Université Claude Bernard Lyon 1

THÈSE de DOCTORAT DE
L'UNIVERSITÉ CLAUDE BERNARD LYON 1

École Doctorale N° 341
Évolution Écosystèmes Microbiologie Modélisation

Discipline : écologie statistique

Soutenue publiquement le 8 juillet 2024 par :

Lisa NICVERT

Statistical methods and software tools to analyze and infer ecological networks and process multi-species data

---

Devant le jury composé de :

ETIENNE Marie-Pierre	Maîtresse de conférences Institut Agro Rennes-Angers	Rapportrice
THÉBAULT Elisa	Chargée de recherche CNRS	Rapportrice
CHAMAILLÉ-JAMMES Simon	Directeur de recherches CNRS	Examinateur
DUFOUR Anne-Béatrice	Maîtresse de conférences UCBL	Examinatrice
DRAY Stéphane	Directeur de recherche CNRS	Directeur de thèse
FRITZ Hervé	Directeur de recherche CNRS	Co-directeur de thèse

This thesis is licensed under CC BY 4.0
Cette thèse est distribuée sous licence CC BY 4.0

This thesis was prepared in the Laboratoire de Biométrie et de Biologie Évolutive (UMR CNRS 5558), 43 Boulevard du 11 Novembre 1918, 69622 Villeurbanne CEDEX.

Statistical methods and software tools to analyze and infer ecological networks and process multi-species data

Abstract

Interactions between species in ecological communities are complex: many species can interact with each other in a variety of ways and at different spatial and temporal scales. Moreover, these interaction networks are the result of multiple causes, generate multiple processes and can have indirect effects transmitted through the structure of the network. This complexity calls for a variety of approaches to understand the determinants of interactions and predict their effects in ecological systems.

This thesis studies several aspects of ecological interaction networks using a methodological approach that focuses on the description, evaluation and development of statistical methods and software tools.

In a first part, I study causes of the structure of interaction networks, focusing on interaction niches and using the notion of trait matching between species. To this end, I use methods from the correspondence analysis family and apply and extend reciprocal scaling methods to the analysis of bipartite networks. I apply these methods to the analysis of a plant-frugivore interaction network in a Peruvian montane forest, and show that species traits can be related to their niche width.

In a second part, I study the consequences of interactions through their influence on the spatio-temporal distribution of species. To this end, I use multivariate Hawkes processes to analyze camera trap data. I illustrate these models on five mammals from the South African savanna, showing attraction and avoidance between several of these species at a short spatio-temporal scale.

In a third part, I consider camera trap data analysis. I develop a R package to clean and standardize camera trap data intended for the Snapshot Safari program, as well as a Shiny application intended for a more general use to visualize data in an interactive and reproducible way.

This thesis presents statistical methods and software tools to analyze complex ecological data and improve our understanding of interaction networks. These results open new perspectives on ecological data analysis and methodological development in ecology.

Keywords— ecological networks; interaction niche; spatio-temporal interaction; reproducibility; correspondence analysis; Hawkes process; R package; Shiny application

Méthodes statistiques et outils logiciels pour l'analyse et l'inférence de réseaux écologiques et le traitement de données multi-espèces

Résumé

Les interactions entre espèces dans les communautés écologiques sont complexes : de nombreuses espèces peuvent interagir les unes avec les autres de façons variées et à différentes échelles spatiales et temporelles. De plus, ces réseaux d'interactions sont la résultante de multiples causes, engendrent de multiples processus et ont des conséquences parfois indirectes transmises au travers de la structure du réseau. Cette complexité nécessite une diversité d'approches pour comprendre les déterminants des interactions et prédire leurs effets dans les systèmes écologiques.

Cette thèse étudie plusieurs aspects des réseaux d'interactions écologiques par une approche méthodologique qui se concentre sur la description, l'évaluation et le développement de méthodes statistiques et d'outils logiciels.

Dans une première partie, j'étudie les causes de la structure des réseaux d'interactions en me concentrant sur les niches d'interactions et en utilisant la notion de concordance des traits entre espèces. Pour cela, j'utilise des méthodes de la famille de l'analyse des correspondances et j'applique et j'étends des méthodes de mise à l'échelle réciproque à l'analyse de réseaux bipartites. J'applique ces méthodes à l'analyse d'un réseau d'interactions plantes-frugivores d'une forêt de montagne péruvienne et je montre que les traits des espèces peuvent être reliés à leur largeur de niche.

Dans une deuxième partie, j'étudie les conséquences des interactions au travers de leur influence sur la répartition spatio-temporelle des espèces. Pour cela, j'utilise des processus de Hawkes multivariés pour analyser des données de pièges photographiques. J'illustre ces modèles sur cinq mammifères de la savane sud-africaine et je montre des attractions et évitements entre plusieurs de ces espèces à courte échelle spatio-temporelle.

Dans une troisième partie, je me penche sur l'analyse de données collectées par pièges photographiques. Je développe un package R pour nettoyer et standardiser ces données à l'usage du programme Snapshot Safari, ainsi qu'une application Shiny destinée à un usage plus général pour visualiser de données de façon interactive et reproductible.

Cette thèse présente des méthodes statistiques et outils logiciels pour analyser des données écologiques complexes et améliorer la compréhension des réseaux d'interactions. Ces résultats ouvrent des perspectives nouvelles concernant l'analyse de données écologiques ainsi que les développements méthodologiques en écologie.

Mots-clés— réseaux écologiques ; niche d'interaction ; interaction spatio-temporelle ; reproductibilité ; analyse factorielle des correspondances ; processus de Hawkes ; package R ; application Shiny

Remerciements

Ce manuscrit est rédigé à la première personne, mais j'ai longuement hésité avant de faire ce choix, car de nombreuses personnes ont contribué à en faire ce que vous lisez. Je voudrais donc les remercier ici pour reconnaître leur rôle dans les pages qui vont suivre.

Tout d'abord, je remercie mes directeurs de thèse, Stéphane et Hervé. Merci de m'avoir fait confiance pour réaliser cette thèse, de la préparation du concours de l'école doctorale à la rédaction du manuscrit. Stéphane, je te remercie en particulier de m'avoir fait découvrir à quel point c'est génial l'écologie statistique en m'invitant à participer au GdR Ecostat ; pour les discussions sur ce que ça implique de n'être ni écologue ni statisticien·ne, mais un peu des deux ; et bien sûr pour le partage de tes connaissances en analyses multivariées et plus largement sur tous types de statistiques et d'approches méthodologiques. Hervé, je te remercie en particulier pour m'avoir fait découvrir les écosystèmes de savanes sud-africaines ; pour les discussions sur l'utilité des recherches en écologie face à la crise de la biodiversité et au changement climatique ; et bien sûr pour avoir partagé tes connaissances en écologie concernant les écosystèmes de savane mais aussi de nombreux autres sujets grâce à ton impressionnante culture générale en écologie.

Un grand merci aux membres du jury qui ont accepté d'évaluer ce travail, Marie-Pierre Etienne, Élisa Thébault, Simon Chamaillé-Jammes et Anne-Béatrice Dufour. Savoir que vous alliez évaluer mon travail m'a mis beaucoup de pression mais m'a aussi énormément motivée pour écrire le meilleur manuscrit possible et j'espère que vous prendrez plaisir à le lire.

Je remercie également les nombreuses personnes qui m'ont apporté des conseils scientifiques pendant la durée de cette thèse. Tout d'abord, je remercie les membres de mon comité de suivi, Stéphane Robin, Vincent Miele, Mahendra Mariadassou, Sara Puijalon et Jan Venter pour vos conseils scientifiques et vos encouragements. Je remercie aussi les personnes que j'ai rencontrées en Afrique du Sud et avec qui j'ai pu avoir de passionnants échanges scientifiques, en particulier Alice Bernard, Élie Pédarros, Lain Pardo et Sarah Huebner. Merci également à la team Snapshot pour avoir répondu à toutes mes questions sur les pièges photographiques et sur Snapshot Safari. En particulier, merci à Lain Pardo pour avoir répondu toujours avec bienveillance à mes (nombreuses) questions par email, et même en vrai quand on a eu la chance

de se croiser, et merci à Sarah Huebner pour avoir partagé ses connaissances sur Snapshot Safari et sur les pièges photos. Merci aux membres de l'ANR Econet pour vos conseils scientifiques et pour m'avoir fait découvrir plein de choses sur les réseaux écologiques. Un grand merci aux membres du pôle info qui m'ont aidée à maîtriser le cluster, Docker, Shiny et autres outils informatiques, en particulier à Stéphane Delmotte, Adil El Filali, Aurélie Siberchicot et Bruno Spataro. Enfin, merci à mes collègues de bureau et de pauses midi qui m'ont aussi apporté leurs conseils et leurs avis, en particulier Blandine, Émilie, Florian, Rémi, Alice G., Léa et Emma.

Des remerciements tous particuliers aux personnes qui m'ont accueillie et aidée lors de ma visite en Afrique du Sud, tout spécialement à Alice B. et Alexia L. F. pour m'avoir accueillie chez vous, à toutes les personnes du Wildlife Ecology lab qui m'ont accueillie dans leurs locaux et aux nombreuses personnes qui m'ont fait visiter les environs et qui m'ont incluse dans les moments de partage alors que je ne connaissais personne là-bas, en particulier Alice B., Jérémy, Alexia L. F., Elizabeth et Markus.

Si j'ai autant apprécié réaliser cette thèse, c'est en grande partie grâce à mes collègues de bureau et de pauses de midi avec qui j'ai pu causer science, mener des projets scientifiques sur le café et bien sûr avoir de longs débats sur tout et n'importe quoi aux pauses de midi : Blandine, Rémi, Léa, Emma, Florian, Émilie, Mary, Marine, Alice G., Alexia N. T., Barbara et Marie M.. Un merci tout particulier aux membres permanents ou passagers du bureau des plantes pour m'avoir supportée (avec mes moucherons) dans le bureau.

Enfin, réaliser une thèse c'est parfois difficile et le soutien de mes proches a été très très précieux. Je remercie les personnes avec qui j'ai partagé la coloc de Gratte-ciel pour tous les bons moments passés ensemble, Alice G., Tiphaine, Ludivine et Cyprien. Merci à la bande des lapins, Lucie, Val, Cyprien et Pierre pour les (délicieux) repas partagés ensemble et les (passionnantes) discussions qu'on a partagées. Merci à mes ami·es de l'INSA, Océane, Tiphaine, Ludivine, Guillaume et Morgane pour m'avoir patiemment écoutée raconter alternativement à quel point ma thèse est fatigante mais aussi à quel point les interactions interspécifiques et les processus de Hawkes sont formidables. Merci à Zélie d'être toujours là pour moi depuis tout ce temps. Merci à ma famille d'avoir cru en moi et de m'avoir soutenue, votre soutien a été très précieux. Merci enfin aux nouveaux ami·es que je me suis fait au LBBE, Florian, Blandine, Rémi, Émilie, Léa, Emma, Alexia N. T. et Mary. Avec vous, j'ai fait les meilleures soirées crêpes, et promis, dès que j'ai fini de rédiger ce manuscrit, je vous invite chez moi pour la soirée mojito tant attendue.

Acknowledgments

This manuscript is written in the first person, but I hesitated for a long time before making this choice, as many people have contributed to make it the text you are reading. I would like to thank them here to acknowledge of their role in the pages that follow.

First of all, I would like to thank my thesis supervisors, Stéphane and Hervé. Thank you for trusting me to carry out this thesis, from the preparation of the doctoral school exam to the writing of the manuscript. Stéphane, I would like to thank you in particular for helping me discover how great statistical ecology is, by inviting me to take part in the GdR Ecostat; for the discussions on what it means to be neither an ecologist nor a statistician, but a bit of both; and of course for sharing your knowledge on multivariate analysis and, more broadly, of all types of statistics and methodological approaches. Hervé, I would like to thank you in particular for introducing me to South African savanna ecosystems; for the discussions on the usefulness of ecological research in the face of the biodiversity crisis and climate change; and of course for sharing your knowledge of savanna ecology, but also in many other subjects thanks to your impressive general knowledge of ecology.

Many thanks to the members of the jury who agreed to evaluate this work, Marie-Pierre Etienne, Élisa Thébault, Simon Chamaillé-Jammes and Anne-Béatrice Dufour. Knowing that you were going to assess my work put me under a lot of pressure, but also tremendously motivated me to write the best manuscript I could, and I hope you will enjoy reading it.

I would also like to thank the many people who provided me with scientific advice during the course of this thesis. First of all, I would like to thank the members of my thesis committee, Stéphane Robin, Vincent Miele, Mahendra Mariadassou, Sara Puijalon and Jan Venter, for your scientific advice and encouragements. I would also like to thank the people I met in South Africa and with whom I had captivating scientific exchanges, in particular Alice Bernard, Élie Pédarros, Lain Pardo and Sarah Huebner. Many thanks to the Snapshot team for answering all my questions about camera traps and Snapshot Safari. In particular, thanks to Lain Pardo for always kindly answering my (many) questions by email, and even in person when we were lucky enough to meet in-person, and thanks to Sarah Huebner for sharing her knowledge of Snapshot Safari and photo traps. Thank you to the members of ANR Econet for your scientific advice and for making me discover so many things about ecological networks. Many thanks

to the members of the pôle info who helped me master the cluster, Docker, Shiny and other tools, in particular Stéphane Delmotte, Adil El Filali, Aurélie Siberchicot and Bruno Spataro. Finally, thanks to my office and lunchtime colleagues who also gave me their advice and opinions, especially Blandine, Émilie, Florian, Rémi, Alice G., Léa and Emma.

Special thanks to the people who welcomed and helped me during my visit to South Africa, especially Alice B. and Alexia L. F. for welcoming me into your home, to all the people in the Wildlife Ecology lab who welcomed me into the lab premises, and to the many people who showed me around and included me in moments of sharing when I did not know anyone there, especially Alice B., Jérémy, Alexia L. F., Elizabeth and Markus.

If I enjoyed doing this thesis so much, it is largely thanks to my office and lunchtime colleagues, with whom I was able to talk science, carry out scientific projects on coffee and, of course, have long debates on anything and everything at lunchtimes: Blandine, Rémi, Léa, Emma, Florian, Émilie, Mary, Marine, Alice G., Alexia N. T., Barbara and Marie M.. Special thanks to the permanent and temporary members of the Bureau des Plantes for putting up with me (and my midges) in the office.

Finally, writing a thesis is sometimes difficult, and the support of people close to me has been invaluable. I would like to thank the people with whom I shared the coloc de Gratte-Ciel for all the good times spent together, Alice G., Tiphaine, Ludivine and Cyprien. Thanks to the lapins, Lucie, Val, Cyprien and Pierre for the (delicious) meals we shared and the (fascinating) discussions we had. Thanks to my INSA friends, Océane, Tiphaine, Ludivine, Guillaume and Morgane for patiently listening to me alternately recount how tiring my PhD is, but also how great interspecific interactions and Hawkes processes are. Thanks to Zélie for always being there for me all this time. Thank you to my family for believing in me and supporting me, your support has been invaluable. Finally, thank you to the new friends I have made at LBBE, Florian, Blandine, Rémi, Émilie, Léa, Emma, Alexia N. T. and Mary. I had the best crêpe parties with you all, and I promise, as soon as I have finished writing this manuscript, I will invite you to my place for the long-awaited mojito party.

Contents

Abstract/Résumé 3

Remerciements/Acknowledgments 6

List of Figures 16

List of Tables 19

List of online resources 20

1 General introduction 21

Introduction 23

1 Ecology . . . 23
2 Ecological communities . . . 26
3 Interspecific interactions . . . 31
3.1 Definitions . . . 31
3.2 Types of pairwise interactions . . . 33
3.3 Indirect effects . . . 37
4 Ecological interaction networks . . . 40
4.1 Network science . . . 40
4.2 Generalities on interaction networks . . . 44
4.3 The structure of interaction networks . . . 47
5 Questions related to ecological networks . . . 50
5.1 Processes and models explaining network patterns . . . 50
5.2 Consequences of network structure on the community . . . 52
6 Sampling and analyzing ecological networks . . . 54
6.1 Sampling interactions . . . 54
6.2 Analyzing and inferring interactions . . . 56
7 Thesis objectives . . . 58

Bibliography 60

2 Investigate trait matching in interaction networks 73

Introduction 75
1 Preamble: PhD context . . . 75
2 Introduction . . . 76
2.1 The ecological niche . . . 77
2.2 Multivariate ordination methods . . . 79
2.3 Plan . . . 80
2.4 Dataset . . . 81

2.1 Trait matching without traits: using CA to analyze interaction networks 83
1 Introduction . . . 85
2 Material and methods . . . 89
2.1 Notations . . . 89
2.2 Quantify trait matching . . . 89
2.3 Correspondence analysis (CA) . . . 90
2.4 Reciprocal scaling . . . 92
2.5 Simulation . . . 93
2.6 Real data analysis . . . 98
3 Results . . . 100
3.1 Simulation . . . 100
3.2 Data analysis . . . 102
4 Discussion . . . 105
4.1 Simulation . . . 106
4.2 Data analysis . . . 107
4.3 Perspectives . . . 111
Appendices . . . 112
A Effect of the relative importance of matching on the recovery of niche parameters . . . 112
B Realized and fundamental niches . . . 114
C Networks visualization using CA . . . 116
D Reciprocal scaling formulas . . . 117
E Species names . . . 119

2.2 CCA and dc-CA to analyze interaction networks, or trait matching with traits 123
1 Introduction . . . 124
2 Material and methods . . . 126
2.1 Canonical correspondence analysis (CCA) . . . 126
2.2 Double constrained correspondence analysis (dc-CA) . . . 128
2.3 Variation partitioning . . . 131

2.4 Data analysis . . . 133
3 Results and discussion . . . 133
3.1 CCA constrained with bird traits . . . 134
3.2 dc-CA constrained with bird and plant traits . . . 135
3.3 Variation partitioning . . . 139
4 Conclusion and perspectives . . . 140
Appendices . . . 142
A Traits distributions . . . 142
B CCA constrained with plant traits . . . 143

2.3 Constrained reciprocal scaling to measure interaction niche breadth 145
1 Introduction . . . 146
2 Theory . . . 148
2.1 Add dispersion to CCA . . . 148
2.2 Add dispersion to dc-CA . . . 154
2.3 Data analysis . . . 157
3 Results and discussion . . . 158
3.1 CCA constrained with bird traits . . . 158
3.2 dc-CA constrained with bird and plant traits . . . 159
4 Conclusion and perspectives . . . 161
Appendices . . . 163
A Demonstration of the equivalence between CCA and CCorA niche measures . . . 163
B CCA constrained with plant traits . . . 164

Discussion 165
1 Quantify the niche with CA methods . . . 165
2 Comparison of the analyses . . . 167
3 Characteristics and limits of CA approaches . . . 169
4 Define the interaction niche . . . 169
5 What structures networks and how much? . . . 171
6 Conclusion and perspectives . . . 172

Bibliography 173

3 Infer attractions and repulsions with Hawkes processes 181

Introduction 183
1 Preamble: PhD context . . . 183
2 Introduction . . . 184
2.1 Dataset . . . 186

3.1 Linear multivariate Hawkes process **189**
1 Introduction . . . 191
2 Material and methods . . . 192
2.1 Model: The multivariate Hawkes process (MHP) . . . 192
2.2 Model inference . . . 193
2.3 Simulation study . . . 195
2.4 Application: Analysis of interactions between five species in the African savanna . . . 196
3 Results . . . 197
3.1 Simulation study . . . 197
3.2 Analysis of real data . . . 197
4 Discussion . . . 198
5 References . . . 202
Appendices . . . 205
S1 Bias in the inference methods . . . 205
S2 Background rates inferred for real data . . . 206
S3 Influence of circadian rhythms . . . 207
S4 Influence of bin width . . . 208

3.2 Non-linear multivariate Hawkes process with a temporal covariate **211**
1 Introduction and motivation . . . 212
2 Material and methods . . . 213
2.1 Model: the non-linear multivariate Hawkes process . . . 213
2.2 Inference . . . 215
2.3 Simulation . . . 216
2.4 Data analysis . . . 217
3 Results . . . 218
3.1 Simulation . . . 218
3.2 Data analysis . . . 219
4 Discussion . . . 222
4.1 Simulations . . . 222
4.2 Real data analysis . . . 223
4.3 Conclusions and perspectives . . . 224

Discussion **227**
1 Compare models outputs . . . 227
2 From the data to the model . . . 229
3 Measuring interaction strength . . . 230
4 Including more covariates? . . . 231

Bibliography **233**

4 Software development 237

Introduction **239**

1 Preamble: PhD context 239
2 Introduction 240

4.1 standardizeSnapshot **243**

1 Introduction 244
2 The data standard 245
3 Main functionalities 247
4 Software development 248
5 Conclusion and perspectives 250

4.2 camtrapviz **251**

1 Introduction 252
2 Preliminary work 253
3 Camera trap data terminology 253
4 Shiny application 254
4.1 Modules 255
4.2 Other features 257
5 Software development 258
6 Conclusion and perspectives 259
Appendices 261
A Import module 261
B Filter module 263
C Overview module 264
D All species module 266
E One species module 267

Bibliography **269**

5 General discussion 273

Discussion **275**

1 Synthesis of the main results 275
1.1 Investigate trait matching in interaction networks 275
1.2 Infer attractions and repulsions 277
1.3 Software development 278
2 Open science and reproducibility 279
2.1 Replicability and reproducibility 279
2.2 Coding practices to improve reproducibility 281

2.3 Open science, code and data 282
2.4 In this thesis 283
2.5 Additional remarks 283
3 Model validation: with simulations or real data? 284
3.1 Validation with a simulation approach 284
3.2 Validation using real data 285
3.3 Two complementary approaches and other perspectives 286
4 Modeling ecological data 287
5 Generalism and specialization in science 290

Bibliography **293**

List of Figures

1.1 Ecological scales . . . 25
1.2 Kelp forest community . . . 28
1.3 Classification of interactions following their effect on the interacting species 34
1.4 Schematic representation of indirect interactions . . . 37
1.5 Indirect effects of ants *Formica yessensis* on oak trees *Quercus dentata* described by Ito and Higashi (1991) . . . 39
1.6 Examples of networks generated with different models . . . 42
1.7 Examples of networks representing different effects . . . 45
1.8 Matrix representations of perfectly modular and nested networks . . . 47
1.9 Different causes affecting the structure of ecological networks . . . 51

2.I.1 Illustration of trait matching . . . 77
2.I.2 Original illustration of the Hutchinsonian niche concept . . . 78
2.I.3 Interaction matrix for the bird-fruit interaction network . . . 81
2.I.4 Overview of some bird and plant species sampled in the network . . . 82

2.1.1 Visualization of the interaction niche in the multivariate plane . . . 93
2.1.2 Model used to simulate interactions between consumer and resource species . . . 94
2.1.3 Results of the simulation study . . . 101
2.1.4 Reciprocal scaling of the birds-plants interaction network . . . 102
2.1.5 Niche breadth versus niche optimum on the 2 multivariate axes . . . 104
2.1.6 Correlation circle measured a posteriori between species traits and multivariate axes . . . 105
A.1 Effect of the relative importance of matching on the inference of niche parameters . . . 112
A.2 Eigenvalues evolution with sampling intensity for different values of δ . . . 113
B.1 Performance of the model to infer fundamental niche parameters . . . 114
C.1 Interaction matrix reordered using correspondence analysis . . . 116

2.2.1 Variance partitioning of the interaction table . . . 131
2.2.2 Correlation circle and plant niches optima from CCA constrained with bird traits . . . 134
2.2.3 Correlation circle and niches optima from dc-CA . . . 137

A.1 Bird and plant traits histograms . . . 142
B.1 Correlation circle and bird niches optima from CCA constrained with plant traits . . . 143

2.3.1 Correspondence between original and inflated matrices . . . 149
2.3.2 Graphical effect of the proposed scaling. . . . 154
2.3.3 Correlation circle and plant niches from CCA constrained with bird traits . 159
2.3.4 Correlation circle and niches from dc-CA . . . 160
B.1 Correlation circle and bird niches from CCA constrained with plant traits . 164

2.D.1 Summary of CA methods and niche inference . . . 166
2.D.2 Comparison of the niches inferred with the different methods . . . 167

3.I.1 Capture count for each species . . . 186
3.I.2 Camera trap pictures of focal species . . . 187

1 Example of a realization of a multivariate Hawkes process . . . 194
2 Study sites . . . 196
3 Evaluation of an interevent times method . . . 197
4 Performance of the inference with the multivariate Hawkes model . . . 198
5 Inference of interactions from real data using the multivariate Hawkes model . . . 199
S1 Bias in the inference . . . 205
S2 Background rates inferred with the Hawkes process for each species . . . 206
S3 Inference of spurious interactions when circadian rhythms are present . . 207
S4 Inference of a multivariate Hawkes process with different bin widths . . . 208

3.2.1 Example of a realization of a non-linear multivariate Hawkes process . . . 216
3.2.2 Performance of the inference with the non-linear multivariate Hawkes process . . . 219
3.2.3 Model inferred without covariates . . . 220
3.2.4 Model inferred with a time of day covariate . . . 221

3.D.1 Comparison of the linear and non-linear MHPs . . . 228

4.1.1 Location of the Snapshot Safari camera trap grids . . . 244
4.1.2 Workflow of `standardizeSnapshot` . . . 248

4.2.1 Overview of the `camtrapviz` application . . . 254
4.2.2 Analysis modules in the Shiny application . . . 255
4.2.3 R code display in the Shiny application . . . 258
A.1 Import module with the example datasets . . . 261
A.2 Import module with a user-chosen dataset . . . 262
B.1 Filter module . . . 263

C.1 Overview module summary statistics 264
C.2 Overview module map 264
C.3 Overview module records plot 265
C.4 Overview module summary tables 265
D.1 All species module 266
E.1 One species module overview 267
E.2 One species module analyses 268

5.1 Link between coding practices, reproducibility and replicability 280
5.2 Axiomatic and pragmatic view of models 288
5.3 Real and expected research trajectories 292

List of Tables

1.1 Summary of some direct and indirect methods to sample interaction networks 56
1.2 Thesis objectives for each part 59

2.1.1 Parameter values used for the simulation study 97
E.1 Correspondence between plant codes and species names 120
E.2 Correspondence between bird codes and species names 121

2.2.1 Variance partitioning of the interaction matrix 139

2.3.1 Summary table for the correspondences, mean, variance and covariance of interactions from CCA 153
2.3.2 Summary table for the correspondences, mean, variance and covariance of interactions from dc-CA 157

4.1.1 Standard column names 246

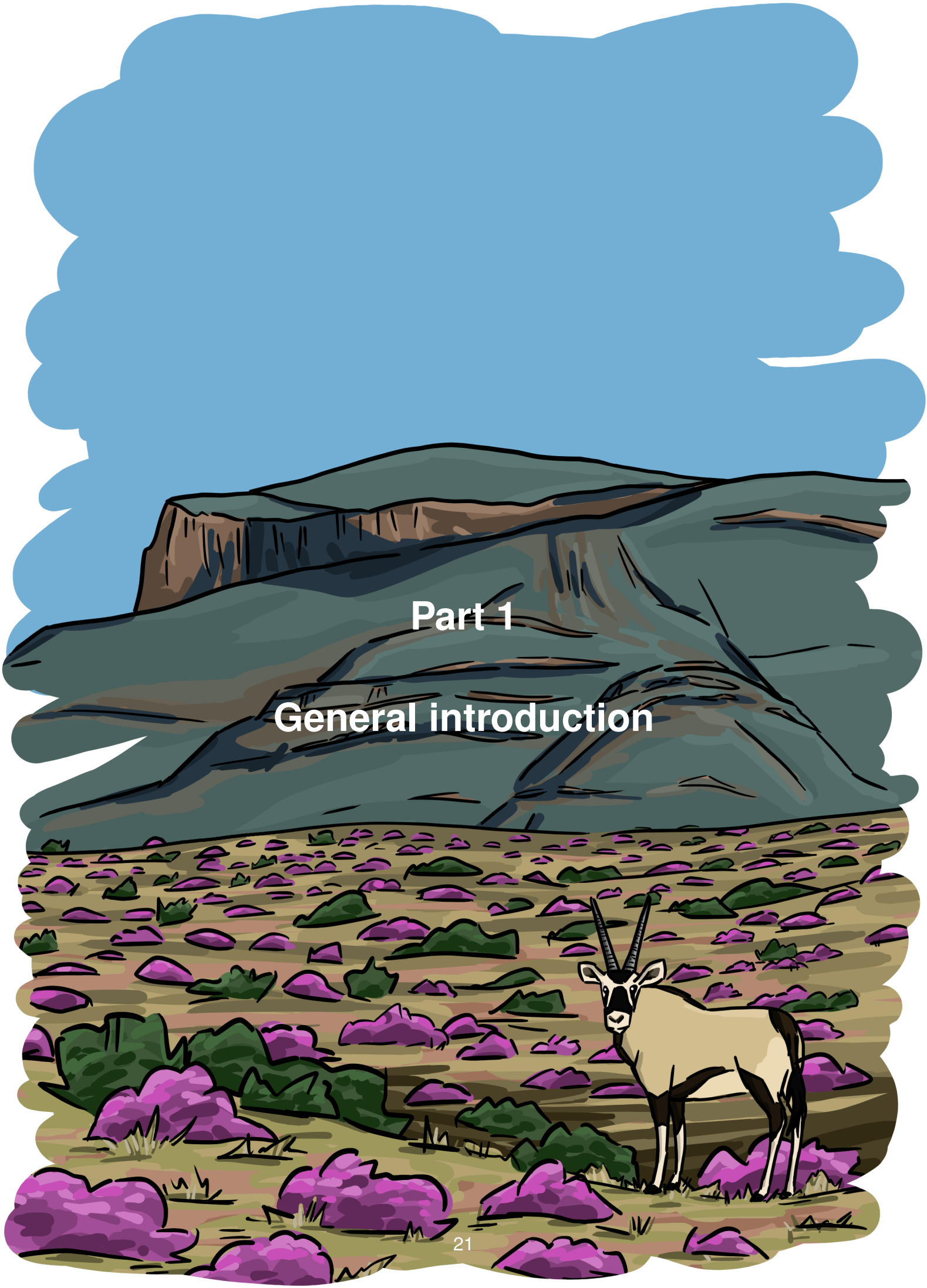
5.1 Synthesis of the main results of the thesis 276

List of online resources

All online resources are free-of-access, except for Part 2.

Part	Chapter	Material	Link
2	All	Code	https://github.com/LisaNicvert/PhDaxis01_CAnetwork/tree/thesis (private repository)[1]
3	All	Code	https://github.com/LisaNicvert/PhDaxis02_camtrapHawkes
	3.1	Code	https://doi.org/10.6084/m9.figshare.24552157.v5
4	4.1	Code	https://github.com/SnapshotSafari/standard-merge
		Documentation	https://snapshotsafari.github.io/standard-merge/
	4.2	Code	https://github.com/LisaNicvert/camtrapviz
		Web application	https://lbbe-shiny.univ-lyon1.fr/camtrapviz/
		Documentation	https://lisanicvert.github.io/camtrapviz/

[1] This repository is private at the time of writing, but can be shared on demand. Moreover, we intend to make it public upon publication of articles related to the work done in chapters 2.1 and 2.2.



Part 1

General introduction

Cover picture: this drawing represents a semi-arid savanna in the Karoo semi-desertic area (South Africa). A gemsbok antelope stands in the foreground and succulent shrubs *Drosanthemum eburneum* are blooming with bright purple flowers. The background features a characteristic Karoo koppie (plateau formed by erosion).

© Own drawing inspired from a picture taken in the Karoo National park.

Introduction

In this general introduction, I outline the theoretical and methodological framework in which this thesis takes place. This introduction takes a broad perspective, and each part of the manuscript is prefaced with a more specific introduction.

I begin by presenting the general framework of ecology, with a focus on the ecological community, which is the scale at which questions are asked in this thesis. Then, I present interspecific interactions, their types and the challenges and questions associated to the study of interspecific interactions. After that, I briefly present the diversity of methods allowing to study interactions in ecological communities and conclude by outlining the questions addressed in this thesis and presenting the structure of the manuscript.

1 Ecology

Ecology is the science studying the distribution and abundance of organisms, and their interactions with the environment (C. R. Townsend et al., 2008; British Ecological Society, 2024). This umbrella definition covers a wide range of questions appealing to many neighboring disciplines like biology, chemistry, ethology or geology. For instance, ecology aims to determine the environmental factors affecting species spatial distribution on the globe (Pan et al., 2013; Pie et al., 2017), predict how species coexistence in the same areas can alter their abundance (Estes et al., 2011), or determine behavioral factors explaining species movements (Nathan et al., 2008).

Ecology was born out of a long tradition of natural history, focused on describing the diversity of the natural world. Aristotle (4th century BC) is often regarded as one of the first scientists of natural history (Matagne, 2002). More recently, in the 1700s-1800s, ecology was conducted by naturalists, and naturalist clubs and societies (especially

botanists) were quite popular (Matagne, 2002). The aim of these naturalists back then was not to understand the functioning of ecosystems, but to describe and inventory species (their characteristics and their distributions). It was not until the early 1800s that botanists made the first attempts to go beyond description of the natural world. One of the first works to study general patterns of species distribution was made by Alexander von Humboldt (Matagne, 2002). In his *Essay on the Geography of Plants* (Humboldt et al., 2008, first published in 1807), he described the distribution of vegetation in the Andes, notably recognizing the importance of altitude gradients in explaining the observed distribution patterns. This publication is one of the first to go beyond the simple description of plant species. The following extract captures the innovative aspect of this book (Humboldt et al., 2008):

> "Botanists usually direct their research towards objects that encompass only a very small part of their science. They are concerned almost exclusively with the discovery of new species of plants, the study of their external structure, their distinguishing characteristics [...]. Even if this knowledge is worthy of occupying a great number of botanists, [...] it is no less important to understand the Geography of Plants, [...] the science that concerns itself with plants in their local association in the various climates."

The term “ecology” was coined in 1866 by the zoologist Ernst Haeckel, from the Ancient Greek *oikos* (house), and *-logia* (study of). Although this term is now widely adopted, it initially coexisted with other appellations before gradually becoming more popular in the early 1900s. An important additional development to ecology as we know it today emerged after the 1920s with the growing mathematization of the discipline: one of the first successes of this mathematical formalization was through the works of the mathematicians Alfred J. Lotka and Vito Volterra, who independently modeled prey-predator population dynamics with differential equations (Lotka, 1920; Volterra, 1926).

Charles Darwin’s work, and especially *On the Origins of species*, published in 1859 (Darwin, 1859), is particularly influential in ecology. The idea of natural selection proposed by Darwin prompted a whole new set of questions, because this mechanism provided the basis for causes leading to observed patterns in species morphology, behavior and distribution. Since all organisms have evolved, and are subject to natural

selection forces, this allows ecologists to answer questions like “why do giraffes have a long neck?”. This makes ecology address final questions, i.e. questions related to the cause of observed patterns (Barbault, 2008; Mossio et al., 2014). This emphasis on final causes contrasts for instance with physics: it would not be relevant to ask “why” gravity exists, but rather “how” it operates, and determining the laws it follows.

Organization scales



Figure 1.1: Ecological scales. Ecological systems are hierarchically organized as individuals, populations, communities and ecosystems. Here, we use the example of a poppy plant individual. All poppy plants individuals make up the population of this species, which grows in a prairie community composed of other plant species, but also interacting with pollinator insects, birds that eat and disperse their seeds and microorganisms living in the soil. This community is included in an urban ecosystem on La Doua campus (Lyon, France). Own pictures, inspired from a figure in Smith and Smith (2015).

Ecological systems are characterized by different levels of organization (see Figure 1.1) (Levin, 1992; C. R. Townsend et al., 2008; Smith & Smith, 2015). The smallest level used in ecology is the individual, and its characteristics (e.g. morphology, personality traits) that are relevant to explain its relation to the environment and to other species. Then, there are populations: an assemblage of individuals of the same species. Different populations coexist within a community, and communities are comprised in ecosystems and biomes. These organization scales are not independent, and focusing on a given scale does not amount to ignoring other levels of organization: if we want to understand processes happening at a given level, it is generally necessary to study the organization level immediately below (Betts et al., 2021). For instance, to study the response of a poppy individual to drought, we need to quantify physiological responses occurring inside the organism, such as stomatal closure, leaf rolling or osmotic adjustment in its cells (Fang & Xiong, 2015). Similarly, to study poppy population dynamics, we need to take into account mechanisms influencing individuals’ death

and reproduction rates. Conversely, a process might be affected by higher organization scales, like competition between poppies individuals of a population that can weaken the individual response to drought (Guo et al., 2020).

In this thesis, I focused on a particular level of organization of ecological systems: the ecological community, which is defined and described below.

2 Ecological communities

Ecological communities are broadly defined as the species that occur together in space (Morin, 2011; Stroud et al., 2015). Examples of ecological communities include coral reefs and resident fishes and mollusks, tropical forests trees and associated mammal, birds and insect species, or lakes and their inhabitant fishes and algae.

This generic definition of ecological communities leaves plenty of space for different interpretations (Morin, 2011; Stroud et al., 2015). First, this definition remains evasive on the operational way to delimit the contours of the "shared space" in practice. To come back to the previous example, if we take a coral reef community, does the shared space stop at the coral edge, or 1, 10, 100 meters from it? Second, defining the species considered in the communities is subject to a similar arbitrary concept. For instance, for a lake community, should the heron fishing on its banks, but nesting in a nearby tree, be included in the community? Finally, another important component of the community is its temporal aspect. We can ask ourselves how long does it take for two species to share the same space before we consider that they are part of the same community, in particular in regard to immigration and emigration dynamics (Stroud et al., 2015) and succession ecology (Clements, 1936): does a community remain the same community even if species change, or are succession dynamics a fundamental alteration of the community? The contours of the communities are not clear-cut, but rather have a blurry frontier.

Several factors have been proposed to delineate communities, and can be classified into four underlying approaches (Morin, 2011). First, communities might be defined based on their dominant species, i.e. species whose abundance is important and thought to structure the community. For instance, this is the case for kelp forests (see Figure 1.2), where kelp (algae of the order *Laminariales*) are an abundant species that

seem to structure the algal community and provide a habitat and food resources for otters, fish, urchins and other invertebrates (Harrold & Reed, 1985). Second, communities can also be characterized using the physical environment, often based on discontinuities in the landscape. For example, bodies of water (lakes, ponds) or anatomical features (e.g. gut or mouth and their associated microbial communities) can be used to characterize communities (Smith & Smith, 2015; Ding & Schloss, 2014). Third, statistical associations between species might be used as indicators of communities: the community is then defined using its species composition quantified using statistical synthetic indices, often defined with ordination methods, which consists in positioning species composition vectors in a multivariate space. Communities are then defined based on the distance between vectors (Morin, 2011). Finally, communities might be defined in terms of the interactions of species sharing the same space (for a detailed definition of interactions, see section 3.1), on the grounds that these interactions have strong consequences on the abundance and maintenance of species and on the functions they perform. For instance, we might define the plant-pollinator community composed of a set of plant species and insects that pollinate these plants (Memmott, 1999), or the host-parasitoid community, like lepidopteran (butterfly) species and their tachinid (files) parasitoids (Stireman III & Singer, 2003).

No matter the definition used, ecological communities are typically made up of tens to hundreds of species, sometimes rare or difficult to observe. This complexity of the community makes it difficult to accurately sample the entire community: moreover, the sampling of communities is often limited by practical considerations (Morin, 2011). Therefore, although ecologists recognize a broader definition of the community, practical considerations often constrain them to restrict themselves to an observable subset of this community. These limitations are to keep in mind when interpreting community ecology studies, because it is probable that some species of the community are missing due to sampling constraints.

The importance of interactions as a structuring factor of ecological communities has been the subject of debates. Schematically, one view holds that interactions are a fundamental factor structuring communities, and the other that communities are defined primarily by the physical environment, and that interactions are a consequence of this (Smith & Smith, 2015). This debate has its origins in the early 1900s in the con-



Figure 1.2: Kelp forest community. Kelp algae (shown on the picture) are considered foundational species and sometimes used to define kelp communities because they provide a habitat for a wide array of species, among which otters, fish (shown on the picture), algae, urchins and other vertebrates and invertebrates. Own picture taken at the Cape Town aquarium (CC-BY 4.0).

text of plant ecology, with the organismic versus individualistic debate. At that time, the dominant view among botanists, notably championed by Frederic Clements (Clements, 1936), was that plant communities were composed of tightly linked species, structured by their interactions, that formed an operational unit akin to an organism: this is the organismic view of communities. A major argument supporting organicism was that plant communities are fundamentally distinct units (e.g. deciduous forest and prairie). Thus, to explain these fundamental differences, organicism relies on internal regulating processes occurring inside communities, and these processes are mediated by interactions between species (Nicolson, 1990). In 1926, Henry Allan Gleason challenged this view and proposed the individualistic concept to explain the structure of plant communities (Gleason, 1926). This paradigm views the emergence of a plant community as resulting of individual colonization processes for different species, determined by their specific environmental requirements (which would later be developed by the concept of ecological niche). The individualistic view explained vegetation gradients much better than organicism: for instance, consider an altitude gradient, where communities dominated by coniferous species gradually give way to deciduous trees dominated communities. In that case, the distinction between communities are not clear-cut, thus challenging the organismic view.

The contemporary view of ecological communities has become wider and includes other taxa than plants, and recognizes both interactions and the environment as structuring factors of communities. More recently, the relevance of community ecology has been challenged by some and defended by others (see Box 1.1). The organismic view of communities has gradually become less popular, notably due to the invocation of final causes to explain observed patterns, in particular with the vision of successional changes in a community as a path towards its stable state (climax) (Pickett et al., 2009). However, studying processes regulating communities is still relevant today (Pickett et al., 2009). In particular, the study of a global functioning of communities has been revived through the study of ecological networks (see section 4 for a definition and discussion) (Elton, 1927; Proulx et al., 2005; Valiente-Banuet et al., 2015).

All definitions of communities presented above can be relevant, depending on the context: ultimately, the concept of community resides in the eye of the observer, and defining its contours depends on what they want to study (Levin, 1992). In this thesis, I define communities so that they encompass interacting species. Because of limitations imposed by sampling, I also restrict this definition to some phyla. In part 2 of the thesis, I use a community of birds and fruits that interact through seed dispersal. In part 3, I study the spatio-temporal distribution of preys and a predator from the African savanna and for practical reasons, I restrict the community to a subset of 5 mammal species.

Box 1.1: The scientific status of community ecology

In 1999, John H. Lawton published an article entitled “Are there general laws in ecology?” (Lawton, 1999). In this article, he criticized the study of ecological systems at the scale of communities, and called community ecology a “mess”, because community dynamics are essentially contingent on the species and the environment involved. Lawton deems the community scale as unworkable and impossible to study scientifically, and asks “why ecologists continue to devote so much time and effort to traditional studies in community ecology”, stating that “in [his] view, the time has come to move on.”

Daniel Simberloff answered Lawton (Simberloff, 2004) and agreed that although communities are largely contingent, in the context of climate change and increasing anthropic pressure on ecological systems, many contemporary chal-

lenges require a good understanding of communities to protect biodiversity and ensure humanity's survival. In his response, he also states that general laws are not necessarily a good indicator of the value of the scientific knowledge. Indeed, Lawton's critique of community ecology as a science lacking laws is shared by many ecologists criticizing ecology as a whole, in particular when they compare ecological knowledge to the laws of physics, and exhibit a so-called "physics envy" (Travassos-Britto et al., 2021; Simberloff, 2004; Egler, 1986).

Another addition to the debate is brought by McGill et al. (2006), who challenge the idea that no generalizations are possible in community ecology. They proposed to shift the focus from species as fundamental units of communities to their functional traits. Indeed, studying species through their traits could allow to transpose knowledge between communities of species with similar traits, thereby effectively reaching community ecology laws.

More recently, Bruno Travassos-Britto and colleagues (Travassos-Britto et al., 2021) reflected on the epistemological status of ecology. In their article, they point out that there are several ways to generate knowledge that do not necessarily require the formulation of universal laws. They propose that the knowledge generation process in ecology is best described using the pragmatic view, by which knowledge emerge from the confrontation of multiple models with data. Interestingly, models with different underlying assumptions can still both be informative under this view: Travassos-Britto et al. (2021) cite the example of Hubbell's neutral theory (Hubbell, 2001) and niche models of community assembly, two models that inform us about different facets of the niche assembly process.

I am aware and deeply concerned about global changes affecting ecological systems. When I read the article of Lawton, it really made me doubt that community ecology could be useful to protect natural systems. Reading the argument of Travassos-Britto and colleagues was an epiphany, because I realized that even if ecology does not find universal laws, it can still generate useful knowledge.

3 Interspecific interactions

3.1 Definitions

In their broader sense, interspecific interactions are generally defined as the effect that a species has on another one (Strydom et al., 2021). In the "Community ecology" article of the Encyclopedia Brittanica, interspecific interactions are defined as "the interactive relationships that arise between populations of different species" (Thompson, 2024). Definitions are rarely expressed with more details (see Box 1.2), and these generic definitions do not precise the nature of the effects exerted on species (Strydom et al., 2021; Landi et al., 2018). In the literature, these effects have been characterized in different ways. Below, we list three ways to define interactions, without the ambition to be exhaustive: through plastic, demographic and co-occurrence effects of interactions.

