Interpretability tools for characterizing plant generalization in flower-visiting interactions

Application on citizen science data: SPIPOLL monitoring program

supervised by Colin Fontaine (CESCO – MNHN) & Pierre Barbillon (MIA Paris-Saclay - APT)

Jean Cohen • Aussois 2026 • June 16 2026

MUSÉUM N^AL H^IST NATURELLE

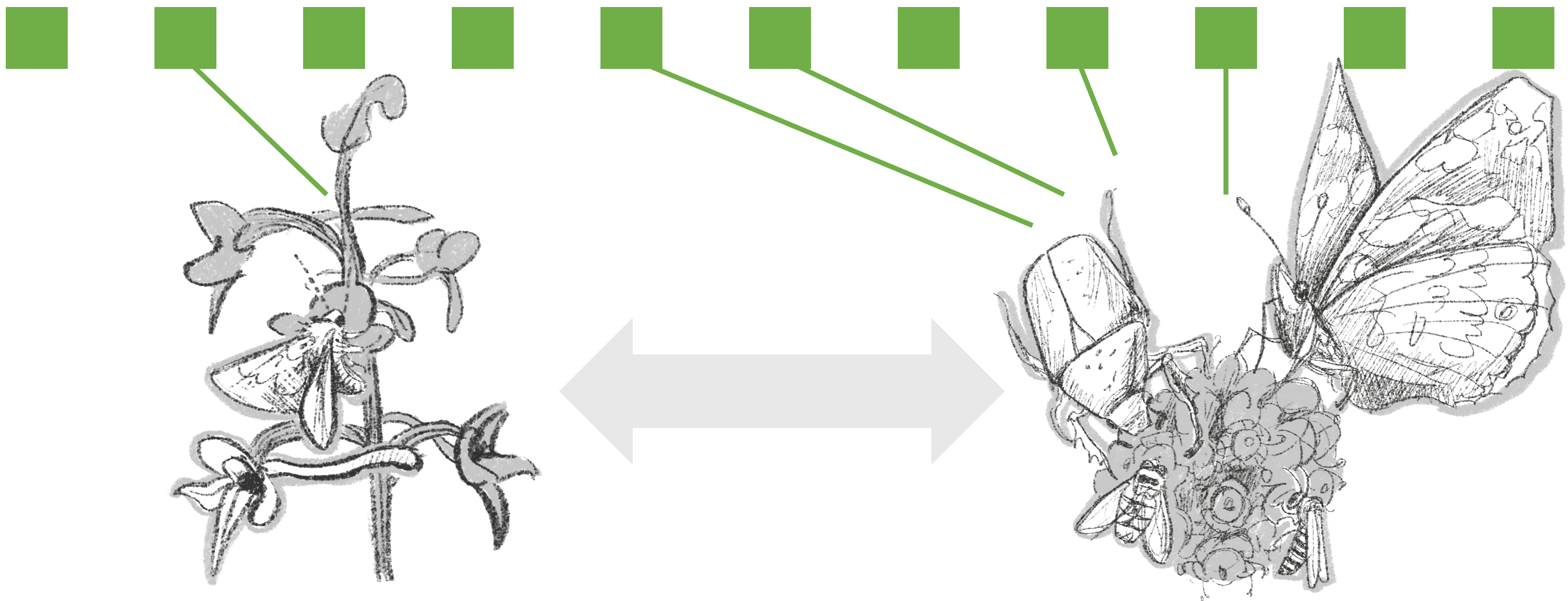
MINISTÈRES TRANSITION ÉCOLOGIQUE AMÉNAGEMENT DU TERRITOIRE TRANSPORTS VILLE ET LOGEMENT
Liberté
Égalité
Fraternité

CESCO Centre d'Écologie et des Sciences de la Conservation

MIA PARIS-SACLAY

Generalization, specialization

A spectra of strategies in plant-pollinator networks



specialist
ex. *Platanthera* (orchids)

generalist
ex. *Hedera* (ivy)

limited number of interaction partners
(low species interaction richness)

various interaction partners
(high species interaction richness)

