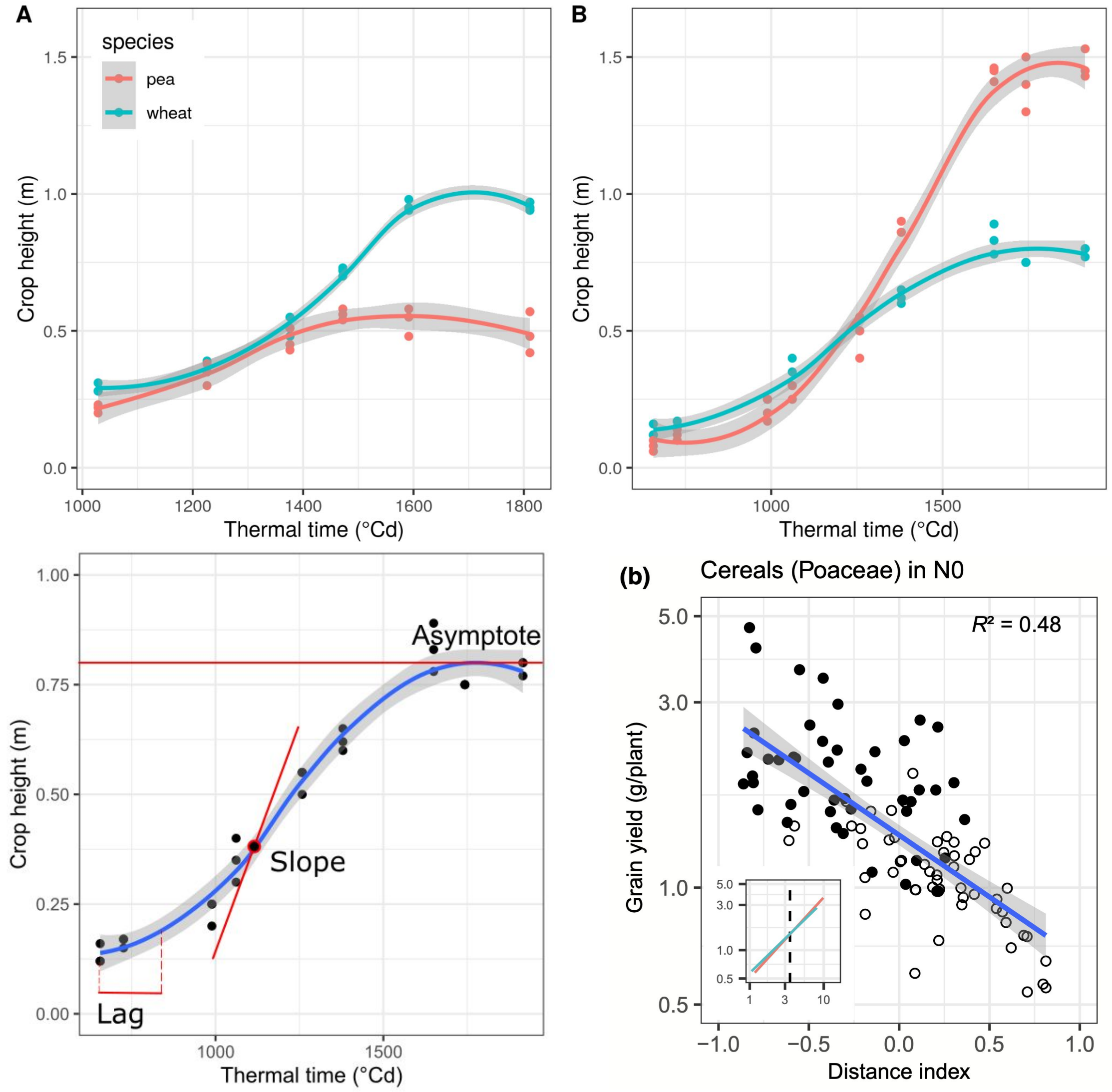
Concepts



[Mahaut2023]

Practice

[Gaudio2021]



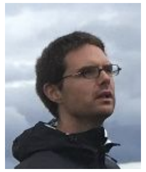
Aims and methods for modeling crop mixtures

> *Process-based models do not scale well from sole crop to less controlled and more species-varied crop stands [Gaudio2019].*

Understand the determinants of canopy performance.

- Statistical learning on features derived from our knowledge on crop ecology.
- Sort and prioritize relevant predictors.