Intuitively, interactions are generally conceived as the physical encounter between two individuals of different species: the effect of an interaction is to trigger behavioral or physiological responses (Wootton & Emmerson, 2005). In this view, interactions are characterized by the effect on the plastic response of species when there is a direct encounter. For example, we could view the predation interaction as the encounter of an individual of a predator species (say, a wolf) with an individual of a prey species (deer), and predation would then be described as the resulting sequence of pursuit, and potentially catching and eating of the prey by the predator. Another example involving plants is the response of a plant species to another plant competitor, where one plant grows of deeper roots to evade the competition for nutrients. However, the plastic responses of a species to another one encompass a much wider variety than reactions to a physical encounters (Montgomery et al., 2019). In particular, some behaviors might be displayed in anticipation of the encounter, like preys behavioral strategies to mitigate predation, including grouping, sound-signaling or vigilance: this is the notion of "landscape of fear" (Palmer & Packer, 2021; Palmer et al., 2022). Similarly, the search for resources of a pollinator species, or bees' waggle dance to communicate the position of flowers, might be considered as behaviors arising from the interaction of pollination.

Another way to characterize the effects of interactions is through demography, by

quantifying the effect of interactions through the influence that the abundance of a species has on the abundance of another one. This conception of interactions is very popular in ecology, especially in the context of food webs (Wootton & Emmerson, 2005; Berlow et al., 2004). Demographic effects are mediated by finer-scale behavioral processes as described above, so they can be viewed as a more integrative scale. However, as Say-Sallaz et al. (2019) point out in the context of predation interactions, demographic effects do not encompass the complete array of preys responses to interactions. Moreover, these demographic effects are more easily conceivable in the context of trophic interactions, and the correspondence between behavior and demography is less evident in the context of mutualistic or competitive interactions (Wootton & Emmerson, 2005).

A third definition of interaction is related to species spatial and/or temporal distribution. This conception rests upon the concept of species co-occurrences, i.e. the joint presence of two species in the same place. In this context, interactions are defined by Blanchet et al. (2020) as "the presence of a species [that] has some influence on the occurrence of another". Like the demographic view, this definition also relies on smaller scale responses, like movement for mobile species or demographic effects leading to the loss of individuals in a given area (e.g. due to predation). Species co-occurrence can be defined at various spatial and temporal scales, and the choice of the scale at which co-occurrence is defined impacts the ecological interpretation of co-occurrences (Levin, 1992; Araujo et al., 2010). For instance, if two animal species co-occur in a 10 square kilometers area, we should not exclude finer-scale spatial avoidance mechanisms inside this larger unit (Palmer et al., 2022). Conversely, two species might avoid each other at fine scale, but their range can be restricted to the same biome.

These different effects of interactions are not mutually exclusive, and probably occur jointly in most communities. The focus on one of these effects is dictated by the ecological question of interest, but recognizing that interactions may have several effects (on behavior, demography and co-occurrences) allows to view interactions as an integrative process that affects many characteristics of the community.

In this thesis, I use two definitions of interactions. In part 2, I define bird-fruit interactions as the observable behavior of birds seen eating fruits. In part 3, I define interspecific interactions as the attraction-avoidance between species at a small spatio-

temporal scale.

Box 1.2: Definitions of interspecific interactions in the literature

Although interspecific interactions are an important ecological concept, it is rarely defined in the literature other than by examples. This lack of comprehensive definition became apparent to me when I searched for definitions in academic articles mentioning interspecific interactions, as only two articles gave an operational definition of what the authors meant by interaction (Strydom et al., 2021; Blanchet et al., 2020). Similarly, in the ecology textbooks that I could read, interactions are not defined in a generic way, but instead exemplified through several instances of interaction types (C. R. Townsend et al., 2008; Morin, 2011; Smith & Smith, 2015). This is maybe due to the perception of the definition of interactions as intuitive, and comparable to its dictionary definition: "a mutual or reciprocal action or influence" (Collins English Dictionary, 2024).

This lack of definition does not prevent the practical study of interactions in ecological systems, but it precludes a more general view of what interactions are. I also think that employing a vague definition can weaken the study of interactions, because without a clear characterization of interactions in mind, we might miss some of their effects or causes when studying them.

3.2 Types of pairwise interactions

Interspecific interactions are generally considered between pairs of species and classified following the influence of the interactions on each species (see Figure 1.3). Interactions can have a positive (+), negative (-) or neutral (0) effect on the fitness of each species, i.e. its survival and/or reproduction abilities (Araújo & Rozenfeld, 2014; Mathis & Bronstein, 2020).

First, mutualistic (+, +) interactions are beneficial for both interacting species. For example, pollination of plant species by insect, birds or mammals can be characterized as mutualistic interactions, like the orchid species *Disa uniflora* pollinated by the butterfly *Meneris tulbaghia*. The butterfly benefits the orchid by fecundating the flowers and allowing the plant to produce fruits, while the flower provides nectar resources

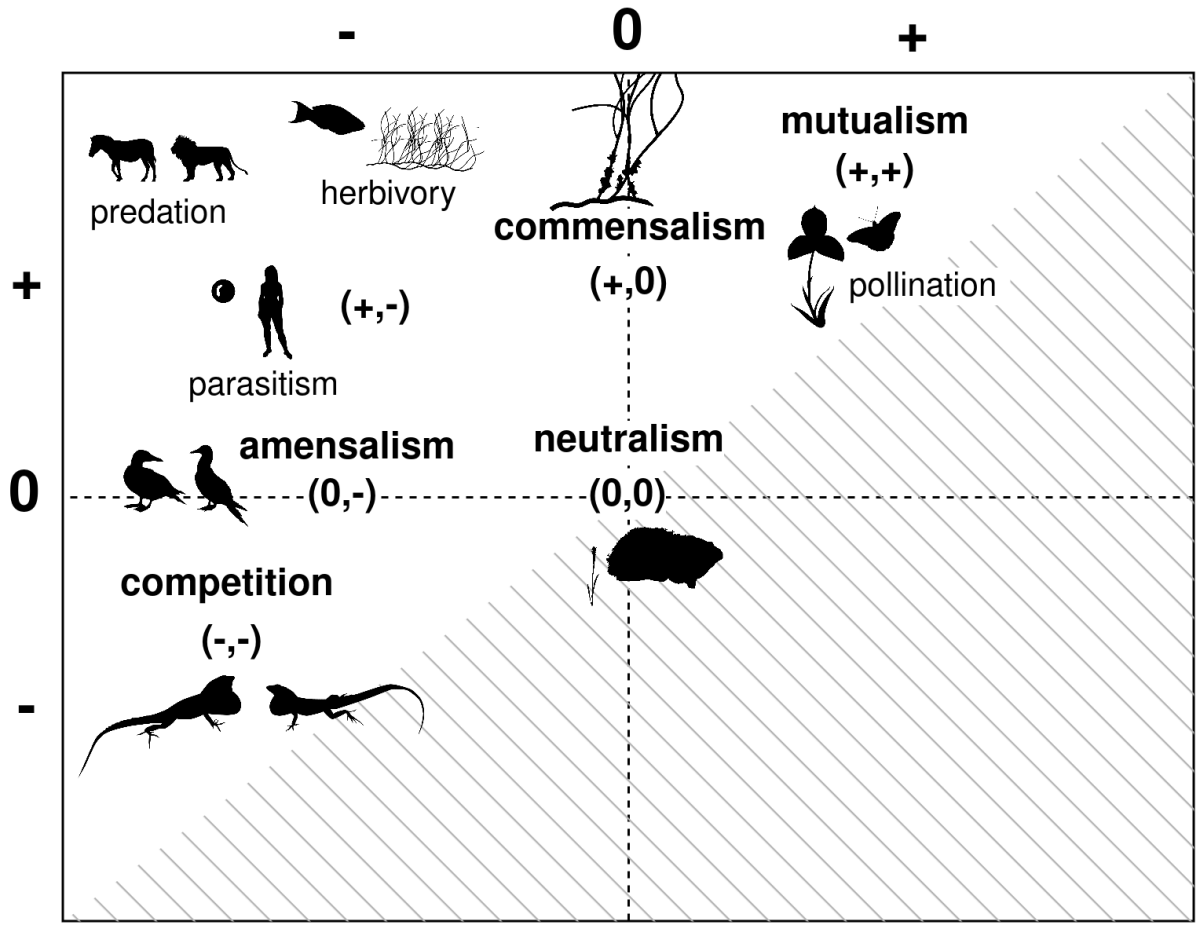


Figure 1.3: Classification of interactions following their effect on the interacting species. Interactions can be classified along a positive-negative effect gradient with two dimensions, one for each species. The x-axis describes the effect of the interaction on the first species, and the y-axis on the second species: positive (+), negative (-) or neutral (0) effect. Each zone corresponds to one interaction type and species silhouettes illustrate examples of interactions described in the main text. The hatched area covers symmetric interactions types to those already represented on the figure. Own figure (CC BY-SA 4.0) inspired by Figure II in Morales-Castilla et al. (2015). Silhouette images from Phylopic by Andy Wilson (*Meneris tulbaghia*), Beth Reinke (Nazca booby), Jon Hill (blue-footed booby), Jonathan Wells (*Plasmodium falciparum*), and NASA (Homo sapiens sapiens), others are in the public domain.

to the butterfly (Johnson & Bond, 1992). Commensal interactions (+,0) are beneficial for one species only, while the effect on the other species is neutral (Mathis & Bronstein, 2020). For example, the liana species *Schisandra repanda* and *Schizophragma hydrangeoides* grow on trees, that provide them the necessary structure to reach the canopy and photosynthesize more efficiently, while leaving their host trees unaffected by their presence (Ichihashi & Tateno, 2011). Another example is the relationship between a sponge species *Halichondria melanodocia*, and a seagrass species *Thalassia testudinum* (Archer et al., 2015). Sponges benefit from the seagrass because they provide a substrate on which they can grow, namely the base of the seagrass shoots. For the seagrass, the cost incurred by sponges blocking photosynthesis balances out with

the benefit conferred by the presence of sponges that produce bioavailable nutrients for the plant (Archer et al., 2015). When interactions are positive for one partner, but negative for the other (+,-), 3 main subtypes of interaction have been identified: predation, herbivory and parasitism. Predation occurs when one animal species eats another, thereby killing it. For instance, lions are iconic predators of many species of the African savanna, including gemsbok, buffalo and zebra (Hayward & Kerley, 2008). Herbivory is an interaction where one animal species eats a plant (or some part of the plant): for example, the herbivorous parrotfish *Calotomus spinidens* predominantly grazes on seagrass, species of marine angiosperms (flowering plants) (Vonk et al., 2008; Heck et al., 2008). Parasitism is an interaction involving one species that takes advantage of the other, but needs to maintain this species alive to do so. During a parasitic interaction, the two species often live in close association, like the protozoan *Plasmodium falciparum*, a parasite of humans that benefits from the protected environment of liver cells to accomplish a step of their life cycle. When doing so, they considerably affect their human hosts, as *P. falciparum* is one of the *Plasmodium* species causing malaria in humans, with severe symptoms including fever, headaches and sometimes leading to death (Ménard et al., 2013; Paul et al., 2003). Another instance of parasitism is parasitoidism, where the host is eventually killed by the parasite: most parasitoids are Hymenoptera insects which develop inside an arthropod host in the larval stage (Santos & Quicke, 2011). Neutral interactions (0,0) occur when two species interact, but the interaction has no effect on either species. For instance, in mild environmental conditions, the cushion plant *Silene acaulis* seems to have no detrimental or positive effect on the common alpine forb growth *Bistorta vivipara* (Kjær et al., 2018). Amensalism (0,-) occurs when one species is negatively affected by the interaction while the other gains no apparent benefit. For example, H. M. Townsend et al. (2002) describe the relationship between two bird species, blue-footed and Nazca boobies. Non-breeding Nazca boobies have been observed attacking blue-footed boobies nestlings, thus negatively affecting this species, while apparently retiring no benefit from this interaction because they do not predate on the nestling or acquire the nestling site. Amensalism can be considered a form of asymmetric competitive interaction. Competition (-,-) arises when the effect of the interaction is detrimental to both species. Two species can compete for resources like food or territory. Examples of direct or interference

competition include aggression between animal species. For instance, the two species of lizards *Anolis sagrei* and *Anolis cristatellus* exhibit a stereotyped behavior when the other species enter their territory, by doing pushups, head-bobs and dewlap displays (Grether et al., 2013). This type of behavior is potentially costly to the defender species, because they consume energy to display the aggressive behavior, and to the intruder species, because they are prevented the access to the territory.

All these examples highlight the diversity of outcomes of interactions types, but also the range of behaviors that constitute interactions. Many other forms of interactions are possible, and they might be more difficult to classify than the simple examples given above. Indeed, in ecological communities, the delineation between interactions' outcomes is a gradient (see Figure 1.3). For instance, some mutualistic interactions might be only marginally positive for both species, and in practice closer to neutralism. Interactions might also be asymmetric: thus, competition (-,-) might negatively affect a species more strongly than the other. The outcome of interactions may also vary depending on the context. To give an example, the neutral interactions between the cushion plants and the common alpine forb described above becomes commensal for the forb when environmental conditions are harsher (closer to the glacier), in accordance with the stress-gradient hypothesis (Kjær et al., 2018). Within the same species, inter- and intra-individual variations might also change the outcome of the interaction. For example, several pollinator species are capable of nectar robbing, a behavior consisting in perforating the flower's corolla to access nectar instead of pollinating the flower by entering from the top. Individuals from the same species can use both tactics to access nectar (Bronstein et al., 2017): this behavioral variability has the potential to change the interspecific interaction from mutualistic to parasitic/commensal. Finally, accurately quantifying the whole range of positive or negative effects of a species interaction is extremely difficult, and we may fail to accurately measure the effect of the interaction on fitness (Mathis & Bronstein, 2020). For example, the shark-remora association, by which remoras attach to marine megafauna by their suction disk, has long been considered mutualistic or commensal, because of the benefits of such an association for remoras. But recent evidence nuance this conclusion by highlighting a hydrodynamical burden of remoras attachment for sharks, shifting the relationship to parasitic in some contexts (Gayford, 2024).

3.3 Indirect effects

Above, we described the effect of species pairwise interactions in isolation. But inside communities, species coexist with a variety of other species, and a single species interacts with many other species at once. These complex webs of interacting species have long been recognized in ecology, especially in the context of predator-prey interactions gathered in food webs (Ings et al., 2009). For example, in 1927, Charles Elton wrote that "there are [...] chains of animals linked together by food" (Elton, 1927). Considering the whole network of interactions occurring within a community can substantially affect the outcome of the interactions involving only two species. Indeed, the interaction between two species can be modified by a third species through indirect effects (Strauss, 1991). There are two fundamental mechanisms by which such effects can appear (Wootton & Emmerson, 2005): density-mediated effects and higher-order interactions (see Figure 1.4), presented in the two following paragraphs.

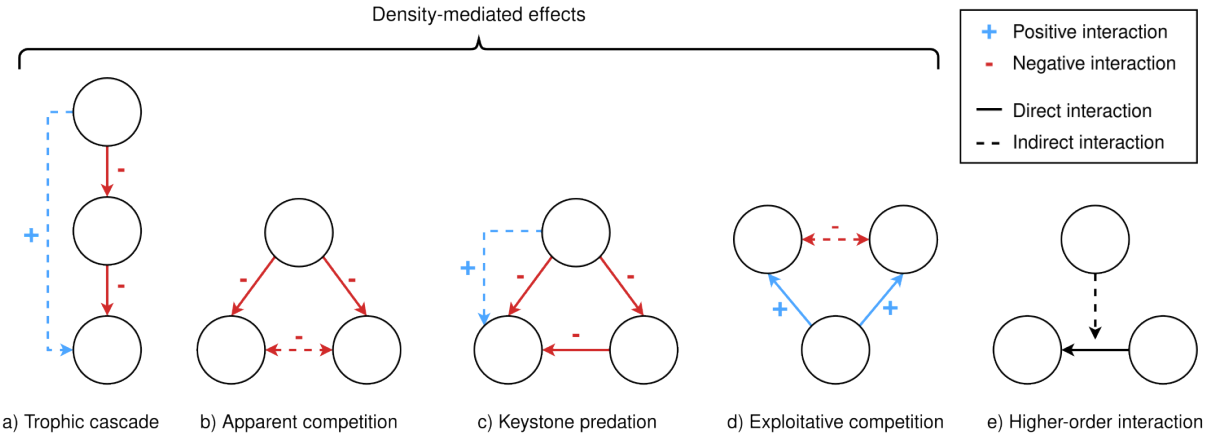


Figure 1.4: Schematic representation of indirect interactions. Species are depicted as circles and interactions as arrows. The arrow points to the species receiving a positive or negative outcome. Only the relevant interactions are represented (e.g. the predation in a) is only symbolized with a negative outcome for the prey). Density-mediated effects include: a) trophic cascade, where a top predator indirectly benefits a species at an inferior trophic level; b) apparent competition, when a predator induces indirect competition between two prey species; c) keystone predation, where a predator indirectly benefits a subordinate competitor; d) exploitative competition, where two predators indirectly compete for the same resource. e) Higher-order interactions occur when a species modifies the nature of the interaction between two other species (on this diagram, the trophic level is irrelevant).

Density-mediated effects (Figure 1.4a-d), sometimes called interaction chains (Wootton, 1994), occur when a species alters the consequence of an interaction by reducing or increasing the abundance of one (or both) species through a direct in-

teraction. Frequently-described density-mediated indirect interactions include trophic cascades, apparent competition, keystone predator effect, and exploitative competition (Werner & Peacor, 2003; Wootton, 1994). First, trophic cascades can occur when a top predator is added to a dyadic interaction. For instance, the recently introduced brown trout present in New Zealand streams exerts a strong predation pressure on small invertebrates that graze on small algae growing on the riverbed (Huryn, 1998). In streams where the trout is present, algal biomass is much higher than in the absence of trout, due to the release of grazing by invertebrates caused by higher predation pressure (McIntosh & Townsend, 1996; Huryn, 1998). Apparent competition occurs when a shared predator is added to the system and fosters indirect competition between these species. For instance, the important abundance of soybean aphids in the Midwestern United States allows to maintain a substantial population of predaceous coccinellid beetles, which then turn to pea aphids. Hence, soy aphids exert an indirect negative effect on pea aphids through a shared predator, and an (asymmetric) competition arises, even though they forage on distinct resources (Kaplan & Denno, 2007). Another outcome of shared predators is keystone predation, when a predator holds a dominant competitor in check and allows species coexistence, thus being indirectly beneficial to the subordinate competitor. One example of keystone predation is the classical study by Robert Paine on an intertidal marine community composed of a species of starfish, *Pisaster ochraceus*, as well as several smaller species preyed upon by the starfish, like bivalves, barnacles and chitons, that all compete for space to grow on the rocks (Paine, 1966). The experimental removal of starfish lead to an increase of the population of mussels *Mytilus californianus*, free from the predation pressure; this species then outcompeted other species, resulting in a decrease of specific diversity that shifted from 15 species when the starfish was present to 8 species in its absence. In this experiment, the starfish presence then indirectly benefits the competitors of mussels. Exploitative competition occurs when two species compete for the same resource (Wootton, 1994; Strauss, 1991). For instance, Teder et al. (2013) describe three species of parasitoid wasps competing for the access to resources, in that case, the larvae of their moth hosts in which they lay eggs. In their study, they show evidence of indirect competition because when a parasitoid species laid more eggs, the other two laid less.

Another type of indirect interactions are higher-order interactions (Figure 1.4e), also

called trait-mediated interactions. They occur when the presence of a third species affects the strength of the interaction between two other species, potentially even modifying the interaction outcome, mediated by plastic changes (Levine et al., 2017; Werner & Peacor, 2003). To illustrate, we can take the example of wolves interactions with other species, which could be affected by humans in various ways (Kuijper et al., 2024). Notably, wolves predation on ungulates might be modified when humans are present, because humans modify wolves behavior by making them avoiding human-dominated areas. As wolves tend to avoid humans more than their preys, this could lead to a "human-shield" effect for ungulates, thus releasing wolves predation pressure even when wolf density remains the same. As higher-order effects have been little studied, and are difficult to quantify, few ecological examples exist yet (Levine et al., 2017).

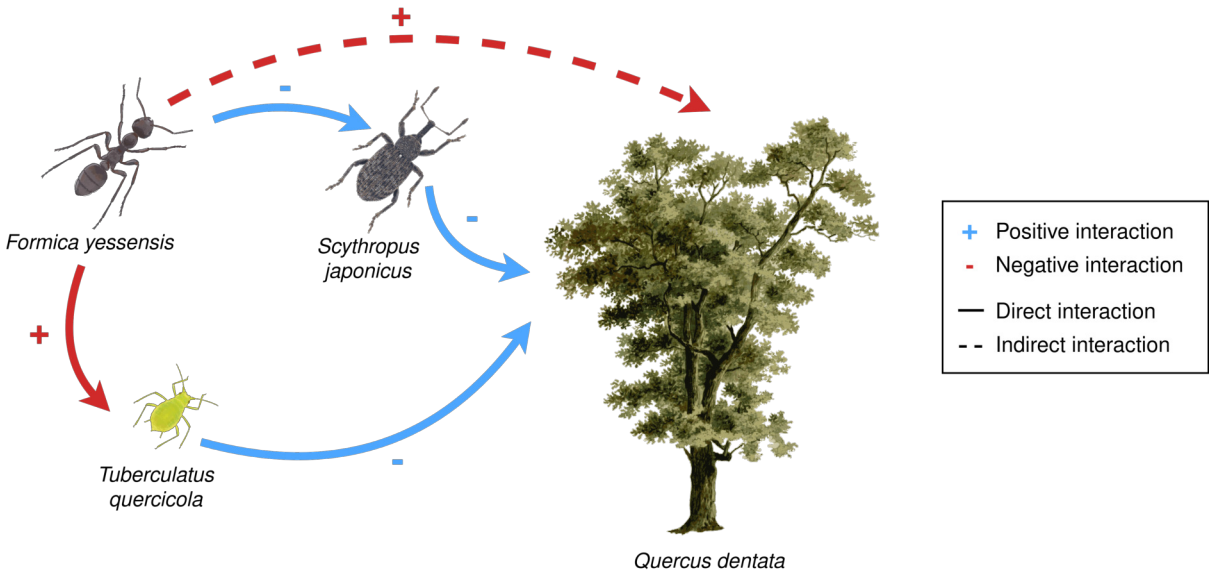


Figure 1.5: Indirect effects of ants *Formica yessensis* on oak trees *Quercus dentata* described by Ito and Higashi (1991). This diagram summarizes the interactions between ants *Tuberculatus quercicola*, aphids *Tuberculatus quercicola*, acorn borers exemplified by the weevil *Scythropus japonicus* and oak trees *Quercus dentata*. Ant and aphid icons by DBCLS (CC-BY 4.0), weevil adapted from a picture of Udo Schmidt (CC-BY-SA 2.0) and tree adapted from a drawing of Carl Theodor Reiffenstein (public domain).

These examples illustrate classic cases of indirect interactions. However, depending on the combination of interactions considered, and of the number of species involved, many types of indirect effects can arise. For example, we can consider the relationship between ants, aphids, oak trees and acorns insect parasites described by Ito and Higashi (1991) (see Figure 1.5). Some ants and aphid species undergo a mutualistic relationship, where aphids produce nutritious honeydew consumed by ants, while ants protect aphids from predators. Aphids feed on plants sap, thus dam-

aging and weakening plants, so their effect on plants is expected to be negative. At the same time, ants are predators of various insects, including larvae of acorn borers. Ito and Higashi (1991) observed the ants *Formica yessensis* and aphids *Tuberculatus quercicola* provided a net benefit for oak trees *Quercus dentata* on which they were installed. Indeed, although acorn production was the same between trees with and without ants/aphids associations, the parasitism rate was higher on trees that did not harbor ants and aphids. This was explained by ant predation on insect larvae and aggression towards adult weevils and lepidopterans. This study suggests an indirect benefit of ants for oaks, despite the negative effect aphids have on oaks, because it is outweighed by ants' predation on oaks parasites. With this example, we can see how complex the effects of interaction networks can be inside communities when multiple interactions are considered.

4 Ecological interaction networks

To study these complex interaction networks, ecology is increasingly using the tools of network science (Miranda et al., 2013). In this section, I briefly present network science and describe how networks are used to describe and analyze of ecological interactions.

4.1 Network science

Networks (or graphs) are mathematical objects describing the relationships between entities, represented by a set of nodes (or vertices) linked by links (or edges). Networks have been used to describe a variety of systems: for instance, social networks can be modeled with people (nodes) linked by relations (links), such as friendships; metabolic networks can be modeled as molecules linked by biochemical reactions; or the World Wide Web can be represented as webpages connected by hyperlinks (Albert & Barabási, 2002).

Network science, which is the study of networks representing real phenomena, rests upon the mathematical tools of graph theory (see Box 1.3), which traces back to the mid-18th century and the bridges of Königsberg problem, asking if one can find a path going through all seven bridges of the town of Königsberg without crossing one

bridge twice. In 1735, Leonard Euler solved the problem by representing it as a graph, with nodes representing land masses and links representing bridges, thus demonstrating that such a path did not exist (Euler, 1741). This work is considered to be the first mathematical proof using graph theory (Barabási & Pósfai, 2016). Network science gained traction in the beginning of the 21st century: this growth was notably motivated by computational advances making it easier to map, store and analyze networks (Albert & Barabási, 2002; Barabási & Pósfai, 2016).

Box 1.3: Graph theory and network science

In this section, I discuss graph theory and network science, which are two connected fields: graph theory refers to the fundamental aspect of the study of networks and often involves mathematical or numerical developments, while network science refers to the study of networks with a much more empirical focus, applying the principles of graph theory to real-life networks.

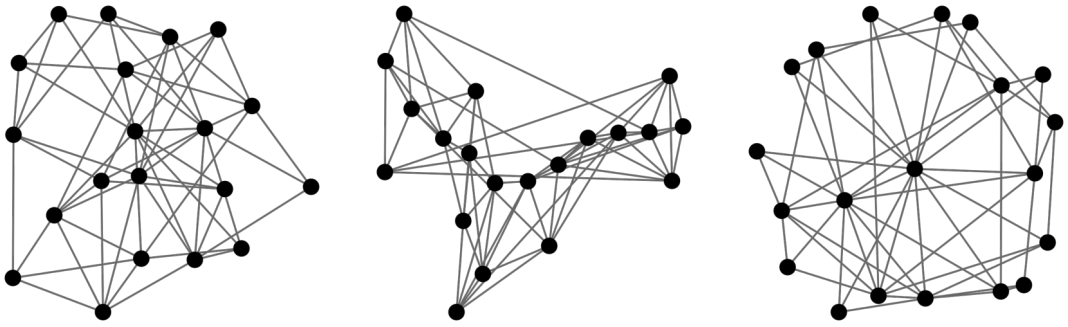
Graph theory and network science use different terminologies to refer to similar concepts: hence, in network science, people tend to use the terms “network”, “link” and “node”, while in graph theory they use “graph”, “edge” and “vertex” to refer to the same concepts. In this manuscript, I mainly use the terminology of network science.

Networks can be characterized by a variety of summary statistics. Some properties are global: for instance, a network can be characterized by its size, defined as its number of nodes, or its connectance, defined as the proportion of realized links among all theoretically possible links. Other indices are more local: for example, the node degree defines how many links a given node engages in, and the clustering coefficient characterizes the degree to which neighbors of a node (nodes connected to it) are connected to each other.

In the beginning of the 21st century, it became increasingly clear that several empirical networks describing very different systems tend to share common structural characteristics (Albert & Barabási, 2002). For instance, real networks tend to be sparse: that is, they have a low connectance (Barabási & Pósfai, 2016). They also tend to be “small-world”, which means that few links are needed to join two nodes in comparison with the size of the graph (Watts & Strogatz, 1998). Moreover, real networks

tend to display clusters (i.e. subsets of highly connected nodes) (Albert & Barabási, 2002). Networks have also been found to have a scale-free degree distribution, which means that the distribution of node degree tends to follow a power law (Barabási & Albert, 1999) with a few well-connected nodes and many weakly connected nodes. However, more recently, this property has been contested: in particular, Broido and Clauset (2019) analyzed close to 1000 real networks from various domains and found that less than 4% exhibited a strong evidence for a scale-free degree distribution.

In parallel to the discovery of these shared structural patterns, several modeling paradigms have been proposed to generate networks with given structural properties, and in particular to reproduce structures observed in real networks (see Figure 1.6). One of the earliest probabilistic graph model is the Erdős-Rényi model (Gilbert, 1959; Erdős & Rényi, 1959). This model starts from N nodes, between which links are formed with probability p (see Figure 1.6a). The resulting networks are also called "random graphs" and often used as a null model for graphs, because the links formation process is random. Erdős-Rényi networks can be sparse, and they typically have a small-world property, but they fail to capture the clustering property commonly observed in real networks.



(a) Erdős-Rényi (b) Watts-Strogatz (c) Barabási-Albert

Figure 1.6: Examples of networks generated with different models. (a) Erdős-Rényi network with $N = 20$ nodes and linkage probability $p = 0.3$. (b) Watts-Strogatz network with $N = 20$ nodes, $K = 3$ links per node and rewiring probability $p = 0.1$. (c) Barabási-Albert network with $N = 20$ nodes, $m_0 = 1$ initial nodes and $m = 3$ links per new node.

To reproduce typically clustered real networks, Watts and Strogatz (1998) proposed a model allowing to generate small-world networks with high clustering coefficients (see Figure 1.6b). This model starts with a regular network where N nodes are disposed as a ring, and each node is connected to its first K neighbors. Each link is then randomly rewired with probability p. This model rests on the idea that a node is more

likely to be connected to its neighbors, but with time, some long-distance links may develop: for instance, in social systems, people tend to form links with people that are spatially close, but some may also be connected to people that live far away (Albert & Barabási, 2002).

The Barabási-Albert model (Barabási & Albert, 1999) importantly proposed a dynamical mechanism to explain the emergence of scale-free graphs (Albert & Barabási, 2002). Contrary to the Erdős-Rényi and the Watts-Strogatz models, the Barabási-Albert generation process does not start with its final number of nodes N, but with a subset $m_0 < N$, thus allowing to describe the dynamical formation process of a network. The network grows by sequentially adding new nodes connected to m old nodes. These links are preferentially attached to nodes that are already highly connected. This preferential attachment property is thought to be realistic in many systems, where already highly connected agents are more likely to form new links (e.g. due to popularity in social networks).

The simplicity and versatility of these three models have made them classical starting points for modeling networks. However, some features of real networks fail to be accurately captured with these models (Amaral et al., 2000; Albert & Barabási, 2002; Broido & Clauset, 2019) and other models have been proposed to take into account specific constraints of the system. For instance, the network of actors collaborations in movies from the Internet Movie Database exhibits a truncated scale-free distribution, with the most connected nodes (actors) having a lower degree than predicted by the Barabási-Albert model: Amaral et al. (2000) proposed that this could be explained by aging, as when actors get older, they become less likely to act in new movies, and this imposes a threshold on the maximal number of collaborations.

In ecology and evolution, networks have been used to describe a variety of systems (Lau et al., 2017; Gosak et al., 2018), among which gene regulatory networks (Banf & Rhee, 2017), fluxes of biomass and energy (Ings et al., 2009) or spatial connectivity between patches (Fletcher et al., 2013). In this thesis, I will focus on networks representing interaction between species, which are described in the following section.

4.2 Generalities on interaction networks

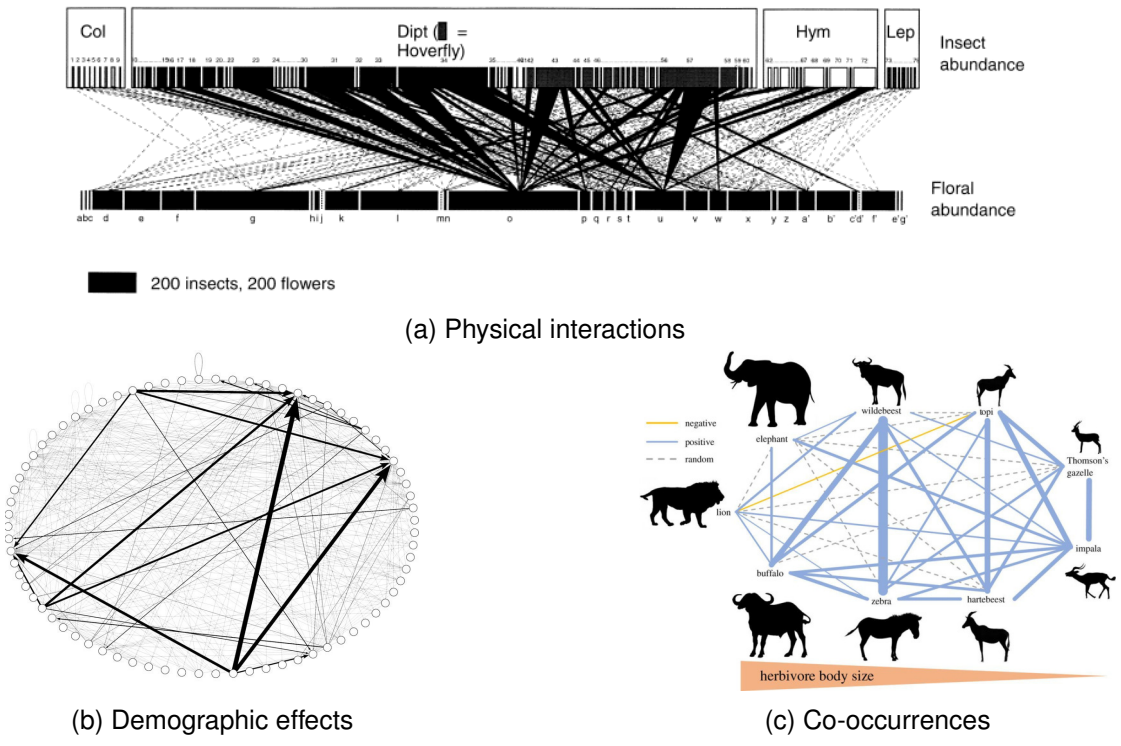
In community ecology, networks are commonly used to describe interactions between species: nodes represent species, and links represent their ecological interactions (Ings et al., 2009). Values might be associated to links and/or nodes, for instance to represent the interaction strength (links) or species abundance (nodes).

Interaction networks have been used to describe a variety of interaction types: the most commonly described are predation, mutualism and parasitism (Ings et al., 2009). Historically, food webs have been the earliest and most studied network type. They describe feeding relationships between species in a community and are typically composed of multiple trophic levels. These levels can be blurred, for example by mutual predation of two species, or by cannibalism. For this reason, food webs can be described as unipartite networks, i.e. networks whose nodes belong to the same category (Montoya & Solé, 2002; Jordano et al., 2003). However, strict prey-predator networks have also been described as bipartite networks (networks which have two classes of nodes, with links only between the two classes) (Bascompte et al., 2003). Host-parasitoid networks describe relationships between parasitoids (e.g. parasitoid wasps species) and their hosts (e.g. insect species), and are most generally described as bipartite networks. Mutualistic networks describe interactions such as pollination (often between insects and flowering plants) and frugivory (often between birds and plants producing fruits dispersed by birds). These networks are bipartite as well.

Even though these multiple interaction types coexist within communities, for practical reasons, they have been studied mostly separately. However, statistical tools and data allowing to integrate multiple interaction types are beginning to emerge, and some studies have used multiplex networks, i.e. networks integrating multiple interaction types (Kéfi et al., 2016; Melián et al., 2009): these exciting studies pave the way for a deeper understanding of complex ecological systems.

Since interactions can be defined in multiple ways (section 3.1), interaction networks can represent physical, demographic or co-occurrence interactions (see Figure 1.7). For instance, Memmott (1999) (Figure 1.7a) construct a plant visitation network between flowering plants and insects of an English meadow, where links represent the number of observed insect visits, defined as an insect being observed on the flower. Food webs frequently represent demographic effects of predators on prey, like the net-

work by Bascompte et al. (2005) (Figure 1.7b), who consider trophic interactions from a marine community, where edges represent the proportion of prey biomass consumed per unit of predator biomass per day. Another example is the co-occurrence network of eight herbivore mammals and lion analyzed by Anderson et al. (2024) (Figure 1.7c), where links represent positive or negative spatial co-occurrence estimated from an occupancy model.



(a) Physical interactions

(b) Demographic effects

(c) Co-occurrences

Figure 1.7: Examples of networks representing different effects. (a) Physical interactions defined as visits of flowers by insect pollinators (Figure 1 from Memmott, 1999). (b) Demographic interactions defined as per-capita effect of predators on prey (Figure 2a from Bascompte et al., 2005). (c) Co-occurrence network defined as the non-random spatial association between species (Figure 5 from Anderson et al., 2016). a) © 1999 Blackwell Science Ltd/CNRS; b) © (2005) National Academy of Sciences, U.S.A. c) © The Author(s) Published by the Royal Society.

Inside communities, tens of species, and thousands of individuals might interact in diverse ways, possibly dependent on many individual, spatial and temporal variables. Therefore, ecological networks can be classified as complex systems, systems that are made up of multiple interacting components. One of the characteristics of complex systems is that their dynamics are hard to predict, due to the multiple interactions between their components and to their structure that can give rise to non-linearity or feedback loops (Barabási & Pósfai, 2016; Pessoa, 2022). Complex systems can also display emergence, which refers to a property that arises only when all components of the system interact, but cannot be deduced from the behavior of its subcomponents.

A prime example of emergence in ecological systems is the higher-order interactions mechanism described above. Indeed, if a species modifies the interaction between other species, it is hard to predict this modification based only on each species' pairwise interactions, and we need to observe the whole system for this higher-order interaction to arise. Emergence is a philosophically debated concept: it is not clear if emergent properties could theoretically be predictable with enough information, or if they are intrinsically unpredictable by breaking down the system (Pessoa, 2022). For instance, if we consider the impact of humans on the interactions between wolves and other species (such as the system described by Kuijper et al., 2024, and discussed in section 3.3), it might be that with enough knowledge on wolves' behavioral response to humans, we could predict the dynamics of the system. However, emergence can be defined pragmatically in a weaker sense, as a property that is not predictable with the knowledge available at a given time (Pessoa, 2022).

Box 1.4: Holism and reductionism in ecology

In ecology, there is a long-standing debate between tenants of a holistic and a reductionist view of ecological systems (Lévêque, 2001). Holistic supporters believe that to understand ecological systems, they should be studied as a whole because their global functioning is fundamentally different than what we could observe by cutting down the problem. This way of thinking is well summarized by the saying "the whole is more than the sum of its parts".

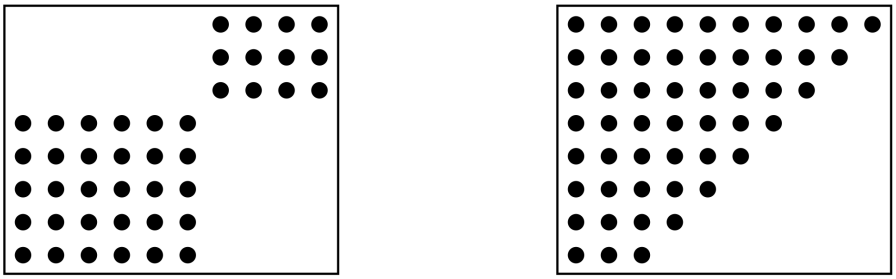
On the contrary, reductionists believe that the best way to understand complex ecological systems is to study manageable subsets of its components one at a time. This position can be summarized by the following quote from the biologist Edward O. Wilson: "The cutting edge of science is reductionism, the breaking apart of nature into its natural constituents" (Wilson, 1998).

These two approaches have the potential to complement each other, by back-and-forth movements between detailed study of isolated components (reductionism) and overview of the system (holism). I think this process can be especially useful to study interaction networks, and link the observed network-wide patterns to physiological or behavioral constraints of species.

4.3 The structure of interaction networks

Several network patterns have attracted interest in the context of ecological network analysis. I describe a few of these patterns and their occurrence in the context of ecological networks below, but more exhaustive reviews can be found in Delmas et al. (2019), Landi et al. (2018) or Lau et al. (2017).

Some patterns describe properties at the whole network scale. One of these patterns is the connectance of the network, i.e. the proportion of realized interactions among all possible interactions. Notably, the seminal article by May (1972) yielded a theoretical result stating that the probability that a network will be stable decreases with its connectance, constrained by the interaction strengths. Although this work made a number of simplifying hypotheses, like considering a random network structure, it was the starting point for a new research topic investigating ecological network stability (Pimm, 1979; Thébault & Fontaine, 2010; Mougi & Kondoh, 2012; Landi et al., 2018).



(a) Modular (b) Nested

Figure 1.8: Matrix representations of perfectly modular (a) and nested (b) networks. Rows and columns of the matrices represent respectively the two types of nodes of a bipartite network, and non-null values in the matrix (dots) represent a link.

Two other network-scale patterns have been repeatedly highlighted in ecological networks: modularity and nestedness (see Figure 1.8) (Ings et al., 2009; Thébault & Fontaine, 2010). Modularity (see Figure 1.8a) quantifies the degree to which a given network can be described as subnetworks of distinct densely connected components, called modules. Real networks are rarely perfectly nested (i.e. composed of subnetworks that are completely disjoint and in which all nodes are connected), and several ways to quantify and detect imperfect modules have been proposed (Delmas et al., 2019). Modules have been found in particular in food webs (Krause et al., 2003; Ings et al., 2009), but also more recently in some mutualistic networks (Olesen et al., 2007).

Nestedness (see Figure 1.8b) is a measure used to describe bipartite networks (with two classes of nodes): a bipartite network is said to be nested when the interacting partners of specialist species are a subset of those of more generalist species (Bascompte et al., 2003). Real networks are rarely perfectly nested, and imperfect nestedness can be detected and quantified using several approaches (Landi et al., 2018). Nestedness has been highlighted in particular in mutualistic networks (Bascompte et al., 2003; Bascompte & Jordano, 2007; Thébault & Fontaine, 2010), but some degree of nestedness has also been shown in bipartite prey-predator food webs (Ings et al., 2009), although it has been found to be lower than in mutualistic networks (Bascompte et al., 2003).

Modularity and nestedness are notably linked to network stability: modularity has been found to increase stability of food webs, but to decrease stability of mutualistic networks (Thébault & Fontaine, 2010), while nestedness seems to destabilize food webs (Thébault & Fontaine, 2010) and has varying effects on the stability of mutualistic networks (Landi et al., 2018).

Network properties discovered in several real networks in the beginning of the 21st century (discussed in section 4.1 above) have also been studied in ecological networks. For instance, studies have investigated small-world properties in food webs (Montoya & Solé, 2002; Dunne et al., 2002) and mutualistic networks (Olesen et al., 2006). These studies have found that mutualistic networks tend to have stronger small-word properties than food webs (Olesen et al., 2006; Dunne et al., 2002); in food webs in particular, the clustering property necessary for small-world networks does not seem universal (Dunne et al., 2002). The degree distribution of ecological networks has also been studied. Regarding food webs, several studies have found degree distributions deviating from the Poisson distribution typically expected under Erdős-Rényi random networks (Landi et al., 2018; Camacho et al., 2002; Dunne et al., 2002). However, the actual degree distribution is debated; for instance, Montoya and Solé (2002) found a good fit with a scale-free distribution for three well-resolved food webs; Camacho et al. (2002) found a better fit to a distribution predicted by another model for seven food webs, and Dunne et al. (2002) found varying degree distributions for 16 food webs. Regarding mutualistic networks, Jordano et al. (2003) found that most of them have a truncated scale-free distribution.

Other network metrics center on a unique node (species). For instance, the node degree quantifies the number of links connected to this node, which can also be weighted (using links weights). In the context of ecological networks, the node degree quantifies the number of interaction partners of a species, possibly weighted by interaction frequencies. Weighted or unweighted degree has often been used to quantify species' specialization or generalization (Landi et al., 2018; Devictor et al., 2010). However, it has been criticized for being biased (Blüthgen et al., 2008) and other indices integrating species availability have been proposed (notably by Blüthgen et al., 2006). Another measure is the centrality of a node, defined as its "importance" (Delmas et al., 2019): in ecological networks, centrality is used in relation with the notion of keystone species (i.e. a species which disproportionately affects the system in relation to its abundance). Several centrality measures exist to quantify a node's importance: for instance, betweenness centrality quantifies the number of paths in the network that go through a given species (Delmas et al., 2019).

Finally, network metrics can be defined at an intermediate scale, for a subset of the network. Metrics defined at this scale notably include motifs, which describe the arrangements of links between a given number of nodes (Milo et al., 2002). Graph motifs are central to the concept of indirect interactions discussed before in section 3.3: indirect interactions depicted in Figure 1.4 can be seen as graphs motifs, and seeking such motifs in ecological networks can inform us about indirect interactions between species (Simmons et al., 2019). Another example of intermediate-scale metrics concerns bipartite networks, where it can be relevant to study the properties of one class of nodes, e.g. compare resources and consumers (Bascompte & Jordano, 2007): for instance, Blüthgen et al. (2006) compare the specialization of plants and of their pollinators.