Monitoring plant-insect interactions

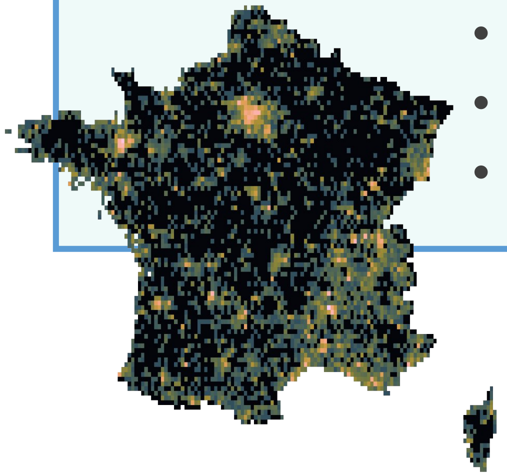
The SPIPOLL citizen science program

making a Spipoll collection

- 20 minutes of photographic observation
- enter and identify each morphospecies
- online community validation

measure the
interaction richness

the extent of the citizen science monitoring program



- since 2010: more than 15 years of data
- 90 000 entries from 5000 participants
- covering of the french hexagonal territory

log-density of Spipoll collection (2010-2023)

SPIPOLL

Variations of plant generalization

Intrinsic and environmental drivers

landscape composition

season

user
characterization

climate

weather anomalies

plant genus

Statistical modeling of the interaction richness

Groups of variables

landscape composition — 47

land use categories (CORINE Land Cover):

- At location (15 variables)
- In a 250 m radius buffer (13 variables)
- In a 1000 m radius buffer (19 variables)

season — 1*

day of the year (cyclic non-linear effect)

user characterization — 4*

- experience (non-linear)
- program persistence
- use of smartphone app
- identity (random effect)

climate — 7

bioclimatic variables 1980-2009 (E-OBS)

sampling conditions — 4*

- shadow
- wind
- nebulosity
- hour (cyclic non-linear)

weather anomalies — 19

- the day of the sampling
- at key moments the year before (development perturbations)

Poisson regression for each plant genus (GAM)

$$R_{int}|X \sim \mathscr{P}\left(\mu_g(X)\right)$$

Ecological questions and challenges

Characterization of plant generalization

- 1 – What is the relative importance of the environmental drivers in explaining the variations of plant generalization throughout France?
- 2 – Among France common flora, which plant genera are intrinsically the most generalist?
- 3 – How does the environmental drivers affect plant generalization?

Challenges:

- high number of variables of different kind (linear, non-linear, random effects) → 82 * 20 plant genera
- shared information between groups of variables (sampling and natural correlations)

analysis with tools used for machine learning model interpretability (xAI)

1. Relative importance of environmental factors

Variation partitioning with Shapley values

Key idea: partition the explained variation (deviance) between groups of variables

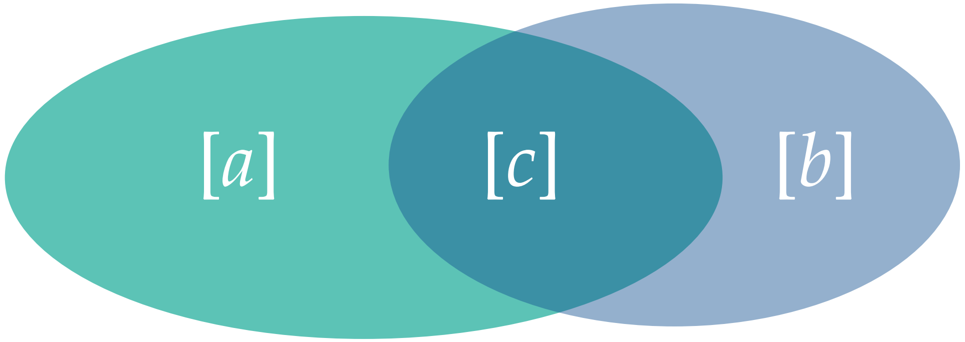
Limitations: correlations between groups

$$pseudo\ R^2 = \sum_{i}^{groups} S_i$$

xAI: seen in Shapley prediction attribution, SHAP (Lundberg & Lee 2017)