Rémi Mahmoud *PhD & teams (INRAE, UT, CNRS)*




- ***Agronomy*** *: Laurent Bedoussac, (Pierre Casadebaig)*
- ***Ecology*** *: Florian Fort, Noémie Gaudio, Cyrille Violle*
- ***Applied Mathematics*** *: Xavier Gendre, Nadine Hilgert*

Overview of the modeling approach

Raw data

Durum wheat / legume intercrops
Wheat / faba bean (n=39); wheat / pea (n=47)
+ sole crops

⚠️ ~15 % of the complete dataset (~ 60 % for simple studies)

Data processing for modeling

Data pre-computation

Plant growth metrics
Times series summary (biomass, height)
→ 2 parameters per time series (onset, inflexion)

Agricultural practices-related features
Nitrogen nutrition index → 1 explanatory variable
Cultivar identity → 2 explanatory variables

▽ Dataset with raw and computed data

Multiple imputation → 10 imputed versions of the dataset

▽

Interspecific indicators
Interspecific differences within intercrops
→ 6 explanatory variables per model

Intraspecific indicators
Intraspecific differences between inter- and sole-cropping conditions → 7 explanatory variables

▼ 16 explanatory variables

Modeling approach

Models development
3 types of model: Linear Mixed-Effect (LME), Random Forest (RF), Mixed-Effect Random Forest (MERF)
→ 3 models (LME, RF, MERF) per species per imputed dataset

Models evaluation
RMSE on training and test datasets

▽

Variable selection (on RF-based models)
Boruta method (sorting and prioritizing variable importance)
→ 4-7 explanatory variables per model

Balance data-driven and concept-driven approaches.

Features

Plant

Traits (3)
biomass, height, area
▼
Features (3)
inflexion, asymptote, lag
▼
Metrics (2)
relative distance in IC,
between IC and SC