The patterns described above are not necessarily static: indeed, ecological networks can change dynamically and spatially (Poisot et al., 2016; Tylianakis et al., 2008; CaraDonna et al., 2021). These changes can be driven by species turnover, due for instance to phenological or environmental constraints. They can also be due to a modification of interaction frequencies, irrespective of species abundances (Tylianakis & Morris, 2017). For instance, Laliberté and Tylianakis (2010) showed that host-parasitoids interaction networks in Ecuador were more homogenous in deforested than in forested sites: to explain these differences, they hypothesized that foraging efficiency could in-

crease in less complex (deforested) environments, thus increasing coupling between species abundances and interaction probabilities, compared to forested habitats. Such patterns of interaction rewiring, i.e. changes in interactions even in the absence of change of community composition, are increasingly recognized in interaction networks (CaraDonna et al., 2017; Martins et al., 2022).

5 Questions related to ecological networks

While early studies on ecological networks concentrated primarily on describing and characterizing their structure, recent research is beginning to address questions related to the underlying causes and broader ecological consequences of these structures. In this section, I describe how mechanisms leading to observed network structures have been investigated, and how the consequences of these structures on ecological communities have been studied.

5.1 Processes and models explaining network patterns

Several processes influence observed patterns in the structure of interaction networks (Vázquez, Blüthgen, et al., 2009; Benadi et al., 2022). Some of these patterns are discussed below and summarized in Figure 1.9.

First, species abundances are thought to play a prominent role by influencing the encounter, and hence the interaction probability between species (Poisot et al., 2015; Dormann et al., 2017). Additionally, as discussed above, these abundances can also vary in space and time, thus inducing spatio-temporal variations in networks structures (CaraDonna et al., 2021; Tylianakis & Morris, 2017).

Species traits may also influence interaction probabilities through trait matching, which posits that two species need to have compatible traits to interact, such as the proboscis length of a pollinator and the corolla depth that determine the likelihood of interaction (Peralta et al., 2024). Trait matching could emerge through trait adaptations between interacting partners, but could also be due to environmental pressures favoring certain sets of traits that incidentally favor species interactions (Dormann et al., 2017). The detection of trait matching from experimental interaction data is investigated in more detail in part 2.

These traits may also be partially constrained by the evolutionary history of species via phylogenetic inertia (Benadi et al., 2022; Dormann et al., 2017). Alternatively, phylogenetic relatedness may also indicate that species evolved with similar pressures, possibly adapting their traits to the same interaction partner or to each other reciprocally (Dormann et al., 2017). These mechanisms may explain why phylogenetically related species tend to display similar interaction patterns (Rezende et al., 2007).

Additionally, network structure may also be shaped by stability constraints: in particular, theoretical works have shown a stabilizing or disrupting effects of patterns like nestedness or modularity (Thébault & Fontaine, 2010; Okuyama & Holland, 2008; Krause et al., 2003). Note that the different mechanisms listed above are not necessarily independent, and might interact with each other to create the observed network structure (Vázquez, Blüthgen, et al., 2009).

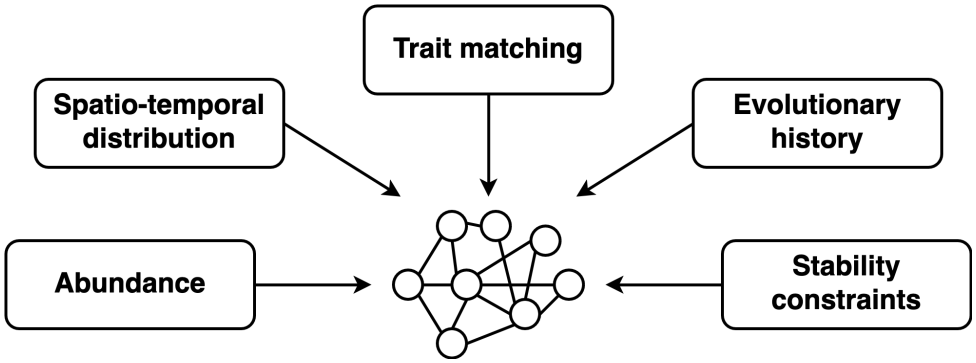


Figure 1.9: Different causes affecting the structure of ecological networks. Own figure inspired from Figure 1A in Peralta et al. (2024).

These factors can explain network patterns at different scales. For instance, network modules may emerge through co-evolution of species leading to increased interactions frequencies (Olesen et al., 2007; Dormann et al., 2017) and nestedness could emerge from power-law species abundances distributions (as shown by the analytical model of Araujo et al., 2010). At a more local scale, pairwise interaction frequencies between species are thought to be influenced by their traits (Fründ et al., 2016; Dehling et al., 2014). At the species level, traits have also been proposed to explain patterns of specialization or generalization (Dehling et al., 2016).

Several mechanistic models integrating these factors have been proposed to explain the assembly of ecological networks (Valdovinos, 2019; Eklöf et al., 2013). First, null models describe interaction frequencies as arising only from species respective abundances (Bartomeus et al., 2016; see for instance Araujo et al., 2010). Other

models integrate trait matching processes to constrain interaction probabilities (e.g. the niche model of Williams & Martinez, 2000, for food webs), often jointly with neutral abundance-based processes (Valdovinos, 2019). These models frequently incorporate additional constraints or mechanisms, like species phylogeny (Benadi et al., 2022), network sampling processes (Fründ et al., 2016) or spatial and temporal overlap of species (Vázquez, Chacoff, & Cagnolo, 2009). Other frameworks use optimal foraging theory to predict interaction frequencies (Petchey et al., 2008). These models have generally been found to reasonably explain network-level patterns, but still need to be improved to predict pairwise species interactions (Olito & Fox, 2015; Benadi et al., 2022), although food webs are generally better predicted, especially using body sizes (Benadi et al., 2022).

5.2 Consequences of network structure on the community

In addition to being shaped by ecological processes, network patterns are also thought to influence various aspects of ecological communities, including population dynamics, community stability, species evolution, ecosystem functions and behavior. The influence of interactions on these processes is discussed below.

The influence of interactions on population dynamics is very well known (Morin & Lawler, 1995), and has been extensively studied both theoretically and empirically. In particular, dynamical models like the Lotka-Volterra equations (Lotka, 1920; Volterra, 1926) have been used to predict oscillations in prey and predator abundances (Bengfort et al., 2017), and such dynamics have been observed in some ecological systems (e.g. Gilg et al., 2003, predicted oscillations in lemming and stoat abundances in a predators-prey system with five species).

A crucial question regards community stability, and the conditions under which interacting species coexist. The seminal work of Robert May (May, 1972) (see section 4.3) showed that strong interactions and a high network connectivity destabilize random networks. Other network structures have been shown to influence network stability, for instance nestedness (Thébault & Fontaine, 2010; Bastolla et al., 2009) and modularity (Thébault & Fontaine, 2010; Pimm, 1979). Interaction types and strengths may also affect the network stability (Mougi & Kondoh, 2012; Kéfi et al., 2016). Network structure can then influence species abundances and even cause species extinctions (Dormann

et al., 2017). Therefore, in addition to the influence of species abundances on network structure (discussed in the previous section), network structure also affects species abundances (this reciprocal relationship has been described as the “chicken-and-egg problem” of abundance and network structure; Dormann et al., 2017).

Interactions also have the potential to influence evolutionary processes in communities. For instance, Munoz et al. (2023) use a theoretical model showing that interspecific interactions can favor the evolution of functionally distinct species. Similarly to species abundances, evolutionary processes can both influence (e.g. through coevolution, discussed in the previous section) and be influenced by interactions.

Ecological interactions are also the backbone of many ecosystem functions, such as seed dispersal or pollination (Schleuning et al., 2015; Jordano, 2016b): interactions are related to these functions through the interaction efficacy (Bartomeus et al., 2016; Vázquez et al., 2015), which can be defined in several ways, but broadly speaking quantifies the efficacy of the interaction in realizing a specific process. For example, the interaction efficacy can be defined as the effect of the interaction on the per capita or total growth rate of an interacting partner (Berlow et al., 2004; Vázquez et al., 2015). Interaction efficacy is difficult to quantify experimentally: it is most often approximated assuming interaction frequency as a proxy (Vázquez et al. (2005) showed that it is a good approximation, if the efficacy is considered at the population scale), but has also been measured for instance using pollen removal or pollen deposition (Vázquez et al., 2015).

Finally, another question regards the degree to which interspecific interactions can influence the plastic responses of interacting species, either in response or in anticipation of a physical interaction (Palmer et al., 2022). For instance, Palmer and Packer (2021) highlighted a diversity of behavioral responses of zebra, wildebeest and impala to the simulated presence of their predators. Their study found that the type and intensity of response was modulated by the predator hunting style (e.g. zebra fled more from coursing than ambush predators). The response of species to interactions is discussed more extensively in part 3, where I investigate the short-term effect of interactions on species spatio-temporal distribution patterns.

6 Sampling and analyzing ecological networks

To address the questions above and assess their importance in real communities, among other things, we need to collect and analyze interaction data. However, accurate sampling of interspecific interactions in communities is often difficult, due to the number of interacting species and the spatial and temporal scales (Dormann et al., 2017). First, a lot of species are often involved in interaction networks, so sampling methods must be adapted to the potentially various characteristics of these species to detect them accurately. A diverse community may contain birds, trees, mosses, fungi and bacteria: if we want to study the interactions between these species, all groups must be accurately sampled. Some species might also be rare or elusive, which makes it even harder to detect those species and their interactions (Jordano, 2016a). Second, the spatial scale of communities may also make it difficult to sample the community accurately, either because it is very large (e.g. tropical rainforests) or difficult of access (e.g. marine communities), or on the contrary, because it is very small and difficult or impossible to observe directly (e.g. microbial communities). Similarly, the temporal scale at which communities dynamics happen can be very long, when studying long-lived species, or it can happen very fast, thus making it difficult to accurately capture the community dynamics (e.g. microbial communities) (CaraDonna et al., 2017). Because of these challenges, ecological networks are generally poorly characterized: the lack of knowledge pertaining to species interactions has been called the Eltonian shortfall (Hortal et al., 2015).

6.1 Sampling interactions

Despite these challenges, there are many methods to collect interaction data (see Table 1.1). These methods can record interspecific interactions directly, by observing physical interactions, or indirectly, using signs of interactions detected on species or species occurrences to infer interactions. Direct observation methods often involve scientists observing focal individuals and waiting to observe interactive behaviors. When the interactions involve plant and animals (pollination, seed dispersal), focal individuals are often plants (Dormann et al., 2017; Jordano, 2016a). For example, Albrecht et al. (2018) collected pollination interactions between insects and plants by doing transect

walks and recording all interactions of insects with plants, defined as all observations of an insect touching a plant's reproductive part. Interactions can also be assessed by automated methods, as the study of Melidonis and Peter (2015), who documented the first instance of rodent pollination on the plant species *Protea foliosa* by using camera traps, i.e. cameras that trigger automatically when movement is detected, focused on *P. foliosa* individuals. Indirect methods to record interactions are very diverse: for example, a popular method to assess trophic interactions is to analyze fecal or stomach content to assess species' diet (Miranda et al., 2013), either by visual examination or using molecular techniques relying on antibodies or DNA (Sheppard & Harwood, 2005). Similarly, it is possible to assess pollination interactions by analyzing pollen grains deposited on pollinators, either visually, using microscopy (possibly assisted with machine learning tools for automated identification) or DNA metabarcoding (Lowe et al., 2022).

Other indirect methods to detect interactions rely on species occurrences, i.e. species detections at a particular time and place. These methods aim at inferring interactions from species occurrences: the rationale is that the observed patterns in species occurrences are due to underlying ecological interactions. For instance, two species that never co-occur can be inferred to have a negative interaction. This idea was popularized by Jared Diamond (J. M. Diamond, 1975): he observed that some bird species in the Bismarck Archipelago occurred only on some islands. These islands were geographically interspersed, forming a checkerboard pattern, and he interpreted this as evidence of interspecific competition. There are diverse data collection methods to gather species occurrence data. Classical methods involve visual observation on the field: for instance, transect consists in moving along a path and recording species seen along this path, and quadrat sampling consists in counting species within defined areas (quadrats) on a given landscape. Automated methods can also be used, such as camera traps, taking pictures or videos, or acoustic telemetry transceivers, allowing to collect movement data from tagged animals. In microbial communities, sequencing approaches have been used to assess species' relative abundances and co-occurrences (Faust & Raes, 2012). More recently, opportunistic data collected via participatory science programs (for example via platforms like iNaturalist) arose as an enormous and promising source of species occurrence data (Binley & Bennett, 2023).

Type	Observation method	References
Direct methods	Direct observation	Albrecht et al. (2018)
	Automated sensors	Melidonis and Peter (2015)
Indirect methods	Fecal or stomach content	Miranda et al. (2013)
	Pollen load of pollinators	Lowe et al. (2022)
	Species occurrences	
	Direct observation	J. M. Diamond (1975)
	Transects	Bao et al. (2017)
	Quadrats	Raath-Krüger et al. (2021)
	Automated sensors	Anderson et al. (2016)
	Sequencing	Faust and Raes (2012)
	Participatory science	Binley and Bennett (2023)

Table 1.1: Summary of some direct and indirect methods to sample interaction networks.

6.2 Analyzing and inferring interactions

This diversity of data collection processes require different models to analyze data and draw conclusions on the interactions. Indeed, each sampling method has specific characteristics and biases that must be accounted for during the analysis. Using adapted models can be useful to mitigate biases, and make the most out of each data type. For example, direct observation methods can suffer bias in the detection of interactions between rare species (Jordano, 2016a) that can greatly affect ecological data interpretation (Blüthgen et al., 2008; Vázquez, Blüthgen, et al., 2009). Opportunistic datasets of species occurrences, collected for example via participatory science, are not homogeneous in space, as sampling intensity is more intense in highly populated or easily accessible areas (Binley & Bennett, 2023). All data collection methods are not equivalent: therefore, they require adapted modeling techniques to be analyzed and provide information on the processes underlying the observed patterns. For instance, statistical approaches exist to account for species detection biases in ecological networks (Jordano, 2016a; Blüthgen et al., 2008) or spatial biases in opportunistic data (Fink et al., 2014).

Among all approaches used to sample species interactions in communities, assessing interactions from occurrence data requires particular attention. Indeed, this approach makes strong hypotheses on the influence of interactions on species dis-

tribution and/or abundance, and particular care is needed when designing sampling and analyzing data (Blanchet et al., 2020). One of the challenges to infer interactions from occurrence data is to choose the spatial scale at which co-occurrences are considered: for example, in a simulation study, Araújo and Rozenfeld (2014) showed that the co-occurrence signal produced by an interaction is not necessarily conserved across scales, depending on the sign of the interaction. Moreover, to infer interactions from occurrence data, sophisticated models are usually required to account for all factors affecting species occurrences besides interactions (Blanchet et al., 2020). For instance, Diamond's foundational idea itself (J. M. Diamond, 1975) led to a considerable debate in the following years, notably regarding his modeling approach and the absence of null model to generate occurrence patterns expected by chance alone (Connor & Simberloff, 1979; Connor et al., 2013; J. Diamond et al., 2015; Connor et al., 2015). Another illustration of these modeling tools is the use of the precision matrix, allowing to tease apart occurrences affected by direct and indirect interactions (Harris, 2016; Popovic et al., 2019; Chiquet et al., 2021). For the reasons described above, several authors have advised to treat co-occurrence data with caution when inferring interactions (Blanchet et al., 2020; Friedman & Alm, 2012).

Another challenge pertaining to all analyses described above regards the data analysis process itself. Here by data analysis, I refer to the stage conduced once data has been collected and the analysis method chosen, and before the interpretation. Data analysis includes formatting data, inputting data in the model and producing figures and summary statistics. Today, these analysis steps are often conduced on a computer, and ecology increasingly relies on programming and coding (Filazzola & Lortie, 2022; Markowetz, 2017). Several challenges come with these computing tools. First, computer code can be considered a product of the research itself, and, as such, the same standards of integrity, data transparency and openness apply (O'Dea et al., 2021). When conducted with a coding language, these analyses steps can easily be saved to a file and shared with other materials for the publication; in fact, more and more journals require or encourage authors to share their code. Second, computing tools allow more and more sophisticated models to be built (Warton et al., 2015; King, 2014), which opens great opportunities for data analyses but also makes room for errors (MacFadyen et al., 2022). In this context, using reproducible and readable

data analyses pipelines is important to prevent data analysis errors (Filazzola & Lortie, 2022). Third, ecology also enters an era of big data, with tools such as DNA metabarcoding or camera trap data producing tremendous amounts of data. As the amount of data grows, it becomes increasingly essential to automate steps of data processing (MacFadyen et al., 2022), especially when integrating multiple data sources that require an additional step of data standardization (Ronquillo et al., 2024).

7 Thesis objectives

This thesis aims at advancing our understanding of ecological communities, and in particular of ecological interactions inside ecological communities. To do so, I undertake a methodological approach and develop statistical or software tools to analyze ecological networks and ecological community data. In each part, presented below, I address different scientific questions and develop methods tailored to these questions and adapted to specific data collection processes. Each part begins with a short introductory chapter presenting its methods and the question, and parts 2 and 3 also have a dedicated discussion chapter. The bibliography is presented separately for each part.

In part 2, I investigate the causes of observed patterns in ecological networks. More specifically, I use and develop multivariate methods to investigate trait matching as a driver of the observed patterns of interactions. In this part, I quantify interaction patterns using the notion of interaction niche (see the introduction chapter 7). In chapter 2.1, I use correspondence analysis and reciprocal scaling to measure species' interaction niches breadth and optima. In this chapter, I do not use species traits per se, but latent traits inferred with correspondence analysis. I intend to submit the results of this chapter for publication in *PCI Ecology*. In chapter 2.2, I use constrained multivariate analyses to include species traits. In this chapter, I quantify species' interaction niche optima taking into account measured traits, and I also partition the total variance of the network. I intend to use the results of this chapter for an upcoming publication. In chapter 2.3, I extend the constrained analyses used in the preceding chapter to quantify niche breadth in addition to niche optima quantified in the preceding chapter. To illustrate the analyses along part 2, I use a bird-fruit interaction network sampled in the Peruvian montane forest (data from Dehling et al., 2021).

In part 3, I investigate the consequences of interspecific interactions on species' plastic response. In particular, I investigate how different animal species respond to the presence of other species by modifying their spatio-temporal distribution patterns at short spatio-temporal scales. In chapter 3.1, I use a multivariate point process model to infer attraction and avoidance between species, from data collected in continuous time using camera traps. This chapter has been published as an article in *Ecology* (Nicvert et al., 2024). In chapter 3.2, I use a multivariate point process model allowing to incorporate temporal covariates to tease apart species' circadian activities from the attraction-avoidance due to other species. In this part, I illustrate the methods on a camera trap dataset with five mammal species from the South Africa savanna (data available at Nicvert et al., 2023), collected with the Snapshot Safari project (Pardo et al., 2021).

In part 4, I present two software tools that I developed to automate data analyses. These tools aim at making data analysis easier and more reproducible when processing species occurrence data obtained with camera traps. Chapter 4.1 describes a R package to standardize and clean camera trap records, targeted to meet the needs of data from the Snapshot Safari project. Chapter 4.2 presents a Shiny application and R package to visualize camera trap data, aiming at a wider audience of camera trap data users.

Part	Broad question	Objectives	Data used	Tool
2	Which causes affect network structure?	Measure the importance of trait matching	Direct observation	Multivariate methods
3	What are the consequences of interactions in communities?	Infer attraction or avoidance between species at short spatio-temporal scales	Camera trap	Multivariate point process models
4	How to improve the reliability and reproducibility of scientific analyses?	Design software tools for data processing and analysis	Camera trap	R packages and Shiny application

Table 1.2: Thesis objectives for each part. This table summarizes the broad scientific question, the more precise objectives, the source of data and the method used in each part of the thesis.

Bibliography

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*(1), 47–97. https://doi.org/10.1103/RevModPhys.74.47

Albrecht, J., Classen, A., Vollstädt, M. G. R., Mayr, A., Mollel, N. P., Schellenberger Costa, D., Dulle, H. I., Fischer, M., Hemp, A., Howell, K. M., Kleyer, M., Nauss, T., Peters, M. K., Tschapka, M., Steffan-Dewenter, I., Böhning-Gaese, K., & Schleuning, M. (2018). Plant and animal functional diversity drive mutualistic network assembly across an elevational gradient. *Nature Communications*, *9*(1), 3177. https://doi.org/10.1038/s41467-018-05610-w

Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences*, *97*(21), 11149–11152. https://doi.org/10.1073/pnas.200327197

Anderson, T. M., Hepler, S. A., Holdo, R. M., Donaldson, J. E., Erhardt, R. J., Hopcraft, J. G. C., Hutchinson, M. C., Huebner, S. E., Morrison, T. A., Muday, J., Munuo, I. N., Palmer, M. S., Pansu, J., Pringle, R. M., Sketch, R., & Packer, C. (2024). Interplay of competition and facilitation in grazing succession by migrant Serengeti herbivores. *Science*, *383*(6684), 782–788. https://doi.org/10.1126/science.adg0744

Anderson, T. M., White, S., Davis, B., Erhardt, R., Palmer, M., Swanson, A., Kosmala, M., & Packer, C. (2016). The spatial distribution of African savannah herbivores: Species associations and habitat occupancy in a landscape context. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1703), 20150314. https://doi.org/10.1098/rstb.2015.0314

Araujo, A. I. L., de Almeida, A. M., Cardoso, M. Z., & Corso, G. (2010). Abundance and nestedness in interaction networks. *Ecological Complexity*, *7*(4), 494–499. https://doi.org/10.1016/j.ecocom.2010.02.004

Araújo, M. B., & Rozenfeld, A. (2014). The geographic scaling of biotic interactions. *Ecography*, *37*(5), 406–415. https://doi.org/10.1111/j.1600-0587.2013.00643.x

Archer, S. K., Stoner, E. W., & Layman, C. A. (2015). A complex interaction between a sponge (*Halichondria melanadocia*) and a seagrass (*Thalassia testudinum*) in a subtropical coastal ecosystem. *Journal of Experimental Marine Biology and Ecology*, *465*, 33–40. https://doi.org/10.1016/j.jembe.2015.01.003

Banf, M., & Rhee, S. Y. (2017). Computational inference of gene regulatory networks: Approaches, limitations and opportunities. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1860*(1), 41–52. https://doi.org/10.1016/j.bbagrm.2016.09.003

Bao, H., Fryxell, J. M., Liu, H., Dou, H., Ma, Y., & Jiang, G. (2017). Effects of interspecific interaction-linked habitat factors on moose resource selection and environmental stress. *Scientific Reports*, *7*(1), 41514. https://doi.org/10.1038/srep41514

Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, *286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Barabási, A.-L., & Pósfai, M. (2016). *Network science* [OCLC: ocn910772793]. Cambridge University Press. http://networksciencebook.com/

Barbault, R. (2008). *Écologie générale: Structure et fonctionnement de la biosphère* (6e éd). Dunod.

Bartomeus, I., Gravel, D., Tylianakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Functional Ecology*, *30*(12), 1894–1903. https://doi.org/10.1111/1365-2435.12666

Bascompte, J., & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 567–593. https://doi.org/10.1146/annurev.ecolsys.38.091206.095818

Bascompte, J., Jordano, P., Melián, C. J., & Olesen, J. M. (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, *100*(16), 9383–9387. https://doi.org/10.1073/pnas.1633576100

Bascompte, J., Melián, C. J., & Sala, E. (2005). Interaction strength combinations and the overfishing of a marine food web. *Proceedings of the National Academy of Sciences*, *102*(15), 5443–5447. https://doi.org/10.1073/pnas.0501562102

Bastolla, U., Fortuna, M. A., Pascual-García, A., Ferrera, A., Luque, B., & Bascompte, J. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, *458*(7241), 1018–1020. https://doi.org/10.1038/nature07950

Benadi, G., Dormann, C. F., Fründ, J., Stephan, R., & Vázquez, D. P. (2022). Quantitative Prediction of Interactions in Bipartite Networks Based on Traits, Abundances, and Phylogeny. *The American Naturalist*, *199*(6), 841–854. https://doi.org/10.1086/714420

Bengfort, M., van Velzen, E., & Gaedke, U. (2017). Slight phenotypic variation in predators and prey causes complex predator-prey oscillations. *Ecological Complexity*, *31*, 115–124. https://doi.org/10.1016/j.ecocom.2017.06.003

Berlow, E. L., Neutel, A.-M., Cohen, J. E., De Ruiter, P. C., Ebenman, B., Emmerson, M., Fox, J. W., Jansen, V. A. A., Jones, J. I., Kokkoris, G. D., Logofet, D. O., McKane, A. J., Montoya, J. M., & Petchey, O. (2004). Interaction Strengths in Food Webs: Issues and Opportunities. *Journal of Animal Ecology*, *73*(3), 585–598. Retrieved April 23, 2024, from https://www.jstor.org/stable/3505669

Betts, M. G., Hadley, A. S., Frey, D. W., Frey, S. J. K., Gannon, D., Harris, S. H., Kim, H., Kormann, U. G., Leimberger, K., Moriarty, K., Northrup, J. M., Phalan, B., Rousseau, J. S., Stokely, T. D., Valente, J. J., Wolf, C., & Zárrate-Charry, D. (2021). When are hypotheses useful in ecology and evolution? *Ecology and Evolution*, *11*(11), 5762–5776. https://doi.org/10.1002/ece3.7365

Binley, A. D., & Bennett, J. R. (2023). The data double standard. *Methods in Ecology and Evolution*, *14*(6), 1389–1397. https://doi.org/10.1111/2041-210X.14110

Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, *23*(7), 1050–1063. https://doi.org/10.1111/ele.13525

Blüthgen, N., Fründ, J., Vázquez, D. P., & Menzel, F. (2008). What do interaction network metrics tell us about specialization and biological traits. *Ecology*, *89*(12), 3387–3399. https://doi.org/10.1890/07-2121.1

Blüthgen, N., Menzel, F., & Blüthgen, N. (2006). Measuring specialization in species interaction networks. *BMC Ecology*, *6*(1), 1–12. https://doi.org/10.1186/1472-6785-6-9

British Ecological Society. (2024). What is ecology? Retrieved April 19, 2024, from https://www.britishecologicalsociety.org/about/what-is-ecology/

Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, *10*, 1017. https://doi.org/10.1038/s41467-019-08746-5

Bronstein, J. L., Barker, J. L., Lichtenberg, E. M., Richardson, L. L., & Irwin, R. E. (2017). The behavioral ecology of nectar robbing: Why be tactic constant? *Current Opinion in Insect Science*, *21*, 14–18. https://doi.org/10.1016/j.cois.2017.05.013

Camacho, J., Guimerà, R., & Nunes Amaral, L. A. (2002). Robust Patterns in Food Web Structure. *Physical Review Letters*, *88*(22), 228102. https://doi.org/10.1103/PhysRevLett.88.228102

CaraDonna, P. J., Burkle, L. A., Schwarz, B., Resasco, J., Knight, T. M., Benadi, G., Blüthgen, N., Dormann, C. F., Fang, Q., Fründ, J., Gauzens, B., Kaiser-Bunbury, C. N., Winfree, R., & Vázquez, D. P. (2021). Seeing through the static: The temporal dimension of plant–animal mutualistic interactions. *Ecology Letters*, *24*(1), 149–161. https://doi.org/https://doi.org/10.1111/ele.13623

CaraDonna, P. J., Petry, W. K., Brennan, R. M., Cunningham, J. L., Bronstein, J. L., Waser, N. M., & Sanders, N. J. (2017). Interaction rewiring and the rapid turnover of plant–pollinator networks. *Ecology Letters*, *20*(3), 385–394. https://doi.org/10.1111/ele.12740

Chiquet, J., Mariadassou, M., & Robin, S. (2021). The Poisson-Lognormal Model as a Versatile Framework for the Joint Analysis of Species Abundances. *Frontiers in Ecology and Evolution*, *9*, 188. https://doi.org/10.3389/fevo.2021.588292

Clements, F. E. (1936). Nature and Structure of the Climax. *Journal of Ecology*, *24*(1), 252–284. https://doi.org/10.2307/2256278

Collins English Dictionary. (2024, April). Interaction. Retrieved April 28, 2024, from https://www.collinsdictionary.com/dictionary/english/interaction

Connor, E. F., Collins, M. D., & Simberloff, D. (2013). The checkered history of checkerboard distributions. *Ecology*, *94*(11), 2403–2414. https://doi.org/10.1890/12-1471.1

Connor, E. F., Collins, M. D., & Simberloff, D. (2015). The checkered history of checkerboard distributions: Reply. *Ecology*, *96*(12), 3388–3389. https://doi.org/10.1890/15-1034.1

Connor, E. F., & Simberloff, D. (1979). The Assembly of Species Communities: Chance or Competition? *Ecology*, *60*(6), 1132–1140. https://doi.org/10.2307/1936961

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* (1st ed.). John Murray.

Dehling, D. M., Bender, I. M. A., Blendinger, P. G., Muñoz, M. C., Quitián, M., Saavedra, F., Santillán, V., Böhning-Gaese, K., Neuschulz, E.-L., & Schleuning, M. (2021). ANDEAN frugivory: Data on plant–bird interactions and functional traits of plant and bird species from montane forests along the Andes. https://doi.org/10.5061/DRYAD.WM37PVMN5

Dehling, D. M., Jordano, P., Schaefer, H. M., Böhning-Gaese, K., & Schleuning, M. (2016). Morphology predicts species' functional roles and their degree of specialization in plant–frugivore interactions. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1823), 20152444. https://doi.org/10.1098/rspb.2015.2444

Dehling, D. M., Töpfer, T., Schaefer, H. M., Jordano, P., Böhning-Gaese, K., & Schleuning, M. (2014). Functional relationships beyond species richness patterns: Trait matching in plant–bird mutualisms across scales. *Global Ecology and Biogeography*, *23*(10), 1085–1093. https://doi.org/10.1111/geb.12193

Delmas, E., Besson, M., Brice, M.-H., Burkle, L. A., Dalla Riva, G. V., Fortin, M.-J., Gravel, D., Guimarães Jr., P. R., Hembry, D. H., Newman, E. A., Olesen, J. M., Pires, M. M., Yeakel, J. D., & Poisot, T.

(2019). Analysing ecological networks of species interactions. *Biological Reviews*, *94*(1), 16–36. https://doi.org/10.1111/brv.12433

Devictor, V., Clavel, J., Julliard, R., Lavergne, S., Mouillot, D., Thuiller, W., Venail, P., Villéger, S., & Mouquet, N. (2010). Defining and measuring ecological specialization. *Journal of Applied Ecology*, *47*(1), 15–25. https://doi.org/10.1111/j.1365-2664.2009.01744.x

Diamond, J., Pimm, S. L., & Sanderson, J. G. (2015). The checkered history of checkerboard distributions: Comment. *Ecology*, *96*(12), 3386–3388. https://doi.org/10.1890/14-1848.1

Diamond, J. M. (1975). Assembly of species communities. In *Ecology and evolution of communities* (pp. 342–444). Harvard Univ. Press.

Ding, T., & Schloss, P. D. (2014). Dynamics and associations of microbial community types across the human body. *Nature*, *509*(7500), 357–360. https://doi.org/10.1038/nature13178

Dormann, C. F., Fründ, J., & Schaefer, H. M. (2017). Identifying Causes of Patterns in Ecological Networks: Opportunities and Limitations. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 559–584. https://doi.org/10.1146/annurev-ecolsys-110316-022928

Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences*, *99*(20), 12917–12922. https://doi.org/10.1073/pnas.192407699

Egler, F. E. (1986). "Physics Envy" in Ecology. *Bulletin of the Ecological Society of America*, *67*(3), 233–235. Retrieved May 6, 2024, from https://www.jstor.org/stable/20166525

Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N. P., Dalsgaard, B., de Sassi, C., Galetti, M., Guimarães, P. R., Lomáscolo, S. B., Martín González, A. M., Pizo, M. A., Rader, R., Rodrigo, A., Tylianakis, J. M., Vázquez, D. P., & Allesina, S. (2013). The dimensionality of ecological networks. *Ecology Letters*, *16*(5), 577–583. https://doi.org/10.1111/ele.12081

Elton, C. (1927). The animal community. In *Animal Ecology*. Sidgwick; Jackson.

Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, *6*(3-4), 290–297. https://doi.org/10.5486/PMD.1959.6.3-4.12

Estes, J. A., Terborgh, J., Brashares, J. S., Power, M. E., Berger, J., Bond, W. J., Carpenter, S. R., Essington, T. E., Holt, R. D., Jackson, J. B. C., Marquis, R. J., Oksanen, L., Oksanen, T., Paine, R. T., Pikitch, E. K., Ripple, W. J., Sandin, S. A., Scheffer, M., Schoener, T. W., … Wardle, D. A. (2011). Trophic Downgrading of Planet Earth. *Science*, *333*(6040), 301–306. https://doi.org/10.1126/science.1205106

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 128–140. https://scholarlycommons.pacific.edu/euler-works/53

Fang, Y., & Xiong, L. (2015). General mechanisms of drought response and their application in drought resistance improvement in plants. *Cellular and Molecular Life Sciences*, *72*(4), 673–689. https://doi.org/10.1007/s00018-014-1767-0

Faust, K., & Raes, J. (2012). Microbial interactions: From networks to models. *Nature Reviews Microbiology*, *10*(8), 538–550. https://doi.org/10.1038/nrmicro2832

Filazzola, A., & Lortie, C. (2022). A call for clean code to effectively communicate science. *Methods in Ecology and Evolution*, *13*(10), 2119–2128. https://doi.org/10.1111/2041-210X.13961

Fink, D., Damoulas, T., Bruns, N. E., La Sorte, F. A., Hochachka, W. M., Gomes, C. P., & Kelling, S. (2014). Crowdsourcing Meets Ecology: Hemispherewide Spatiotemporal Species Distribution Models. *AI Magazine*, *35*(2), 19–30. https://doi.org/10.1609/aimag.v35i2.2533

Fletcher, R. J., Revell, A., Reichert, B. E., Kitchens, W. M., Dixon, J. D., & Austin, J. D. (2013). Network modularity reveals critical scales for connectivity in ecology and evolution. *Nature Communications*, *4*(1), 2572. https://doi.org/10.1038/ncomms3572

Friedman, J., & Alm, E. J. (2012). Inferring Correlation Networks from Genomic Survey Data. *PLoS Computational Biology*, *8*(9). https://doi.org/10.1371/journal.pcbi.1002687

Fründ, J., McCann, K. S., & Williams, N. M. (2016). Sampling bias is a challenge for quantifying specialization and network structure: Lessons from a quantitative niche model. *Oikos*, *125*(4), 502–513. https://doi.org/10.1111/oik.02256

Gayford, J. H. (2024). The multidimensional spectrum of eco-evolutionary relationships between sharks and remoras. *Journal of Fish Biology*, *n/a*(n/a). https://doi.org/10.1111/jfb.15759

Gilbert, E. N. (1959). Random Graphs. *The Annals of Mathematical Statistics*, *30*(4), 1141–1144. https://doi.org/10.1214/aoms/1177706098

Gilg, O., Hanski, I., & Sittler, B. (2003). Cyclic Dynamics in a Simple Vertebrate Predator-Prey Community. *Science*, *302*(5646), 866–868. https://doi.org/10.1126/science.1087509

Gleason, H. A. (1926). The Individualistic Concept of the Plant Association. *Bulletin of the Torrey Botanical Club*, *53*(1), 7–26. https://doi.org/10.2307/2479933

Gosak, M., Markovič, R., Dolenšek, J., Slak Rupnik, M., Marhl, M., Stožer, A., & Perc, M. (2018). Network science of biological systems at different scales: A review. *Physics of Life Reviews*, *24*, 118–135. https://doi.org/10.1016/j.plrev.2017.11.003

Grether, G. F., Anderson, C. N., Drury, J. P., Kirschel, A. N. G., Losin, N., Okamoto, K., & Peiman, K. S. (2013). The evolutionary consequences of interspecific aggression. *Annals of the New York Academy of Sciences*, *1289*(1), 48–68. https://doi.org/10.1111/nyas.12082

Guo, Q., Wu, X., Korpelainen, H., & Li, C. (2020). Stronger intra-specific competition aggravates negative effects of drought on the growth of *Cunninghamia lanceolata*. *Environmental and Experimental Botany*, *175*, 104042. https://doi.org/10.1016/j.envexpbot.2020.104042

Harris, D. J. (2016). Inferring species interactions from co-occurrence data with Markov networks. *Ecology*, *97*(12), 2562–2569.

Harrold, C., & Reed, D. C. (1985). Food Availability, Sea Urchin Grazing, and Kelp Forest Community Structure. *Ecology*, *66*(4), 1160–1169. https://doi.org/10.2307/1939168

Hayward, M. W., & Kerley, G. I. (2008). Prey preferences and dietary overlap amongst Africa's large predators. *African Journal of Wildlife Research*, *38*(2), 93–108. https://doi.org/10.3957/0379-4369-38.2.93

Heck, K. L., Carruthers, T. J. B., Duarte, C. M., Hughes, A. R., Kendrick, G., Orth, R. J., & Williams, S. W. (2008). Trophic Transfers from Seagrass Meadows Subsidize Diverse Marine and Terrestrial Consumers. *Ecosystems*, *11*(7), 1198–1210. https://doi.org/10.1007/s10021-008-9155-y

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.

Humboldt, A. v., Bonpland, A., Jackson, S. T., & Romanowski, S. (2008). *Essay on the geography of plants* [OCLC: ocn247962343]. University of Chicago Press.

Huryn, A. D. (1998). Ecosystem-level evidence for top-down and bottom-up control of production in a grassland stream system. *Oecologia*, *115*(1), 173–183. https://doi.org/10.1007/s004420050505

Ichihashi, R., & Tateno, M. (2011). Strategies to balance between light acquisition and the risk of falls of four temperate liana species: To overtop host canopies or not? *Journal of Ecology*, *99*(4), 1071–1080. https://doi.org/10.1111/j.1365-2745.2011.01808.x

Ings, T. C., Montoya, J. M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C. F., Edwards, F., Figueroa, D., Jacob, U., Jones, J. I., Lauridsen, R. B., Ledger, M. E., Lewis, H. M., Olesen, J. M., Van Veen, F. F., Warren, P. H., & Woodward, G. (2009). Ecological networks – beyond food webs. *Journal of Animal Ecology*, *78*(1), 253–269. https://doi.org/10.1111/j.1365-2656.2008.01460.x

Ito, F., & Higashi, S. (1991). An Indirect Mutualism between Oaks and Wood Ants Via Aphids. *Journal of Animal Ecology*, *60*(2), 463–470. https://doi.org/10.2307/5291

Johnson, S. D., & Bond, W. J. (1992). Habitat dependent pollination success in a Cape orchid. *Oecologia*, *91*(3), 455–456. https://doi.org/10.1007/BF00317637

Jordano, P. (2016a). Sampling networks of ecological interactions. *Functional Ecology*, *30*(12), 1883–1893. https://doi.org/10.1111/1365-2435.12763

Jordano, P. (2016b). Chasing Ecological Interactions. *PLOS Biology*, *14*(9), e1002559. https://doi.org/10.1371/journal.pbio.1002559

Jordano, P., Bascompte, J., & Olesen, J. M. (2003). Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters*, *6*(1), 69–81. https://doi.org/10.1046/j.1461-0248.2003.00403.x

Kaplan, I., & Denno, R. F. (2007). Interspecific interactions in phytophagous insects revisited: A quantitative assessment of competition theory. *Ecology Letters*, *10*(10), 977–994. https://doi.org/10.1111/j.1461-0248.2007.01093.x

Kéfi, S., Miele, V., Wieters, E. A., Navarrete, S. A., & Berlow, E. L. (2016). How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased Persistence and Resilience. *PLOS Biology*, *14*(8), e1002527. https://doi.org/10.1371/journal.pbio.1002527

King, R. (2014). Statistical Ecology. *Annual Review of Statistics and Its Application*, *1*(Volume 1, 2014), 401–426. https://doi.org/10.1146/annurev-statistics-022513-115633

Kjær, U., Olsen, S. L., & Klanderud, K. (2018). Shift from facilitative to neutral interactions by the cushion plant Silene acaulis along a primary succession gradient. *Journal of Vegetation Science*, *29*(1), 42–51. https://doi.org/10.1111/jvs.12584

Krause, A. E., Frank, K. A., Mason, D. M., Ulanowicz, R. E., & Taylor, W. W. (2003). Compartments revealed in food-web structure. *Nature*, *426*(6964), 282–285. https://doi.org/10.1038/nature02115

Kuijper, D. P. J., Diserens, T. A., Say-Sallaz, E., Kasper, K., Szafrańska, P. A., Szewczyk, M., Stępniak, K. M., & Churski, M. (2024). Wolves recolonize novel ecosystems leading to novel interactions. *Journal of Applied Ecology*, *61*(5), 906–921. https://doi.org/10.1111/1365-2664.14602

Laliberté, E., & Tylianakis, J. M. (2010). Deforestation homogenizes tropical parasitoid–host networks. *Ecology*, *91*(6), 1740–1747. https://doi.org/10.1890/09-1328.1

Landi, P., Minoarivelo, H. O., Brännström, Å., Hui, C., & Dieckmann, U. (2018). Complexity and stability of ecological networks: A review of the theory. *Population Ecology*, *60*(4), 319–345. https://doi.org/10.1007/s10144-018-0628-3

Lau, M. K., Borrett, S. R., Baiser, B., Gotelli, N. J., & Ellison, A. M. (2017). Ecological network metrics: Opportunities for synthesis. *Ecosphere*, *8*(8), e01900. https://doi.org/10.1002/ecs2.1900

Lawton, J. H. (1999). Are There General Laws in Ecology? *Oikos*, *84*(2), 177–192. https://doi.org/10.2307/3546712

Lévêque, C. (2001). *Ecologie : De l'écosystème à la biosphère*. Dunod.

Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, *73*(6), 1943–1967. https://doi.org/10.2307/1941447

Levine, J. M., Bascompte, J., Adler, P. B., & Allesina, S. (2017). Beyond pairwise mechanisms of species coexistence in complex communities. *Nature*, *546*(7656), 56–64. https://doi.org/10.1038/nature22898

Lotka, A. J. (1920). Analytical Note on Certain Rhythmic Relations in Organic Systems. *Proceedings of the National Academy of Sciences of the United States of America*, *6*(7), 410–415. Retrieved April 27, 2024, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1084562/

Lowe, A., Jones, L., Witter, L., Creer, S., & de Vere, N. (2022). Using DNA Metabarcoding to Identify Floral Visitation by Pollinators. *Diversity*, *14*(4), 236. https://doi.org/10.3390/d14040236

MacFadyen, S., Allsopp, N., Altwegg, R., Archibald, S., Botha, J., Bradshaw, K., Carruthers, J., De Klerk, H., de Vos, A., Distiller, G., Foord, S., Freitag-Ronaldson, S., Gibbs, R., Hamer, M., Landi, P., MacFadyen, D., Manuel, J., Midgley, G., Moncrieff, G., … Hui, C. (2022). Drowning in data, thirsty for information and starved for understanding: A biodiversity information hub for cooperative environmental monitoring in South Africa. *Biological Conservation*, *274*, 109736. https://doi.org/10.1016/j.biocon.2022.109736

Markowetz, F. (2017). All biology is computational biology. *PLOS Biology*, *15*(3), e2002050. https://doi.org/10.1371/journal.pbio.2002050

Martins, L. P., Stouffer, D. B., Blendinger, P. G., Böhning-Gaese, K., Buitrón-Jurado, G., Correia, M., Costa, J. M., Dehling, D. M., Donatti, C. I., Emer, C., Galetti, M., Heleno, R., Jordano, P., Menezes, Í., Morante-Filho, J. C., Muñoz, M. C., Neuschulz, E. L., Pizo, M. A., Quitián, M., … Tylianakis, J. M. (2022). Global and regional ecological boundaries explain abrupt spatial discontinuities in avian frugivory interactions. *Nature Communications*, *13*(1), 6943. https://doi.org/10.1038/s41467-022-34355-w

Matagne, P. (2002). *Comprendre l'écologie et son histoire: Les origines, les fondateurs et l'évolution d'une science*. Delachaux et Niestlé.