Shapley values (1951) from cooperative game theory:

- **Players:** groups of variables
  [landscape, climate, weather, sampling, season, user]
- **Game score:** deviance explained by the model trained on the groups of variables
- **Objective:** fair attribution of each player contribution to the total score



$$S_1 = [a] + \frac{[c]}{2} \qquad S_2 = [b] + \frac{[c]}{2}$$

6 groups = 64 training combinations

1. Relative importance of environmental factors

Finer-grained analysis with Owen values

Key idea: partition the explained variation (deviance) between groups of variables and detail the partition in sub groups

Limitations: computational cost of retraining with too many groups

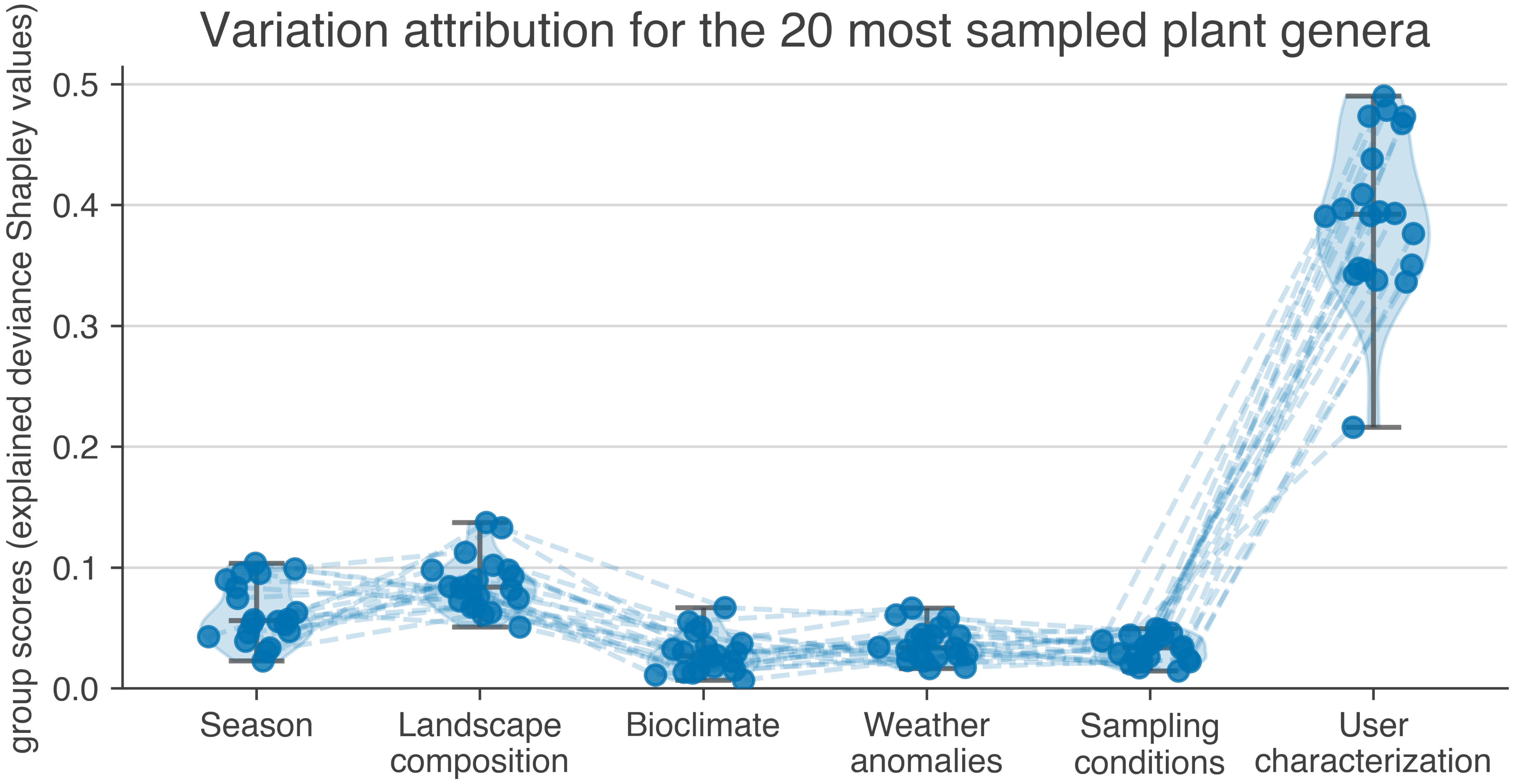
Owen values (1977) eq. of Shapley values for:

- **Players:** sub-groups of variables
  [sampling weather anomalies, past weather anomalies]
- **Game score:** Shapley value of the weather group
- **Objective:** fair attribution of each player contribution to the total score

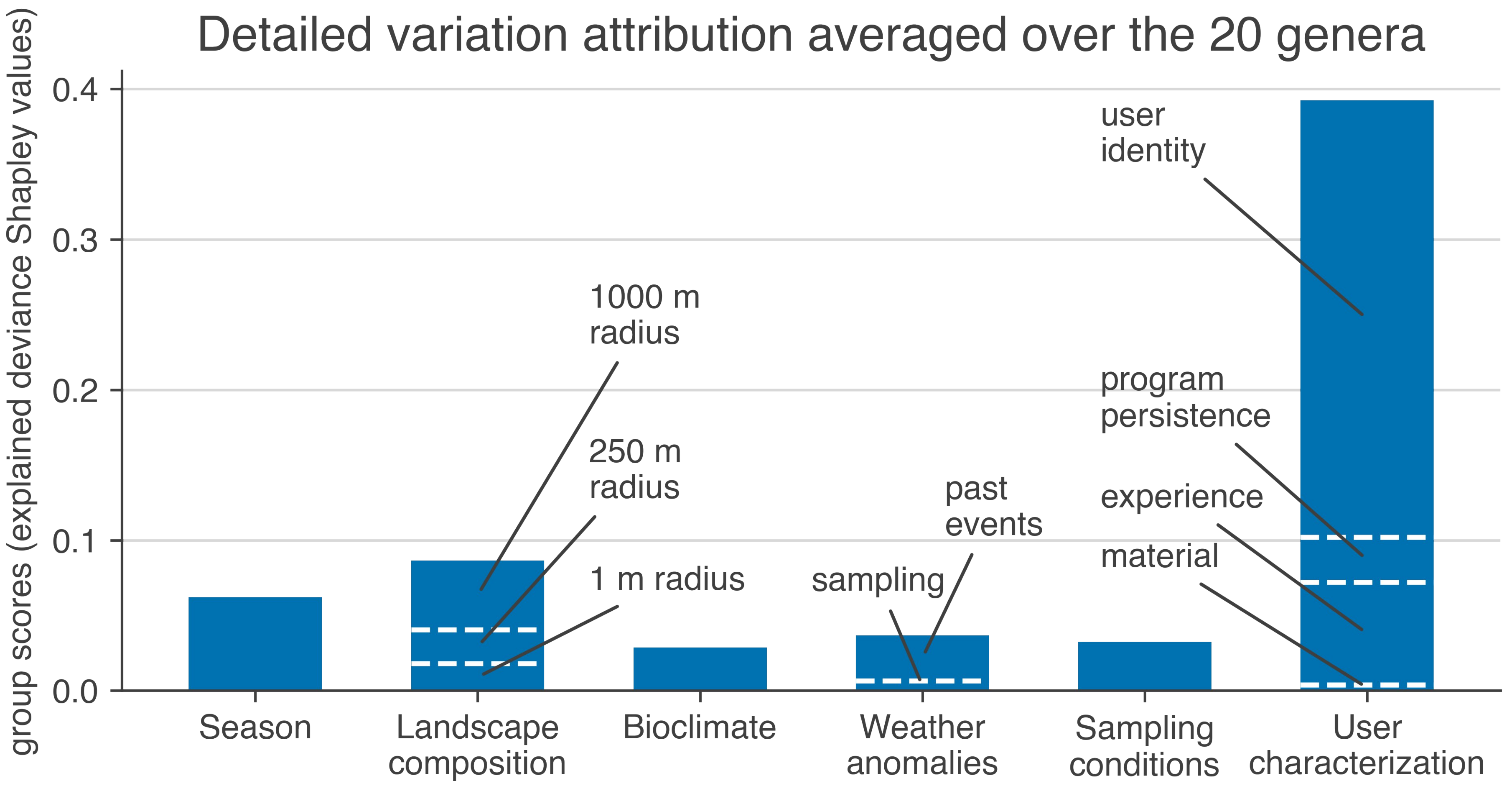
$$pseudo\ R^2 = \sum_{i}^{groups} S_i = \sum_{i}^{groups} \sum_{k}^{sub-g_i} S_{i,k}$$