Management

crop nitrogen status [Louarn2021]
cultivar modality

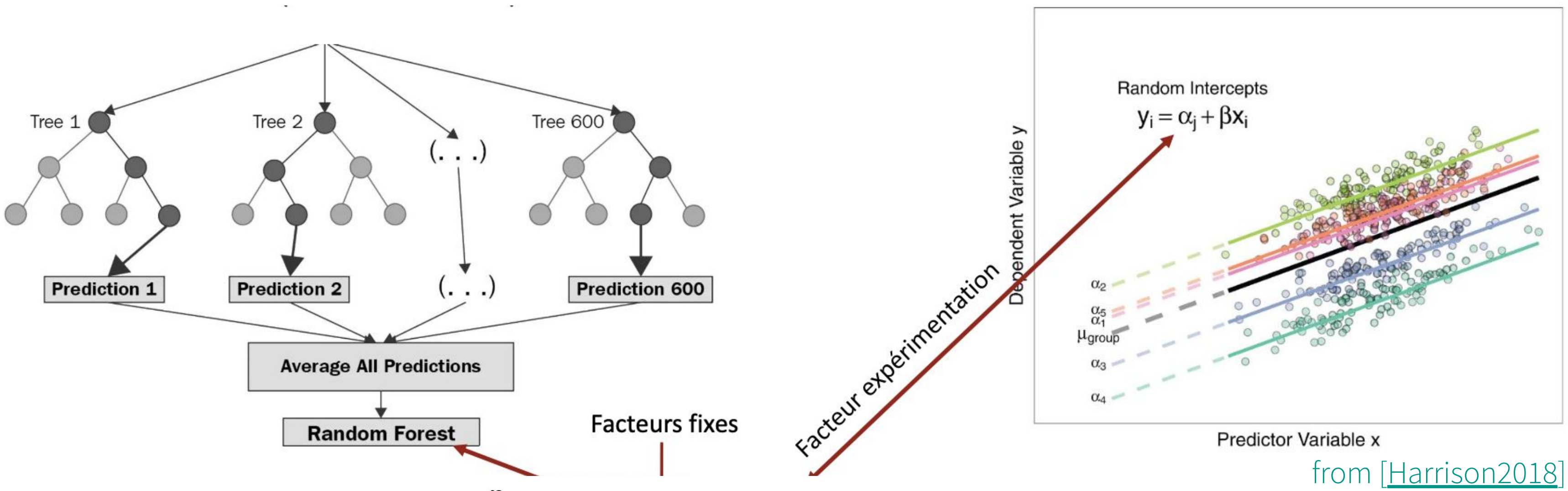
Environment

~~functional regression~~
experiment modality

Model

Random forest ~ crop features

Mixed model ~ environment

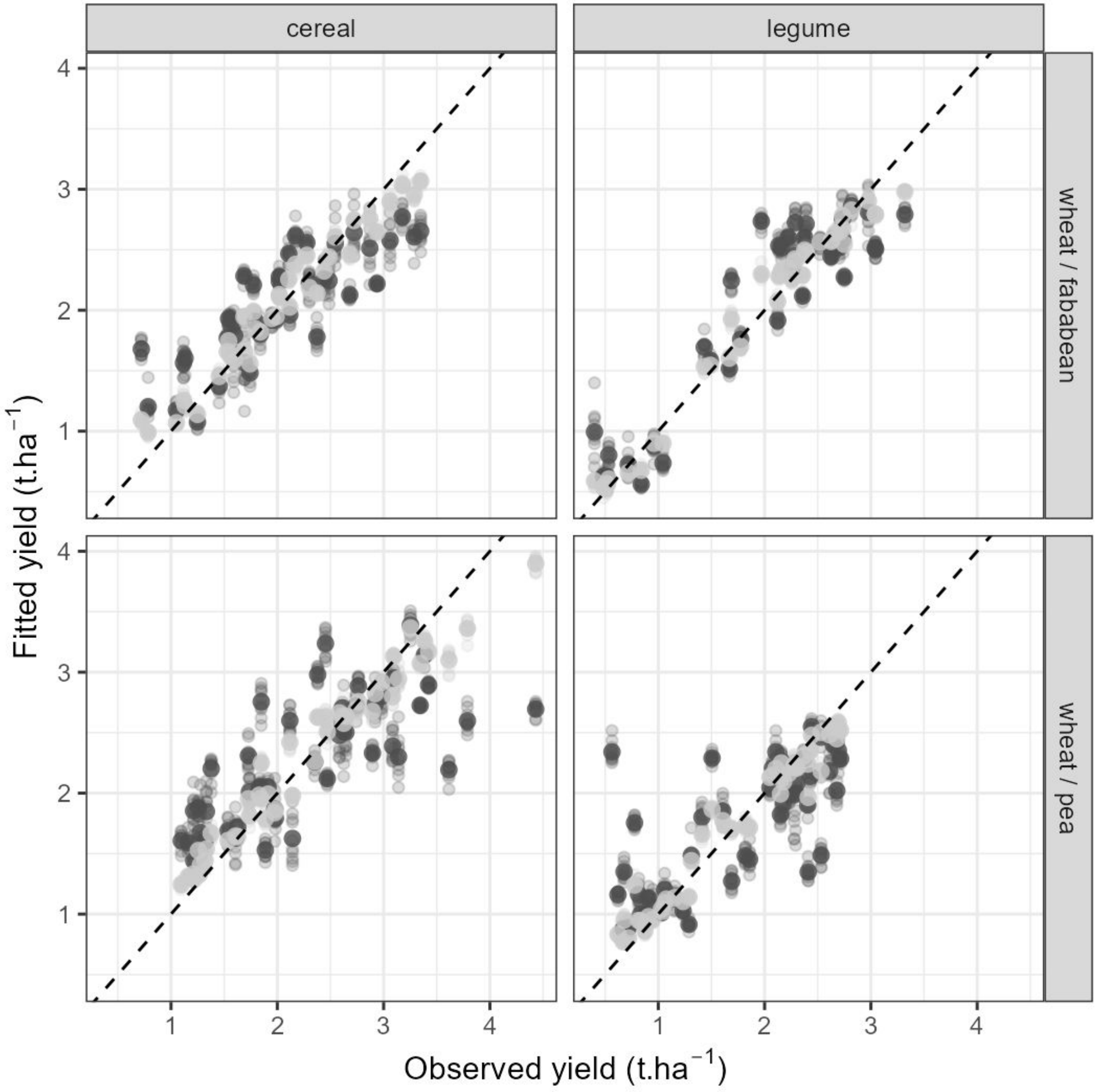


from [Harrison2018]

$$L_{MERF}(\mathcal{T}, u) = \sum_{i=1}^{n} \left(y_i - g_{\mathcal{T}}(x_i) - u^\top z_i\right)^2$$

Combine **random forest** non-linearity with within-experiment error structure from **linear mixed models** [Hajjem2014]