Mathis, K. A., & Bronstein, J. L. (2020). Our Current Understanding of Commensalism. *Annual Review of Ecology, Evolution, and Systematics*, *51*(1), 167–189. https://doi.org/10.1146/annurev-ecolsys-011720-040844

May, R. (1972). Will a large complex system be stable? *Nature*. Retrieved October 6, 2020, from https://www-nature-com.docelec.univ-lyon1.fr/articles/238413a0.pdf?origin=ppub

McGill, B., Enquist, B., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, *21*(4), 178–185. https://doi.org/10.1016/j.tree.2006.02.002

McIntosh, A. R., & Townsend, C. R. (1996). Interactions between fish, grazing invertebrates and algae in a New Zealand stream: A trophic cascade mediated by fish-induced changes to grazer behaviour? *Oecologia*, *108*(1), 174–181. https://doi.org/10.1007/BF00333229

Melián, C. J., Bascompte, J., Jordano, P., & Krivan, V. (2009). Diversity in a complex ecological network with two interaction types. *Oikos*, *118*(1), 122–130. https://doi.org/10.1111/j.1600-0706.2008.16751.x

Melidonis, C. A., & Peter, C. I. (2015). Diurnal pollination, primarily by a single species of rodent, documented in *Protea foliosa* using modified camera traps. *South African Journal of Botany*, *97*, 9–15. https://doi.org/10.1016/j.sajb.2014.12.009

Memmott, J. (1999). The structure of a plant-pollinator food web. *Ecology Letters*, *2*(5), 276–280. https://doi.org/10.1046/j.1461-0248.1999.00087.x

Ménard, R., Tavares, J., Cockburn, I., Markus, M., Zavala, F., & Amino, R. (2013). Looking under the skin: The first steps in malarial infection and immunity. *Nature Reviews Microbiology*, *11*(10), 701–712. https://doi.org/10.1038/nrmicro3111

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science*, *298*(5594), 824–827. https://doi.org/10.1126/science.298.5594.824

Miranda, M., Parrini, F., & Dalerum, F. (2013). A categorization of recent network approaches to analyse trophic interactions. *Methods in Ecology and Evolution*, *4*(10), 897–905. https://doi.org/10.1111/2041-210X.12092

Montgomery, R. A., Moll, R. J., Say-Sallaz, E., Valeix, M., & Prugh, L. R. (2019). A tendency to simplify complex systems. *Biological Conservation*, *233*, 1–11. https://doi.org/10.1016/j.biocon.2019.02.001

Montoya, J. M., & Solé, R. V. (2002). Small World Patterns in Food Webs. *Journal of Theoretical Biology*, *214*(3), 405–412. https://doi.org/10.1006/jtbi.2001.2460

Morales-Castilla, I., Matias, M. G., Gravel, D., & Araújo, M. B. (2015). Inferring biotic interactions from proxies. *Trends in Ecology & Evolution*, *30*(6), 347–356. https://doi.org/10.1016/j.tree.2015.03.014

Morin, P. J., & Lawler, S. P. (1995). Food Web Architecture and Population Dynamics: Theory and Empirical Evidence. *Annual Review of Ecology, Evolution, and Systematics*, *26*(Volume 26,), 505–529. https://doi.org/10.1146/annurev.es.26.110195.002445

Morin, P. J. (2011). *Community ecology* (2nd ed) [OCLC: ocn700735116]. Wiley.

Mossio, M., Bich, L., Varenne, F., & Silberstein, M. (2014). La circularité biologique : Concepts et modèles [Bibliographie_available: 0 Cairndomain: www.cairn-sciences.info Cite Par_available: 0 ISSN: 2425-5661 Section: Informatique, Informatique]. In *Modéliser & Simuler. Epistémologies et pratiques de la modélisation et de la simulation* (pp. 137–169, Vol. 2). Éditions Matériologiques. Retrieved May 19, 2024, from https://www.cairn.sciences.info/modeliser-et-simuler--9782919694730-page-137.htm

Mougi, A., & Kondoh, M. (2012). Diversity of interaction types and ecological community stability. *Science*, *337*(6092), 349–351. https://doi.org/10.1126/science.1220529

Munoz, F., Klausmeier, C. A., Gaüzère, P., Kandlikar, G., Litchman, E., Mouquet, N., Ostling, A., Thuiller, W., Algar, A. C., Auber, A., Cadotte, M. W., Delalandre, L., Denelle, P., Enquist, B. J., Fortunel, C., Grenié, M., Loiseau, N., Mahaut, L., Maire, A., … Kraft, N. J. B. (2023). The ecological causes of functional distinctiveness in communities. *Ecology Letters*, *26*(8). https://doi.org/10.1111/ele.14265

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A Movement Ecology Paradigm for Unifying Organismal Movement Research. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19052–19059. Retrieved July 11, 2022, from http://www.jstor.org/stable/25465597

Nicolson, M. (1990). Henry Allan Gleason and the Individualistic Hypothesis: The Structure of a Botanist's Career. *Botanical Review*, *56*(2), 91–161. Retrieved April 22, 2024, from https://www.jstor.org/stable/4354143

Nicvert, L., Donnet, S., Keith, M., Peel, M., Somers, M. J., Swanepoel, L. H., Venter, J., Fritz, H., & Dray, S. (2024). Using the multivariate Hawkes process to study interactions between multiple species from camera trap data. *Ecology*, *105*(4), e4237. https://doi.org/10.1002/ecy.4237

Nicvert, L., Donnet, S., Keith, M., Peel, M., Somers, M. J., Swanepoel, L. H., & Venter, J. A. (2023). Code and data for: Using the multivariate Hawkes process to study interactions between multiple species from camera trap data [Publisher: figshare]. Retrieved November 13, 2023, from https://doi.org/10.6084/m9.figshare.24552157.v3

O'Dea, R. E., Parker, T. H., Chee, Y. E., Culina, A., Drobniak, S. M., Duncan, D. H., Fidler, F., Gould, E., Ihle, M., Kelly, C. D., Lagisz, M., Roche, D. G., Sánchez-Tójar, A., Wilkinson, D. P., Wintle, B. C., & Nakagawa, S. (2021). Towards open, reliable, and transparent ecology and evolutionary biology. *BMC Biology*, *19*(1), 1–5. https://doi.org/10.1186/s12915-021-01006-3

Okuyama, T., & Holland, J. N. (2008). Network structural properties mediate the stability of mutualistic communities. *Ecology Letters*, *11*(3), 208–216. https://doi.org/10.1111/j.1461-0248.2007.01137.x

Olesen, J. M., Bascompte, J., Dupont, Y. L., & Jordano, P. (2006). The smallest of all worlds: Pollination networks. *Journal of Theoretical Biology*, *240*(2), 270–276. https://doi.org/10.1016/j.jtbi.2005.09.014

Olesen, J. M., Bascompte, J., Dupont, Y. L., & Jordano, P. (2007). The modularity of pollination networks. *Proceedings of the National Academy of Sciences*, *104*(50), 19891–19896. https://doi.org/10.1073/pnas.0706375104

Olito, C., & Fox, J. W. (2015). Species traits and abundances predict metrics of plant–pollinator network structure, but not pairwise interactions. *Oikos*, *124*(4), 428–436. https://doi.org/10.1111/oik.01439

Paine, R. T. (1966). Food Web Complexity and Species Diversity. *The American Naturalist*, *100*(910), 65–75. https://doi.org/10.1086/282400

Palmer, M. S., Gaynor, K. M., Becker, J. A., Abraham, J. O., Mumma, M. A., & Pringle, R. M. (2022). Dynamic landscapes of fear: Understanding spatiotemporal risk. *Trends in Ecology & Evolution*, *37*(10), 911–925. https://doi.org/10.1016/j.tree.2022.06.007

Palmer, M. S., & Packer, C. (2021). Reactive anti-predator behavioral strategy shaped by predator characteristics. *PLoS ONE*, *16*(8), e0256147. https://doi.org/10.1371/journal.pone.0256147

Pan, Y., Birdsey, R. A., Phillips, O. L., & Jackson, R. B. (2013). The Structure, Distribution, and Biomass of the World's Forests. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 593–622. https://doi.org/10.1146/annurev-ecolsys-110512-135914

Pardo, L. E., Bombaci, S. P., Huebner, S., Somers, M. J., Fritz, H., Downs, C., Guthmann, A., Hetem, R. S., Keith, M., le Roux, A., Mgqatsa, N., Packer, C., Palmer, M. S., Parker, D. M., Peel, M., Slotow, R., Strauss, W. M., Swanepoel, L., Tambling, C., … Venter, J. A. (2021). Snapshot Safari: A large-scale collaborative to monitor Africa's remarkable biodiversity. *South African Journal of Science*, *117*(1/2), 1–4. https://doi.org/10.17159/sajs.2021/8134

Paul, R. E. L., Ariey, F., & Robert, V. (2003). The evolutionary ecology of Plasmodium. *Ecology Letters*, *6*(9), 866–880. https://doi.org/10.1046/j.1461-0248.2003.00509.x

Peralta, G., CaraDonna, P. J., Rakosy, D., Fründ, J., Pascual Tudanca, M. P., Dormann, C. F., Burkle, L. A., Kaiser-Bunbury, C. N., Knight, T. M., Resasco, J., Winfree, R., Blüthgen, N., Castillo,

W. J., & Vázquez, D. P. (2024). Predicting plant–pollinator interactions: Concepts, methods, and challenges. *Trends in Ecology & Evolution*, S0169534723003361. https://doi.org/10.1016/j.tree.2023.12.005

Pessoa, L. (2022). Complex Systems: The Science of Interacting Parts [type: article]. In *The Entangled Brain: How Perception, Cognition, and Emotion Are Woven Together*. MIT press. https://doi.org/10.31219/osf.io/635tq

Petchey, O. L., Beckerman, A. P., Riede, J. O., & Warren, P. H. (2008). Size, foraging, and food web structure. *Proceedings of the National Academy of Sciences*, *105*(11), 4191–4196. https://doi.org/10.1073/pnas.0710672105

Pickett, S. T. A., Cadenasso, M. L., & Meiners, S. J. (2009). Ever since Clements: From Succession to Vegetation Dynamics and Understanding to Intervention. *Applied Vegetation Science*, *12*(1), 9–21. Retrieved April 22, 2024, from https://www.jstor.org/stable/27735041

Pie, M. R., Campos, L. L. F., Meyer, A. L. S., & Duran, A. (2017). The evolution of climatic niches in squamate reptiles. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1858), 20170268. https://doi.org/10.1098/rspb.2017.0268

Pimm, S. L. (1979). The structure of food webs. *Theoretical Population Biology*, *16*(2), 144–158. https://doi.org/10.1016/0040-5809(79)90010-8

Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary through space and time. *Oikos*, *124*(3), 243–251. https://doi.org/10.1111/oik.01719

Poisot, T., Stouffer, D. B., & Sonia, K. (2016). Describe , understand and predict : Why do we need networks in ecology ? *Functional Ecology*, *30*, 1878–1882. https://doi.org/10.1111/1365-2435.12799

Popovic, G. C., Warton, D. I., Thomson, F. J., Hui, F. K., & Moles, A. T. (2019). Untangling direct species associations from indirect mediator species effects with graphical models. *Methods in Ecology and Evolution*, *10*(9), 1571–1583. https://doi.org/10.1111/2041-210X.13247

Proulx, S. R., Promislow, D. E. L., & Phillips, P. C. (2005). Network thinking in ecology and evolution. *Trends in Ecology & Evolution*, *20*(6), 345–353. https://doi.org/10.1016/j.tree.2005.04.004

Raath-Krüger, M. J., Schöb, C., McGeoch, M. A., & le Roux, P. C. (2021). Interspecific facilitation mediates the outcome of intraspecific interactions across an elevational gradient. *Ecology*, *102*(1), e03200. https://doi.org/10.1002/ecy.3200

Rezende, E. L., Lavabre, J. E., Guimarães, P. R., Jordano, P., & Bascompte, J. (2007). Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*, *448*(7156), 925–928. https://doi.org/10.1038/nature05956

Ronquillo, C., Stropp, J., & Hortal, J. (2024). OCCUR Shiny application: A user-friendly guide for curating species occurrence records. *Methods in Ecology and Evolution*, *15*(5). https://doi.org/10.1111/2041-210X.14271

Santos, A. M. C., & Quicke, D. L. J. (2011). Large-scale diversity patterns of parasitoid insects. *Entomological Science*, *14*(4), 371–382. https://doi.org/10.1111/j.1479-8298.2011.00481.x

Say-Sallaz, E., Chamaillé-Jammes, S., Fritz, H., & Valeix, M. (2019). Non-consumptive effects of predation in large terrestrial mammals: Mapping our knowledge and revealing the tip of the iceberg. *Biological Conservation*, *235*(October 2018), 36–52. https://doi.org/10.1016/j.biocon.2019.03.044

Schleuning, M., Fründ, J., & García, D. (2015). Predicting ecosystem functions from biodiversity and mutualistic networks: An extension of trait-based concepts to plant–animal interactions. *Ecography*, *38*(4), 380–392. https://doi.org/10.1111/ecog.00983

Sheppard, S. K., & Harwood, J. D. (2005). Advances in molecular ecology: Tracking trophic links through predator–prey food-webs. *Functional Ecology*, *19*(5), 751–762. https://doi.org/10.1111/j.1365-2435.2005.01041.x

Simberloff, D. (2004). Community Ecology: Is It Time to Move On?: (An American Society of Naturalists Presidential Address). *The American Naturalist*, *163*(6), 787–799. https://doi.org/10.1086/420777

Simmons, B. I., Cirtwill, A. R., Baker, N. J., Wauchope, H. S., Dicks, L. V., Stouffer, D. B., & Sutherland, W. J. (2019). Motifs in bipartite ecological networks: Uncovering indirect interactions. *Oikos*, *128*(2), 154–170. https://doi.org/10.1111/oik.05670

Smith, T. M., & Smith, R. L. (2015). *Elements of ecology* (Global ed., 9. ed). Pearson.

Stireman III, J. O., & Singer, M. S. (2003). Determinants of Parasitoid–Host Associations: Insights from a Natural Tachinid–Lepidopteran Community. *Ecology*, *84*(2), 296–310. https://doi.org/10.1890/0012-9658(2003)084[0296:DOPHAI]2.0.CO;2

Strauss, S. Y. (1991). Indirect effects in community ecology: Their definition, study and importance. *Trends in Ecology & Evolution*, *6*(7), 206–210. https://doi.org/10.1016/0169-5347(91)90023-Q

Stroud, J. T., Bush, M. R., Ladd, M. C., Nowicki, R. J., Shantz, A. A., & Sweatman, J. (2015). Is a community still a community? Reviewing definitions of key terms in community ecology. *Ecology and Evolution*, *5*(21), 4757–4765. https://doi.org/10.1002/ece3.1651

Strydom, T., Catchen, M. D., Banville, F., Caron, D., Dansereau, G., Desjardins-Proulx, P., Forero-Muñoz, N. R., Higino, G., Mercier, B., Gonzalez, A., Gravel, D., Pollock, L., & Poisot, T. (2021). A roadmap towards predicting species interaction networks (across space and time). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1837), 20210063. https://doi.org/10.1098/rstb.2021.0063

Teder, T., Tammaru, T., & Kaasik, A. (2013). Exploitative competition and coexistence in a parasitoid assemblage. *Population Ecology*, *55*(1), 77–86. https://doi.org/10.1007/s10144-012-0341-6

Thébault, E., & Fontaine, C. (2010). Stability of Ecological Communities and the Architecture of Mutualistic and Trophic Networks. *Science*, *329*(5993), 853–856. https://doi.org/10.1126/science.1188321

Thompson, J. N. (2024). Community ecology. Retrieved April 23, 2024, from https://www.britannica.com/science/community-ecology

Townsend, C. R., Begon, M., & Harper, J. L. (2008). *Essentials of ecology* (3rd ed). Blackwell Publishing.

Townsend, H. M., Huyvaert, K. P., Hodum, P. J., & Anderson, D. J. (2002). Nesting distributions of Galápagos boobies (Aves: Sulidae): An apparent case of amensalism. *Oecologia*, *132*(3), 419–427. https://doi.org/10.1007/s00442-002-0992-7

Travassos-Britto, B., Pardini, R., El-Hani, C. N., & Prado, P. I. (2021). Towards a pragmatic view of theories in ecology. *Oikos*, *130*(6), 821–830. https://doi.org/https://doi.org/10.1111/oik.07314

Tylianakis, J. M., Didham, R. K., Bascompte, J., & Wardle, D. A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, *11*(12), 1351–1363. https://doi.org/10.1111/j.1461-0248.2008.01250.x

Tylianakis, J. M., & Morris, R. J. (2017). Ecological Networks Across Environmental Gradients. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 25–48. https://doi.org/10.1146/annurev-ecolsys-110316-022821

Valdovinos, F. S. (2019). Mutualistic networks: Moving closer to a predictive theory. *Ecology Letters*, *22*(9), 1517–1534. https://doi.org/10.1111/ele.13279

Valiente-Banuet, A., Aizen, M. A., Alcántara, J. M., Arroyo, J., Cocucci, A., Galetti, M., García, M. B., García, D., Gómez, J. M., Jordano, P., Medel, R., Navarro, L., Obeso, J. R., Oviedo, R., Ramírez, N., Rey, P. J., Traveset, A., Verdú, M., & Zamora, R. (2015). Beyond species loss: The extinction of ecological interactions in a changing world. *Functional Ecology*, *29*(3), 299–307. https://doi.org/10.1111/1365-2435.12356

Vázquez, D. P., Blüthgen, N., Cagnolo, L., & Chacoff, N. P. (2009). Uniting pattern and process in plant–animal mutualistic networks: A review. *Annals of Botany*, *103*(9), 1445–1457. https://doi.org/10.1093/aob/mcp057

Vázquez, D. P., Chacoff, N. P., & Cagnolo, L. (2009). Evaluating multiple determinants of the structure of plant–animal mutualistic networks. *Ecology*, *90*(8), 2039–2046. https://doi.org/10.1890/08-1837.1

Vázquez, D. P., Morris, W. F., & Jordano, P. (2005). Interaction frequency as a surrogate for the total effect of animal mutualists on plants. *Ecology Letters*, *8*(10), 1088–1094. https://doi.org/10.1111/j.1461-0248.2005.00810.x

Vázquez, D. P., Ramos-Jiliberto, R., Urbani, P., & Valdovinos, F. S. (2015). A conceptual framework for studying the strength of plant–animal mutualistic interactions. *Ecology Letters*, *18*(4), 385–400. https://doi.org/10.1111/ele.12411

Volterra, V. (1926). Fluctuations in the Abundance of a Species considered Mathematically (letter). *Nature*, *118*(2972), 558–560. https://doi.org/10.1038/119012b0

Vonk, J. A., Christianen, M. J. A., & Stapel, J. (2008). Redefining the trophic importance of seagrasses for fauna in tropical Indo-Pacific meadows. *Estuarine, Coastal and Shelf Science*, *79*(4), 653–660. https://doi.org/10.1016/j.ecss.2008.06.002

Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology & Evolution*, *30*(12), 766–779. https://doi.org/10.1016/j.tree.2015.09.007

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Werner, E. E., & Peacor, S. D. (2003). A Review of Trait-Mediated Indirect Interactions in Ecological Communities. *Ecology*, *84*(5), 1083–1100. https://doi.org/10.1890/0012-9658(2003)084[1083:AROTII]2.0.CO;2

Williams, R. J., & Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, *404*(6774), 180–183. https://doi.org/10.1038/35004572

Wilson, E. O. (1998). *Consilience: The unity of knowledge* (1st ed). Knopf : Distributed by Random House.

Wootton, J. T. (1994). The nature and consequences of indirect effects in ecological communities. *Annual Review of Ecology, Evolution, and Systematics*, *25*(Volume 25,), 443–466. https://doi.org/10.1146/annurev.es.25.110194.002303

Wootton, J. T., & Emmerson, M. (2005). Measurement of Interaction Strength in Nature. *Annual Review of Ecology, Evolution, and Systematics*, *36*(1), 419–444. https://doi.org/10.1146/annurev.ecolsys.36.091704.175535



Part 2

Investigate trait matching in interaction networks with correspondence analysis and reciprocal scaling

Cover picture: plants and birds from the Peruvian montane forest are linked by a network of frugivorous interactions. Each species has specific interaction partners: their interaction niches are represented with ellipses. From left to right and top to bottom, the plants are *Vismia gracilis*, *Banara arguta*, *Ocotea sp.*, *Ficus coerulescens* and *Anthurium obtusum*, and the birds are *Elaenia pallatangae*, *Anisognathus somptuosus*, *Chlorochrysa calliparaea*, *Trogon personatus*, *Cephalopterus ornatus* and *Rupicola peruvianus*.

© Own drawing.

Introduction

1 Preamble: PhD context

This axis, which I began investigating at the end of my first year of PhD, was not initially planned in the PhD project. Indeed, after one year of PhD working on inference of interactions from camera trap data (presented in part 3), we decided to take on a new approach to study interspecific interactions, because the first part did not yield the results allowing to pursue the original PhD project. We choose this particular research direction because it was a research question that one of my supervisors, Stéphane Dray, had begun to investigate and for which he had seen a scientific potential, and that was less exploratory than the original PhD project, which was reassuring for me at the end of one year of PhD with no publishable results.

Initially, I approached this subject with less interest than the original PhD project, and somewhat reluctantly since I had to give up the time and energy previously invested in another scientific question and study system. But with time, I began to appreciate working on this axis, to find the scientific question really interesting and to understand and appropriate the associated methods and algebra. A small anecdote shows how this part really grew on me. One evening, as I was walking in the streets of Lyon, I saw a group of pigeons trying to eat a marinated olive. The olive probably fell on the floor from a terrace table, and pigeons seized the opportunity. As I saw them struggle to peck tiny bits of this olive, I immediately thought of trait matching: their beaks were too small.

2 Introduction

In ecological networks, many patterns deviating from random networks have been observed (Bascompte and Jordano, 2007; Ings et al., 2009; see also section 5.1 in the general introduction). For instance, ecological networks tend to have a degree distribution (i.e. distribution of the number of links per species) that deviates from the distribution expected for random networks (Dunne et al., 2002; Jordano et al., 2003; Ings et al., 2009) in the sense of the Erdős-Rényi model (Erdős & Rényi, 1959). Ecological networks also have a tendency to be modular, i.e they comport modules of species that tend to interact with each other but not with other species (Ings et al., 2009). Some patterns also depend on the type of network: in particular, food webs include many generalist species, while host-parasite networks include more specialists (Ings et al., 2009). Nestedness is also commonly observed in mutualistic networks, i.e. interactions of specialist species tend to be subsets of interactions of more generalist species (Bascompte et al., 2003; Valdovinos, 2019). While these patterns begin to be relatively well-known, the processes that underpin them are less understood (Benadi et al., 2022).

Identifying causes that determine whether two species inside a community interact or not, and the nature of their interaction, is a broad ecological question of great interest. As discussed more extensively in the general introduction (section 5.1), several factors have been hypothesized to influence the probability of species interactions, among which species abundances (Poisot et al., 2015; Dormann et al., 2017), trait matching (Dormann et al., 2017), or species evolutionary history (Benadi et al., 2022; Dormann et al., 2017), and mechanistic models integrating some of these factors have been proposed.

In this part, I more precisely focus on one of the drivers of interactions in ecological networks: trait matching (Figure 2.I.1), which posits that species interactions are notably driven by the traits complementary of the interacting species. Trait matching may result from species adapting to their interactions over evolutionary time, but may also result from evolutionary processes independent of their interactions (Dormann et al., 2017): either way, studying trait matching remains valuable to understand the proximate drivers of species interactions.

Trait matching has been studied notably through niche-based models, which repre-

sent the effect of species traits on interactions (Williams & Martinez, 2000), possibly modulated by other processes such as species abundances, sampling process or phylogeny (Fründ et al., 2016; Benadi et al., 2022). These models have been useful to study the effect of traits on interactions, but also have shortcomings. First, most niche-based models consider only a single niche dimension (but see Eklöf et al., 2013 and Benadi et al., 2022), thus generating interval networks (i.e. networks in which interactions frequencies are explained by one single trait), even though interaction probabilities are generally thought to be shaped by several traits (Williams & Martinez, 2000; Eklöf et al., 2013). Second, these models generally require knowing which traits are relevant for trait matching beforehand to evaluate its importance (but see Eklöf et al., 2013). Finally, most methods ignore the tolerance of species around their trait matching constraints.



Figure 2.I.1: Illustration of trait matching between a bird and a fruiting plant. Here, trait matching is based on the morphological complementarity between the bird's beak height h and the fruit diameter d (a single trait is involved).

In this part, I consider multivariate methods allowing to model trait matching between multiple and possibly latent traits, and model species' preferences for their interacting partners with their interaction niche, thus taking the variability around matching constraints into account. The concept of niche, its origins and influence on interaction networks are briefly discussed below.

2.1 The ecological niche

The ecological niche is broadly defined as the set of conditions a species can persist in: more formally, the niche has been defined by Hutchinson as an n-dimensional hypervolume in which the species can persist, where each niche axis represents environmental conditions and resources (Hutchinson, 1957) (see Figure 2.I.2). The ecological niche is usually studied considering either environmental conditions or interactions with

other species, although both factors can affect the niche at the same time. The environmental (or Grinnellian) niche (Grinnell, 1924) considers environmental resources and constraints, such as pH, temperature or precipitation. The interaction (or Eltonian) niche (Elton, 1927) considers the niche in terms of the interactions of a species with other species in the community, allowing to study how a species is included in the community, "its relation to food and enemies" (Elton, 1927).

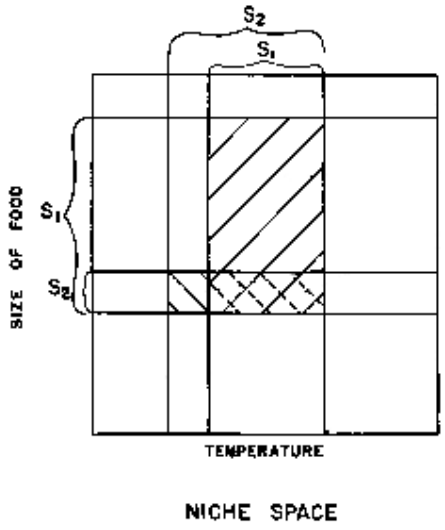


Figure 2.I.2: Original illustration of the Hutchinsonian niche concept. This figure shows two-dimensional niches defined by food size and temperature. The niches of species S_1 and S_2 are depicted as hatched areas and overlap in the middle of the graph. Adapted from Figure 1 in Hutchinson (1957).

The ecological niche is recognized as an important structuring factor of communities (Smith & Smith, 2015). Among other things, niche partitioning between species is thought to be a stabilizing factor of the composition of ecological communities under modern coexistence theory (Chesson, 2000): by utilizing different resources, species can alleviate competition enough to coexist.

More recently, studies have also investigated the stability of species interactions (Valiente-Banuet et al., 2015). Interaction niches are one of the factors influencing the stability of interaction networks (Godoy et al., 2018). For example, the simulation approach used in Valdovinos et al. (2016) shows that niche partitioning in plant-pollinator networks can stabilize the network. Species interaction niches also influence network structure through trait matching. For instance, in most predation networks, especially for aquatic systems, body size is an important factor determining if predation can occur (Ings et al., 2009). Hence, the interaction niche notably determines which interactions are "forbidden", i.e. cannot occur due to incompatible species traits (Jordano et al.,

2003).

The notion of interaction niche is often fuzzy and can be defined in multiple ways (Ferry-Graham et al., 2002). Here, I propose to use the Hutchinsonian conceptualization of the niche as a hypervolume, where we can view other species in the network as resource axes that allow a given species to persist. As species are not quantitative variables, they cannot readily define axes in the sense of the Hutchinsonian niche, which is defined with quantitative axes (Hutchinson, 1957; Vandermeer, 1972). However, viewing species through their traits allow to define quantitative traits axes: then, for a given species, we can define the subset of their interacting partners' trait space (the Hutchinsonian n-dimensional hypervolume) in which this species is able to interact. For instance, if we consider a pollinator, its pollination niche can be defined in terms of floral traits such as corolla depth, nectar quantity or flowering phenology.

Several methods can be used to quantify species niches (see Devictor et al., 2010, for a review): but if we quantify niches in the Hutchinsonian sense, multivariate ordination methods seem particularly suited.

2.2 Multivariate ordination methods

Multivariate ordination methods (hereafter, simply ordination methods) aim at arranging entities on the basis of variables that are associated to them (C. J. F. ter Braak, 1995). In ecology, these methods are classically used to arrange sites, using the associated counts of species occurrences in each site. Ordination methods transform the data to highlight the main patterns of variation with the tools of matrix algebra, where the matrix is classically a species-by-sites matrix.

Ordination methods are also relevant to analyze networks, as they can also be seen as matrices. Here, I consider a particular class of networks representing the interactions between two classes of objects, called bipartite networks. These network can be represented by an incidence matrix, a matrix with one class of objects in rows and the other in columns. In the context of interaction networks, bipartite networks represent interactions between two groups of species (e.g. plant-pollinator). Incidence matrices can also be seen as contingency matrices, i.e. matrices describing the frequency distribution of two crossed factors (here, the counts of interactions between the two species classes).

Contingency matrices are routinely analyzed using methods in the correspondence analysis (CA) family, which is a class of ordination methods. There are three CA methods: simple correspondence analysis (CA, or simple CA) (Hirschfeld, 1935; Hill, 1974), that allows to analyze a contingency table alone; canonical correspondence analysis (CCA), that allows to include a second table of quantitative variables (C. J. F. ter Braak, 1986) and double-constrained correspondence analysis (dc-CA), that allows to include two tables of quantitative variables in the analysis (C. J. F. ter Braak et al., 2018).

Methods of the CA family allow to examine the relationship between qualitative variables (here, species from each class) by associating a score to each species along several axes. With CA, these scores are determined using the identity of the interacting species, and with CCA and dc-CA, these scores take into account a linear combination of the interacting species traits. Species scores are fitted so that species positions on one axis are optimally separated.

Species scores obtained with CA methods can be interpreted as a measure of their niche optima (i.e. their preference), where species niches are approximated by a normal distribution along each axis. CA methods have been mainly used to describe environmental niches in the context of species-environment associations (C. J. F. ter Braak, 1987). Here, on the contrary, I analyze interaction networks, which allows to investigate the interaction niche. Ordination methods thus allow to ordinate qualitative variables, the species, along multivariate axes. These axes can be seen as the niche axes in Hutchinson's framework: ordination methods of the CA family thus extend the definition of the Hutchinsonian niche to interaction niches (Eklöf et al., 2013).

2.3 Plan

In this part, I show how CA methods can be used to study species interaction niches. In chapter 2.1, I describe how CA and reciprocal scaling (Thioulouse & Chessel, 1992) can be used to determine species niche optima and breadth. I validate the method with a simulation approach and illustrate it on a real interaction network (Dehling et al., 2021). In chapter 2.2, I describe how constrained methods, CCA and dc-CA, can be used to determine species niche optima constrained with species traits. I also partition the network variance into parts due to species traits and residual variation, and illustrate the method on the same interaction network. Finally, in a more exploratory

chapter 2.3, I explore how a measure of niche breadth can be added to constrained analyses, by adapting reciprocal scaling to CCA and dc-CA.

2.4 Dataset

For all three chapters described above, I illustrate the method on a bird-fruit interaction network collected by Dehling et al. (2021) in the ANDEAN frugivory dataset. This dataset was collected in Peruvian lower montane forest (1500 meters above sea level) by direct observation of birds removing fruits from fleshy-fruited plants. Further information on the dataset can be found in chapter 2.1, but I also provide a brief overview of the dataset here.

Figure 2.I.3 shows the interaction matrix between 53 bird species (in columns) and 40 plant species (in rows) (after data filtering, see chapter 2.1). In total, the dataset records almost 5000 interactions sampled over 960h.

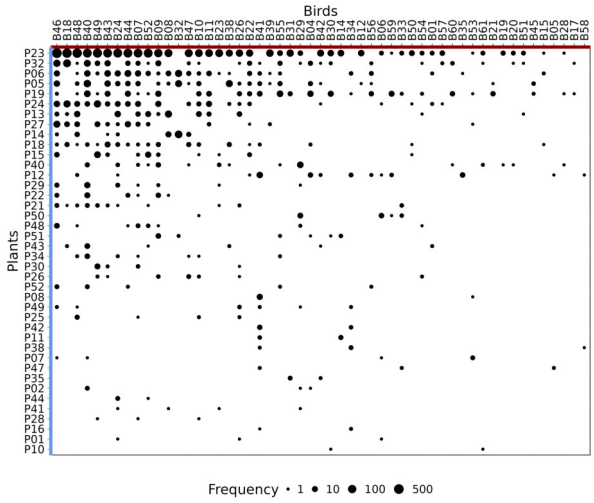


Figure 2.I.3: Interaction matrix for the bird-fruit interaction network. Rows and columns are ordered by increasing number of weighted interactions.

Figure 2.I.4 shows a glimpse of the diversity of birds and plant species. Birds body mass ranges from 8 g (scale-crested pygmy tyrant *Lophotriccus pileatus*) to 700 g (Andean cock-of-the-rock *Rupicola peruvianus*), with most birds being small passerines (80% of the birds weigh less than 100 g). Plants have very diverse heights, ranging from 3 m (*Miconia barbeyana*) to 15 m *Ficus americana*), and correspond to understory or canopy plants. Most of these plants produce small fruits, ranging from 2.2 mm (*Miconia egensis*) to 20.4 mm (*Cecropia polystachya*).



(a) *Lophotriccus pileatus*



(b) *Patagioenas plumbea*



(c) *Rupicola peruvianus*



(d) *Cestrum parqui*



(e) *Miconia affinis*



(f) *Ficus americana*

Figure 2.I.4: Overview of some bird and plant species sampled in the network, ordered from smaller to larger. Credits (from left to right): Chris Jimenez, Félix Uribe, Charles J. Sharp, tayloredtotaylor, Sébastien SANT and sebacixl (CC-BY pictures).

Chapter 2.1

Trait matching without traits: using correspondence analysis to analyze the latent structure of interaction networks

Lisa Nicvert, Hervé Fritz and Stéphane Dray

We intend to submit the results of this chapter for publication in *PCI Ecology*.

Abstract

Species interact with each other within interaction networks. The place of species in the interaction network determines their interaction (or Eltonian) niche and is strongly related to the notion of trait matching, by which a species interacts with partners whose traits are complementary to their own.

Multivariate methods are commonly used to quantify species environmental (or Grinnellian) niche. More recently, multivariate methods have also been used to study the interaction niche, but they consider only the niche optimum and require trait data, which can be tedious to collect and challenging to select.

In this paper, we propose to use correspondence analysis to study interaction networks and investigate trait matching without requiring traits, using the notion of latent traits. We also use reciprocal scaling to estimate the optimum and breadth of the interaction niche. In our framework, niche breadth is defined as the diversity of latent traits of interacting partners. We present the model, test its performance with a simulation approach using a model we designed and analyze a real mutalistic bird-fruits interaction network.

The simulation study shows that the method is able to recover niche breadths and optima for data generated with parameters values typical of ecological networks. The bird-fruit network analysis shows strong relationships between species niche optima and niche breadths: a posteriori correlation with measured species traits suggest that this latent structure is related to traits. In this network, birds and plants of intermediate size tend to have the widest niches. Birds with pointed wings (preferentially foraging in the canopy) have smaller niches than birds with rounded wings (preferentially foraging in the understory).

Correspondence analysis and reciprocal scaling are described as fruitful methods to characterize species interaction profiles. These methods provide an ecologically meaningful graphical representation of interaction niches and allow to explore the effect of latent traits on network structure.

1 Introduction

The ecological niches constitute the pool of environmental conditions and resource required for species persistence (Vandermeer, 1972). These factors can be abiotic (e.g., temperature, precipitation, light) and grouped together in the notion of Grinnellian niche (Grinnell, 1924), which centers on the physical and environmental requirements essential for a species' survival. The niche can also be seen through the lens of species' interactions with other organisms (Eltonian niche) originally formalized in the context of food webs by Charles Elton (Elton, 1927) but that can be extended to other types of interactions like competition and symbiosis. By encapsulating the ecological requirements and constraints of a species, the ecological niche is a crucial concept contributing to enlighten our understanding of species distributions, coexistence or exclusion.

Hutchinson (1957) famously formalized the niche concept as an n-dimensional hypervolume. One common way to describe these hypervolumes is to approximate the niche space with Gaussian curves along each niche axis (Gauch Jr. & Whittaker, 1972) to estimate species niche optimum and breadth along environmental gradients (as mean and standard deviation parameters of a normal distribution). Niche optimum describes the conditions where species growth is maximized, whereas niche breadth defines the range of conditions tolerated by a species. Niche breadth (or tolerance) allows to define the degree of specialization/generalization of species: generalists have wide niches and specialists have narrow niches. Generalization or specialization are usually defined for a given niche dimension, so that a generalist that tolerates a broad range of temperatures might tolerate a very narrow range of precipitations. The factors influencing niche breadth are still being investigated (Sexton et al., 2017; Devictor et al., 2010).

In this paper, we focus on the niche occupied by a species in its community, i.e. its realized niche, by opposition to its fundamental niche, which is the range of conditions a species can tolerate due to physiological or morphological constraints (Hutchinson, 1957; Devictor et al., 2010). The realized niche has often been defined as the niche of a species restricted to a narrower range than its fundamental niche by biotic interactions with other species in the community (Futuyma & Moreno, 1988): however, this characterization has environmental niches in mind and here, we use a broader defi-

nition of the realized niches as the range of resources effectively used by a species, affected by a range of processes including for instance competition with other species but also by availability of resources.

Multiple methods have been developed to study species environmental (or Grinnellian) niches. In particular, weighted averaging is a very simple method that estimates species optimum for a single environmental gradient by averaging the values of the environmental variable over the samples in which a species occurs (weighted by species abundance). It is at the basis of several popular multivariate techniques routinely used when the data at hand are species abundance (or presence) sampled across different sites. They include reciprocal averaging (Hill, 1973) (a.k.a. correspondence analysis (CA) Hill, 1974) to estimate latent gradients when no environmental data are available; canonical correspondence analysis (C. J. F. ter Braak, 1986) when multiple environmental data are recorded; or fourth-corner analysis (Legendre et al., 1997), RLQ analysis (Dolédec et al., 1996) and double-constrained correspondence analysis (C. J. F. ter Braak et al., 2018) to link niche optima to the species characteristics when traits data are also available. Note that these methods focus on species niche optima and ignore the aspects related to niche breadth. However, Thioulouse and Chessel (1992) notably proposed reciprocal scaling as a method to estimate niche breadth in the context of CA using weighted variances.

By contrast, methods used to describe and analyze interaction (or Eltonian) niches often use the formalism of networks (Ings et al., 2009; Bascompte & Jordano, 2007). Networks are mathematical objects composed of a set of nodes (species) which are linked by edges (interactions). In this paper, we focus on bipartite networks, i.e. networks where nodes are separated in two sets (e.g., plants and pollinators) and interactions occur only between the two sets, but not within a set: bipartite networks can describe a variety of interactions like pollination, parasitism or frugivory.

Ecological networks generally display non-random structures (Bascompte et al., 2005; Ings et al., 2009; Jordano et al., 2003) such as nestedness, where interactions of specialist species tend to be subsets of interactions of more generalist species (Bascompte et al., 2003). Various factors can influence the structure of ecological networks (Ings et al., 2009; Vázquez, Blüthgen, et al., 2009; Peralta et al., 2024). Neutral effects imply that a given species has more chance to interact with abundant species than with

the rarest ones (Vázquez, Blüthgen, et al., 2009). Variations in abundance, influenced for instance by environmental conditions or species phenology, can thus induce structures in ecological networks. Species evolutionary history is also thought to influence their interaction patterns (Benadi et al., 2022). Trait matching, wherein species with complementary traits preferentially engage in interactions, is also a well studied mechanism: for instance, body size has been identified as a key factor influencing food web structure (Elton, 1927; Ings et al., 2009). Importantly, these processes are not mutually exclusive and are often considered to act in synergy in shaping the observed network patterns.

More recently, several multivariate methods originally developed to study environmental niches have been applied to study trait matching. Indeed, as networks can be encoded with matrices, many methods applicable to species-by-sites matrices are readily applicable to networks (van Dam et al., 2021). For instance, RLQ and fourth-corner analyses have been used to describe trait matching by measuring the correlation between species traits of two sets of partners weighted by the interaction matrix (Albrecht et al., 2018; Dehling et al., 2014; Bender et al., 2018). Even if niche theory is usually not explicitly integrated in the study of ecological networks (Godoy et al., 2018; Phillips et al., 2020), it is clear that trait matching implicitly relies on the notion of interaction niche (Albrecht et al., 2018). Indeed, trait matching posits that a species with given traits will interact only with species with trait values lying in a range defining its interaction niche. The average value of trait of their partners defines its niche optimum, while their diversity correspond to its niche breadth. Niche breadth, which can be seen as the degree of generalization of a species, has often been defined as the number of interacting partners, potentially considering their abundance in the network (Devictor et al., 2010). Other studies have also defined niche breadth with respect to the traits of interacting species (functional niche, Dehling et al., 2016), considering multiple traits separately (Maglianesi et al., 2015) or simultaneously (Dehling et al., 2016, 2014).

Methods to analyze ecological networks have brought invaluable insights on trait matching and degree of specialization, but also have shortcomings. First, trait matching in ecological networks has mainly been investigated separately from interaction niches. However, trait matching can be seen through the lens of ecological niches (Eklöf et al., 2013): the term "matching" then refers to the alignment of the interacting species

traits with their functional niche optima (as considered in several simulations models, e.g. Fründ et al., 2016; Benadi et al., 2022). Second, trait matching studies only consider the average relationship between species traits across the network. Hence, they generally ignore the variability around the matching, which can be related to the concept of niche breadth. Lastly, methods to study trait matching require the collection of traits data, which has been recognized as a challenge (Vázquez, Blüthgen, et al., 2009), because it is tedious to collect multiple traits and identify a priori which ones are relevant drivers of interactions.

We propose to use correspondence analysis (CA) (Hirschfeld, 1935; Hill, 1974), a method often used in ecology to study species' environmental niches, to study trait matching and specialization/generalization in interaction networks. Indeed, these concepts can be related to the interaction niche, with trait matching defining niche optimum and specialization/generalization defining niche breadth. CA can also be useful to analyze interaction networks, as previously pointed out by Lewinsohn et al. (2006) and van Dam et al. (2021). Here, we combine the niche and the network aspect to show how CA can be used to infer species' interaction niches from interaction network data. Our framework allows to measure the effect of trait matching without using traits data (using the notion of latent traits). Scores produced by CA allow to ordinate species along gradients determined by their interacting partners. We show that these scores estimate the optima of species interaction niches and can be used as proxy for unmeasured traits determining the occurrences of interactions. We also use reciprocal scaling (Thioulouse & Chessel, 1992) to estimate species niche breadths. Lastly, these methods also enable ecologically meaningful graphical representation of bipartite networks and species interaction niches.

In this paper, we first present CA and reciprocal scaling and their interpretation in the context of interaction networks. Then, we evaluate the performance of our framework to estimate species niches (optimum and breadth) using simulated data. Finally, we analyze a real bird-fruit interaction network (Dehling et al., 2021) to show how our framework can help to measure and represent interaction niches, highlight trait matching and explore the drivers of niche breadth.

2 Material and methods

All analyses were performed with R 4.3.3 (R Core Team, 2024) and are stored in a private GitHub repository (https://github.com/LisaNicvert/PhDaxis01_CAnetwork/tree/thesis), and can be shared on demand. It is also intended to be made public upon publication.

2.1 Notations

In this part, we consider a $r \times c$ matrix $\mathbf{Y} = [y_{ij}]$ representing the interactions between r resource species and c consumer species (either abundances or presences). We use the terms "resource" and "consumer" in accordance with the literature, as these terms describe broad categories that encompass diverse networks such as pollination, herbivory or parasitism: but as long the interaction network is bipartite, the two sets of species could belong to any category.

Let $\mathbf{P}$ be the table of relative frequencies ($\mathbf{P} = [y_{ij}/y_{++}]$, where $y_{++} = \sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}$ is the grand total of $\mathbf{Y}$). We define the weight matrices for resources/rows $\mathbf{D_r} = \mathrm{diag}(\mathbf{r})$ and column/consumers $\mathbf{D_c} = \mathrm{diag}(\mathbf{c})$, where the vectors $\mathbf{r} = \mathbf{P1_r} = [p_{1+}, \dots, p_{r+}]^\top$ and $\mathbf{c} = \mathbf{P}^\top\mathbf{1_c} = [p_{+1}, \dots, p_{+c}]^\top$ represent respectively the row and column marginal sums and $p_{i+} = \sum_{j=1}^{c} p_{ij}$ and $p_{+j} = \sum_{i=1}^{r} p_{ij}$.

2.2 Quantify trait matching

We consider two known traits $\mathbf{x}$ and $\mathbf{y}$ measured, respectively, on the resource and consumer species. A simple measure of trait matching is given by the fourth-corner statistic, which measures the correlation between the traits of consumers and resources, weighted by the interaction matrix:

$$\mathrm{cor}_{\mathbf{P}}(\mathbf{x}, \mathbf{y}) = \tilde{\mathbf{x}}^\top \mathbf{P} \tilde{\mathbf{y}} \tag{2.1.1}$$

where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ contains normalized traits using weights ($\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ are centered scaled, respectively using weights $\mathbf{D_r}$ and $\mathbf{D_c}$). When the traits are not known, indirect gradient methods such as correspondence analysis are useful to identify latent structures in ecological networks.