1. Relative importance of environmental factors

Shapley values results



1. Relative importance of environmental factors
Owen values results



2. Comparison of generalization strategies

Local marginal importance of plant identity

Key idea: compare the expected interaction richness between plant once the effect explained by the environment is taken into account

Limitations: plants have different response to the environment

Local predictions:

- plant-blind predictions: $\mu_{blind}(x)$ (model trained on all data without plant identity)
- plant-informed predictions: $\mu_{informed}(x)$ (model trained by plant genus)

prediction ratio: $\alpha(x) = \frac{\mu_{informed}(x)}{\mu_{blind}(x)}$

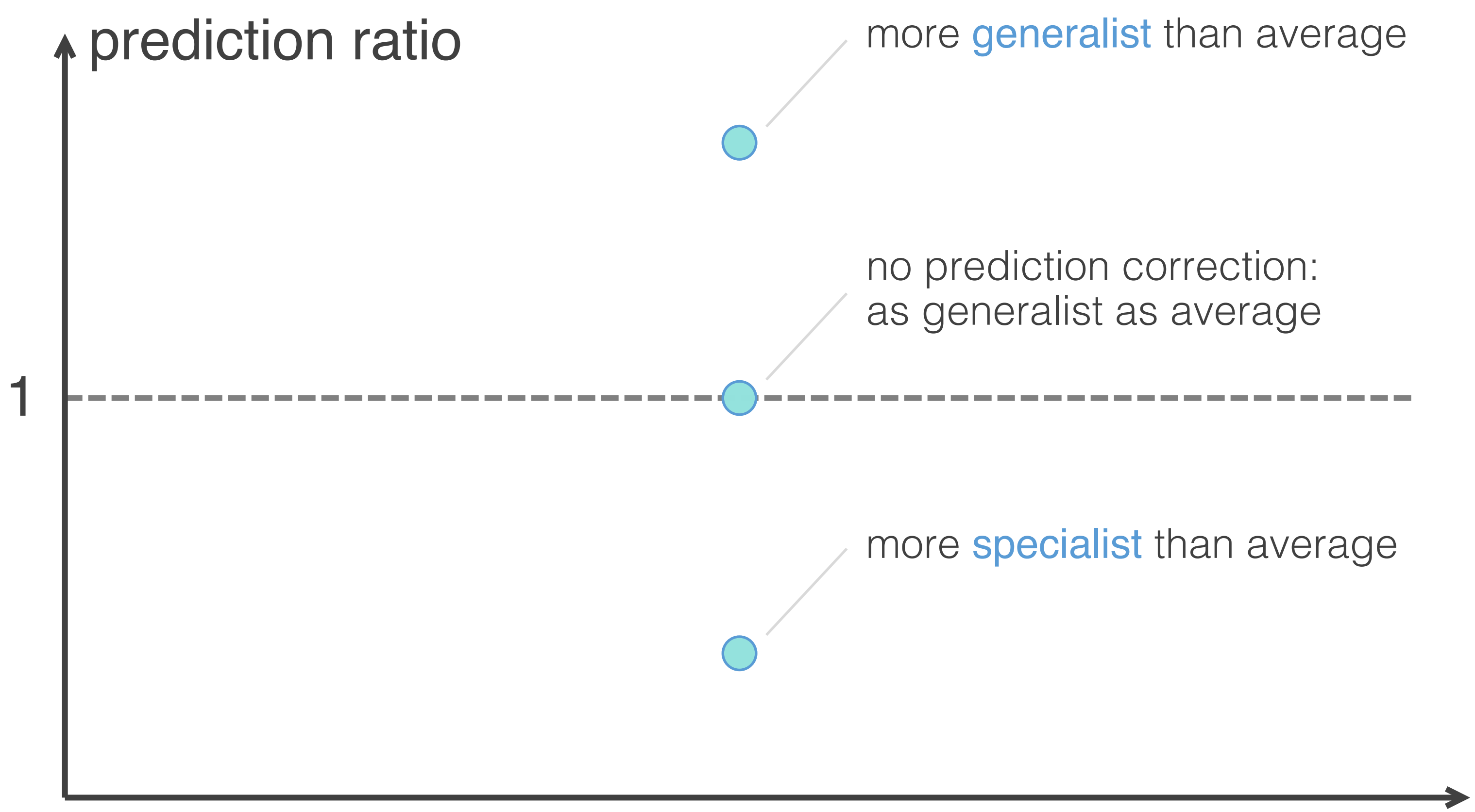
$\rightarrow \alpha(x) > 1$: underestimated blind-prediction
the plant is more **generalist** than expected in those conditions

$\rightarrow \alpha(x) < 1$: overestimate blind prediction
the plant is more **specialist** than expected in those conditions

xAI: seen in Leave-One-Group-Out (LOGO)