ML is an accurate and useful method to study systems



Accuracy

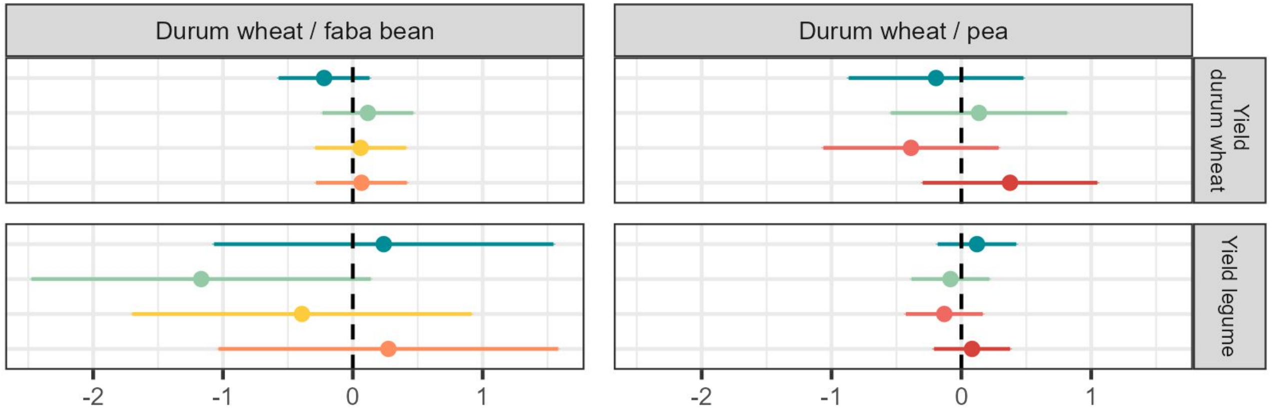
Prediction error (RMSE, test) was twice lower for RF model (0.45 $t.ha^{-1}$) than for linear mixed-model (0.87 $t.ha^{-1}$).

Robustness

Harsh variable selection: variable permutation and binomial testing (Kursa2010). From 16 to 4-7 variables.