2.3 Correspondence analysis (CA)

Correspondence analysis (Hirschfeld, 1935; Hill, 1974) is a multivariate method used to analyze contingency tables. CA is based on the generalized singular value decomposition of the doubly centered frequency interaction matrix $\mathbf{P}$, $\mathbf{P_0} = \mathbf{P} - \mathbf{rc}^\top$:

$$\mathbf{P_0} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^\top \tag{2.1.2}$$

The matrices $\mathbf{U}$ and $\mathbf{V}$ are orthonormal with respect to weights $\mathbf{D_r}$ and $\mathbf{D_c}$ ($\mathbf{U}^\top\mathbf{D_r}\mathbf{U} = \mathbf{I}$ and $\mathbf{V}^\top\mathbf{D_c}\mathbf{V} = \mathbf{I}$) and contain the left (resource) and right (consumer) generalized singular vectors (respectively). $\boldsymbol{\Delta}$ is the diagonal matrix containing ordered singular values. The matrix of CA eigenvalues $\boldsymbol{\Lambda}$ is equal to $\boldsymbol{\Delta}^2$.

If we consider a given dimension k, rewriting the previous equation shows that CA finds singular vectors $\mathbf{u}_k$ and $\mathbf{v}_k$ so that the singular value δ_k is maximized:

$$\delta_k = \mathbf{u}_k^\top\mathbf{P_0}\mathbf{v}_k \tag{2.1.3}$$

Equation (2.1.3) is analogous to the trait matching equation (2.1.1) (except for the use of $\mathbf{P_0}$ instead of $\mathbf{P}$, which does not affect results if traits data are centered), where known traits are replaced with the generalized singular vectors $\mathbf{u}_k$ and $\mathbf{v}_k$. CA amounts to finding the singular vectors with maximal correlation: the square-root of CA eigenvalue λ_k represents the absolute value of the correlation between $\mathbf{u}_k$ and $\mathbf{v}_k$ ($\sqrt{\lambda_k} = |\mathrm{cor}_\mathbf{P}(\mathbf{u}_k, \mathbf{v}_k)|$). Thus, scores $\mathbf{u}_k$ and $\mathbf{v}_k$ can be interpreted as latent traits, associated to resources (respectively, consumers), that maximizes trait matching. The eigenvalues λ_k give an indication on the agreement of consumer and resource species scores on dimension k: if eigenvalues are high, it indicates that latent traits are able to explain well the probability of interaction.

CA has the advantage to evaluate the effect of trait matching, without using traits data, using these latent scores. However, latent traits can be interpreted in a broader sense as proxies for unmeasured properties for species, not only traits (for instance as phylogenetic signal as in Benadi et al., 2022). To help the interpretation, it is possible to link a posteriori latent scores identified by CA to known properties such as traits or phylogeny. CA is well suited to exploratory analyses, to identify structures in the network and estimate the part that the measured traits or phylogeny succeed, or fail,

to explain.

CA assigns similar scores to species with the same interaction profile. In other words, two consumer species interacting with the same resources will be positioned nearby in the multivariate space (and vice versa for resource species). Species scores can thus be used to reorder rows and columns of the interaction matrix and highlight the structure of the network (Lewinsohn et al., 2006) (see Appendix C).

Standards outputs of CA consist of two biplots where both consumers and resources are displayed. For that, we use the following transition formulas:

$$\mathbf{U}^\star = {\mathbf{D_r}}^{-1}\mathbf{P}\mathbf{V} \qquad \text{(resources)} \qquad (2.1.4a)$$

$$\mathbf{V}^\star = {\mathbf{D_c}}^{-1}\mathbf{P}^\top\mathbf{U} \qquad \text{(consumers)} \qquad (2.1.4b)$$

Equation (2.1.4a) expresses resources scores as a weighted mean of the consumers scores they interact with. Reciprocally, Equation (2.1.4b) expresses consumers scores as a weighted mean of the resources scores they interact with. The transformation described in Equations (2.1.4) is called weighted averaging.

A first biplot can be drawn by displaying the resources scores given by the latent trait ($\mathbf{U}$) and consumers scores given by weighted averaging ($\mathbf{V}^\star$). On this plot, a consumer is located at the centroid of the resource it uses, and a resource is located according to its latent traits.

On the second biplot, scores are inverted so that consumers are represented with scores $\mathbf{V}$ (given by the latent traits) and resources with scores $\mathbf{U}^\star$ (given by weighted averaging). On this plot, a resource is located at the centroid of the consumers that use it, and a consumer is located according to its latent traits.

The coordinates of $\mathbf{U}^\star$ and $\mathbf{V}^\star$ can be interpreted as the niche optima of resource (respectively, consumer) species in the interaction network, measured with its interacting partners. This interpretation of the interaction niche is similar to the niche quantified with the method of Dehling and Stouffer (2018), but in the context of CA (as discussed by C. J. ter Braak and Prentice, 2004 and C. J. F. ter Braak and Verdonschot, 1995 for environmental niches).

CA seems particularly adapted to the analysis of interaction networks, as it allows to

display structures associated to trait matching and represent species niches by estimating their optima by weighted averaging. A major drawback of CA is that it completely ignores the diversity of interacting partners, and thus the niche breadth. Moreover, niche optima can only be represented on two different biplots.

2.4 Reciprocal scaling

Reciprocal scaling (Thioulouse & Chessel, 1992) is an extension of CA that solves the two issues mentioned above by providing a simultaneous display of niche optima and breadths of consumers and resources.

Reciprocal scaling uses the interpretation of CA as a special case of canonical correlation analysis and shift the focus of the analysis by considering the interactions, rather than the species, as statistical individuals (for complete mathematical development, see Thioulouse & Chessel, 1992). For the k-th dimension, reciprocal scaling defines a score $\mathbf{h}_k$ for each resource-consumer interaction in the matrix: these interactions, represented by non-empty cells in $\mathbf{Y}$, are called correspondence by Thioulouse and Chessel (1992). $\mathbf{h}_k$ is of length $n_{\bar{0}}$, where $n_{\bar{0}}$ is the number of unique interactions in the matrix (i.e. non-empty cells). $\mathbf{h}_k$ can be easily computed using the CA scores for the k-th dimension:

$$h_k(i,j) = \frac{u^\star_{ik} + v^\star_{jk}}{\sqrt{2\lambda_k \mu_k}} \tag{2.1.5}$$

where λ_k is the CA eigenvalue for axis k and $\mu_k = 1 + \sqrt{\lambda_k}$. Here, a double indexing is used for the elements of $\boldsymbol{h}_k$, which is a one-dimensional vector: $h_k(i,j)$ corresponds to the x-th element of $\boldsymbol{h}_k$ ($x = 1 \dots n_{\bar{0}}$), containing the interaction between consumer i and resource j. Note that the score $h_k(i,j)$ is defined only when species i and j interact.

The interaction score $h_k(i,j)$ allows to display each resource-consumer interaction in the multivariate space. Then, it is possible to display a consumer (or a resource) as a cloud of individual interactions in the multivariate space. This cloud represents the interaction niche of a species, and can be summarized by estimating its optimum and its breadth. To do so, ellipses corresponding to the bivariate normal distribution summarizing species niches can be drawn in the multivariate plane (see Figure 2.1.1).

If we consider species s in the plane given by axes k and l, the center of its ellipse is given by the niche optima of species s in axes k and l (weighted averages of interaction scores), the semi-minor and -major axes are given by its niche breadths in axes k and l (weighted standard deviations of interactions scores) and the orientation of the ellipse represents the weighted covariance of interaction scores for axes k and l. Moreover, we can show that the niche optima computed from these scores correspond to the scores $\mathbf{U}^\star$ and $\mathbf{V}^\star$ defined above, with a scaling factor. Formulas to compute niche optima, breadths and covariances from Thioulouse and Chessel (1992) adapted with the notation used in this article are given in Appendix D.

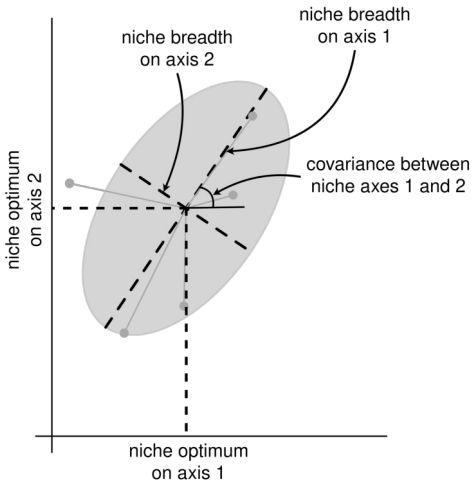


Figure 2.1.1: Visualization of the interaction niche in the multivariate plane. The niche of is depicted by an ellipse whose center represents niche optima and axes represent niche breadths on each axis. The angle of the ellipse represents the covariance between niche axes.

2.5 Simulation

We simulated interaction data to evaluate the performance of CA and reciprocal scaling to estimate parameters of species interaction niches. Our simulation model takes into account species traits (via trait matching) and their abundances (via neutral effects) and combines these two processes to generate an observed interaction matrix (see Figure 2.1.2).

Simulation model

Our model is inspired from the models of Fründ et al. (2016) and Benadi et al. (2022) to simulate trait-based interaction networks and of the models of Dray and Legendre

(2008) and Minchin (1987) for niche modeling in the context of species/sites associations.

We consider the interactions between the i-th resource ($i = 1 \dots r$) and the j-th consumer species ($j = 1 \dots c$). The simulation procedure consists in different steps described below and illustrated in Figure 2.1.2:

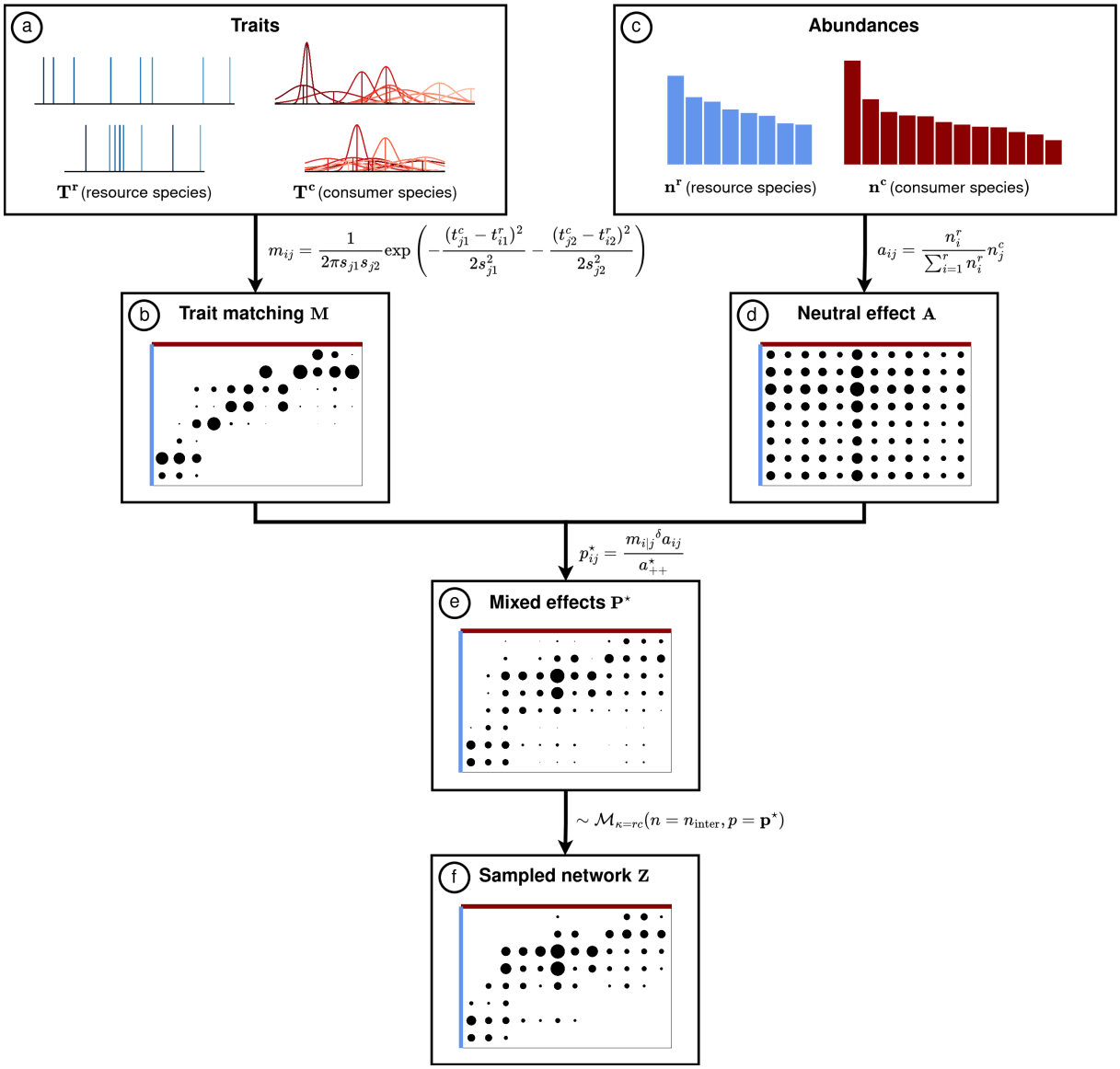


Figure 2.1.2: Model used to simulate interactions between consumer and resource species. (a) Two traits are generated for each species corresponding to their niche optima. For consumers, standard deviations are also associated to traits to generate the niche breadths. (b) Interaction probabilities due to trait matching are computed using a multivariate normal distribution. (c) In parallel, resource and consumers abundances are randomly generated. (d) Interactions counts representing the neutral effect of abundances on interaction probability are computed from abundances. (e) Matching and abundance-driven interactions are combined to get the interactions counts based on both processes. (f) Observed interactions are then sampled from the mixed interaction probabilities. Equations are explained in the main text.

Step (a): we simulate two traits for both consumer and resource species (Figure 2.1.2a). Resource species traits are stored in matrix $\mathbf{T^r} = [t^r_{ik}]$ $(r \times 2)$ and consumer traits are stored in matrix $\mathbf{T^c} = [t^c_{jk}]$ $(c \times 2)$. The first trait for both resources and consumers is drawn from a uniform distribution between 0 and 100, and the second trait is generated similarly but in the interval 15-85 (see length of the traits axes on Figure 2.1.2a), so that the second trait has less weight in driving matching. For consumers, we associate the two columns vectors of $\mathbf{T^c}$ to those of $\mathbf{S} = [s_{jk}]$ that allow to define the degree of specialization (i.e., niche breadth) for consumers (values of $\mathbf{S}$ are represented on Figure 2.1.2a as the standard deviation of a Gaussian curve). The elements of $\mathbf{S}$ contain the absolute values of numbers drawn from a normal distribution with mean μ_{breadth} and standard deviation σ_{breadth}.

Step (b): traits generated in the previous step are used to compute an interaction probability due to matching (Figure 2.1.2b). Species traits are assumed to define the interaction niche with a bivariate normal distribution: the probability of interaction due to matching m_{ij} follows a normal distribution influenced by the difference between the trait values of resource species i and consumer species j $(t^c_{jk} - t^r_{ik})$ and the degree of generalization of consumer species (s_{jk}) as follows:

$$\mathbf{M} = [m_{ij}] = \frac{1}{2\pi s_{j1} s_{j2}} \exp\left(-\frac{(t^c_{j1} - t^r_{i1})^2}{2s^2_{j1}} - \frac{(t^c_{j2} - t^r_{i2})^2}{2s^2_{j2}} \right) \tag{2.1.6}$$

Step (c): we generate species abundances for consumers and resource species independently (Figure 2.1.2c) from a uniform or a log-normal distribution. These abundances are stored in vectors $\mathbf{n^c}$ $(c \times 1)$ for consumers and $\mathbf{n^r}$ $(r \times 1)$ for resources.

Step (d): we then compute a matrix of predicted interaction counts based solely on the abundances (Figure 2.1.2c):

$$\mathbf{A} = [a_{ij}] = \left[\frac{n^r_i}{\sum_{i=1}^{r} n^r_i} n^c_j \right] \tag{2.1.7}$$

Here, we model neutral effects where interactions are the result of consumer choices influenced only by the relative abundance of resources (representing the availability of resource species in the landscape). The values in matrix $\mathbf{A}$ represent the predicted interaction counts, for each species pair, based solely on their abundances.

Step (e): we compute a composite interaction probability resulting from the com-

bined effects of trait matching ($\mathbf{M}$, equation (2.1.6)) and neutral abundance effects ($\mathbf{A}$, equation (2.1.7)) (Figure 2.1.2e). A factor δ ($0 \leq \delta \leq 1$) allows to control the relative weight of the matching and the abundance processes: if $\delta = 0$, then the strength of the matching is null, and interactions are driven only by abundances. If $\delta = 1$, abundances and matching concur to produce the observed pattern, and the strength of matching is maximal compared to other values of δ. The mixed interaction frequency matrix $\mathbf{A}^\star = [a_{ij}^\star]$ is computed as:

$$a_{ij}^\star = (m_{i|j})^\delta a_{ij} \tag{2.1.8}$$

where $m_{i|j} = m_{ij}/m_{+j}$ represents the probability of interaction with resources per consumer species and $m_{+j} = \sum_{i=1}^{r} m_{ij}$ is the marginal probability of interaction for species j.

The interaction probability matrix $\mathbf{P}^\star = [p_{ij}^\star]$ is then computed from the abundance matrix:

$$p_{ij}^\star = \frac{a_{ij}^\star}{a_{++}^\star} \tag{2.1.9}$$

where $a_{++}^\star$ is the grand total of $\mathbf{A}^\star$.

Step (f): finally, we sample observed interactions counts from the interaction probability matrix $\mathbf{P}^*$ (Figure 2.1.2f). To do so, we sample n_{inter} interactions from a multinomial distribution with $\kappa = rc$ outcomes corresponding to the pairwise interactions. The probability vector corresponds to the flattened matrix $\mathbf{P}^\star$, noted $\mathbf{p}^\star$ (of length κ).

$$\mathbf{z} \sim \mathcal{M}_{\kappa=rc}(n = n_{\text{inter}}, p = \mathbf{p}^\star) \tag{2.1.10}$$

The interactions counts obtained after sampling in the vector $\mathbf{z}$ are finally rearranged in the matrix $\mathbf{Z}$ ($r \times c$) which corresponds to the sampled interaction network.

Simulation parameters

To evaluate our method, we simulate data under the model described above. We conduct 4 experiments to study the influence of 4 parameters (the detailed parameter values are presented in Table 2.1.1):

Experiment	n_{inter}	Distribution of $\mathbf{n^r}$ and $\mathbf{n^c}$	μ_{breadth}	σ_{breadth}
1. n_{inter}	200 500 5000 1000 10 000	$\text{Lognormal}(\mu = \ln(3), \sigma = \ln(1.5))$	10	5
2. $\mathbf{n^r}$ and $\mathbf{n^c}$	5000	$\mathcal{U}(a = 0, b = 100)$ $\text{Lognormal}(\mu = \ln(10), \sigma = \ln(1.5))$ $\text{Lognormal}(\mu = \ln(3), \sigma = \ln(1.5))$ $\text{Lognormal}(\mu = 0, \sigma = \ln(10))$	10	5
3. μ_{breadth}	5000	$\text{Lognormal}(\mu = \ln(3), \sigma = \ln(1.5))$	2 10 20 50	1 5 10 25
4. σ_{breadth}	5000	$\text{Lognormal}(\mu = \ln(3), \sigma = \ln(1.5))$	10	0.1 1 5 10 50

Table 2.1.1: Parameter values used for the simulation study. Each experiment (first column) consists of simulations with different values for one of the parameters. Experiment 1 (n_{inter}) varies the total number of interactions. Experiment 2 ($\mathbf{n^r}$,$\mathbf{n^c}$) progressively increases the skewness of the species abundances. Experiment 3 (μ_{breadth}) changes the mean value of the consumers' niche breadth. Experiment 4 (σ_{breadth}) changes the standard deviation of the consumers' niche breadth. For Experiment 3 (μ_{breadth}), the standard deviation is also changed to keep a constant ratio of $\sigma_{\text{breadth}} = \mu_{\text{breadth}}/2$.

1. Sampling (n_{inter}): we vary the total number of interactions in the matrix n_{inter}, which represents the sampling intensity and allows to evaluate the robustness to incomplete sampling.

2. Species abundance distributions ($\mathbf{n^r}$ and $\mathbf{n^c}$): we sample $\mathbf{n^r}$ and $\mathbf{n^c}$ from a uniform distribution, and log-normal distributions either mildly, moderately- or very skewed. The aim is to evaluate the robustness of the method to skewed abundances typically encountered in ecological communities.

3. Average degree of specialization of consumers (μ_{breadth}): we vary the mean of the normal distribution in which the consumers traits variance s^2_{jk} ($k = 1, 2$) are sampled. This allows to contrast generalists consumers (high variance) with specialized consumers (low variance).

4. Heterogeneity of the degree of specialization (σ_{breadth}): we vary the variance of

the normal distribution in which the consumers traits variance s^2_{jk} ($k = 1, 2$) are sampled. This allows to contrast homogeneous consumers (same degree of specialization) with heterogeneous consumers (diverse degrees of specialization).

For each of the four experiments described above, we vary one of the simulation parameters, corresponding to the effect we study. The other parameters are fixed to a default value (see Table 2.1.1). For all simulations, we generated interactions between $r = 50$ resources and $c = 60$ consumers. We set the matching strength to $\delta = 0.2$, a value that was fixed empirically because it generated data where the relative strengths of abundance- and matching-driven processes seemed to both structure the observed matrix (results with other values of δ are presented in Appendix A).

For each experiment, we generated 100 datasets. For each dataset, we computed the true niche optimum as the mean of the traits of the interacting partners of a species, weighted by their interaction frequency. Similarly, the true niche breadth is computed as the variance of the traits of the interacting partners of a species weighted by the interaction frequency. These true values quantify species' realized niche (inference of the fundamental niche is discussed in Appendix B). Then, we performed CA and reciprocal scaling on each dataset and obtained estimates of the realized niches optima and breadths with these methods. To evaluate the performance, we measured the correlation between the niches optima and breadths estimated by reciprocal scaling and the true values of the realized niche. In order to choose which trait/axis pair correlation to measure, we chose the trait for which the correlation of optima is maximal on the first axis, and the other trait is correlated to axis 2.

2.6 Real data analysis

Dataset

To illustrate CA and reciprocal scaling on real ecological data, we used a bird-fruit interaction network from the ANDEAN frugivory dataset (Dehling et al., 2021) (the corresponding interaction matrix is shown in Appendix C). In this dataset, consumers are bird species and resources are the fruiting plants they feed on and whose seeds they disperse. This network was sampled in the lower montane rainforest in Peru (data Peru1 in the ANDEAN dataset). We chose this network as this is the one with the most

observations and sampling effort (960h) in the ANDEAN dataset. Data were collected by direct repeated observation conduced throughout an entire year (4 times between 2009 and 2010) in 3 plots of 100 m × 30 m. A transect was used to determine focus fruiting plants inside each plot and seed removal by birds were recorded on these plants.

Along with interactions, this dataset also includes species traits which are thought to have an important role in trait matching processes. Although CA and reciprocal scaling do not require traits, we will use these traits a posteriori to interpret the analysis. Plant traits were collected in the field, and bird traits on museum specimens. Four traits were collected for plants: fruit diameter, fruit length, crop mass (mean number of fruits per plant $\times$ mean fruit mass) and plant height. For birds, four traits were measured as well: bill width, bill length, body mass and Kipp's index. The Kipp's index is a measure of the pointedness of the wing: it is the Kipp's distance (distance from the tip of the first secondary to the wing tip) divided by wing length (Dehling et al., 2014). A low Kipp's index indicates rounded wings, and a high Kipp's index indicates pointed wings. In the following analyses, to ease interpretation, we discard fruit length and beak length, as these traits are functionally similar to fruit diameter and bill width.

Before data analysis, we filter out the birds or plants that interact only with one other species (8 birds and 12 plants). Indeed, since CA is based on species ordination, these species are problematic from a methodological point of view (Greenacre, 2013). These very specialized species can be seen as a part of an interaction module, as indicated in van Dam et al. (2021), who also advise analyzing modules separately. The final data has 40 plants and 53 birds (the list of these species, their traits and their corresponding codes can be found in Appendix E).

Analyses

We performed CA and reciprocal scaling on this dataset using the R package `ade4` (Thioulouse et al., 2018). This allows us to visualize species niches in the multivariate space and to compute their niche breadths and optima.

To show how our method can answer ecological questions, we investigated the drivers of niche breadth using the latent variables given by the multivariate axes. Therefore, we modeled the relationship between niche breadths and optima obtained by re-

ciprocal scaling for the first two axes. In total, we model 4 relationships (two axes for resources' and for consumers' niches). We consider linear and quadratic regression models and choose the best model with a likelihood ratio test.

Finally, to interpret CA and reciprocal scaling outputs, we computed a posteriori correlations between the coordinates of the species on each CA axis and measured species traits.

3 Results

3.1 Simulation

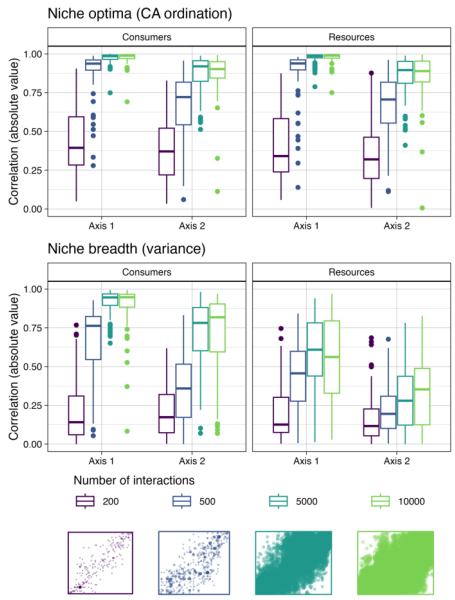
We carried out simulations studies to evaluate the capacity of CA and reciprocal scaling to infer species latent niches. We performed 4 experiments to test the influence of sampling intensity, species abundance distribution, consumer niche breadth and heterogeneity. The results are presented in Figure 2.1.3.

First, as expected, performance consistently improves with sampling completeness for all niche measures (Figure 2.1.3a). Note however that the niche breadth for resources on axis 2 is always the least well estimated. Second, in all reasonable cases and when there are enough data (Figure 2.1.3a), the niche optima are recovered correctly on both axes (median correlation value above 0.875), and niche breadths are recovered less precisely. Niche parameters are always better recovered on the first axis, which is consistent with the fact that, by construction, trait matching with the first trait explains more structure. Finally, consumers niche breadths are always better recovered than resource niche breadths.

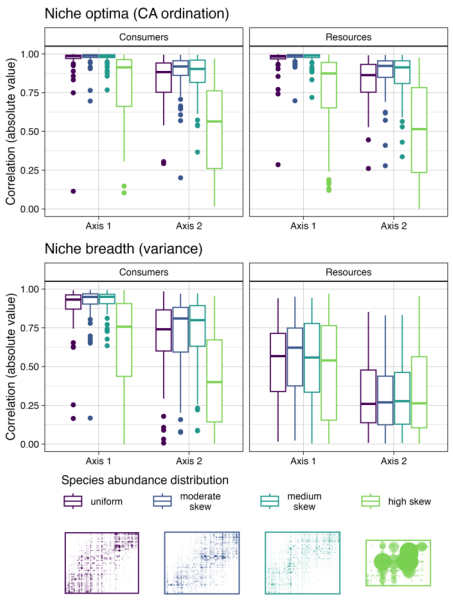
Regarding the effect of abundance distribution (Figure 2.1.3b), the model is robust to skewed abundance distributions, although performance drops a little with very skewed distributions.

Regarding consumer niche breadth (Figure 2.1.3c), the performance is better for niche breadth of intermediate size (the optimal performance is reached with a niche breadth of 10, representing 10% of the total length of the gradient in our setting).

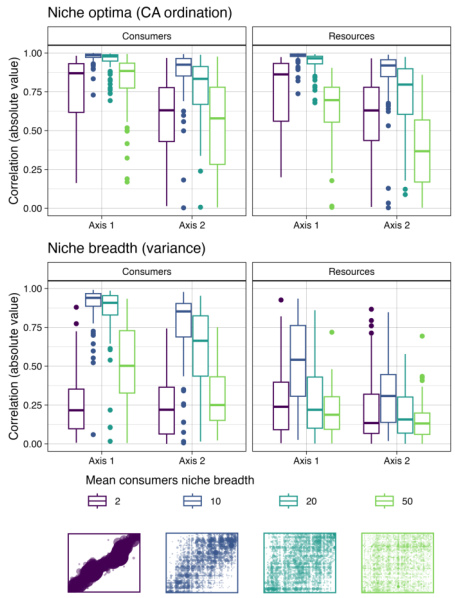
Finally, regarding the heterogeneity of consumers' niche breadths (Figure 2.1.3d), homogeneous niche breadths improve the performance. In particular, the niche



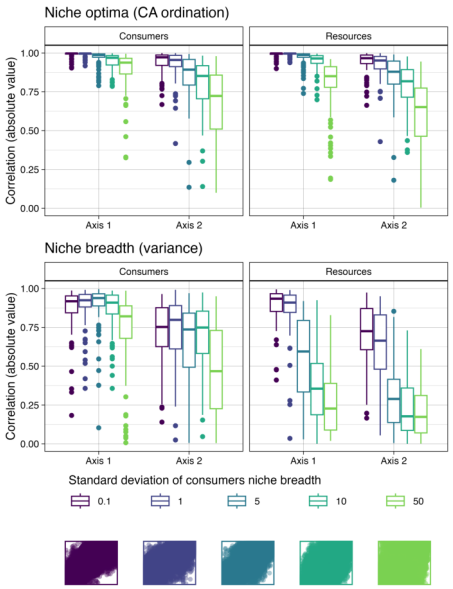
(a) Sampling intensity



(b) Abundance distribution



(c) Consumers niche breadth



(d) Heterogeneity of consumers niche breadth

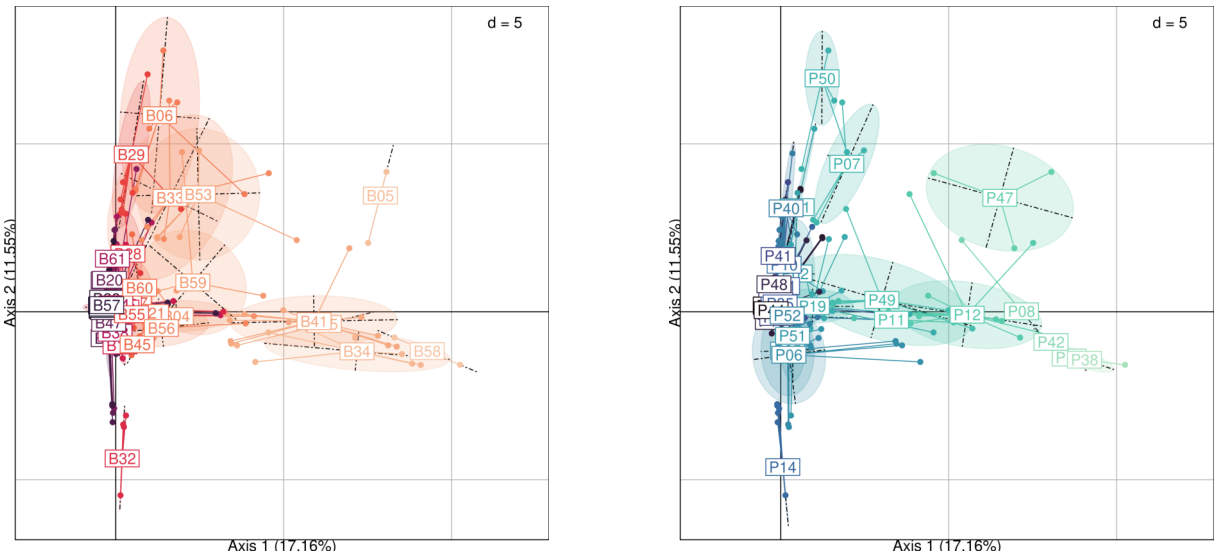
Figure 2.1.3: Results of the simulation study. Each subfigure explores the influence of one parameter on the model performance. (a) Effect of sampling intensity. (b) Effect of species abundance distributions. (c) Effect of the mean consumers niche breadths. (d) Effect of heterogeneity of consumers niche breadths. The y-axis is the absolute value of the correlation between true and estimated values of niche optima and niche breadths (respectively top and bottom of each subplot). Below the legend, matrices exemplify network data generated with the corresponding parameter value (with the size of points proportional to the number of observed interactions censored at the 1st and 99th percentile).

breadth of resources is correctly recovered only with very homogeneous consumer niches.

3.2 Data analysis

The permutation chi-squared test performed on the interaction table shows that there is a non-random structure in the network (χ^2 = 17784, 2000 permutations, p-value = 5.0×10^{-4}). The effect size computed with Cramér's V ($V = 0.29$, $IC_{95} = [0.26, 1.00]$), suggests a medium effect (Ellis, 2010).

We performed correspondence analysis and reciprocal scaling. CA eigenvalues suggest that matching between latent traits is quite strong. The square-root of the CA eigenvalues correspond to the absolute value of the correlation between latent traits: here, for the first three eigenvalues, we have $\sqrt{\lambda_1} = 0.79$, $\sqrt{\lambda_2} = 0.65$ and $\sqrt{\lambda_3} = 0.63$.



(a) Bird niches

(b) Plant niches

Figure 2.1.4: Reciprocal scaling of the birds-plants interaction network. Points correspond to the correspondences $h_k(i, j)$ in the first 2 dimensions. The correspondences are grouped by bird species (a) or plant species (b). The species label are placed at the reciprocal scaling mean, and the ellipses correspond to the bivariate normal distribution of variances and covariances given by reciprocal scaling (with a scaling factor of 1.5, i.e. the ellipse axes lengths are equal to $1.5\sqrt{\lambda_k}$ on axis k, corresponding to around 67% of the points contained in the ellipse). Species are colored according to their position on the first axis.

With reciprocal scaling, we can position interactions in the multivariate space, and visualize species niches as the mean and variance of these interaction scores. On Figure 2.1.4, interactions (points) are grouped by interacting partner (either birds or plants). Here, we separate the two groupings to avoid a cluttered representation, but the two graphs show niches in the same multivariate space. Figures 2.1.4a and 2.1.4b show interactions grouped by bird and plant species, respectively. The first 2 axes together explain 28.7% of the variance in the data (17.2% on axis 1 and 11.5% on axis

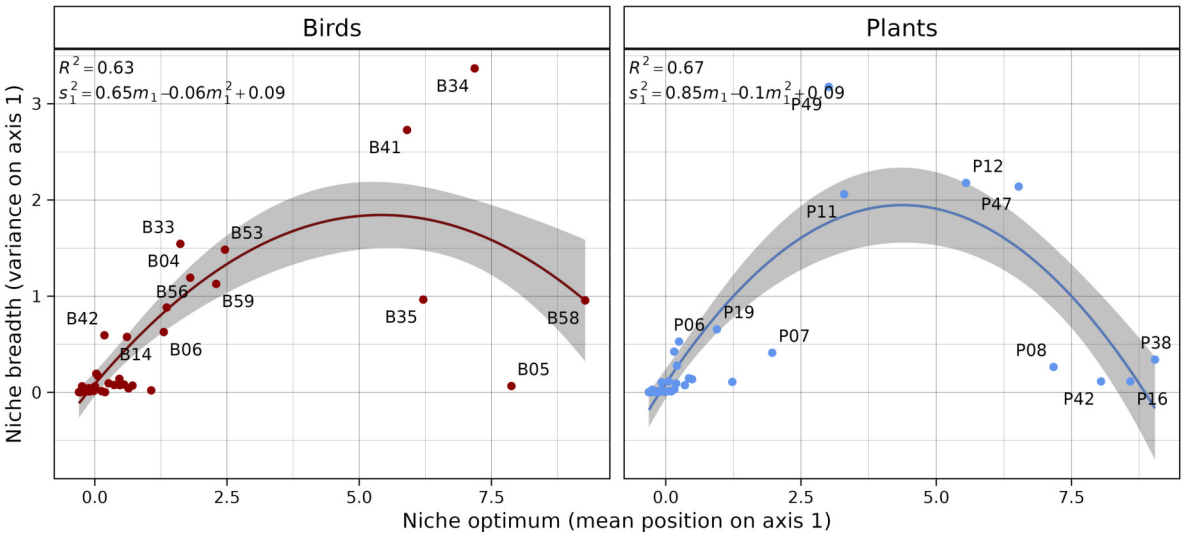
2). Eigenvalues suggest that the third axis also holds structure (10.9% variability), but here we do not show results on the third axis for concision.

On these figures, two birds (or plants) whose means are located nearby have similar niche optima. Moreover, a bird and a plant that interact preferentially have their niche optima nearby. For instance, if we consider B05 (*Cephalopterus ornatus*) and P47 (*Symplocos arechea*), their interaction with each other represent respectively 66% and 33% of their interactions. Finally, the length of the ellipses' axes correspond to the niche width along each multivariate axis (latent trait), and the inclination of the ellipse represents the covariance between the niche axes.

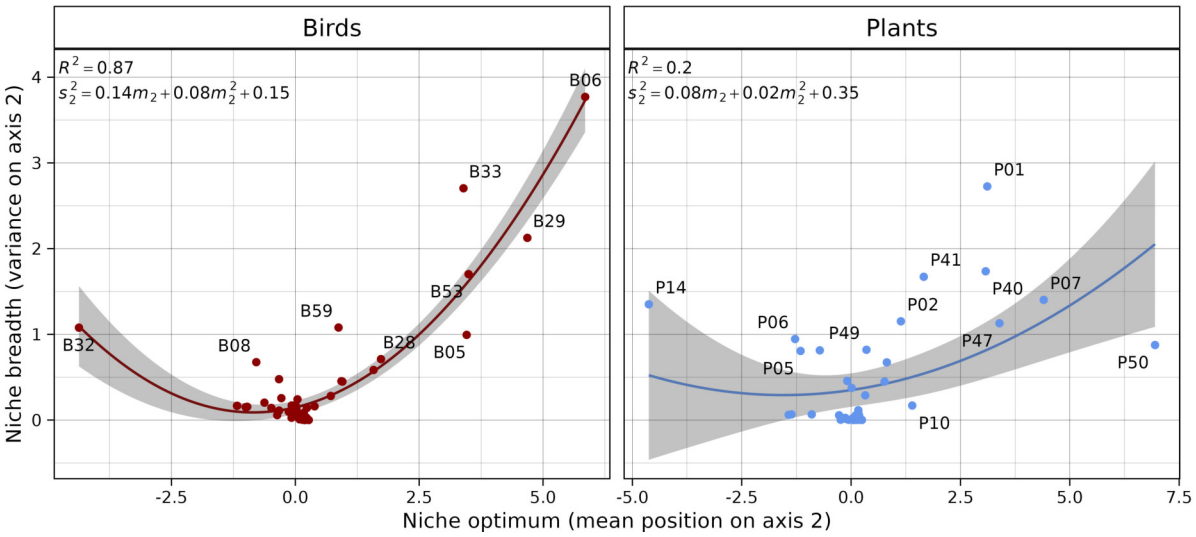
These figures also show how niches are distributed in the latent traits space. For birds (Figure 2.1.4a), many have a niche located around the origin (they interact with average plants, in the sense of their latent traits). Then, we can distinguish two clusters along the positive portion of the two main axes, which show two strategies of birds along the two latent plant traits axes: either interact with plants with above-average values on axis 1, and small values on axis 2, or the reverse. This open triangle is completed by a few bird species around the diagonal (e.g. B33 *Penelope montagnii*, B53 *Thraupis bonariensis* and B59 *Turdus chiguanco*) that seem to have an intermediate niche optimum and a high niche breadth.

For plants, many also have a niche optimum close to the origin (they interact with average birds, in the sense of their latent traits). As for birds, two clusters emerge along the two main axes, which also reflects two plant strategies. However, contrary to birds, we cannot clearly see plant species around the diagonal, except P47 (*Symplocos arechea*) that stands out against the two clusters. Species further from the origin also seem to have wider niches for both birds and plants; but the cluttered niches around the origin make it hard to investigate this pattern on these plots.

To better visualize the relationship between axis position and niche breadth, we plot the relationship between these two variables for each axis in Figure 2.1.5. We also model these relationships as described in the material and methods. On axis 1, there is a concave relationship with a positive linear component between niche optimum and niche breadth for birds, but the variance of the residuals increases along axis 1. For plants, there is a concave relationship only. On axis 2, we have a convex relationship for birds, with a strong linear positive component (the convex form seems to be driven



(a) Axis 1



(b) Axis 2

Figure 2.1.5: Niche breadth versus niche optimum on the 2 multivariate axes. Niche breadths (variances) and niche optima (means) were computed with reciprocal scaling. The top graphs (a) shows the relationship on axis 1 and the bottom graphs (b) on axis 2. The left and right panels represent bird and plant species, respectively. The solid line is the prediction of the best linear model and the gray shading represents the 95% confidence interval around the predicted mean value. The coefficient of determination (R^2) and the linear model equation are shown in the top left corner.

mainly by B32 (*Patagioenas plumbea*). For plants, we have a very weak relationship ($R^2 = 0.2$) with convex and linear positive components.

To interpret the latent trait axes, we correlated measured traits with plant and birds CA coordinates a posteriori. Figure 2.1.6 shows how traits are correlated with the multivariate axis and with each other. All traits correlate positively with axis 1, which strongly suggests a size effect along this axis. Axis 2 is characterized by its negative correla-

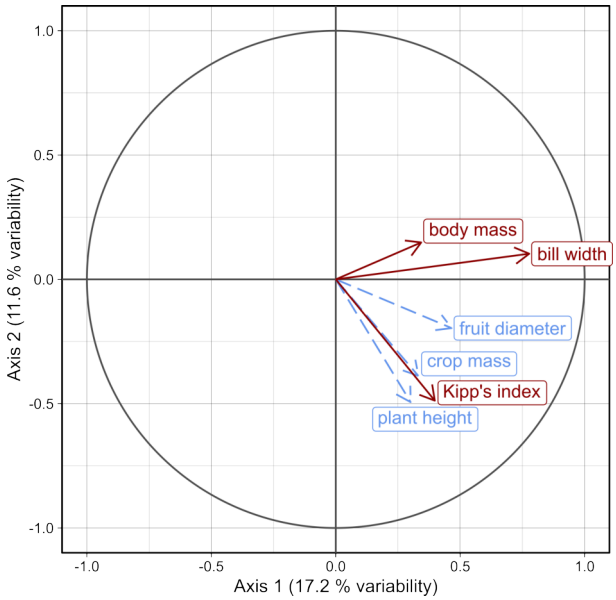


Figure 2.1.6: Correlation circle measured a posteriori between species traits and multivariate axes. Bird traits are shown in solid red lines and plant traits in dashed blue lines. Axis 1 shows a size effect while axis 2 is defined by the matching between plant height and Kipp's index.

tion with plant height, Kipp's index and crop mass. All plant traits are also strongly correlated with each other. For birds, body mass and bill width are highly correlated. Regarding cross-trophic level traits, Kipp's index is strongly positively correlated with crop mass and plant height.

4 Discussion

We evaluated CA and reciprocal scaling on both simulated and real interaction data to quantify species interaction niches. We also showed how this approach can be used to test hypotheses on the drivers of species position on the specialist/generalist gradient with the case study.

Reciprocal scaling provides a quantitative measure of the interaction niches. Species niche optima are approximated using the centroid, and species niche breadths using the variance of species interactions in the multivariate space. The resulting niche is then a hyperellipsoid in n dimensions, corresponding to the isocontours of a multivariate normal distribution in the same hyperspace. This definition of the niche bears a strong parallel with Hutchinson's definition as a hypervolume (Hutchinson, 1957). The latent approach of CA also has the advantage not to be limited to collected traits to assess niche patterns. However, latent gradients are also harder to interpret, especially

without external knowledge on the studied system.

4.1 Simulation

The simulation study shows good performance of this method to recover niche optima, but mixed performance regarding niche breadth. In particular, the niche breadths of resource species are poorly recovered overall, except in a few particularly favorable cases. Unsurprisingly, performance increases with sampling intensity (when n_{inter} increases, Figure 2.1.3a). When the proportion of variance explained by latent traits is smaller (axis 2), performance drops. The estimation is robust to skewed data, but performance decreases with extremely skewed data: in that case, we can hypothesize that the niche of rare species is estimated less precisely because they have very few observed interactions.

Consumers' niche breadth has an effect on the performance. This was expected, because of the way CA orders individuals on a latent gradient. For instance, if we consider consumers, they are ordered so that two consumers interacting with similar resource species are positioned nearby. Hence, CA needs some overlap in interactions to position consumers on the gradient, but not too much or the signal weakens. Therefore, there is a trade-off between narrow and wide niches: they need to overlap enough to order adjacent species, but not too much so that there is still a structure in the network. This trade-off is well reflected in the performance of the inference in Figure 2.1.3c.

The heterogeneity of consumers niche breadth decreases the performance of the estimation for both niche optima and niche breadth. We can hypothesize that this is due to insufficient overlap with other species' interactions for specialist species, which might bias the placement of niche optima and the estimation of niche breadths for specialist species and their interacting partners.