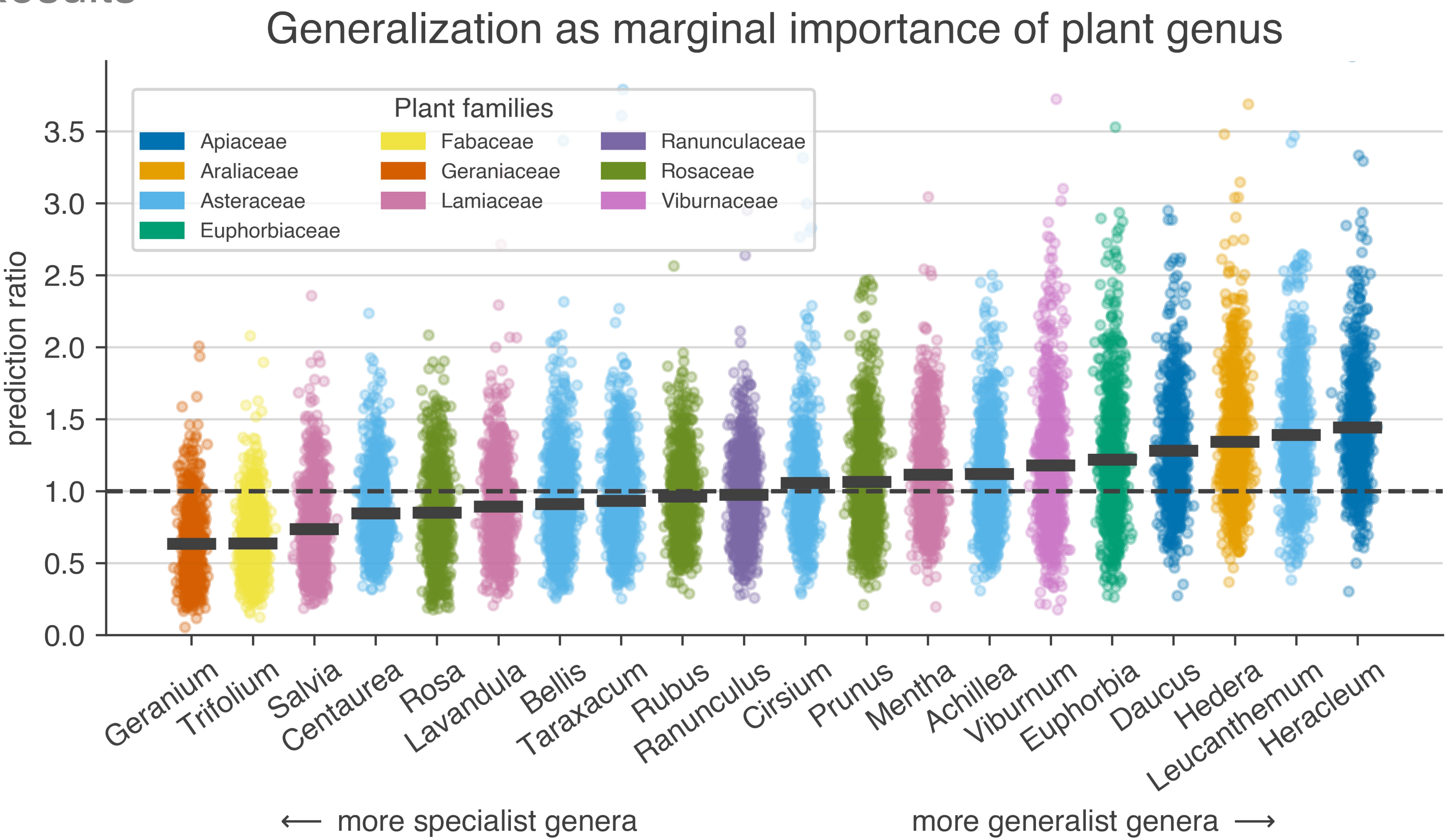
2. Comparison of generalization strategies

Reading results



2. Comparison of generalization strategies

Results



3. Directions of effects

Reading marginal importance ratios with interpretative frameworks

1st step

Local predictions:

- **[group]-blind predictions:** how many insect species will I see in these conditions, without knowing the [group] information?
- **[group]-informed predictions:** how many insect species will I see in these conditions, knowing the [group] information?

prediction ratio: $\alpha_{[g]}(x) = \frac{\mu_{[g]-informed}(x)}{\mu_{[g]-blind}(x)}$