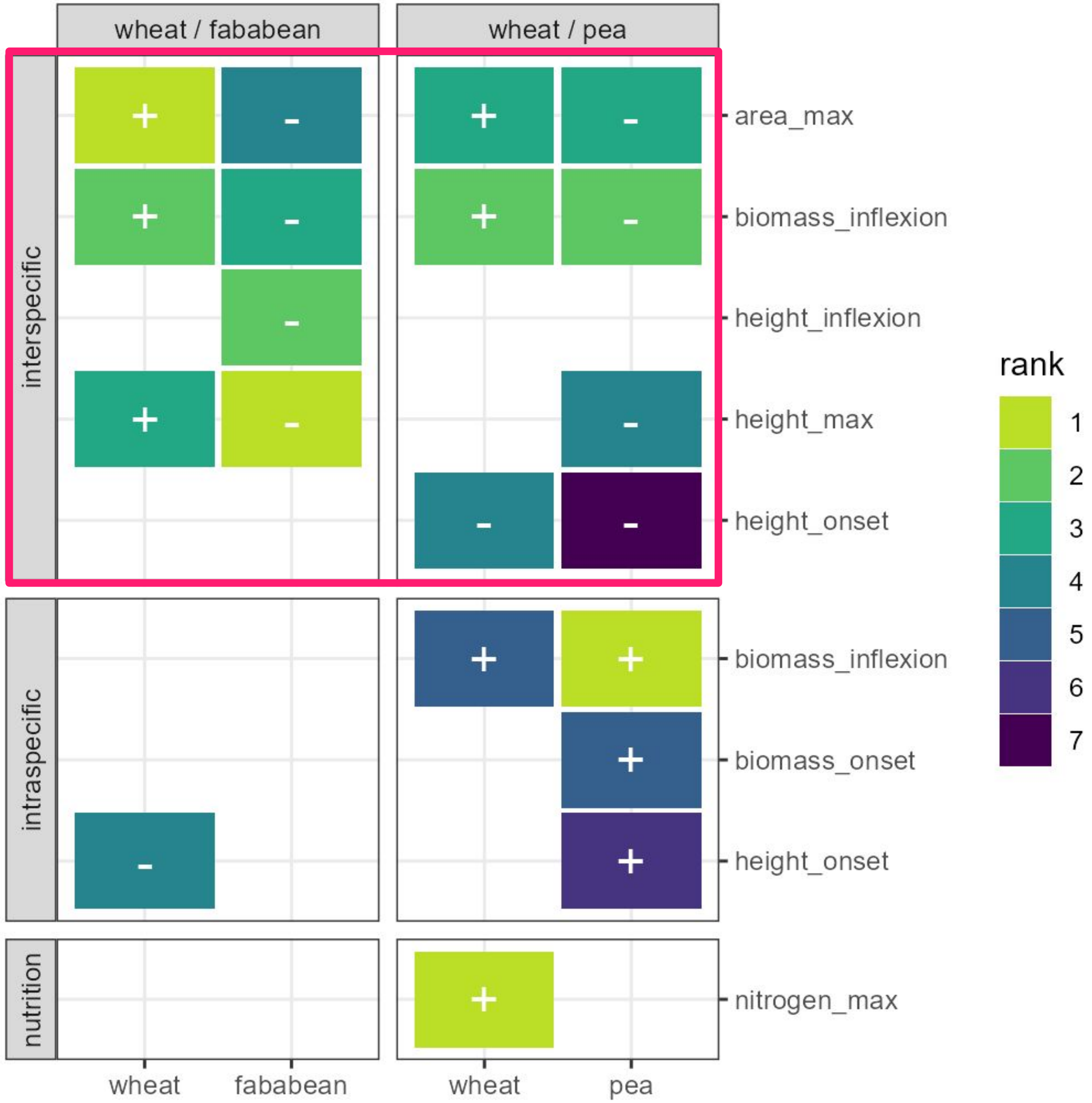
Random effects

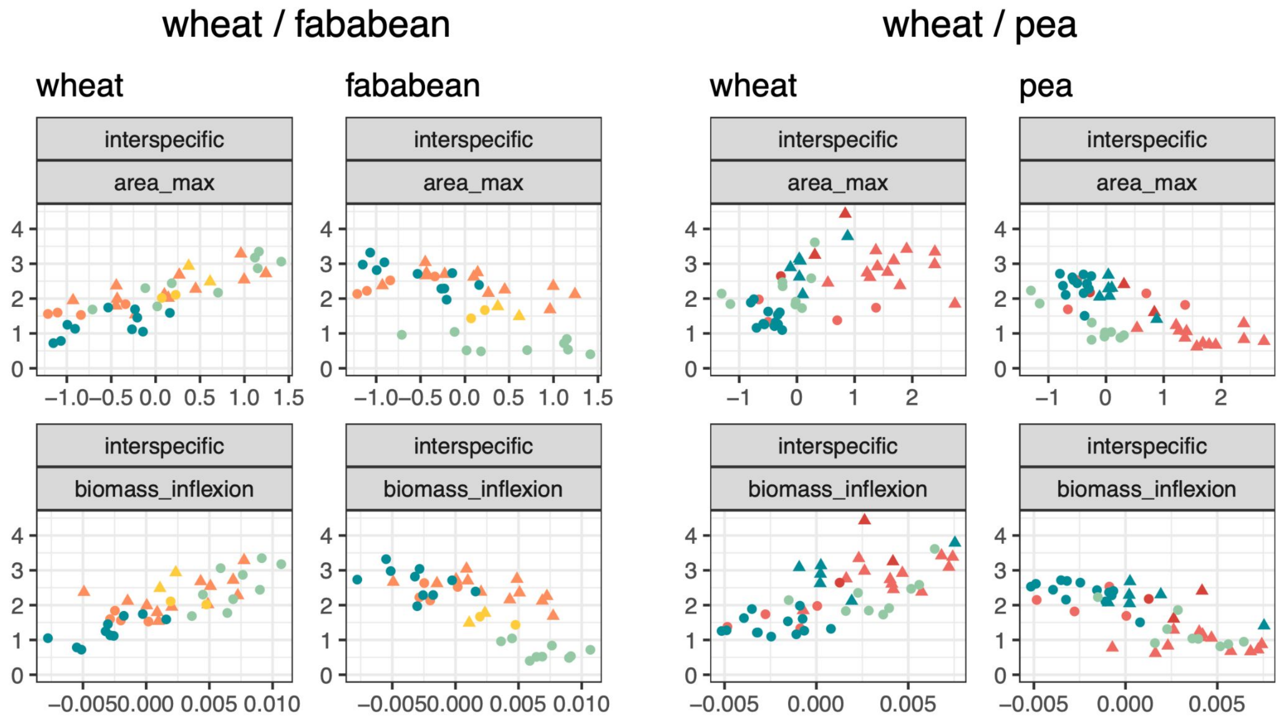
E effects partly accounted in fixed effects.