Finally, niche parameters are better recovered for consumers than for resources. While niche optima are recovered fairly well for both trophic levels, niche breadths are significantly less well recovered for resources. Differences might arise from the asymmetric way niche breadths are specified in our simulation model: indeed, niche breadths for resources are not controlled and driven entirely by the niche breadth of the consumers they interact with.

CA is able to recover a signal of matching rather accurately for any trait matching strength, which we can evaluate by examining eigenvalues and chi-squared tests (see Appendix A). Paradoxically, a stronger matching does not translate to a better recovery of the latent traits, as discussed in Appendix A. Indeed, if the matching is too strong, each species becomes increasingly separated of the others on the ordination, and CA is unable to find the common structuring driver without external information.

Regarding the simulation model, it makes a number of simplifying hypotheses on interaction networks. First, in our model, only consumers have a preferred niche breadth. This makes sense in our case study with birds (as consumers) and plants (as resources), because birds' cognitive processes and movement abilities allow them to choose the plants with which they interact: but on an evolutionary timescale, plants might evolve to attract more diverse or more similar birds, thereby effectively evolving a preferred niche breadth. However, specifying a model taking into account both preferred niche breadths to determine matching (i.e. adapting Equation (2.1.6)) is difficult, because it requires to combine both niche breadths in the trait matching step. In fact, to our knowledge, no such model exists in the literature. The model also specifies trait matching with a multivariate normal distribution, which is a reasonable choice if the real-world interactions are at least symmetrical. Despite theory and observations supporting this hypothesis for environmental niches, especially for plants (e.g. Shelford's law of tolerance, Erofeeva, 2021; Shelford, 1931 and observations of Gauch Jr. and Whittaker, 1972), for interaction networks this hypothesis is debated (Blonder et al., 2014; Benadi et al., 2022). Despite these limitations, we think that the simulation model used here was useful to get an idea of the method's ability to recover niche parameters, before applying it to real data.

4.2 Data analysis

Regarding real data analysis, a permutation chi-squared test suggests that there is a non-random structure in the network. However, the network is also very noisy, as suggested by visual examination of the network and medium effect size (Cramér's $V = 0.29$, $IC_{95} = [0.26, 1.00]$). This non-random structure is quantified as interaction counts that deviate from their expected value in the sense of the chi-squared statistic. However, to assess the expected number of interactions, the chi-squared statistic uses

the matrix margins, thus assuming that the number of times a species is observed in the network is a surrogate of its abundance. This hypothesis might not be valid for real data due to sampling effects. In particular, the interactions range of the most abundant species might be better sampled, thus artificially increasing their generality, and reciprocally under-sampling of rare species might artificially increase their specialization. This is a well-known problem highlighted in simulation studies (Blüthgen et al., 2008; Fründ et al., 2016). Finally, this network represents a subsample of species, and some species in the network were probably not sampled, which may affect the network structure.

To interpret latent traits given by the multivariate axes, we correlated species coordinates on the multivariate axes (giving their latent traits) with their measured traits. Because of allometry, traits are strongly correlated within each trophic level. Tall plants also tend to produce larger fruits and yield a higher crop mass, while large birds also tend to have a wide beak. We can also understand the (smaller) positive correlation between Kipp's index (wing pointedness) and other bird traits, because large birds tend to reside in the more open canopy where they can circulate more freely (Pearson, 1971). Traits of birds and plants are also generally positively correlated, which suggests some degree of trait matching.

In the results, we explored the niche patterns for the first two axes (or latent traits). We find relationships between niche optimum and niche breadth for birds and plants. There is a concave relationship between niche breadth and the first latent trait for birds and plants, and a convex relationship between birds niche breadth and the second latent trait for birds.

The first axis is mostly correlated with birds and plants size. For birds (Figure 2.1.5a, left), there is a convex relationship with a positive linear component between niche breadth and axis 1, which suggests that the larger birds are, the wider their niches are, but that the birds with the widest niches are medium-sized birds. Here, a wide niche means a high diversity of fruits sizes eaten. To interpret this result biologically, we can hypothesize that birds are limited by beak size. Small-billed birds may not be able to grasp large fruits, while bigger-billed birds can swallow even small fruits (Wheelwright, 1985). However, birds that are too large do not eat very small fruits because they need to satisfy their high absolute energy expenses and cannot spend too much time and en-

ergy foraging, so they focus on fruits that are larger or found in higher abundance on the plant (crop mass) that minimize foraging time (Schoener, 1971; Albrecht et al., 2018). This limiting effect of energy requirements is weak: the graph rather shows a diversity of strategies for large birds, with for instance B34 (*Pharomachrus antisianus*) and B41 (*Rupicola peruvianus*) eating more diverse-sized fruits and B05 (*Cephalopterus ornatus*) and B58 (*Trogon personatus*) choosing fruits of more uniform sizes. In addition, the correlation between birds' body mass and indices of plant productivity such as crop mass or fruit diameter is weak (Figure 2.1.6). This weak effect of energy traits may be due to the fact that birds foraging on fruits is partially modulated by whether they are fruit-gulpers or pulp-feeders (Palacio et al., 2017). These large birds can also have diverse ways of handling fruits, as well as other food sources (Foster, 1987).

For plants (Figure 2.1.5a, right), the convex relationship between niche breadth and axis 1 is clearer. It suggests that plants of intermediate size attract more diverse-sized birds, while small fruits are only consumed by small birds, and large plants are reserved to big birds. To characterize plants with intermediate coordinates and wide niches on the first axis, we examine the reciprocal scaling plot 2.1.4b. We can see that P11 (*Elaeagia mariae*) and P49 (*Turpinia occidentalis*) predominantly interact with small and medium-sized birds, while P12 (*Endlicheria sp.*) and P47 (*Symplocos arechea*) interact with medium to large birds. A closer examination of their respective traits reveal that the large size of P11 and P49 on the first axis is predominantly driven by their height, so it does not impose a harsh barrier for interactions. On the contrary, P12 and P47 produce large fruits, in particular larger than the mean bird bill size, so the interaction barrier for small birds is stronger.

On axis 2, we interpret the relationship for birds only, because the relationship between the latent trait and niche breadth for plants is very weak. Figure 2.1.5b (left) shows a concave effect with a large positive linear component. In this discussion, we will focus on this linear component, because the convex form seems to be driven by a single species (*Patagioenas plumbea* B32). The second axis is negatively correlated to Kipp's index (wing pointedness) and plant height. So birds with pointed wings have smaller niches on the second axis, i.e. they consume fruits from plants of homogeneous heights, while birds with rounded wings have wider niches (i.e. they consume fruits from plants of heterogeneous heights). We can interpret these results as prefer-

ences driven by foraging space. Indeed, rounded wings confer more maneuverability in dense vegetation (Thiel et al., 2023), so birds with rounded wings primarily feed on small plants from the dense understory. On the contrary, birds with pointed wings preferentially forage on tall plants from the more open canopy. Here, we find a higher specialization for birds with a high Kipp's index (pointed wings). We hypothesize that these birds might find it difficult to navigate dense understory where vegetation is denser, while birds with rounded wings might move to the less dense canopy at a lesser cost. It is also possible that top canopy bird focus on highly productive fruit patches in the canopy by choosing plants with high crops (as suggested by the high correlation between crop mass and Kipp's index in Figure 2.1.6), whereas birds with rounded wings also supplement their diet with other food sources such as insects (Foster, 1990; Thiel et al., 2023). The higher specialization of canopy birds seems in contradiction with the findings of Schleuning et al. (2011): however, in this study, the authors measure specialization with an index considering the diversity of interacting partners and not the diversity of their (latent) traits.

Finally, we must keep in mind that we observe realized niches here, so these niches might not reflect the "true" fundamental preference of birds. In Appendix B, we explore the performance of CA and reciprocal scaling to infer species fundamental niches. We show that, in our simulation setting, the method can also recover fundamental niche parameters, although not as well as the realized niches. The realized niches of birds and plants and might be constrained by two external factors. First, the distribution of species traits in the community might constrain species to settle for suboptimal interacting partners. For instance, here most plants species yield small fruits and have a small crop mass, so large birds might not have a choice with regard to the fruit size they ingest and might be constrained to also eat small fruits. Second, competition may also constrain species to differentiate their interaction niche from other species in the community.

Although the analysis produces sensible results, the two axes discussed above explain roughly 30% of the variability, so these patterns cover only a small fraction of the non-random network structure. In particular, discussing patterns on the third axis would be important, as it explains almost the same variability as the second axis. Moreover, the eigenvalues' distribution shows a gap between the 3rd and 4th and 6th

and 7th eigenvalues, thus suggesting that underlying structures are found up to the 6th axis: considering the first 6 axes explains more than 60% of the variation of the network.

4.3 Perspectives

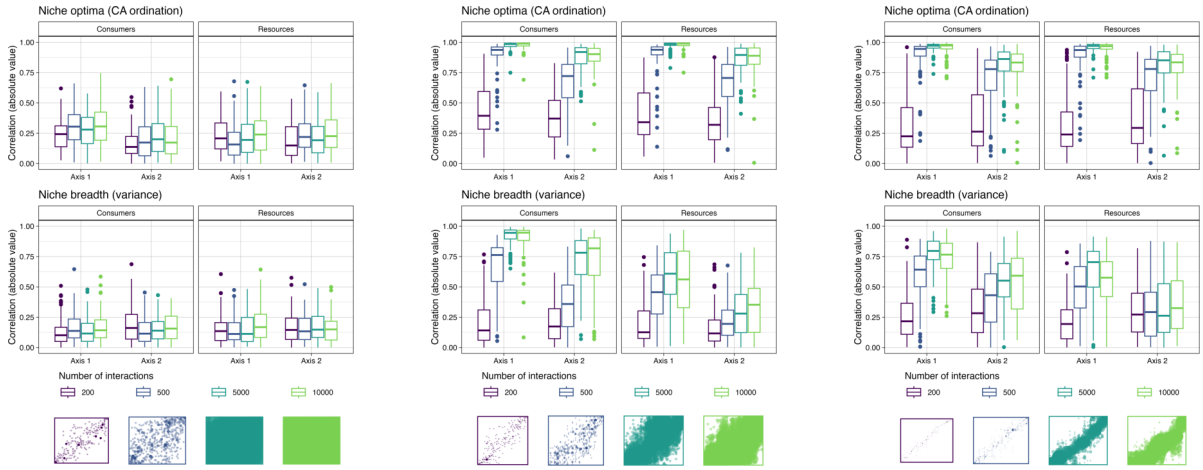
The method and analyses presented above address important questions and open avenues to investigate the structure of interaction networks. CA and reciprocal scaling make a conceptual and quantitative link between species interaction niches, trait matching and specialization/generalization. Using latent traits allows the analysis to be data-driven, not limited by costly trait collection and suited for exploratory analyses. However, when traits are present, it would also be interesting to constrain the multivariate axes with bird, plant or both trophic levels traits. Such constrained analyses have been performed with RLQ or fourth-corner in the context of network data (Albrecht et al., 2018; Dehling et al., 2014; Bender et al., 2018). However, fourth-corner does not allow to visualize data, and RLQ is covariance-based and does not allow to partition the variation explained by traits in interaction networks.

An alternative would be to use constrained versions of CA, like canonical correspondence analysis (using one trait table) (C. J. F. ter Braak, 1986) or double-constrained correspondence analysis (using both trait tables) (C. J. F. ter Braak et al., 2018). This would allow to examine the relationship with traits and the multivariate axes, partition variance explained by traits and measure trait matching (with double-constrained correspondence analysis). Moreover, canonical methods would prevent the separation of specialized species on different axes of the multivariate space when a common trait explains the network structure (results described in Appendix A and discussed above). To our knowledge, constrained versions of correspondence analyses have never been applied to interaction networks. Moreover, for all methods above, there is not a known way to measure niche breadth because reciprocal scaling is restricted to CA. Further methodological developments in multivariate analyses could therefore benefit network analysis.

Appendices

A Effect of the relative importance of matching on the recovery of niche parameters

Here, we investigate how the model performs when the network is more or less structured by trait matching compared to neutral processes. For that, we simulate networks with different values of the parameter δ from Equation (2.1.8): $\delta = 0$ (network structured only by neutral processes), $\delta = 0.2$ (intermediate strength of trait matching) or $\delta = 1$ (maximum strength of trait matching). For each value of δ, we generate networks with different sampling intensities (corresponding to experiment n_{inter} in the main text). Results are shown in Figure A.1 below.



(a) $\delta = 0$ (b) $\delta = 0.2$ (c) $\delta = 1$

Figure A.1: Effect of the relative importance of matching on the inference of niche parameters. Each subfigure shows the result of the inference for a different value of the trait matching strength δ. For a more complete description of the figure axes, panels and example matrices, refer to Figure 2.1.3 from the main text.

As expected, when there is no trait matching ($\delta = 0$), niche parameters are poorly recovered by CA and reciprocal scaling, because the true niche is random. Surprisingly, the performance is best when the matching strength is intermediate ($\delta = 0.2$), and decreases with a stronger matching ($\delta = 1$).

If we examine the associated eigenvalues in Figure A.2, we can see that, as expected, the matching signal is better recovered when $\delta = 1$. This plot shows the square root of eigenvalues (correspond to the absolute value of correlation between

latent traits) for different values of matching strength δ and different sampling intensities. For a given sampling intensity, the eigenvalues are closer to one as δ increases. When $\delta = 0$, the (spurious) signal detected at low sampling intensity vanishes for all axes when the sampling improves. No matching is detected, so the model fails to recover niche parameters that are random and cannot be inferred from the matrix when $\delta = 0$. When $\delta = 0.2$, the two first eigenvalues become increasingly separated from other eigenvalues as sampling intensity increases. This means that two latent traits are inferred, that are well correlated to the two true traits used in the simulation, so the niche parameters are better recovered. When $\delta = 1$, we have high eigenvalues, but on all axes, thus indicating that the model recovers structuring latent traits on all axes. Therefore, the latent traits on the first two axes correspond less closely to the two true traits used in the simulation, thus yielding less accurate estimations of the niche parameters.

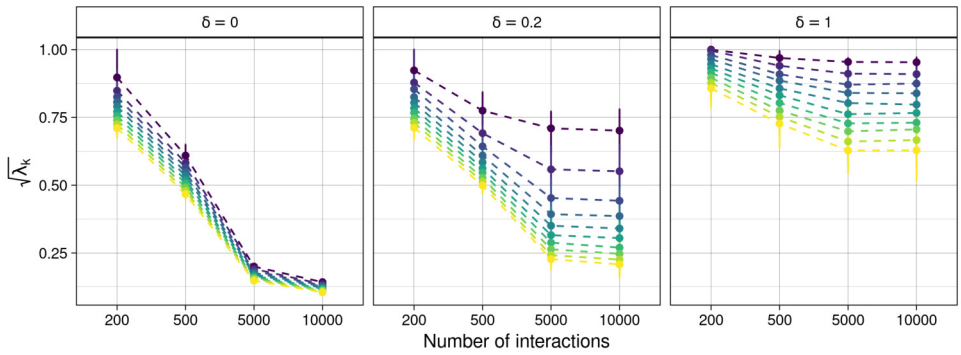
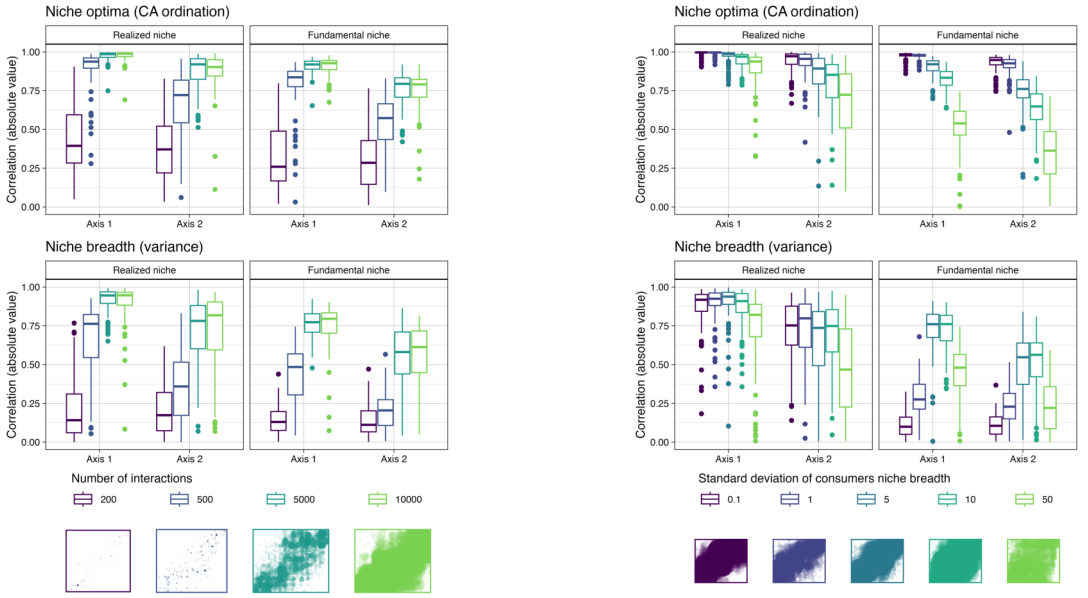


Figure A.2: Eigenvalues evolution with sampling intensity for different values of δ. This plot shows how eigenvalues evolve with sampling strength, for different values of δ (panels). Colored points and dotted line represent an eigenvalue, and error bars represent the 2.5 and 97.5-th percentiles on the eigenvalues over 100 repetitions.

B Realized and fundamental niches

In the article, we measured the model performance by comparing the inferred niche parameters to the realized niches. However, we could also compare the inferred niche parameters to the fundamental niches, i.e. the trait values (and niche breadths, for consumers) fixed in the model. Below, we present the performance of the model when we consider the fundamental niches as the ground truth (results are shown only for consumers). For simulation parameters, we choose the same parameters as described in 2.1.1.



(a) Sampling intensity

(b) Heterogeneity of consumers niche breadth

Figure B.1: Performance of the model to infer fundamental niche parameters. (a) Shows the effect of the sampling intensity and (b) shows the effect of the heterogeneity of consumers niche breadths on the model performance. y-axis and matrices have the same meaning as in Figure 2.1.3. The correlation is presented separately for realized and fundamental niches (respectively left and right panels) on axes 1 and 2.

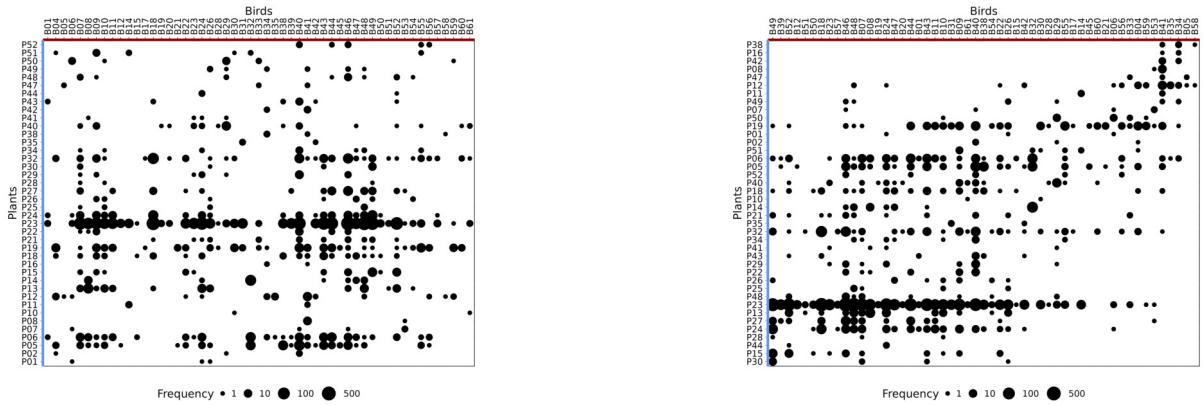
In most cases, the performance of the model to recover fundamental niche is a reflection of the realized niche performance, but with slightly less good performance. As an example, we present the model performance for realized and fundamental niches depending on sampling intensity in Figure B.1a.

However, the performance pattern is different when the true niche breadth varies: whereas the realized niche is less well recovered as the variance of niche breadth increases, the fundamental niche is better recovered for intermediate values of niche breadths (Figure B.1b). We can see it as an artifact of the performance metric used

here: when niches breadths are very homogeneous, the model is able to recover the absolute niche breadths, but not their relative order because they are very homogeneous.

C Networks visualization using CA

Correspondence analysis can also be useful to reorder rows and columns of the analyzed interaction matrix. Many ways to visualize interaction networks have been proposed, with all of them highlighting a specific aspect of network structure like centrality, modularity or nestedness (Araujo et al., 2010; Miele et al., 2019). CA can also serve this purpose. The figure below shows the interaction matrix corresponding to the bipartite birds-plants interaction network analyzed in this article. In Figure C.1a, rows and columns are ordered according to the alphabetical order, while in Figure C.1b they are ordered according to their rank on the first axis of the CA. This means that species are positioned in an ecologically meaningful way, according to their interaction niche optima on the first latent trait axis. Here, we can see two weakly connected modules, corresponding to the bulk of small plant and bird species that interact together (bottom-left corner), and fewer large birds and plants that also interact together (top-right corner). A few species interact with species from both modules (like P12 (*Endlicheria sp.*) and B41 (*Rupicola peruvianus*)), and correspond to the generalist intermediate species from Figure 2.1.5.



(a) Raw matrix

(b) Reordered matrix

Figure C.1: Interaction matrix reordered using correspondence analysis. This figure shows how an interaction matrix can be reordered using CA. Rows and columns of matrix (a) are ordered alphabetically and for matrix (b) they are ordered following the coordinates of species on the first axis of CA.

D Reciprocal scaling formulas

Reciprocal scaling allows to compute niche optima and niche breadths on each axis k, as well as the covariance between two niche axes k and l. Each of these quantities can be expressed either with the scores $\mathbf{h}_k$ derived from reciprocal scaling or equivalently using the scores $\mathbf{u}_k^\star$ and $\mathbf{v}_k^\star$ of correspondence analysis, so each quantity has two equivalent expressions. The corresponding formulas are written below, adapted from Thioulouse and Chessel (1992) to match our notation.

Niche optima

Niche optima can be expressed either as a weighted mean of the correspondences scores, or from the CA scores. For the resource species i, we have:

$$m_{ik} = \begin{cases} \frac{1}{y_{i+}} \sum_{j=1}^{c} y_{ij} h_k(i,j) \\ \frac{\sqrt{\mu_k}}{\sqrt{2\lambda_k}} u_{ik}^\star \end{cases} \tag{2.1.11}$$

where $\mu_k = 1 + \sqrt{\lambda_k}$.

Similarly, for consumer species j, we have:

$$m_{jk} = \begin{cases} \frac{1}{y_{+j}} \sum_{i=1}^{r} y_{ij} h_k(i,j) \\ \frac{\sqrt{\mu_k}}{\sqrt{2\lambda_k}} v_{jk}^\star \end{cases} \tag{2.1.12}$$

Niche breadths

The niche breadth can be expressed as the weighted variance of the correspondences scores or from the CA scores as well. For resource species i, we have:

$${s_{ik}}^2 = \begin{cases} \frac{1}{y_{i+}} \sum_{j=1}^{c} y_{ij} \left(h_k(i,j) - m_{ik}\right)^2 \\ \frac{1}{2\lambda_k \mu_k} \left(\frac{1}{y_{i+}} \sum_{j=1}^{c} \left(y_{ij} {v_{jk}^\star}^2 \right) - \lambda_k {u_{ik}^\star}^2 \right) \end{cases} \tag{2.1.13}$$

Similarly, for consumer species j, we have:

$${s_{jk}}^2 = \begin{cases} \frac{1}{y_{+j}} \sum_{i=1}^{r} y_{ij} \left(h_k(i,j) - m_{jk}\right)^2 \\ \frac{1}{2\lambda_k \mu_k} \left(\frac{1}{y_{+j}} \sum_{i=1}^{r} \left(y_{ij} {u_{ik}^\star}^2 \right) - \lambda_k {v_{jk}^\star}^2 \right) \end{cases} \tag{2.1.14}$$

Covariances

We can express the covariance between two niche axes from the reciprocal scaling of the CA scores as well. The covariance between niche axes k and l for resource species i is written:

$$c_{kl}(i) = \begin{cases} \frac{1}{y_{i+}} \sum_{j=1}^{c} y_{ij} h_k(i,j) h_l(i,j) - m_{ik} m_{il} \\ \frac{1}{2\sqrt{\lambda_k \lambda_l}\sqrt{\mu_k \mu_l}} \left(\frac{1}{y_{i+}} \sum_{j=1}^{c} y_{ij} v_{jk}^{\star} v_{jl}^{\star} - \sqrt{\lambda_k \lambda_l} u_{ik}^{\star} u_{il}^{\star} \right) \end{cases} \tag{2.1.15}$$

For consumer species j, the covariance is written as:

$$c_{kl}(j) = \begin{cases} \frac{1}{y_{+j}} \sum_{i=1}^{r} y_{ij} h_k(i,j) h_l(i,j) - m_{jk} m_{jl} \\ \frac{1}{2\sqrt{\lambda_k \lambda_l}\sqrt{\mu_k \mu_l}} \left(\frac{1}{y_{+j}} \sum_{i=1}^{r} y_{ij} u_{ik}^{\star} u_{il}^{\star} - \sqrt{\lambda_k \lambda_l} v_{jk}^{\star} v_{jl}^{\star} \right) \end{cases} \tag{2.1.16}$$

E Species names

Throughout the article, we use codes to describe species. Below are the correspondence between these codes and the species names, alongside with their traits.

Code	Species name	Fruit diameter (mm)	Plant height (m)	Crop mass (g)
P01	*Anthurium obtusum*	2.9	5	0.8
P02	*Banara arguta*	9	5.2	241.5
P05	*Cecropia polystachya*	20.4	13	6813.4
P06	*Cecropia strigosa*	9.2	12.1	3813.1
P07	*Cestrum parqui*	5.9	3.9	470.8
P08	*Cestrum stipulatum*	6.9	4	225
P10	*Croton rusbyi*	6.5	6.8	7.2
P11	*Elaeagia mariae*	2.8	9.8	22.4
P12	*Endlicheria sp*	12.3	13.8	3986.1
P13	*Ficus americana*	3.7	15	606.7
P14	*Ficus coerulescens*	6.2	15	5400
P15	*Graffenrieda cucullata*	4.7	8.3	58.1
P16	*Guatteria duodecima*	17.4	13.3	7372.2
P18	*Guettarda crispiflora*	8.5	10.5	1448.6
P19	*Hedyosmum tepuiense*	7.7	5.3	583.5
P21	*Isertia laevis*	11.6	10.1	1100.5
P22	*Miconia affinis*	5	6.3	222.6
P23	*Miconia aprica*	3.2	7	333.7
P24	*Miconia barbeyana*	4	3.3	32.6
P25	*Miconia calvescens*	4.2	4.7	121.4
P26	*Miconia cyanocarpa*	3.6	4.1	416.3
P27	*Miconia egensis*	2.2	6.5	76.3
P28	*Miconia hygrophila*	3.3	4.2	18.4
P29	*Miconia longifolia*	4.2	12.3	293.3
P30	*Miconia myriantha*	2.7	7.3	7.9
P32	*Miconia sp3*	10.1	4	488.1
P34	*Miconia spennerostachya*	5.1	5.2	632.4
P35	*Myrcia fallax*	6.2	10.7	544.6
P38	*Ocotea sp*	8.5	12	850
P40	*Peristethium polystachyum*	3.7	3.5	139.4
P41	*Phoradendron chrysocladon*	3.2	7	16.4
P42	*Pleurothyrium trianae*	10.7	10.8	574.5
P43	*Rubus floribundus*	9.9	5.4	851
P44	*Schefflera mathewsii*	5.2	7.5	1951.3

Code	Species name	Fruit diameter (mm)	Plant height (m)	Crop mass (g)
P47	*Symplocos arechea*	12.6	9.8	3650.6
P48	*Trema micrantha*	2.2	9.1	1.5
P49	*Turpinia occidentalis*	8.3	7.2	235.9
P50	*Viburnum hallii*	6.1	5.5	249.3
P51	*Vismia gracilis*	10.4	9.5	1087
P52	*Zanthoxylum mantaro*	3.8	6.1	27.9

Table E.1: Correspondence between plant codes and species names.

Code	Species name	Kipp's index	Bill width (mm)	Body mass (g)
B01	*Anisognathus somptuosus*	0.19	10.6	42
B04	*Catharus ustulatus*	0.32	9.5	30.3
B05	*Cephalopterus ornatus*	0.17	30.2	380
B06	*Chiroxiphia boliviana*	0.16	9.2	17.2
B07	*Chlorochrysa calliparaea*	0.24	7.5	17
B08	*Chlorophonia cyanea*	0.28	6.5	14
B09	*Chlorospingus flavigularis*	0.17	10.1	25.6
B10	*Chlorospingus ophthalmicus*	0.16	7.7	15.7
B11	*Chlorospingus parvirostris*	0.16	8.5	24
B12	*Colaptes rubiginosus*	0.23	11.2	65.4
B14	*Cyanocorax yncas*	0.08	15	78.5
B15	*Dacnis cayana*	0.25	7.7	13
B17	*Diglossa cyanea*	0.17	7.2	17.1
B18	*Diglossa glauca*	0.19	5.4	12
B19	*Elaenia albiceps*	0.23	8.4	15.5
B20	*Elaenia pallatangae*	0.21	7.1	17.8
B21	*Entomodestes leucotis*	0.21	12.9	61.7
B22	*Eubucco versicolor*	0.14	12.6	32.8
B23	*Euphonia mesochrysa*	0.21	7.4	13
B24	*Euphonia xanthogaster*	0.22	7.4	13
B26	*Iridosornis analis*	0.16	8.4	26
B28	*Lophotriccus pileatus*	0.13	7.6	8.1
B29	*Mionectes striaticollis*	0.15	8.8	15
B30	*Myadestes ralloides*	0.21	12	29.1
B31	*Ortalis guttata*	0.15	13.5	550
B32	*Patagioenas plumbea*	0.35	8.1	179
B33	*Penelope montagnii*	0.15	16.5	706

Code	Species name	Kipp's index	Bill width (mm)	Body mass (g)
B34	*Pharomachrus antisianus*	0.42	21.8	153
B35	*Pharomachrus auriceps*	0.37	22.1	180
B38	*Psarocolius angustifrons*	0.25	17.3	218.5
B39	*Psarocolius atrovirens*	0.25	15.9	152
B40	*Ramphocelus carbo*	0.15	11.4	28
B41	*Rupicola peruvianus*	0.12	20.9	244
B42	*Saltator maximus*	0.18	13	47.7
B43	*Tangara arthus*	0.22	8.5	22
B44	*Tangara chilensis*	0.23	7.4	23
B45	*Tangara chrysotis*	0.26	8.2	24
B46	*Tangara cyanicollis*	0.22	7.7	17
B47	*Tangara gyrola*	0.2	7.7	21
B48	*Tangara nigroviridis*	0.23	7.8	17
B49	*Tangara punctata*	0.21	6.8	15
B50	*Tangara ruficervix*	0.24	7.6	19
B51	*Tangara vassorii*	0.23	8	18
B52	*Tangara xanthocephala*	0.22	7.3	19
B53	*Thraupis bonariensis*	0.2	11	36
B54	*Thraupis cyanocephala*	0.16	10.7	36
B55	*Thraupis episcopus*	0.23	10.2	35
B56	*Thraupis palmarum*	0.22	9.6	39
B57	*Trichothraupis melanops*	0.19	9.2	24.3
B58	*Trogon personatus*	0.37	16	63.4
B59	*Turdus chiguanco*	0.22	13.5	93.3
B60	*Xenopipo unicolor*	0.19	10	15.5
B61	*Zimmerius bolivianus*	0.14	6.5	11

Table E.2: Correspondence between bird codes and species names.

Chapter 2.2

Constrained correspondence analyses to analyze the structure of interaction networks, or trait matching with traits

I intend to use the results of this chapter for an upcoming publication.

1 Introduction

In the previous chapter (2.1), we used an interaction network to infer species interaction niches using the notion of latent traits. To do so, we used correspondence analysis to associate latent traits to species that best explained their interaction patterns. This method is useful to explore data, but does not allow to test the effect of specific traits on network structure. In the present chapter, I describe methods allowing to constrain multivariate axes with measured traits, and to quantify the part of variation due to these traits.

Multivariate methods are very popular to analyze the link between species abundance and environmental factors, allowing to investigate the notion of species environmental (or Grinnellian) niche (Green, 1971, 1974; C. J. F. ter Braak & Verdonschot, 1995; Dolédec et al., 2000). In the previous chapter, we presented correspondence analysis (CA), which is an indirect gradient analysis method, i.e. a method in which multivariate axes are not constrained with external variables. In addition to CA, constrained (or canonical) methods such as canonical correspondence analysis (CCA) and double-constrained correspondence analysis (dc-CA) have also been used to characterize the environmental niche (Dolédec et al., 2000; C. J. F. ter Braak & Verdonschot, 1995; C. J. ter Braak & Prentice, 2004).

In this chapter, I investigate the ecological implications of analyzing species interactions instead of species abundance data, from a niche point of view. If we consider bipartite interaction networks (i.e. networks with two sets of species interacting between sets but not within sets), we can use the same algebra that is classically used to analyze species-environment relationships to understand how traits shape interaction networks. However, this implies a different biological interpretation: instead of defining environmental niches (see box 2.1), we quantify interaction (Eltonian) niches. Here, I use the same convention as in chapter 2.1 and call the two sets of species resources and consumers.

To investigate interaction niches, we can use canonical correspondence analysis (CCA) (C. J. F. ter Braak, 1986). Depending on the table used to constrain the analysis, CCA will define resource or consumer niches: resource niches are defined when the analysis is constrained with consumer traits and conversely. To investigate interaction niches with double constrained correspondence analysis (dc-CA) (C. J. F. ter Braak

et al., 2018), we use both resource and consumer traits. dc-CA defines resource and consumer species niches constrained with their traits. We can draw a parallel between these interaction niches and species environmental niches defined by CCA and dc-CA in the context of species-environment analyses (see box 2.1 and C. J. F. ter Braak and Verdonschot, 1995).

Box 2.1: Environmental niches from constrained analyses

Multivariate analyses are often used to analyze species-by-sites tables and obtain estimates of species environmental niches optima. Here, I explain how two constrained analyses, CCA and dc-CA, can be used to this end.

CCA classically analyzes a species-by-sites table constrained with a table of environmental variables measured for these sites. CCA defines species positions as weighted averages (WA) of sites scores, which are expressed as linear combinations (LC) of environmental variables. These species WA scores can be seen as species environmental niche optima, as they characterize the mean score of sites in which the species is present. The sites LC scores maximize the separation of species WA scores (niches) (C. J. F. ter Braak & Verdonschot, 1995).

dc-CA analyzes a species-by-sites table constrained with a table of environmental variables and a table of species traits. dc-CA defines two sets of LC scores: for sites and for species. These two sets of scores are defined so that their correlation is maximized. In other words, we search the combination of traits that best respond to environmental variation, and the environmental gradient that is the most important to explain observed species traits. Then, species WA scores can be computed from sites LC scores to represent their niche optima (as for CCA).

These multivariate methods (CA, CCA and dc-CA) can also be used to quantify the part of inertia explained by resource or consumer traits, or matching between resource and consumer traits (trait matching), thus realizing variance decomposition. This allows to quantify the part of structure in the network due to resource traits, consumer traits, or trait matching between resource and consumers.

This chapter describes constrained analyses allowing to incorporate traits into the analysis: either the traits of one set of species (CCA) or the traits for both species sets

(dc-CA). First, I present CCA and dc-CA in the context of interaction networks analysis. Then, I exemplify these methods by analyzing a bird-fruit interaction network from the ANDEAN frugivory dataset also used in chapter 2.1 (Dehling et al., 2021).

2 Material and methods

All analyses were performed with R 4.3.3 (R Core Team, 2024) and are stored in a private GitHub repository (https://github.com/LisaNicvert/PhDaxis01_CAnetwork/tree/thesis) but can be shared on demand and are intended to be made public upon publication.

As in the previous chapter, we consider an interaction matrix $\mathbf{Y}$ ($r \times c$). Additionally, we suppose that l traits were collected for the resource species, stored in matrix $\mathbf{L}$ ($r \times l$), and k traits were collected for consumer species, stored in matrix $\mathbf{Q}$ ($c \times k$).

Like in chapter 2.1, we define the following matrices and vectors: the table of relative frequencies $\mathbf{P} = [y_{ij}/y_{++}]$ (where y_{++} is the grand total of $\mathbf{Y}$); the weight matrices for resources $\mathbf{D_r} = \mathrm{diag}(\mathbf{r})$ and consumers $\mathbf{D_c} = \mathrm{diag}(\mathbf{c})$, where the vectors $\mathbf{r} = \mathbf{P1_r} = [p_{1+}, \dots, p_{r+}]^\top$ and $\mathbf{c} = \mathbf{P}^\top\mathbf{1_c} = [p_{+1}, \dots, p_{+c}]^\top$ represent respectively the row and column marginal sums ($p_{i+} = \sum_{j=1}^{c} p_{ij}$ and $p_{+j} = \sum_{i=1}^{r} p_{ij}$); matrix $\mathbf{P_0} = \mathbf{P} - \mathbf{rc}^\top$. Additionally, we define matrices $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{Q}}$ as the centered scaled version of $\mathbf{R}$ and $\mathbf{Q}$, respectively weighted by $\mathbf{r}$ and $\mathbf{c}$.

2.1 Canonical correspondence analysis (CCA)

Canonical correspondence analysis (CCA) (C. J. F. ter Braak, 1986) considers the interaction matrix $\mathbf{Y}$ and one of the matrices $\mathbf{R}$ or $\mathbf{Q}$. CCA allows to ordinate resource and consumer species in a multivariate space constrained with resource species traits $\mathbf{R}$ (or consumer species traits $\mathbf{Q}$). From an interaction niche perspective, CCA allows to quantify meaningfully the niche optima of one species type (resource of consumers). Indeed, if we use the *resource traits* $\mathbf{R}$, only *consumer niches* can be quantified meaningfully. Conversely, if we use *consumer traits* $\mathbf{Q}$, only *resource niches* can be quantified meaningfully.

Below, I present the method considering that we use matrix $\mathbf{R}$ (resource traits) to constrain the analysis: to use consumer traits, we would need to replace $\mathbf{Y}$ by $\mathbf{Y}^\top$ and

$\mathbf{R}$ by $\mathbf{Q}$ in the equations below.

Algebra

First, we perform a weighted multiple linear regression of $\mathbf{P_0}$ by $\tilde{\mathbf{R}}$:

$$\hat{\mathbf{P}}_\mathbf{0} = {\mathbf{D_r}}^{1/2}\tilde{\mathbf{R}}\mathbf{B} \tag{2.2.1}$$

where $\mathbf{B}$ $(r \times c)$ is the matrix of regression coefficients and each coefficient b_{kj} corresponds to the contribution of resource trait k on the frequency of interaction of consumer species j. $\mathbf{B}$ is defined as (Legendre & Legendre, 2012):

$$\mathbf{B} = \left[\tilde{\mathbf{R}}^\top \mathbf{D_r} \tilde{\mathbf{R}}\right]^{-1} \tilde{\mathbf{R}}^\top {\mathbf{D_r}}^{1/2} \mathbf{P_0} \tag{2.2.2}$$

Then, we diagonalize the covariance matrix of the predicted matrix $\hat{\mathbf{P}}_\mathbf{0}$, noted $\hat{\mathbf{S}} = \hat{\mathbf{P}}_\mathbf{0}^\top \hat{\mathbf{P}}_\mathbf{0}$.

$$\hat{\mathbf{S}} = \mathbf{V_0} \mathbf{\Lambda} {\mathbf{V_0}}^{-1} \tag{2.2.3}$$

$\mathbf{V_0}$ contains the consumer species scores and is orthonormal. $\mathbf{\Lambda}$ is the matrix of eigenvalues of dimension $\min(r-1, c, l)$.

For resource species, we define the linear combination scores (LC scores) $\mathbf{Z_0}$, which represent scores predicted by the multiple linear regression. They are computed as (Legendre & Legendre, 2012):

$$\mathbf{Z_0} = {\mathbf{D_r}}^{-1/2} \hat{\mathbf{P}}_\mathbf{0} \mathbf{V_0} \tag{2.2.4}$$

We can also define observed scores $\mathbf{U_0}$ for resource species (Legendre & Legendre, 2012). These scores are computed from the coordinates of consumer species $\mathbf{V_0}$, and contrary to $\mathbf{Z_0}$, they have no relation to resource species traits.

$$\mathbf{U_0} = \mathbf{P_0} \mathbf{V_0} \mathbf{\Lambda}^{-1/2} \tag{2.2.5}$$

Finally, the coordinates of explanatory variables (here, resource species traits) are given by:

$$\mathbf{C} = \tilde{\mathbf{R}}^\top \mathbf{D_r} \tilde{\mathbf{Z}}_\mathbf{0} \tag{2.2.6}$$

where $\tilde{\mathbf{Z}}_\mathbf{0}$ is a centered scaled version of $\mathbf{Z_0}$ using weights $\mathbf{r}$. These coordinates represent the correlations of resource species traits with the multivariate axes. $\mathbf{C}$ is often drawn on the correlation circle, where the angles between elements of $\mathbf{C}$ give correlations between traits, and the angle between elements of $\mathbf{C}$ and multivariate axes also give the correlation between traits and multivariate axes.

Biplot

Standard outputs of CCA consist of biplots where both resource and consumers scores are represented. In the context of interaction niches, we are interested only in the biplot in the consumer species space (scaling type 2). For that, we define the following scaling for resource species LC scores: $\mathbf{Z} = \mathbf{Z_0}\mathbf{\Lambda}^{-1/2}$ (the scores $\mathbf{Z}$ are normed to one). Then, the consumer weighted average (WA) score $\mathbf{V}^\star$ is defined as (Greenacre, 2010):

$$\mathbf{V}^\star = {\mathbf{D_c}}^{-1}\mathbf{P}^\top\mathbf{Z} \tag{2.2.7}$$

Equation (2.2.7) is analogous to the CA transition formula in for consumers in the previous chapter (Equation (2.1.4b)), except that resource scores $\mathbf{Z}$ are defined as a linear combination of their traits.

We can then represent resource LC scores $\mathbf{Z}$ and consumer WA scores $\mathbf{V}^\star$ in the same space. On this biplot, resource species LC scores are positioned as a linear combination of their traits with variance 1 on each axis. Consumer species WA scores are at the centroid of the resource species LC scores they interact with, so they are positioned at their interaction niche optimum. Additionally, consumer species scores have maximal variance.

Other biplots (or triplots, if we plot variables scores $\mathbf{C}$ as well) are possible, but here we will focus on this biplot as it conveys the position of the consumer species in the network by displaying their niche optima.

2.2 Double constrained correspondence analysis (dc-CA)

Double constrained correspondence analysis (dc-CA) (C. J. F. ter Braak et al., 2018) considers the 3 tables defined above: interactions $\mathbf{Y}$, resource traits $\mathbf{R}$ and consumer traits $\mathbf{Q}$. dc-CA finds resource and consumer species scores defined by linear combi-

nation of their respective traits, so that the correlation between these scores is maximized.

From an interaction niche perspective, dc-CA allows to define the niche optima of resource species using the traits of consumer species they interact with, and conversely, the niche optima of consumer species using the traits of resource species they interact with.

Algebra

We search resource scores $\mathbf{U}$ and consumer scores $\mathbf{V}$, each defined as a linear combination of the traits tables. $\mathbf{U}$ and $\mathbf{V}$ are analogous to the LC score defined with CCA noted $\mathbf{Z}$: we could view $\mathbf{U}$ as $\mathbf{Z_R}$ and $\mathbf{V}$ as $\mathbf{Z_Q}$, but we write these scores $\mathbf{U}$ and $\mathbf{V}$ to ease the notation. These scores maximize the following correlation:

$$\max_{\mathbf{B},\mathbf{C}}(\mathbf{U}^\top \mathbf{P}\mathbf{V}) = \max_{\mathbf{B},\mathbf{C}} \left(\left[\tilde{\mathbf{R}}\mathbf{B}\right]^\top \mathbf{P}\tilde{\mathbf{Q}}\mathbf{C} \right) \tag{2.2.8}$$

This equation is analogous to Equation 2.1.3 defined for CA in the previous chapter, except the scores that maximize the correlation are linear combinations of species traits.