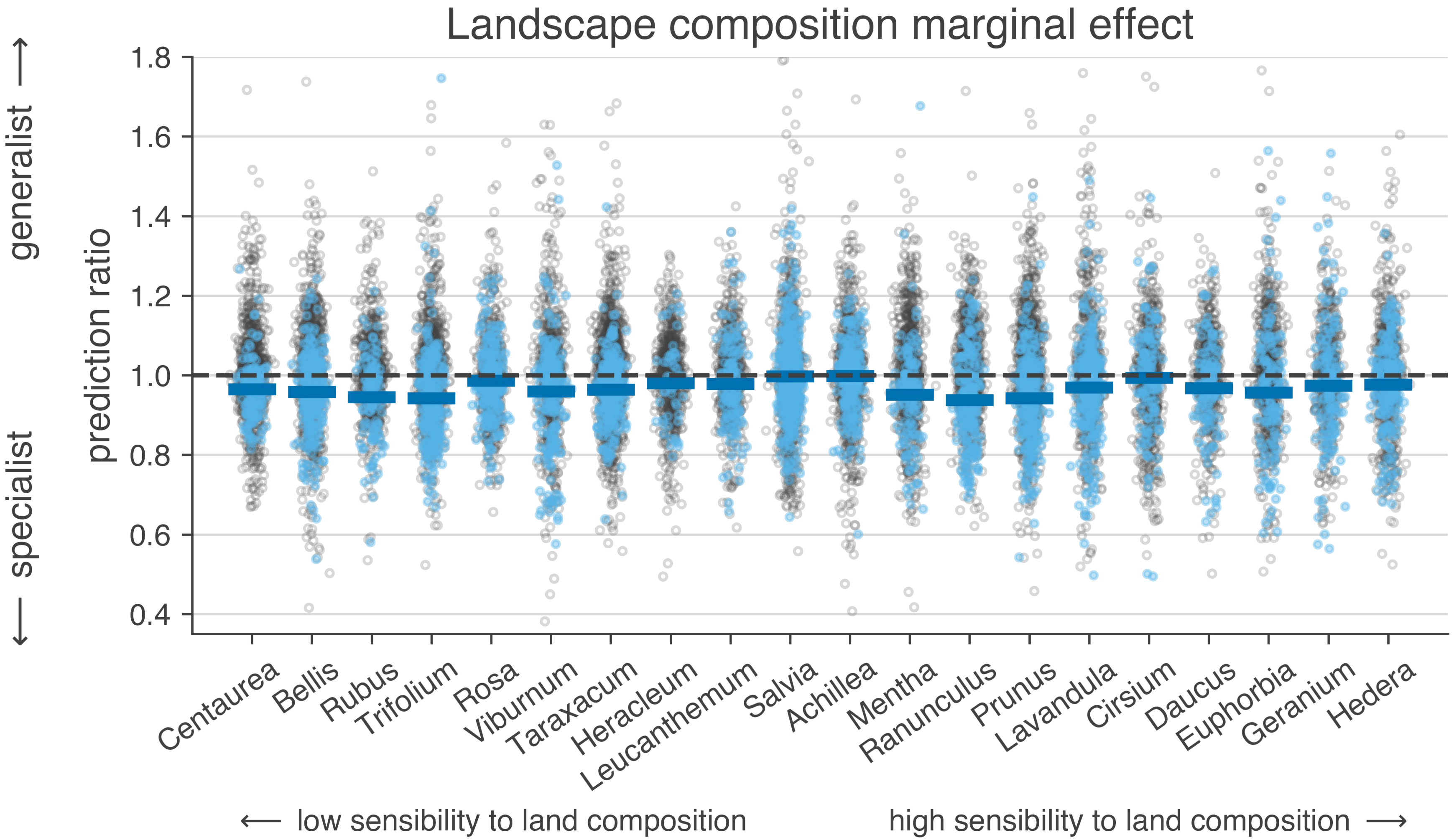
2nd step

- read $\alpha_{[g]}$ distributions through an interpretative framework (one dimensional, categorical, ...)
- ! intra-group correlations

xAI: seen in Leave-One-Group-Out (LOGO), SHAP

3. Directions of effects

Application on urban areas



Conclusions and perspectives

- Season and landscape composition are the most explanatory drivers besides user characterization
- We proposed a ranking of plant genera strategies from the most specialist to the most generalist
- We suggest a framework to explore marginal effects
- xAI: analysis at different level of details, limited by computation time

Next steps:

- Explain generalization scores with a trait-based analysis
- Explore a more detailed reaction to landscape composition
- Finer analysis with an interaction-level model (graph variational auto encoder)

Thank you for
your attention