Values of the coefficients (with their confidence intervals) of the random effects

Differences between species in mixture are key predictors.





Competition

Most predictors had opposite effects on focal species, yet some asymmetry remains.

Genericity

Species and cultivar modality were not selected

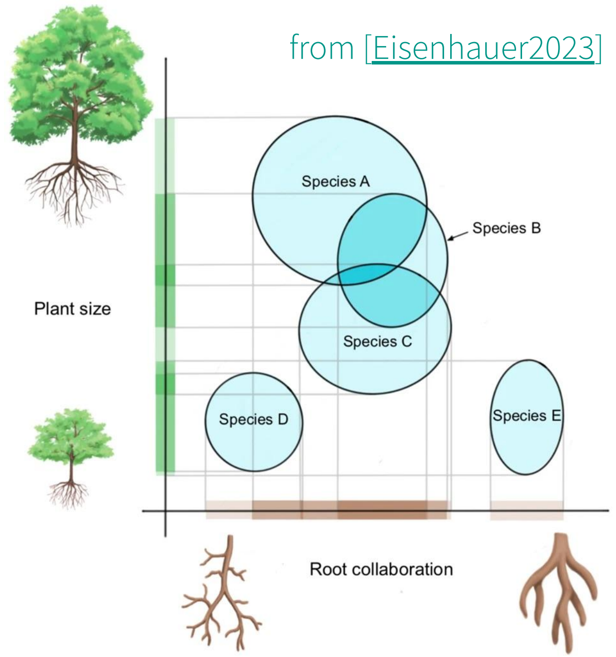
Data science for studying diversified agrosystems

Modelling to ease the tension between generalisation and specialisation

Diversification is easy, mechanisms are not [McGuire2021]. Broad frameworks report **context-dependent** positive effects of diversification on productivity or stability [Dee2023, Lipoma2024].

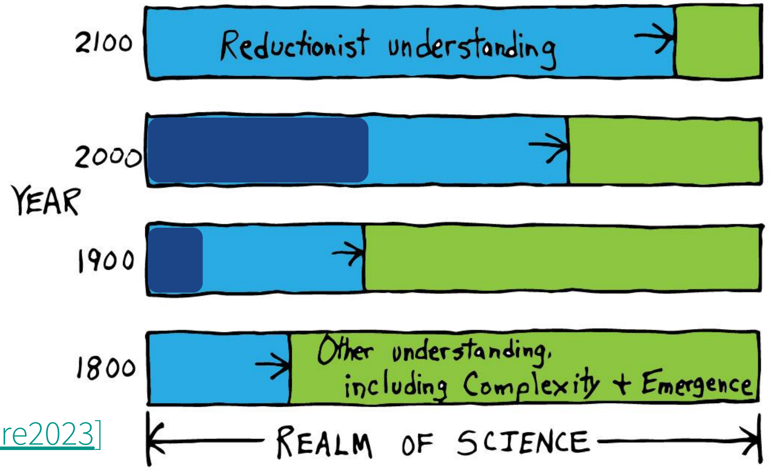
For plants, traits-derived variables used as a signature of processes driving mixture functioning is a solid base for predictive models.
For environment, we might need variables indicating stress patterns rather than practices and geography.

from [Eisenhauer2023]



Work on the integration of scientific cultures [Enquist2024]

Separate (1) what can be explained through **processes**, (2) what can be explained with **data**, (3) what cannot be explained with a reductionist approach (**expertise**)



from [McGuire2023]

Thanks for the shared expertise and trials!

Design and experimentation

Laurent Bedoussac, Eric Justes, Etienne-Pascal Journet, Christophe Naudin, Henrik Hauggaard-Nielsen, Erik Steen Jensen, Elise Pelzer, Guénaëlle Corre-Hellou, Bochra Kammoun, Loic Viguier, Romain Barillot, Antoine Couëdel, Philippe Hinsinger

We're open!

Joint analysis of multiple experiments - a key consideration given the pressing need for consolidating results in the context of an increasingly variable and changing climate.

Thank you for you attention! Any questions ?

INRAE la science pour la vie, l'humain, la terre

#DigitAg