In order to find the coefficient matrices $\mathbf{B}$ and $\mathbf{C}$, we define the following matrix $\mathbf{D}$ (similarly to C. J. F. ter Braak et al., 2018, but using centered scaled traits instead of centered traits and $\mathbf{P}$ instead of $\mathbf{Y}$):

$$\mathbf{D} = [\underbrace{\tilde{\mathbf{R}}^\top \mathbf{D_r} \tilde{\mathbf{R}}}_{\mathbf{L}}]^{-1/2} \tilde{\mathbf{R}}^\top \mathbf{P} \tilde{\mathbf{Q}} [\underbrace{\tilde{\mathbf{Q}}^\top \mathbf{D_c} \tilde{\mathbf{Q}}}_{\mathbf{K}}]^{-1/2} \tag{2.2.9}$$

Matrices $\mathbf{L}$ and $\mathbf{K}$ represent weights matrices for resource (respectively, consumer) trait tables. Then, we perform the generalized SVD of $\mathbf{D}$:

$$\mathbf{D} = \mathbf{B}\boldsymbol{\Delta}\mathbf{C}^\top \tag{2.2.10}$$

where $\mathbf{B}$ ($l \times d$) and $\mathbf{C}$ ($k \times d$) are the matrices of coefficients for resource (respectively consumer) species traits, and $d = \min(k, l)$. $\mathbf{B}$ and $\mathbf{C}$ are orthonormal with respect to weights $\mathbf{L}$ and $\mathbf{K}$ ($\mathbf{B}^\top\mathbf{L}\mathbf{B} = \mathbf{I}$ and $\mathbf{C}^\top\mathbf{K}\mathbf{C} = \mathbf{I}$). $\boldsymbol{\Delta}$ is the diagonal matrix of singular values. dc-CA eigenvalues $\boldsymbol{\Lambda}$ are computed as $\boldsymbol{\Delta}^2$, and there are d non-null eigenvalues.

We can derive resource and consumer LC scores using coefficient matrices $\mathbf{B}$ and $\mathbf{C}$:

$$\mathbf{U} = \tilde{\mathbf{R}}\mathbf{B} \qquad \text{(resources)} \qquad (2.2.11a)$$

$$\mathbf{V} = \tilde{\mathbf{Q}}\mathbf{C} \qquad \text{(consumers)} \qquad (2.2.11b)$$

Resource species scores $\mathbf{U}$ are defined as a linear combination of their traits, and reciprocally consumer scores $\mathbf{V}$ are defined as a linear combination of their traits.

Like for CA, the square root of the dc-CA eigenvalues represent the absolute value of the correlation between resource and consumer scores:

$$\sqrt{\lambda_k} = |\mathrm{cor}(\mathbf{u}_k, \mathbf{v}_k)| \qquad (2.2.12)$$

Variables scores are often represented on the correlation circle using the following coordinates: resource species traits are represented with $\mathbf{C_R} = \tilde{\mathbf{L}}^\top \mathbf{D_c} \mathbf{V}$ and consumer species traits with $\mathbf{C_Q} = \tilde{\mathbf{Q}}^\top \mathbf{D_r} \mathbf{U}$.

Biplots

To represent resource and consumers on the same biplots, we use the following transition formulas giving the weighted averaging scores (WA scores):

$$\mathbf{U}^\star = {\mathbf{D_r}}^{-1} \mathbf{P} \mathbf{V} \qquad \text{(resources)} \qquad (2.2.13a)$$

$$\mathbf{V}^\star = {\mathbf{D_c}}^{-1} \mathbf{P}^\top \mathbf{U} \qquad \text{(consumers)} \qquad (2.2.13b)$$

Equation (2.2.13a) expresses resources scores as the weighted mean of the consumers scores they interact with and Equation (2.2.13b) expresses consumers scores as the weighted mean of the resources scores they interact with. Equation (2.2.13) is analogous to the CA transition formula in the previous chapter (Equation (2.1.4)), except that resource and consumer scores $\mathbf{U}$ and $\mathbf{V}$ are defined as linear combinations of their traits.

To visualize the niche optima of resource species, we plot consumer LC scores $\mathbf{V}$ jointly with resource WA scores $\mathbf{U}^\star$, so that resource species are at the centroid of the consumer species they interact with. Reciprocally, to visualize the niche optima of

consumer species, we represent the consumer WA scores $\mathbf{V}^\star$ jointly with the resource LC scores $\mathbf{U}$, so that consumer species are at the centroid of the resource species they interact with.

It can also be interesting to represent the WA scores jointly with the LC scores normed to the eigenvalues (instead of one), i.e., $\mathbf{U}^\star$ with $\mathbf{U_1} = \mathbf{U}\mathbf{\Lambda}^{1/2}$, and $\mathbf{V}^\star$ with $\mathbf{V_1} = \mathbf{V}\mathbf{\Lambda}^{1/2}$. The closer the scores are on these biplots, the stronger the matching is between species niches and their traits, so this biplot can then visually represent trait matching strength. Other biplots (or tri- and quadri-plots) are possible (see C. J. F. ter Braak et al., 2018), but they are not all detailed here in the context of niche representation.

2.3 Variation partitioning

CCA and dc-CA are regression-based methods: therefore, they allow to decompose the inertia of the interaction table into parts explained by resource traits, consumer traits or resource-consumer trait matching (Figure 2.2.1) (Borcard et al., 1992; Peres-Neto et al., 2006; Peng et al., 2021).

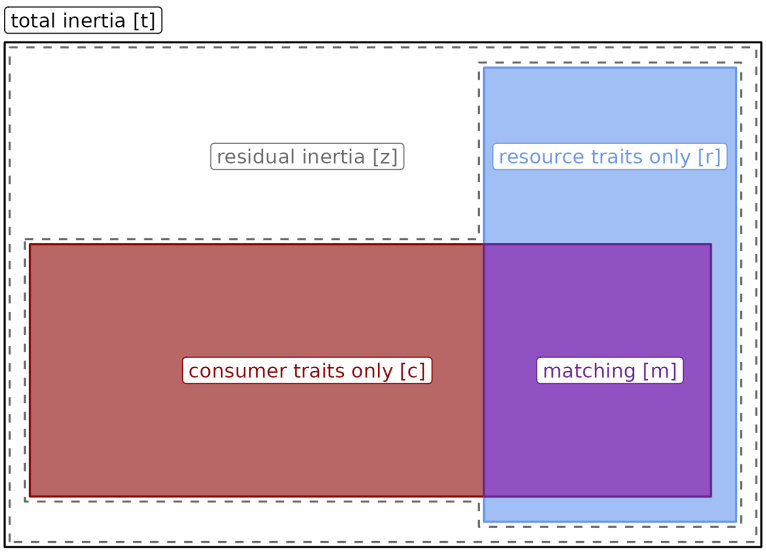


Figure 2.2.1: Variance partitioning of the interaction table. This schematic representation shows how the total inertia of the interaction table [t] can be partitioned into inertia explained by resource and consumer traits in isolation (respectively [r] and [c]), inertia due to trait matching [m] and residual inertia [z] ([t] = [r + c + m + z]). In the general case, when there is some degree of trait matching (i.e. [m] $\neq \varnothing$), the total variance explained by resource ([r + m]) and consumer traits ([c + m]) overlap. Own figure inspired from Figure 1 of Sîrbu et al. (2021).

First, we can compute the total inertia of the interaction table due to non-random in-

teractions ([t] in Figure 2.2.1). Non-random patterns in the interaction matrix are quantified as counts that deviate from their expected value in the sense of the χ^2 statistic, i.e. counts that deviate from the neutral abundance effects (as defined in chapter 2.1). The total inertia is given by the sum of the CA eigenvalues. Equivalently, [t] can be computed with the trace of the variance-covariance matrix $\mathbf{P_0}\mathbf{P_0}^\top$.

The variation explained by resource or consumer traits can be quantified using CCA constrained with either of those traits (Borcard et al., 1992; Peres-Neto et al., 2006; Peng et al., 2021). Quantitatively, this corresponds to the inertia of $\mathbf{P_0}$ explained by the regression ($\hat{\mathbf{P}}_\mathbf{0}$). The variation explained by resource (respectively, consumer) traits correspond to [r + m] and [c + m] in Figure 2.2.1. They can be computed as the sum of the eigenvalues of each corresponding CCA. Equivalently, the inertia can be computed using matrices $\mathbf{P_0}$ and $\hat{\mathbf{P}}_\mathbf{0}$ (see Appendix A from Peres-Neto et al., 2006). This inertia is also commonly expressed as a fraction of the total inertia [t] defined above. This fraction is noted $R^2_{\mathbf{Y}|\mathbf{R}}$ and $R^2_{\mathbf{Y}|\mathbf{Q}}$ for resource and consumer traits respectively. It is akin to a coefficient of determination in univariate linear regression.

Finally, we can quantify the variation explained by trait matching ([m] in Figure 2.2.1). [m] corresponds to the inertia explained by resource and consumer traits that covary, which biologically corresponds to interactions constrained by trait matching. [m] can be quantified as the sum of the dc-CA eigenvalues. It can also be expressed as a fraction of the total inertia, noted $R^2_{\mathbf{Y}|\mathbf{R},\mathbf{Q}}$.

Variation partitioning can be performed using simple subtractions to get the inertia due to resource and consumer traits only ([c] and [r]), knowing [t], [r + m], [c + m] and [m] from CA, CCA and dc-CA.

Corrected R^2

Peres-Neto et al. (2006) showed that the inertia of CCA computed as above is a biased estimator of the true inertia. They propose a corrected estimator $R^2_{(\mathbf{Y}|\mathbf{X})\text{perm}}$ (in our case, $\mathbf{X}$ denotes the traits matrix $\mathbf{R}$ or $\mathbf{Q}$):

$$R^2_{(\mathbf{Y}|\mathbf{X})\text{perm}} = 1 - \frac{1}{1 - \bar{R}^2_{(\mathbf{Y}|\mathbf{X})\text{perm}}} \left(1 - R^2_{\mathbf{Y}|\mathbf{X}}\right) \tag{2.2.14}$$

where $\bar{R}^2_{(\mathbf{Y}|\mathbf{X})\text{perm}}$ is the mean coefficient obtained with permuted matrices (see

Peres-Neto et al., 2006, for details).

By analogy, for dc-CA, we can think that the estimator for the inertia is biased too. To my knowledge, this hypothesis has not been formally tested and an unbiased estimator of the dc-CA inertia has never been proposed. Here, I propose a correction for the dc-CA estimator inspired from Peres-Neto et al. (2006) using the works of Dray and Legendre (2008) in the case of fourth-corner analysis. I use the formula of Equation (2.2.14), where $\mathbf{X}$ is replaced by $\mathbf{R}, \mathbf{Q}$. To define $R^2_{(\mathbf{Y}|\mathbf{R},\mathbf{Q})\text{perm}}$, I use the permutation approach combining model 2 with model 4 proposed by Dray and Legendre (2008) to randomize matrices in the context of fourth-corner analysis (see Dray & Legendre, 2008, for more details).

2.4 Data analysis

I reanalyze the bird-fruit interaction data from the ANDEAN frugivory dataset used in chapter 2.1 (Dehling et al., 2021). I perform three analyses of the interaction table: a CCA constrained by resource (plant species) traits, a CCA constrained by consumer (bird species) traits, and a dc-CA constrained by both resource and consumer (plant and bird species) traits. This allows to partition the variation of the interaction table, and thus to quantify the impact of traits on the structure of the interaction network. Moreover, these analyses allow to visualize and compare species niches optima defined with different metrics.

3 Results and discussion

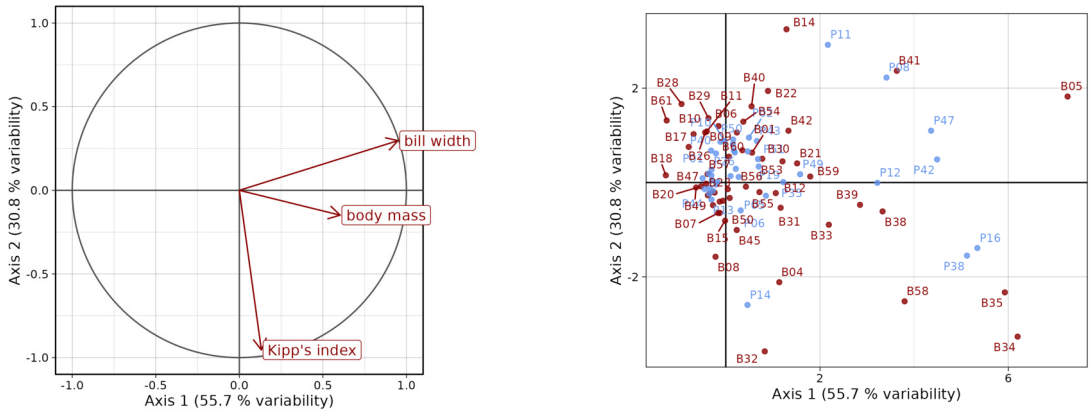
In this section, I present the results of CCA and dc-CA applied to the bird-fruit interaction network. By concision, I present biplots only for two of the three analyses: CCA constrained with bird traits and dc-CA. The biplot for CCA constrained with plant traits is left in Appendix B.

To interpret these biplots, I compare species traits to the general distribution of traits in the community. An overview of traits distribution in the community is presented in Appendix A with histograms of trait values, and the traits values per species are presented in Appendix E of the previous chapter.

3.1 CCA constrained with bird traits

First, I investigate how well the plant interaction niches can be explained using the following bird traits: body mass, bill width and Kipp's index. These traits are thought to be related to bird-plants interactions (Albrecht et al., 2018; Bender et al., 2018). Indeed, bird body mass is an energetic constraint, because large birds must be able to get enough calories from fruits they feed on; bill width is a morphological constraint, as birds with a small beak cannot grasp large fruits easily; and Kipp's index is related to foraging strategy, since birds with a large Kipp's index (pointed wings) tend to forage in the canopy, as opposed to the understory.

I first test the significance of the association between bird traits and network structure. A permutation test suggests that bird traits explains a significant part of the network structure ($R^2_{(\mathbf{Y}|\mathbf{Q})}$ = 17.39%, p-value = 0.001, 999 permutations). After correction (see Section 2.3), the part of explained variation $R^2_{(\mathbf{Y}|\mathbf{Q})\text{perm}}$ is 12.93%.



(a) Correlation circle ($\mathbf{C}$)

(b) Plant niches ($\mathbf{V}^\star$ and $\mathbf{Z}$, from $\mathbf{Y}^\top$)

Figure 2.2.2: Correlation circle and plant niches optima from CCA constrained with bird traits. (a) Correlation circle of the CCA constrained with bird traits. (b) Biplot showing plant niches optima ($\mathbf{V}^\star$) in relation to bird traits ($\mathbf{Z}$).

Figure 2.2.2 shows the correlation circle and the biplot obtained from the CCA constrained with bird traits for the first two axes. These two axes account for 86% of the inertia of the constrained analysis, so they are rather representative of the effect of bird traits on the network structure. The correlation circle (Figure 2.2.2a) allows us to interpret the axes. The first axis is positively correlated to bill width and body mass, and can be seen as a size effect. The second axis is negatively correlated to Kipp's index (birds with pointed wings that tend to forage in the canopy).

Figure 2.2.2b shows the biplot representing birds and plants individuals scores in

the multivariate space. This is the result of the CCA of $\mathbf{Y}^\top$ constrained by bird traits $\mathbf{Q}$. Therefore, notations are inverted compared to the methods section: birds scores are denoted by $\mathbf{Z}$ and are linear combinations of their traits, and plant scores are denoted as $\mathbf{V}^\star$ and are positioned at the centroid of bird scores.

First, we can investigate the position of birds in the multivariate space. Most species are located around the origin, so they have average size and Kipp's index (e.g. B55, *Thraupis episcopus*). Birds that have a small score on the first axis also tend to be located in the top-left corner of the graph, which means that small birds also tend to have a small Kipp's index (e.g. B28, *Lophotriccus pileatus*). Then, some species have low scores on the second axis, and tend to have a medium to high score on the first axis. These birds all have a high Kipp's index (pointed wings adapted to the canopy) and their size ranges from average (e.g. B32 *Patagioenas plumbea*, which is a large bird with a medium beak) to large (e.g. B34 *Pharomachrus antisianus*, a large bird with a wide beak). Some birds are also located in the top-right corner (e.g. B41 *Rupicola peruvianus*), so they tend to be large and have a low Kipp's index (rounded wings adapted to the understory).

Then, we can investigate the position of plant species, as their niche optimum. Most plants' niches are located around the origin, where most birds are as well. This implies that most plants have an interaction niche optimum corresponding to average birds. A few plants also have a niche optimum corresponding to birds with medium to high scores on the first axis, and a wide range of bird scores on the second axis. Interestingly, as their niche optimum moves to larger birds, plants tend to have a niche optimum favoring birds with a higher Kipp's index as well. For instance, P11 *Elaeagia mariae* tends to interact with medium birds with a low Kipp's index, and P16 *Guatteria duodecima* tends to interact with large birds with a high Kipp's index. This suggests that a latent trait of plants matches the birds size-Kipp's index trait syndrome. Finally, P14 *Ficus coerulescens* stands out as a plant that interacts with average-sized birds that also have a high Kipp's index.

3.2 dc-CA constrained with bird and plant traits

We use dc-CA to constrain the analysis with bird and plant traits. For that, we add the following plant traits to the analysis: fruit diameter, plant height and crop mass.

These traits are thought to be related to interactions with birds (Albrecht et al., 2018; Bender et al., 2018), as plants' crop mass can correspond to the energy that plants invest in fruit production, supposedly reflected in the preferences of birds with high energy expenditure; fruit diameter can constrain morphological matching with birds beak size; and plant height can correspond to birds' foraging strata. Therefore, the following pairs of traits are hypothesized to be positively correlated under trait matching: body mass/crop mass, beak width/fruit diameter and Kipp's index/plant height (Albrecht et al., 2018; Dehling et al., 2014).

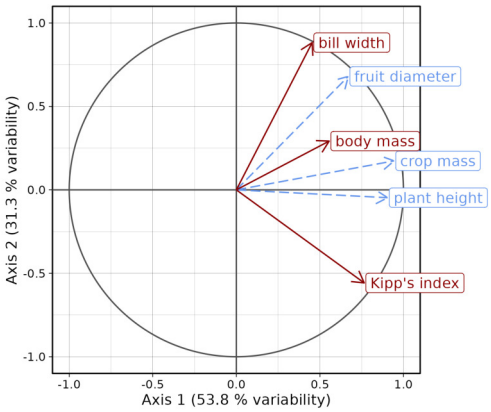
We test the significance of the association between bird traits and network structure, using the permutation approach described in Section 2.3. This test suggests that the association between bird and plant traits is significant ($R^2_{(\mathbf{Y}|\mathbf{R},\mathbf{Q})}$ = 6.65%, p-value = 0.001, 999 permutations). The corrected part of explained variation due to trait matching $R^2_{(\mathbf{Y}|\mathbf{R},\mathbf{Q})\text{perm}}$ is 5.97%.

The eigenvalues can also inform us on the matching strength. Here, I consider the first 2 eigenvalues (on a total of 3 eigenvalues) and consider their square-root, which represent the absolute value of the correlation of traits with each axis. Here, we have $\sqrt{\lambda_1} = 0.39$ and $\sqrt{\lambda_2} = 0.29$, which suggest a limited strength of matching. In total, these 2 axes account for 99.5% of the trait matching inertia.

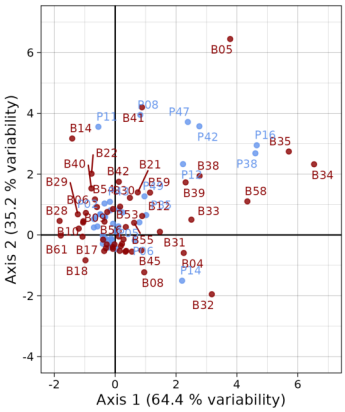
Figure 2.2.3a shows how bird and plant traits are correlated. The hypotheses on trait matching above are verified to some extent: first, the energy traits (crop mass and body mass) are well positively correlated. Second, morphology traits (bill width and fruit diameter) are well positively correlated too. Finally, foraging traits (Kipp's index and plant height) are the least correlated. Another notable trait relationship is the positive correlation between plant height and crop mass, probably due to the fact that large plants tend to produce pore fruits. Comparatively, the correlation between bodymbodymassass and bill width for birds is less strong.

Regarding axes interpretation, axis 1 is characterized by the matching of energy traits between birds and plants, because it is positively correlated to plant height, crop mass and bird body mass. Axis 2 is positively correlated to the morphological matching between birds and plants (bill width and fruit diameter). To a lesser extent, it is also negatively correlated to Kipp's index.

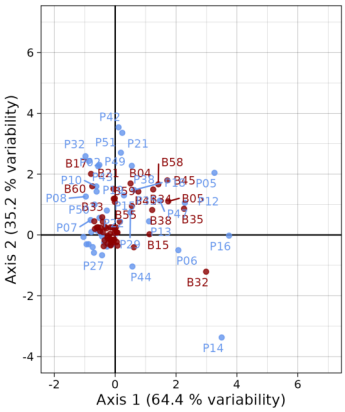
Regarding biplots interpretation, we first focus on Figure 2.2.3b that shows plant



(a) Correlation circle ($\mathbf{C_R}$ and $\mathbf{C_Q}$)



(b) Plant niches ($\mathbf{U}^\star$ and $\mathbf{V}$)



(c) Bird niches ($\mathbf{V}^\star$ and $\mathbf{U}$)

Figure 2.2.3: Correlation circle and niches optima from dc-CA. (a) Correlation circle with plant and bird traits scores (respectively $\mathbf{C_R}$ and $\mathbf{C_Q}$). Bird traits are shown in solid red lines and plant traits in dashed blue lines. (b) Biplot showing birds positioned as linear combination of their traits ($\mathbf{V}$) and plants at their interaction niche optimum ($\mathbf{U}^\star$). (c) Biplot showing plants positioned as linear combination of their traits ($\mathbf{U}$) and birds at their interaction niche optimum ($\mathbf{V}^\star$).

niches in relation to bird traits. On this figure, if we first interpret birds' position, most birds are located near the origin (average birds). Birds on the left of the first axis (small birds) also tend to have medium to high scores on the second axis (i.e. medium to wide beaks and small Kipp's index), like B14 *Cyanocorax yncas*, which is an average bird having the smallest Kipp's index of the dataset. In parallel, larger birds tend to be located in the top-right corner of the plot (large birds with wide beaks), (e.g. B05 *Cephalopterus ornatus*, which has the widest beak of the dataset and is the third largest bird). The only bird that deviates in that respect is B32 (*Patagioenas plumbea*), which seems functionally distinct from other large birds that tend to have a large beak as well, since this bird is among the largest of the dataset but has an average beak.

Regarding plants positions (which indicate their niche optimum), most plants have a niche optimum near the origin (they interact preferentially with average birds). However, a group of plants in the top-right corner of the plot stands out as plants that interact with large-billed birds, and medium to large-bodied birds (e.g. P11 *Elaeagia mariae* and P16 *Guatteria duodecima*). We can note that these plants were already distinct from other plants on the CCA constrained with bird traits. If we categorize these plants with their traits using the second biplot (2.2.3c), we can see that most of these plants whose fruits are eaten preferentially by large birds are located in the top-right corner of the plot: they tend to be large plants that produce large fruits (except for P08, *Cestrum stipulatum*, which is the fourth smallest plant of the dataset but yields medium-sized fruits and a medium crop mass). However, by comparing plants positions on the two biplots, these plants are much less distinct from other plants based on their traits only (Figure 2.2.3c) than based on their niche optimum (Figure 2.2.3b). P14 *Ficus coerulescens* seems to have a particular niche position, as a plant interacting with average-sized birds that have small beaks. If we characterize this species with its traits using the second biplot, its distinctiveness in interaction niches is mirrored in its functional distinctiveness, because it is a large plant (the tallest from the dataset) that yields average-sized fruits.

Next, we focus on the interpretation of Figure 2.2.3c that shows bird niches in relation to plant traits. First, if we interpret plants position, they are well spread out on the multivariate plane. There seems to be a negative relationship between plant score on the first and the second axis (larger plants tend to produce smaller fruits). As noted before, P14 (*Ficus coerulescens*) has traits that are very distinctive from other plants.

Regarding birds positions (niche optima), most birds interact with average plants. A group of distinct birds that have high scores on axis 2 and dispersed scores on axis 1 emerges (e.g. B17 *Diglossa cyanea* and B05 *Cephalopterus ornatus*): these birds tend to eat large fruits from plants of various sizes and crop yields. If we characterize these birds with their traits on the first biplot (Figure 2.2.3b), we can see that the birds with the highest scores on axis 1 (preference for large plants with high yield) are the largest birds with the widest beaks (B05 *Cephalopterus ornatus*, B35 *Pharomachrus auriceps*, B34 *Pharomachrus antisianus* or B58 *Trogon personatus*). B32 (*Patagioenas plumbea*) stands out as a bird consuming small fruits from tall plants with high crop

mass. As discussed before, this can be explained by the fact that it is a large bird, which means that it needs more energy, but it has an average beak size. The distinctiveness of its niche optimum can then be explained by morphological restrictions on matching, that prevents this bird from eating large fruits because of its small beak, but this bird still needs to consume large quantities of fruits, and for that it focuses on plants with high crop mass to optimize its foraging time.

3.3 Variation partitioning

Using CA, CCA and dc-CA, we can partition the variance of the interaction matrix in order to attribute the non-random interactions to either plant traits only, birds traits only, or matching between birds and plant traits (respectively components [r], [c] and [m] from Figure 2.2.1). Results are presented in Table 2.2.1, for the total inertia, the uncorrected and the corrected part of total inertia. Below, I interpret results for the corrected part of inertia (last column of the table).

	Inertia (uncorrected)	R^2 (%)	R^2_{perm} (%)
Plant traits [r]	0.346	9.58	1.06
Bird traits [c]	0.387	10.74	6.96
Trait matching [m]	0.240	6.65	5.97
Residual inertia [z]	2.635	73.03	86.01
Total [t]	3.608	100	100

Table 2.2.1: Variance partitioning of the interaction matrix. In rows, the different sources of variation and their corresponding abbreviation from Figure 2.2.1. In columns, different measures of the inertia: raw inertia (uncorrected), uncorrected part of inertia relatively to the total (R^2), and corrected part of inertia (R^2_{perm}).

First, an extremely large part of the inertia is not attributable to any trait: indeed, 86.01% of total inertia is explained neither by the traits used in the analyses, nor by their matching. This suggests that important factors driving network structure are missing from the analyses. These factors can include unmeasured species traits, for instance, fruit energetic caloric content or nutrient composition, of birds' digestive system, which could play a role in trait matching (Moermond & Denslow, 1985; Palacio et al., 2017) but have not been used in the analysis. This residual inertia could also be due to sampling effects. Indeed, to assess the expected number of interactions, CCA and dc-CA use

the matrix margins, thus assuming that the number of times a species is observed in the network is a surrogate of its abundance: but as discussed in the previous chapter, this might not be the case (see e.g. Blüthgen et al., 2008).

Second, the principal mechanism of interest studied here, trait matching, accounts for only 5.97% of the total inertia. If we compare this to the part of inertia explained by bird or plants traits including matching ([c + m] and [r + m]), which are respectively 12.93 and 7.03%, we can see that plant traits are much more related to bird traits via trait matching than the reverse. Indeed, when taking the matching part into account, the part of inertia attributable to plant traits drops to only 1.06%, while the part of variation due to bird traits is still 6.96%. Ecologically, we would expect most of the inertia explained by consumer and resource traits to be driven by their matching, i.e. consumers will forage on resources with specific traits that match their traits, and reciprocally, resources will be used by consumers because these resources have some traits that benefit the consumer. But here, a substantial part of the inertia due to bird traits is not explained by trait matching (6.96%, which is of the same magnitude as the variation due to trait matching). This suggests that some plant traits that are matched by bird traits used here (body mass, bill width and Kipp's index) are missing in the analysis. In contrast, the plant traits used here (crop mass, fruit diameter and plant height) are accurately matched by these bird traits, since their part of explained variation is only 1.06%.

4 Conclusion and perspectives

In this chapter, I presented constrained correspondences analyses, namely CCA and dc-CA, to visualize the network structure and niche optima and decompose the sources of variation in the network. Visualization of species in the multivariate space allows to highlight patterns of trait matching related to species (latent) traits, as well as the structure of species within the community. This allows to identify species that are distinct from the community in terms of interaction niche or of morphology. Here, for instance, in the dc-CA analysis, the traits of *Ficus coerulescens* P14 stand out from the rest of the community (particularly large plant with small fruits), and the niche of *Patagioenas plumbea* B32 stands out (bird consuming small fruits from large plants). Moreover,

these two species tend to appear as distinct from the community both for their interaction niche and their traits: this supports the findings of Dehling et al. (2016), conducted partly on the same dataset, who showed that morphological specialists also tend to be interaction specialists.

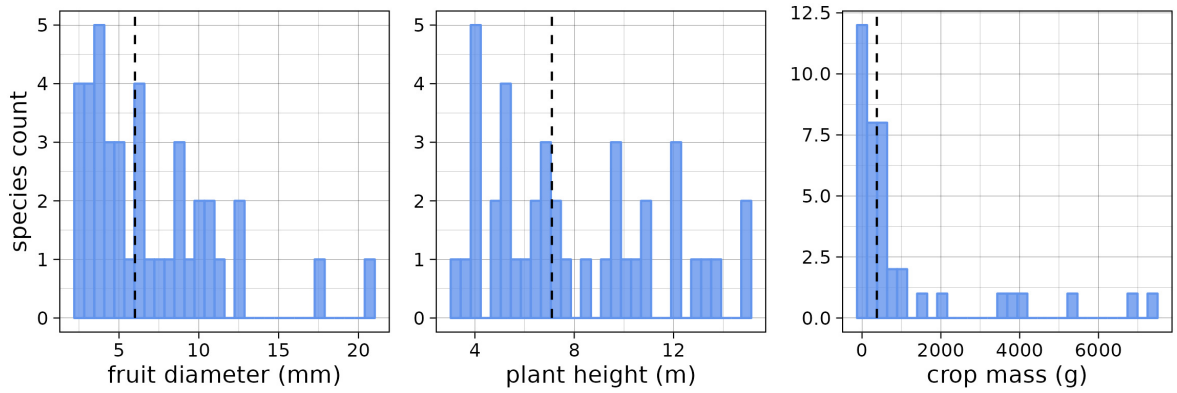
I find relatively weak associations between the traits used in the analyses and network structure. Indeed, more than 85% of the structure is not attributed to those traits. Moreover, trait matching, which is the main hypothesis relating traits to the network structure, explains only 5.97% of the total variation.

To fully characterize the interaction niches, we would also need to quantify niche breadth. But here, the niche scores $\mathbf{U}^\star$ and $\mathbf{V}^\star$ consider only the niche optimum, and there is no measure of the dispersion around these optima (niche breadth). Without a measure of niche breadth, it is difficult to investigate specialization and generalization (although central tendencies can be inferred from the niche optima, as discussed above or in Dehling et al., 2016). In the next chapter, I introduce a measure of niche breadth by extending reciprocal scaling (Thioulouse & Chessel, 1992), presented in the context of CA (chapter 2.1), to the constrained analyses presented in the present chapter (CCA and dc-CA).

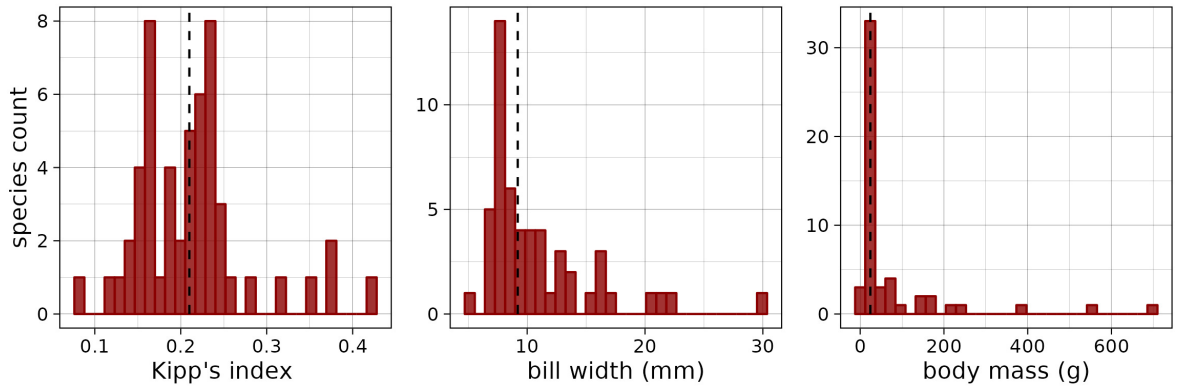
Appendices

A Traits distributions

This figure shows the trait distribution for birds and plants in the observed community sampled through the interaction network.



(a) Plant traits

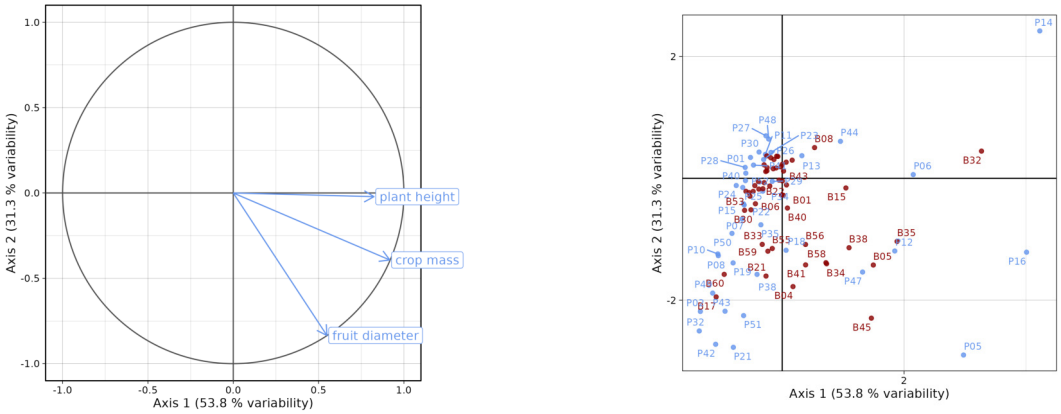


(b) Bird traits

Figure A.1: Bird and plant traits histograms. a) represents the distribution of plant traits in the community and b) for the birds traits. The vertical dashed line represents the median and the y-axis represents species count.

B CCA constrained with plant traits

This appendix shows the result of the CCA constrained with plant traits. First, we test the significance of the association between plant traits and network structure. A permutation test suggests that the association between plant traits and birds latent traits is significant ($R^2_{(\mathbf{Y}|\mathbf{R})}$ = 16.23%, p-value = 0.003, 999 permutations). The corrected part of explained variation $R^2_{(\mathbf{Y}|\mathbf{R})\text{perm}}$ is 6.96%.



(a) Correlation circle ($\mathbf{C}$)

(b) Bird niches ($\mathbf{V}^\star$ and $\mathbf{Z}$)

Figure B.1: Correlation circle and bird niches optima from CCA constrained with plant traits. (a) Correlation circle of the CCA constrained with plant traits. (b) Biplot showing bird niches optima ($\mathbf{V}^\star$) in relation to plant traits ($\mathbf{Z}$).

Figure B.1 below shows the results of the analysis on the correlation circle and biplot for the first two axes accounting for 85% of the explained inertia. The correlation circle (Figure B.1a) shows that the first axis is positively correlated to the three plant traits, so it is a size effect mainly related to plant height and productivity (crop mass). Axis 2 is negatively correlated to all traits, but mainly to fruit diameter.

The biplot (Figure B.1b) shows plants positioned with their traits and bird niches optima. Regarding plant species, most plants have small scores on the first axis, i.e. they are small and low-productivity. A few very large productive plant species exist (e.g. P05 *Cecropia polystachya* and P14 *Ficus coerulescens*) and they have diverse scores on axis 2, which reflects diverse strategies in terms of fruit size. For instance, P05 is the plant that produce the largest fruits while P14 produces average-sized fruits. There is a concave pattern between the score on axis 1 and on axis 2 (except for P14), which means that medium-sized and medium-productivity plants tend to produce the smallest fruits.

Regarding bird niches optima, most bird species are located around the origin (birds tend to interact with average plants). Birds that interact with plants producing large fruits (small score on axis 2) interact with a wide variety of plant sizes and productivity (dispersed scores on axis 1): so birds that eat large fruits tend to interact with plants of any size and productivity. On the contrary, most birds that interact with plants producing small to medium fruits (medium to large scores on axis 2) also interact with small, low productivity plants (small scores on axis 1), so birds consuming small to medium fruits also tend to interact with small and low productivity plants. *Patagioenas plumbea* B32 stands out in this respect, because whereas it tends to consume small fruits (high score on axis 2), it interacts with large, productive plants (very high score on axis 1).

Chapter 2.3

Using constrained reciprocal scaling to measure interaction niche breadth taking species traits into account

1 Introduction

In the previous chapter (2.2), I examined the relationships between species traits and the structure of interaction networks. For that, I used constrained multivariate analyses, namely canonical correspondence analysis (CCA) and double-constrained correspondence analysis (dc-CA). These methods allow to incorporate species traits in the measure of their niche optimum, but do not quantify niche breadth.

Niche breadth has multiple definitions (Ferry-Graham et al., 2002), and can be quantified with numerous indices. In the context of interaction niches, niche breadths are more often defined using only the number of interacting partners: this is the realized niche breadth (as defined in Ferry-Graham et al., 2002 or Devictor et al., 2010). Metrics that quantify realized niche breadth include the number of interacting partners (degree), weighted indices taking into account their respective abundances (e.g. Shannon diversity of interacting partners), or indices taking into account species relative availability (e.g. Blüthgen et al., 2006). However, these measures consider only species identity and not their attributes. The focus on species identity in ecology has been described as the "curse of the Latin binomial" by Raffaelli (2007): this concept argues that by focusing on species identity rather than other features, we might miss important determinants of ecological processes (McGill et al., 2006).

To overcome the "curse of the Latin binomial", the niche breadth of species can be defined with their own traits, using the notion of morphological specialization and assuming that distinct traits reflect specialization. Morphological specialization (as defined by Ferry-Graham et al., 2002) is usually evaluated by quantifying the distinctiveness of species traits compared to the community. Distinctiveness is usually measured on several traits, for instance using multivariate ordination methods like principal component analysis to position species in a functional trait space (Bellwood et al., 2005; Dehling et al., 2014). Specialization is then quantified using the Euclidean distance of a species from the mean of the community (Bellwood et al., 2005).

More recent works have used the traits of species' interacting partners instead of their own traits to define their interaction niche. For instance, this would mean defining the niche of a pollinator with visited flowers' corolla depth instead of proboscis length. Measuring the niche of a species using its interacting partners is arguably better suited to measuring its functional role (Dehling & Stouffer, 2018), and is also in better agree-

ment with the definition of the niche as conceptualized by Hutchinson (1957), defining niche axes using resources. These methods follow the same approach as the ones using species own traits, but they describe specialization using the distinctiveness of interacting partners traits. A method using this approach is described in Dehling et al. (2016) (see also Dolédec et al. (2000) in the context of the environmental niche). Compared to methods using species own traits, these techniques require an additional step to project species in the trait space of its interaction partners (e.g. Dehling et al. (2016) use two PCAs). Using the traits of interacting partners traits instead of species own traits is another way to measure specialization, but under trait matching, the two measures should be equivalent.

Recently, Dehling and Stouffer (2018) defined niche breadth as the trait diversity of a species interacting partners. For that, they projected species in the trait space of their interacting partners, and used the convex hull defined with a species' interacting partners' position in the multivariate space to define niche breadth. This approach is fundamentally different from the approaches described above, because specialization in this context is not defined as distinctiveness, but as diversity of traits of the interacting partners. It is better in line with Hutchinson's conceptualization of the niche as a hypervolume (Hutchinson, 1957). With this last approach, distinctiveness does not measure niche specialization, but niche differentiation, and is assimilated to the niche optimum of species.

Here, I introduce a new method to quantify interaction niche breadth, that is philosophically similar to the approach of Dehling and Stouffer (2018) described above. This method introduces a measure of niche breadth determined by the diversity of interacting partners' traits. To this end, I extend the reciprocal scaling method of Thioulouse and Chessel (1992) to CCA and dc-CA. In this chapter, I line up the theory underpinning the method allowing to visualize the dispersion around niche optima defined in constrained analyses (CCA or dc-CA) and describe the method with interaction matrices constrained with species traits in mind. Then, I exemplify the method on the bird-fruit interaction network analyzed in the two previous chapters (Dehling et al., 2021) and briefly discuss the ecological significance of the results.

2 Theory

This section presents the algebra used to add a measure of dispersion to CCA and dc-CA, thus extending reciprocal scaling for simple correspondence analysis (Thioulouse & Chessel, 1992) described in chapter 2.1. To do so, we use the link between correspondence analyses and canonical correlation analysis (CcorA) (Hotelling, 1936). Indeed, correspondence analysis (CA), CCA and dc-CA can be seen as a special case of CCorA where the analyzed variables are qualitative (see Hill (1974) for CA, Pélissier et al. (2002) and Gimaret-Carpentier et al. (2003) for CCA and C. J. F. ter Braak et al. (2018) for dc-CA). Here, we consider a $r \times c$ interaction matrix $\mathbf{Y}$, along with resource traits $\mathbf{R}$ ($r \times l$) and consumer traits $\mathbf{Q}$ ($c \times k$).

2.1 Add dispersion to CCA

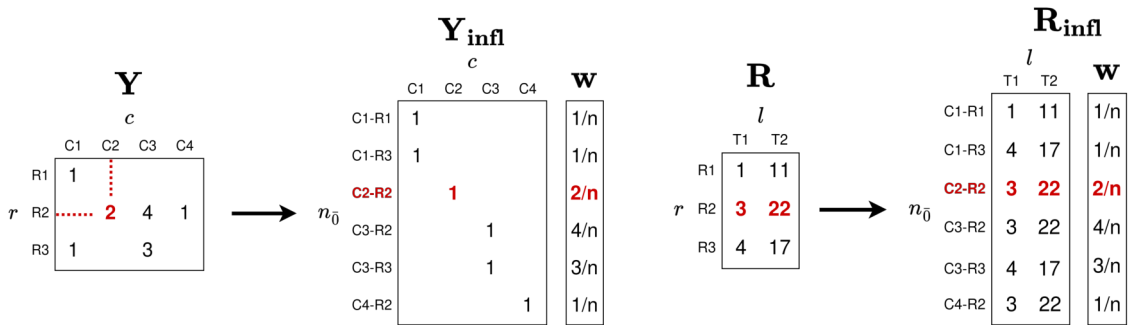
To add a measure of dispersion to CCA, we take advantages of the strong links between canonical correlation analysis (CCorA) and CCA. Indeed, canonical variates obtained from CCorA can be related to the CCA coordinates (Pélissier et al., 2002; Gimaret-Carpentier et al., 2003), but provide scores for interactions (also called correspondences), whereas CCA provides scores for species. Hence, in addition to the niche optima provided by CCA (average of scores computed for species, see chapter 2.2), CCorA provides a measure of the dispersion of interactions (variance of the interactions for each species), thus allowing to quantify niche breadth.

We present the analysis considering the interaction matrix $\mathbf{Y}$ along with resource traits $\mathbf{R}$. To perform the analysis with consumer traits, the developments are the same but using $\mathbf{Y}^\top$ and $\mathbf{Q}$.

Canonical correlation analysis (CCorA)

To highlight the similarities between CCA and CCorA, we inflate tables $\mathbf{R}$ and $\mathbf{Y}$, i.e. we reorganize the data in these matrices so that they have one row per non-null interaction. Therefore, we can see each row of the inflated matrices as representing a pairwise interaction. These rows are called correspondences, as they describe the correspondence between a row and a column of the original matrix: here, correspondences correspond to the interactions between a resource and a consumer species.

Figure 2.3.1 shows the correspondence between original and inflated matrices, using a small example dataset.



(a) Incidence matrix $\mathbf{Y}$ (b) Traits matrix $\mathbf{R}$

Figure 2.3.1: Correspondence between original and inflated matrices. (a) Correspondence between an incidence matrix $\mathbf{Y}$ and the inflated matrix $\mathbf{Y_{infl}}$. (b) Correspondence between a traits matrix $\mathbf{R}$ and the inflated matrix $\mathbf{R_{infl}}$. In this example, $\mathbf{Y}$ represents the interactions between $r = 3$ resources and $c = 4$ consumers. $\mathbf{R}$ represents $l = 2$ traits for resource species. For clarity, zeroes are omitted in the matrices. In the inflated matrices, each row corresponds to an interaction type from the original matrix ($n_{\bar{0}} = 6$ interaction types). The weight vector $\mathbf{w}$ contains interaction weights, computed as the weight of the interaction from the original matrix (where n is the grand total of $\mathbf{Y}$). The interaction between R2 and C2, and the corresponding traits for R2 are highlighted in all matrices to show their correspondence.

The inflated matrices $\mathbf{Y_{infl}}$ and $\mathbf{R_{infl}}$ are of dimensions $n_{\bar{0}} \times c$ and $n_{\bar{0}} \times l$ (respectively) where $n_{\bar{0}}$ denotes the number of non-empty interactions in matrix $\mathbf{Y}$. We also define a weight vector $\mathbf{w}$ of length $n_{\bar{0}}$, which contains the relative frequency of each correspondence from the original matrix, and the associated diagonal matrix $\mathbf{D_w} = \mathrm{diag}(\mathbf{w})$ ($n_{\bar{0}} \times n_{\bar{0}}$).

Then, we perform the weighted CCorA between $\mathbf{R_{infl}}$ and $\mathbf{Y_{infl}}$. CCorA involves the computation of the matrix of correlations $\mathbf{K}$ between $\mathbf{R_{infl}}$ and $\mathbf{Y_{infl}}$, computed as:

$$\mathbf{K} = \mathbf{S_{R_{infl}R_{infl}}}^{\top -0.5} \mathbf{S_{\bar{Y}_{infl}\bar{R}_{infl}}} \mathbf{S_{R_{infl}Y_{infl}}}^{-0.5} \tag{2.3.1}$$

where $\mathbf{\bar{Y}_{infl}}$ and $\mathbf{\bar{R}_{infl}}$ denote the centered matrices using weights $\mathbf{D_w}$. $\mathbf{S_{XZ}} = \mathbf{X}^\top \mathbf{D_w} \mathbf{Z}$ denotes the variance-covariance matrix between $\mathbf{X}$ and $\mathbf{Z}$ weighted with $\mathbf{D_w}$ and $\mathbf{S}^{-0.5}$ is the inverse of the Cholesky root of $\mathbf{S}$ (where the Cholesky root of matrix $\mathbf{A}$ is defined as the upper triangular matrix $\mathbf{L}$ so that $\mathbf{L}^\top \mathbf{L} = \mathbf{A}$). Matrix $\mathbf{K}$ ($l \times c$) contains the correlation coefficients between the columns of $\mathbf{R_{infl}}$ and $\mathbf{Y_{infl}}$ weighted by $\mathbf{D_w}$.

We perform the SDV of $\mathbf{K}$:

$$\mathbf{K} = \mathbf{R_0 \Delta Y_0} \tag{2.3.2}$$

$\mathbf{R_0}$ ($l \times k$) gives the eigenvectors associated to $\mathbf{R_{infl}}$, $\mathbf{Y_0}$ ($c \times k$) gives the eigenvectors associated to $\mathbf{Y_{infl}}$ and $\mathbf{\Delta}$ gives the canonical correlations. The CCorA eigenvalues matrix $\mathbf{\Lambda}$ is equal to $\mathbf{\Delta}^2$, and there are $k \leq \min(l, c)$ non-null eigenvalues.

The canonical coefficients, which give the contribution of $\mathbf{R_{infl}}$ and $\mathbf{Y_{infl}}$ to the multivariate axes, are computed as:

$$\mathbf{C_R} = \mathbf{S_{RR}}^{-0.5}\mathbf{R_0} \tag{2.3.3a}$$

$$\mathbf{C_Y} = \mathbf{S_{YY}}^{-0.5}\mathbf{Y_0} \tag{2.3.3b}$$

Finally, the scores of the objects (resource and consumers correspondences), called canonical variates, are computed from the original tables using the canonical coefficients:

$$\mathbf{S_R} = \mathbf{R_{infl}}\mathbf{C_R} \tag{2.3.4a}$$

$$\mathbf{S_Y} = \mathbf{Y_{infl}}\mathbf{C_Y} \tag{2.3.4b}$$

$\mathbf{S_R}$ and $\mathbf{S_Y}$ are of dimension $n_{\bar{0}} \times k$. These scores are orthonormal with respect to weights $\mathbf{D_w}$ ($\mathbf{S_R}^\top \mathbf{D_w} \mathbf{S_R} = \mathbf{I}$ and $\mathbf{S_Y}^\top \mathbf{D_w} \mathbf{S_Y} = \mathbf{I}$).

Compute niche measures from CCorA scores

Using the scores $\mathbf{S_R}$ computed above, we can define measures corresponding to consumer species niche optima, breadth and to the covariances between niche axes. These measures are defined in a meaningful way for consumers only, because here we consider the analysis constrained with resource species traits. Below, we consider the niche of consumer species j on axis k. Resource species are denoted by i.

The weighted means of the scores $\mathbf{S_R}$ (associated to resources) grouped by consumers can be seen as the consumer niche optima (Equation (2.3.5)). Indeed, this corresponds to the weighted mean of the interactions (or correspondences) in which a given consumer is involved.

$$m_{jk} = \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij}\ s_{Rk}(i, j) \tag{2.3.5}$$

Here, a double indexing is used for the elements of $\mathbf{s_{R}}_k$, which is a one-dimensional vector: $s_{Rk}(i,j)$ corresponds to the x-th element of $\mathbf{s_{R}}_k$ ($x = 1 \dots n_{\bar{0}}$), containing the score for the interaction between consumer i and resource j.

Similarly, we can compute weighted variances from the scores $\mathbf{S_R}$, that represent niche breadths of consumer species.

$$s_{jk}^2 = \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij}\ s_{Rk}(i,j)^2 - m_{jk}^2 \tag{2.3.6}$$

Finally, the weighted covariance can be computed to obtain the covariance between two niche axes k and l:

$$c_{kl}(j) = \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij} s_{Rk}(i,j) s_{Rl}(i,j) - m_{jk} m_{jl} \tag{2.3.7}$$

These measures (mean, variance, covariance) allow to draw multivariate normal ellipses to visualize species niches in the multivariate plan, like reciprocal scaling with CA in chapter 2.1.

Link CCorA scores with CCA

Above, I have established definitions of consumer niche optimum, breadth and covariance between niche axes computed from the scores $\mathbf{S_R}$ (Equations (2.3.5), (2.3.6) and (2.3.7)).

Another definition of the consumer species niche optimum was established in chapter 2.2 using CCA scores: below, I show that these two definitions are equivalent. We consider the CCA constrained with resource traits: this CCA yields resource LC scores $\mathbf{Z}$ (linear constraints) as predicted with their traits, consumer WA scores $\mathbf{V}^\star$ (weighted averages) which represent their niche optimum. Additionally, we can define consumer scores $\mathbf{V} = {\mathbf{D_c}}^{-1/2}\mathbf{V_0}$ positioned with their latent traits.

The scores $\mathbf{S_R}$ and $\mathbf{S_Y}$ computed above have a strong relationship, respectively, with the CCA scores $\mathbf{Z}$ and $\mathbf{V}$. We can even say, in a sense, that they are equivalent. Indeed, some of the $n_{\bar{0}}$ rows of $\mathbf{S_R}$ are duplicated: these duplicated rows are the correspondences *for the same resource species* and are repeated as many times as the number of interactions for this resource species. Similarly, duplicated rows of $\mathbf{S_Y}$

are the correspondences *for the same consumer species*, duplicated as many times as the number of interactions for this consumer species. We can show that the unique scores $\mathbf{S_R}$ given by CCorA correspond to the resource species scores given by CCA $\mathbf{Z}$. Similarly, the unique scores $\mathbf{S_Y}$ correspond to the consumer species scores $\mathbf{V}$. This equivalence between CCorA scores and CCA scores allows us to link the mean, variance and covariances defined from CCorA scores with CCA scores (computed below).

The niche optima computed from the CCorA scores (Equation (2.3.5)) can be expressed from the CCA scores (see Appendix A for the demonstration). For consumer j on axis k, the consumer niche optimum computed from CCorA scores correspond to the niche optimum defined with CCA using the score $\mathbf{V}^\star$.

$$m_{jk} = v^\star_{jk} \tag{2.3.8}$$

We can also express the conditional variance corresponding to niche breadth directly from the CCA scores as follows (see Appendix A for the demonstration):

$$s^2_{jk} = \frac{1}{p_{+j}} \sum_{i=1}^{r} \left(p_{ij}\ z^2_{ik}\right) - {v^\star_{jk}}^2 \tag{2.3.9}$$

It is also possible to express covariances from CCA scores (see Appendix A for the demonstration). For axes k and l, the covariance is written as:

$$c_{kl}(j) = \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij} z_{ik} z_{il} - v^\star_{jk} v^\star_{jl} \tag{2.3.10}$$

The computations above show that in practice, performing CCorA is not necessary to obtain niche measures as they can be obtained directly from CCA scores. These formulas are summarized in Table 2.3.1 (columns 3 and 4).

Proposed scaling

To improve the graphical representation of consumer niches, we propose the following scaling for the correspondences:

$$h^r_k(i,j) = \frac{s_{Rk}(i,j) + m_{jk}}{2} \tag{2.3.11}$$

$h^r_k(i,j)$ represents the scaled interaction scores computed in the resource space (r superscript).

Species	Value	From CCorA	From CCA	Scaled scores
Consumer	Correspondences	$\mathbf{S_R}$ (2.3.4a)		$\mathbf{H^r}$ (2.3.11)
	Niche optimum m_{jk}	(2.3.5)	$v^\star_{jk}$ (2.3.8)	Idem (2.3.13)
	Niche breadth s^2_{jk}	(2.3.6)	(2.3.9)	Divided by 4 (2.3.14)
	Covariance $c_{kl}(j)$	(2.3.7)	(2.3.10)	Divided by 4 (2.3.15)

Table 2.3.1: Summary table for the correspondences, mean, variance and covariance of interactions from CCA. Values give the corresponding equation numbers in the main text. The formulas for non-scaled scores (columns 3 and 4) are equivalent. The relation of scaled scores with these non-scaled scores is indicated in the table.

Using the properties of $s_{Rk}(i,j)$ and m_{jk}, this scaled score can be expressed from the CCA scores:

$$h^r_k(i,j) = \frac{z_{ik} + v^\star_{jk}}{2} \tag{2.3.12}$$

This formula shows that the scaled correspondences are at the mean between the resource LC scores and consumer WA scores.

Means, variances and covariances can then be computed by grouping scaled scores by consumer (as in Equations (2.3.5), (2.3.6) and (2.3.7)). We show below that the scaled scores also have strong links with CCA scores. With this new scaling, niche optima stay the same:

$$m_{jk} = v^\star_{jk} \tag{2.3.13}$$

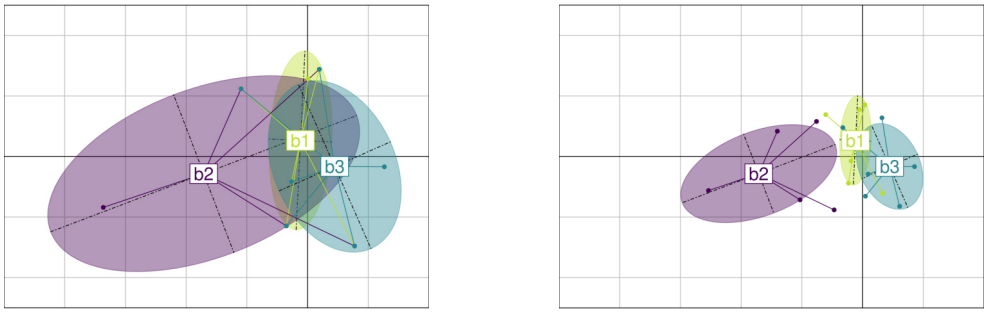
Regarding variances and covariances, they are simply divided by 4 compared with the non-scaled variances/covariances:

$$s^2_{jk} = \frac{1}{4}\left(\frac{1}{p_{+j}}\sum_{i=1}^{r}\left(p_{ij}\ {z_{ik}}^2\right) - {v^\star_{jk}}^2\right) \tag{2.3.14}$$

Covariances between axes k and l are written as:

$$c_{kl}(j) = \frac{1}{4}\left(\frac{1}{p_{+j}}\sum_{i=1}^{r} p_{ij} z_{ik} z_{il} - v^\star_{jk} v^\star_{jl}\right) \tag{2.3.15}$$

These formulas and their correspondence with non-scaled scores are summarized in Table 2.3.1.



(a) No scaling ($\mathbf{H}$)

(b) Proposed scaling ($\mathbf{H^r}$)

Figure 2.3.2: Graphical effect of the proposed scaling. This figure illustrates how the proposed scaling changes the graphical representation of species niches on a synthetic dataset where 3 consumer species niches are represented.

The proposed scaling reduces the variance and covariance compared to the non-scaled score. On the biplots, this has the effect to declutter the representation a little by separating the density ellipses better. Figure 2.3.2 illustrates the difference between niches represented with and without the proposed scaling on a synthetic example: I simulated the interactions between 3 bird species and 10 plant species and plotted the bird niches with and without scaling. Without the proposed scaling (Figure 2.3.2a), the bird-plant interactions have the same position for the same plant, and the ellipses are bigger. With the proposed scaling (Figure 2.3.2b), bird-plant interactions are not confounded and ellipses are smaller.

2.2 Add dispersion to dc-CA

We can use a similar approach to add dispersion measures to dc-CA, using its link with CCorA. In this section, we consider the interaction matrix $\mathbf{Y}$ along with resource and consumer traits $\mathbf{R}$ and $\mathbf{Q}$.

First, we perform a weighted CCorA of the inflated matrices $\mathbf{R_{infl}}$ and $\mathbf{Q_{infl}}$ (with weights $\mathbf{D_w}$). Using the same developments described for CCA, CCorA allows to define the consumer scores $\mathbf{S_Q}$ and the resource scores $\mathbf{S_R}$.

Compute niche measures from CCorA scores

Using the scores $\mathbf{S_Q}$ and $\mathbf{S_R}$ computed with CCorA, we can define means, variances and covariances, which correspond respectively to niche optima, breadths and covariances between niche dimensions. With dc-CA, niche scores can be meaningfully defined for both resources and consumers.

The weighted means of scores $\mathbf{S_Q}$ (associated to consumers) grouped by resources can be seen as consumers niche optima. Reciprocally, the weighted means of scores $\mathbf{S_R}$ (associated to resources) grouped by consumers can be seen as resources niche optima.

$$m_{ik} = \frac{1}{p_{i+}} \sum_{j=1}^{c} p_{ij} \; s_{Q_k}(i,j) \qquad \text{(resources)} \qquad (2.3.16a)$$

$$m_{jk} = \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij} \; s_{Rk}(i,j) \qquad \text{(consumers)} \qquad (2.3.16b)$$

Similarly, we can compute weighted variances from the scores $\mathbf{S_Q}$ and $\mathbf{S_R}$, that represent the niche breadth of the resource (respectively, consumer) species.

$$s^2_{ik} = \frac{1}{p_{i+}} \sum_{j=1}^{c} p_{ij} \; s_{Q_k}(i,j)^2 - m^2_{ik} \qquad \text{(resources)} \qquad (2.3.17a)$$

$$s^2_{jk} = \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij} \; s_{Rk}(i,j)^2 - m^2_{jk} \qquad \text{(consumers)} \qquad (2.3.17b)$$

Finally, we can compute the covariance between two niche axes k and l:

$$c_{kl}(i) = \frac{1}{p_{i+}} \sum_{j=1}^{c} p_{ij} s_{Q_k}(i,j) s_{Q_l}(i,j) - m_{ik} m_{il} \qquad \text{(resources)} \qquad (2.3.18a)$$

$$c_{kl}(j) = \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij} s_{Rk}(i,j) s_{Rl}(i,j) - m_{jk} m_{jl} \qquad \text{(consumers)} \qquad (2.3.18b)$$

Link CCorA scores with dc-CA

The equations above define niche optima, bot another definition was previously proposed from dc-CA in chapter 2.2: but as for CCA, the two definitions are equivalent. dc-CA yields two sets of scores for resource and consumers, used in two biplots. The first biplot uses resource scores $\mathbf{U}$ as predicted with their traits, and consumer scores

$\mathbf{V}^\star$ which represent their niche optimum. The second biplot uses consumers scores $\mathbf{V}$ predicted with their traits, jointly with resource scores $\mathbf{U}^\star$ which represent their niche optimum.

We can show that unique values of $\mathbf{S_Q}$ correspond to the dc-CA scores $\mathbf{V}$. Similarly, the unique values of $\mathbf{S_R}$ correspond scores $\mathbf{U}$. Using this equivalence between CCorA and dc-CA scores, we can express niche optima defined as a weighted mean of the CCorA scores (Equation (2.3.16)) with the dc-CA scores. For resource i or consumer j on axis k:

$$m_{ik} = u^\star_{ik} \qquad \text{(resources)} \qquad (2.3.19a)$$

$$m_{jk} = v^\star_{jk} \qquad \text{(consumers)} \qquad (2.3.19b)$$

We can also express the conditional variances corresponding to niche breadth (Equation (2.3.17)) directly from the dc-CA scores as follows:

$$s^2_{ik} = \frac{1}{p_{i+}} \sum_{j=1}^{c} p_{ij}\ v^2_{jk} - {u^\star_{ik}}^2 \qquad \text{(resources)} \qquad (2.3.20a)$$

$$s^2_{jk} = \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij}\ u^2_{ik} - {v^\star_{jk}}^2 \qquad \text{(consumers)} \qquad (2.3.20b)$$

Finally, covariances (Equation (2.3.18)) expressed from dc-CA scores for axes k and l are written as:

$$c_{kl}(i) = \frac{1}{p_{i+}} \sum_{j=1}^{c} p_{ij} v_{jk} v_{jl} - u^\star_{ik} u^\star_{il} \qquad \text{(resources)} \qquad (2.3.21a)$$

$$c_{kl}(j) = \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij} u_{ik} u_{il} - v^\star_{jk} v^\star_{jl} \qquad \text{(consumers)} \qquad (2.3.21b)$$

Scaling

We can use the same scaling defined for CCA (section 2.1) to represent correspondences in the multivariate space. Here, we define two sets of scores that will be used to define niches for resource and consumers.

$$h_k^c(i,j) = \frac{m_{ik} + s_{Q_k}(i,j)}{2} \qquad \text{for biplot 1 (resources niches)} \qquad (2.3.22a)$$

$$h_k^r(i,j) = \frac{s_{Rk}(i,j) + m_{jk}}{2} \qquad \text{for biplot 2 (consumers niches)} \qquad (2.3.22b)$$

$h_k^r(i,j)$ and $h_k^c(i,j)$ represent the scaled interaction scores computed respectively in the resource (r superscript) and consumer (c superscript) space.

This scaling has the same effect as for CCA, i.e. niche optima remain the same as described in the equations above, while variances and covariances are divided by 4. The correspondence between those scores are summarized in Table 2.3.2.

Species	Value	From CCorA	From CCA	Scaled scores
Resource	Correspondences	$\mathbf{S_Q}$		$\mathbf{H^c}$ (2.3.22a)
	Niche optimum m_{ik}	(2.3.16a)	$u^\star_{ik}$	Idem
	Niche breadth s^2_{ik}	(2.3.17a)	(2.3.20a)	Divided by 4
	Covariance $c_{kl}(j)$	(2.3.18a)	(2.3.21a)	Divided by 4
Consumer	Correspondences	$\mathbf{S_R}$		$\mathbf{H^r}$ (2.3.22b)
	Niche optimum m_{jk}	(2.3.16b)	$v^\star_{jk}$	Idem
	Niche breadth s^2_{jk}	(2.3.17b)	(2.3.20b)	Divided by 4
	Covariance $c_{kl}(j)$	(2.3.18b)	(2.3.21b)	Divided by 4

Table 2.3.2: Summary table for the correspondences, mean, variance and covariance of interactions from dc-CA. Values give the corresponding equation numbers in the main text. The formulas for non-scaled scores (columns 3 and 4) are equivalent. The relation of scaled scores with these non-scaled scores is indicated in the table.

2.3 Data analysis

To illustrate the method, I reanalyze the bird-fruit interaction network from the ANDEAN frugivory dataset (Dehling et al., 2021). All analyses were performed with R 4.3.3 (R Core Team, 2024) and are stored in a private GitHub repository (https://github.com/LisaNicvert/PhDaxis01_CAnetwork/tree/thesis) that can be shared on demand and is intended to be made public upon publication.

I perform two analyses of the interaction table: a CCA constrained by consumer (bird species) traits, and a dc-CA constrained by both resource and consumer (plant

and bird species) traits. For both analyses, I added a measure of niche breadth computed with the method presented above. The CCA constrained with plant traits with niche breadths is presented in Appendix B.

3 Results and discussion

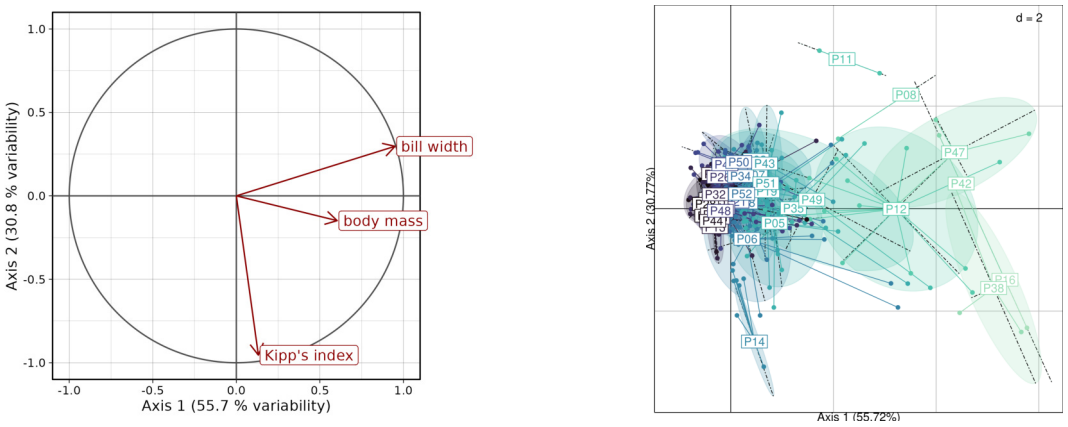
In this chapter, I presented the theoretical framework allowing to add a measure of dispersion to CCA and dc-CA by extending reciprocal scaling, originally designed for CA (Thioulouse & Chessel, 1992). This framework allows to visualize and quantify niche breadths for constrained analyses of interaction networks. In this section, I present the application to a bird-fruit interaction network (Dehling et al., 2021) and discuss the results.

3.1 CCA constrained with bird traits

First, I examine plant niches constrained with bird traits. Figure 2.3.3 shows the correlation circle of bird traits and plant niches in the two first dimensions of the multivariate space. The correlation circle is the same as presented in the previous chapter and shows a size effect on axis 1, and Kipp's index (pointed wings linked with foraging in the canopy) negatively correlated to axis 2. These two axes account for 86% of the inertia of the constrained analysis, but as discussed in the previous chapter, bird traits account for only 13% of the total inertia of the network (see section 3.1). Here, I analyzed $\mathbf{Y}^\top$ constrained with $\mathbf{Q}$, so the notations are inverted compared to the methods section.

Figure 2.3.3b represents the proposed scaling presented in the material and methods: points represent the scores $h_k^r(i,j)$ defined in Equation (2.3.11). The niche optima (corresponding to labels' positions) are computed with Equation (2.3.13), and variances and covariances represented on the ellipses are computed with Equations (2.3.14) and (2.3.15).

This figure complements Figure 2.2.2, that presented plant niche optima computed from the CCA constrained with bird traits (chapter 2.2), by adding a visual representation of niche breadths. In chapter 2.2, we could see that most plants have their optimum around the origin, implying that they tend to interact with average birds. This



(a) Correlation circle ($\mathbf{C}$)

(b) Plant niches ($h_k^c(i,j)$ grouped by plants)

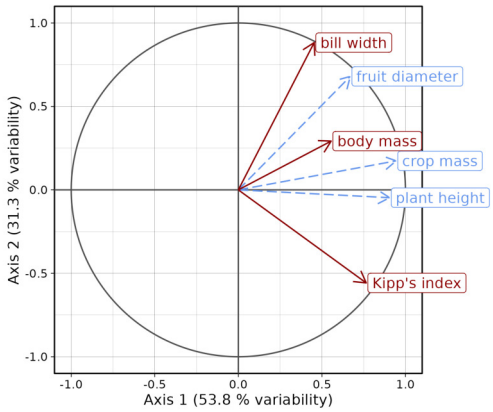
Figure 2.3.3: Correlation circle and plant niches from CCA constrained with bird traits. (a) Correlation circle of the CCA constrained with bird traits. (b) Plant niches. On this figure, each point represents a bird-fruit pairwise interaction. The scores $h_k^c(i,j)$ are computed from $\mathbf{Y}^\top$ constrained with $\mathbf{Q}$. Plant labels are positioned at the optimum of their niche and the ellipse correspond to the bivariate normal distribution of variances and covariances given by reciprocal scaling (with a scaling factor of 1.5, i.e. the ellipse axes lengths are equal to $1.5\sqrt{\lambda_k}$ on axis k, corresponding to around 67% of the points contained in the ellipse). Species are colored according to the position of their niche optima on the first axis.

figure allows to say that these species also tend to have small niches, as their ellipses are small. On this graph, we can also note that P12 *Endlicheria sp.* has a very wide niche on both axes, and that P11 *Elaeagia mariae* has a small niche that is very distinct from other species' niches. Finally, visually, plants that interact with medium-sized birds seem to have large niches on the first axis (i.e. they interact with birds of diverse sizes). This is consistent with the pattern found with the linear model from chapter 2.1 with the unconstrained analysis.

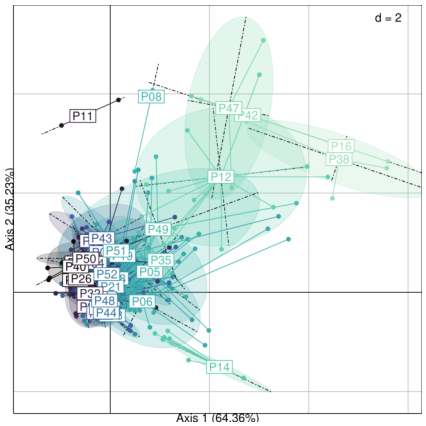
3.2 dc-CA constrained with bird and plant traits

I analyzed niches constrained with both bird and plant traits. Figure 2.3.4 shows the correlation circle (identical to the one presented in the previous chapter) and the species niches on the two first multivariate axes (accounting for 99.5% of the trait matching inertia). As discussed in the previous chapter, trait matching accounts for only about 6% of the total inertia, so these results do not capture the complete network structure.

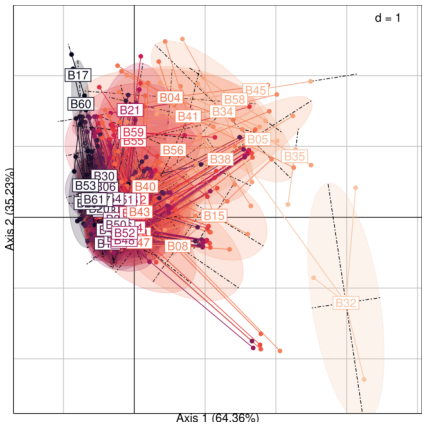
We recall the interpretation of the correlation circle (Figure 2.3.4a) from the previous chapter: axis 1 is characterized by the matching of energy traits, and axis 2 is related



(a) Correlation circle ($\mathbf{C_R}$ and $\mathbf{C_Q}$)



(b) Plant niches ($h_k^c(i,j)$ grouped by plants)



(c) Bird niches ($h_k^r(i,j)$ grouped by birds)

Figure 2.3.4: Correlation circle and niches from dc-CA. (a) Correlation circle with plant and bird traits scores (respectively $\mathbf{C_R}$ and $\mathbf{C_Q}$). Bird traits are shown in solid red lines and plant traits in dashed blue lines. (b) Plant niches. On this figure, each point represents a pairwise interaction (scores $h_k^c(i,j)$). (c) Bird niches. On this figure, each point represents a pairwise interaction (scores $h_k^r(i,j)$). Species are colored according to the position of their niche optima on the first axis. The interpretation of the labels positions and ellipses is the same as in Figure 2.3.3.

to morphological matching.

Figures 2.3.4b and 2.3.4c complete the biplots presented in chapter 2.2 (Figure 2.2.3) by adding species niche breadths. Regarding plant niches (Figure 2.3.4b), in the previous chapter, based on their optima, we could see that most plants interact with average birds, and that a small group of large plants with large fruits (in the top-right corner) tend to interact with large-billed birds and medium- to large-bodied birds. With this analysis, we can add that plants interacting with small birds tend to have small niches, and that the top-right corner group tends to have larger niches. On this analysis, P11 *Elaeagia mariae* stands out again as a plant with a very distinct niche from other plants.

Regarding birds niches (Figure 2.3.4c), in the previous chapter, we could see that most birds interact with average plants, but there was a group of large birds with wide beaks tending to eat large fruits from any plant size (top half of the plot). With this analysis, we can add that there seems to be a high overlap between all birds' niches. However, B32 (*Patagioenas plumbea*) seems to have a distinct niche compared to other species. In chapter 2.2, I hypothesized that its small beak constrained this bird to eat small fruits. With this new analysis, we can nuance this, because although this bird eats mostly small fruits, it has a wide niche on axis 2, which suggests that it consumes fruits of various diameters. If we examine the characteristics of the fruits eaten by this species, they range from 6.20 (average fruit size) to 20.40mm (which is the largest fruit in the dataset): however, the smallest fruits are their favorite, consumed 47% of the times. On this plot, we can also distinguish two opposed types of niches. Birds in the top-left corner, i.e. birds that interact with small to medium plants bearing medium to large fruits have a wider niche on axis 2 than on axis 1: they interact with plants of similar sizes, but these plants have diverse fruits sizes. Conversely, birds in the bottom-right corner (except the B32), i.e. interacting with medium to large plants bearing small to medium fruits have a wider niche on axis 1 than on axis 2: they interact with plants bearing fruits of similar size that have diverse sizes. Between these two extremes are birds with wide niches, interacting with various plant types.

4 Conclusion and perspectives

In this chapter, I developed a method to associate a measure of niche breadth taking into account multiple traits of interacting species. For that, I used the link between canonical correlation analysis (CCorA) and constrained correspondence analyses methods (CCA and dc-CA). I briefly illustrated the methods on an example network, by showing niches in the multivariate space and discussing their interpretation.

These methods allow to quantify the niche optima and niche breadths of a given species related to the traits of their interacting partners. We can use these niche measures to compare the niche optima or breadth of a given species to other species in the network to characterize this species. We could also go further in the interpretation by quantifying the niche overlap between species, for instance to inform us about po-

tential competition for resources between birds. Further perspectives are discussed in the discussion for this part (chapter 4).

Appendices

A Demonstration of the equivalence between CCA and CCorA niche measures

Here, we shortly demonstrate how means, variances and covariances computed with CCorA scores can be related to the CCA scores.

For the mean, we start from the definition of the niche optimum of computed as a weighted mean of the resources canonical variates $\mathbf{S_R}$ (Equation (2.3.5)):

$$\begin{aligned} m_{jk} &= \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij}\ s_{Rk}(i,j) \\ &= \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij} z_{ik} \qquad \text{(by property of } s_{Rk}(i,j)\text{)} \\ m_{jk} &= v^{\star}_{jk} \qquad \text{(by definition of } v^{\star}_{jk}\text{)} \end{aligned}$$

For the variance, we start from the definition of the niche breadth as a weighted variance of $\mathbf{S_R}$ (Equation (2.3.6)):

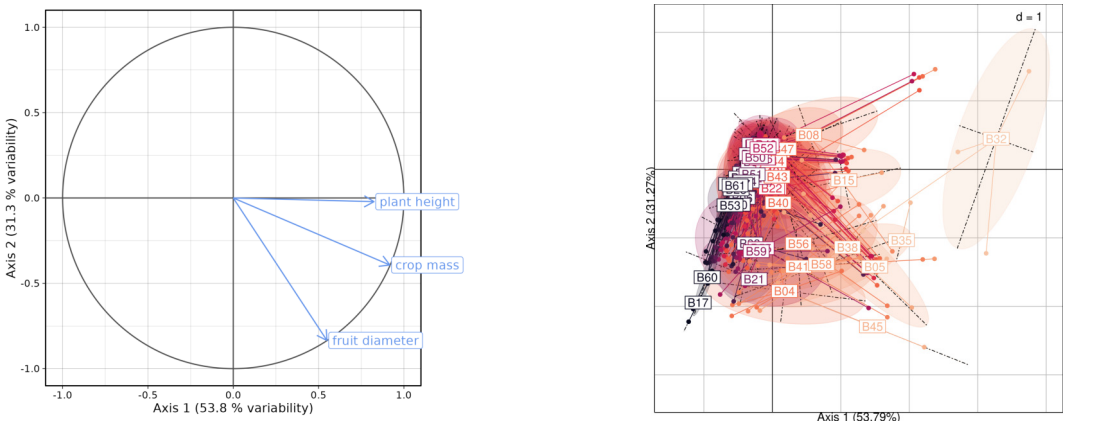
$$\begin{aligned} s^2_{jk} &= \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij}\ s_{Rk}(i,j)^2 - m^2_{jk} \\ s^2_{jk} &= \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij}\ {z_{ik}}^2 - {v^{\star}_{jk}}^2 \qquad \text{(by property of } s_{Rk}(i,j) \text{ and } m_{jk}\text{)} \end{aligned}$$

For the covariance, we start from the definition of the weighted covariance of $\mathbf{S_R}$ (Equation (2.3.7)):

$$\begin{aligned} c_{kl}(j) &= \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij} s_{Rk}(i,j) s_{Rl}(i,j) - m_{jk} m_{jl} \\ c_{kl}(j) &= \frac{1}{p_{+j}} \sum_{i=1}^{r} p_{ij}\ z_{jk} z_{jl} - v^{\star}_{jk} v^{\star}_{jl} \qquad \text{(by property of } s_{Rk,l}(i,j) \text{ and } m_{jk,l}\text{)} \end{aligned}$$

B CCA constrained with plant traits

I constrain the analysis with plant traits to measure bird niche optima and breadth. The figure below shows bird niches in the 2 first dimensions of the multivariate space, where the axes are constrained by plant traits. These 2 axes account for 85% of the inertia of plant traits. However, as discussed in the previous chapter, overall plant traits account for around 7% of the total inertia of the network (see Appendix B).



(a) Correlation circle ($\mathbf{C}$)

(b) Bird niches ($h_k^r(i,j)$ grouped by birds)

Figure B.1: Correlation circle and bird niches from CCA constrained with plant traits. (a) Correlation circle of the CCA constrained with plant traits. (b) Plant niches. On this figure, each point represents a bird-fruit pairwise interaction (scores $h_k^r(i,j)$). Species are colored according to the position of their niche optima on the first axis. The interpretation of the labels positions and ellipses is the same as in Figure 2.3.3.

The correlation circle (identical to the one in Appendix B) shows that the first axis is positively correlated with tall and productive plants, and the second axis is weakly negatively correlated to fruits diameter. In the previous chapter, we could use niche optima to see that most birds interact with small plants. Some birds also tend to interact with plants producing large fruits that can be of any size (bottom part of the graph). B32 (*Patagioenas plumbea*) stands out as a bird eating fruits from large plants bearing small fruits. With this graph, we can say that birds niches overlap a lot, apart from the plumbeous pigeon (B32).

Discussion

In this part, I used multivariate methods of the correspondence analysis family to quantify the interaction niche of species. In chapter 2.1, I employed an indirect gradient approach, using correspondence analysis (CA) and reciprocal scaling to measure species niche optima and breadths without using traits information. In chapter 2.2, I extended this approach to constrained analyses, canonical CA (CCA) and double-constrained CA (dc-CA) to measure species niche optima constrained with their interacting partners' traits and to partition the variation of the network. Finally, in chapter 2.3, I extended reciprocal scaling to constrained analyses, allowing to measure niche breadths constrained with species traits. In all chapters, I exemplified the method on a real bird-fruit interaction dataset, and for chapter 2.1, I also used a simulation approach to validate the method.

1 Quantify the niche with CA methods

In this part, I used several CA methods to define species niches. CA methods use observed interaction networks, so they quantify the realized interaction niches of species. Below and in Figure 2.D.1, I briefly describe the rationale of each method:

CA positions resource species scores as weighted averages (WA) of the consumers scores, so that they are maximally separated, and reciprocally (Figure 2.D.1a). Both scores can be seen as niche optima. In addition to species WA scores (niche optima), we can measure the variance of species interacting partners (niche breadths).

CCA is akin to CA, but we use the traits of resource or consumer species to constrain their scores. If CCA is constrained with resource species traits, it defines resource species scores that are a linear combination (LC) of their traits (Figure 2.D.1b). Consumers WA scores computed from these LC scores can be seen as niche optima.

A weighted variance of species scores (niche breadth) can also be computed using the link between CCA and canonical correlation analysis. If CCA is constrained with consumer traits, it defines resource WA scores (niche optima) and variances (niche breadths) using consumer LC scores constrained with their traits (Figure 2.D.1c).

dc-CA is constrained with both resource and consumer traits. It finds LC scores for resource and consumers, defined as linear combinations of their respective traits. Then, resource species WA scores are defined using consumers LC scores, and reciprocally. These WA scores can be seen as species niche optima, and a weighted variance can be computed using the link between dc-CA and canonical correlation analysis to define niche breadth (Figure 2.D.1b).

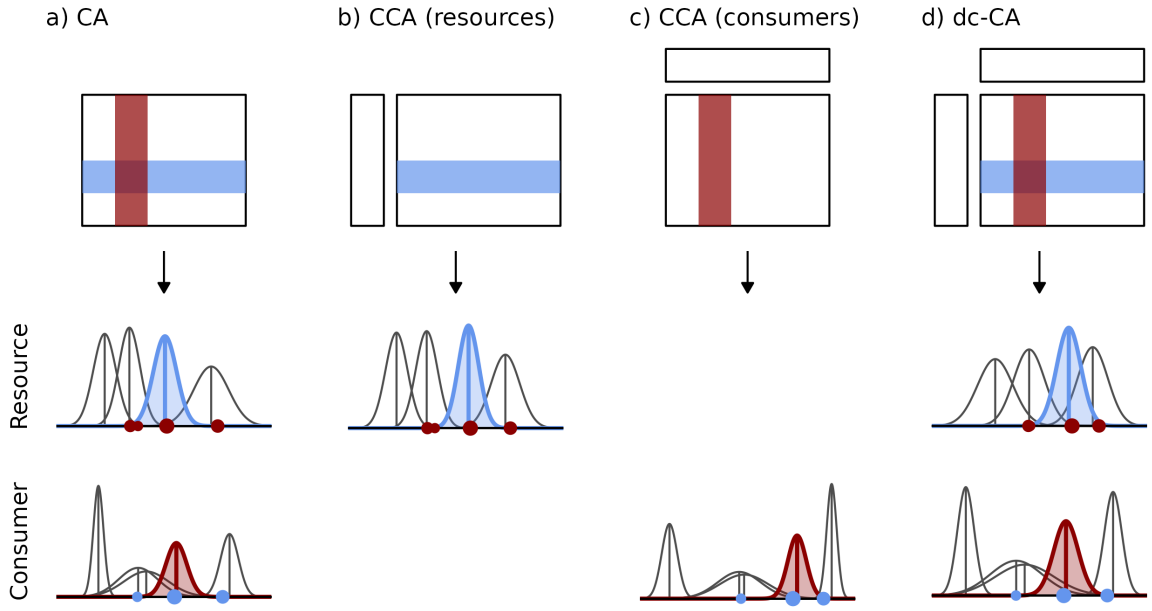
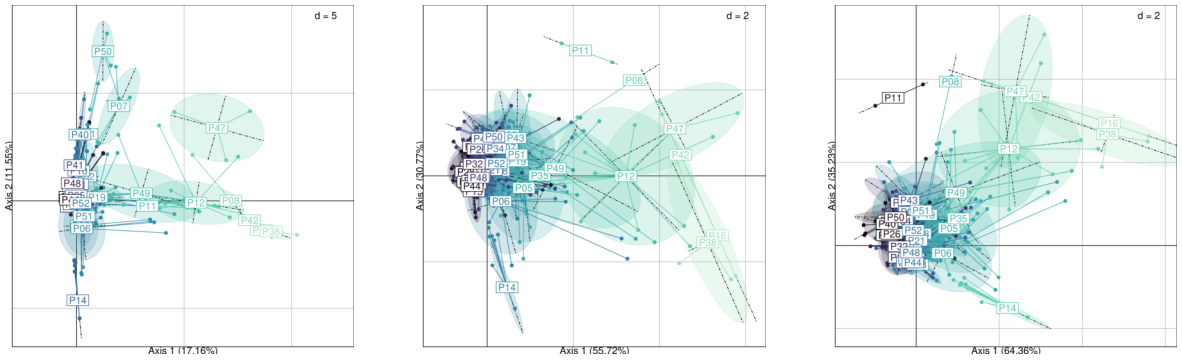


Figure 2.D.1: Summary of CA methods and niche inference. Each column represents a CA method. The first row represents the matrices inputs, and the second and third rows represent the inferred species niches in one dimension. In each case, one species for which the niche is computed is highlighted in blue (resource) or in red (consumers). The interacting partners of the species are represented as colored points on the niche axis (red for consumers and blue for resources), with their size proportional to their abundance.

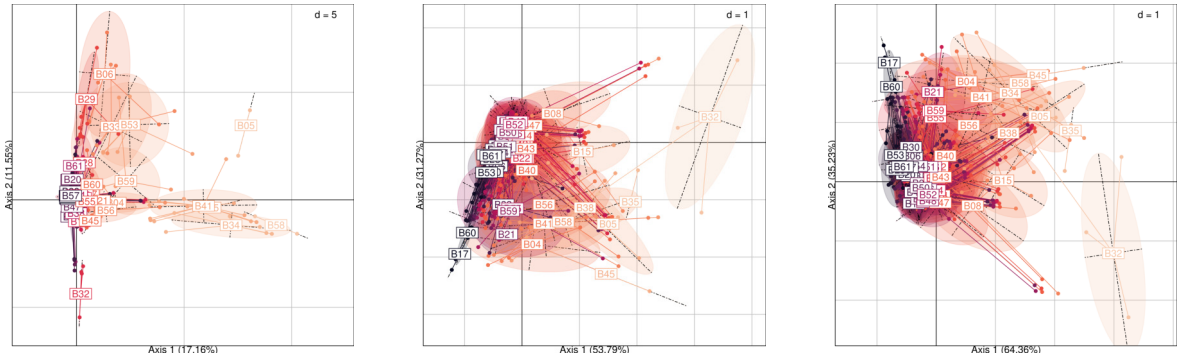
To summarize, CA defines species niche using the similarity of their interacting partners using latent traits and CCA and dc-CA define species niche using the traits of their interacting partners.

2 Comparison of the analyses

In this section, I compare the niches inferred on the bird-fruit interaction network with the different methods and discuss the implications of their differences and similarities. Figure 2.D.2 gathers the niches inferred with each method, represented in the first two axes of each multivariate space (same figures as in the chapters).



(a) CA (plants niches) (b) CCA (plants niches) (c) dc-CA (plants niches)



(d) CA (birds niches) (e) CCA (birds niches) (f) dc-CA (birds niches)

Figure 2.D.2: Comparison of the niches inferred with the different methods. The top row (a-c) shows plant niches and the bottom row (d-e) bird niches. The first column (a and d) shows CA, the second column CCA constrained with bird traits (b) and CCA constrained with plant traits (e) and the last column shows dc-CA (c and f).

First, as expected, the unconstrained method (CA) is the one that seems to separate species niches best (see Figures 2.D.2d and 2.D.2a). Indeed, with simple CA there are no linear constraints on the axes, and they are defined to maximize the variance of birds and plants scores. In other words, CA constructs birds' latent traits that maximize plants niche separation, and conversely.

Second, we can investigate the effect of adding species traits, compared to the analysis using latent traits. Plant niches computed with constrained methods (Figures 2.D.2b and 2.D.2c) are visually very similar to niches computed with CA (Figure 2.D.2a).

In particular, a group of species has consistently high scores on axis 1 (notably P49, P11, P12, P47, P42 and P38). This suggests that bird traits used in the analysis (Kipp's index, bill width and body mass) accurately capture plant niches measured with latent bird traits. On the contrary, there is a visual difference between bird niches computed with constrained methods (Figures 2.D.2e and 2.D.2f) and with CA (Figure 2.D.2d). This suggests that the plant traits (fruit diameter, crop mass and plant height) do not accurately capture birds' niche measured with latent plant traits. These results are consistent with the variation partitioning from chapter 2.2, which found that bird traits explain a larger part of the network structure than plant traits (respectively, 13% and 7%).

Third, we can compare niches computed with CA and dc-CA. For both plants and birds, they seem rather consistent (although axis 2 is flipped for birds). In other words, adding a species' own traits does not change its niche a lot compared to the niche constrained with their interacting partners' traits. Overall, this observation comforts the conclusion from the dc-CA analysis (chapter 2.2) that there is some degree of matching between the available traits. This also suggests that the available bird and plant traits accurately capture latent traits that match available plant (respectively, bird) traits. The consistency of bird niches is in contradiction with the variation partitioning analysis from chapter 2.2, which suggested that some plant traits matched by the available bird traits are missing from the analysis.

Finally, these plots of species niches in the multivariate space can inform us about their niche separation. While birds and plants niches computed with unconstrained methods are similarly separated (Figures 2.D.2a and 2.D.2d), plant niches computed with constrained methods (Figures 2.D.2b and 2.D.2c) seem to be better separated than bird niches (Figures 2.D.2e and 2.D.2f). This could indicate that bird species display a higher niche overlap than plant species, thus potentially being in competition more than plants. However, this conclusion can be nuanced by the fact that birds niches using latent traits are more separated, potentially indicating missing plant traits that separate the niches better.

3 Characteristics and limits of CA approaches

Methods in the correspondence family are useful to study species niches, but they come with a set of assumptions that must be kept in mind. First, constrained methods (CCA and dc-CA) work best when few traits are used to constrain the analyses. Indeed, there is no selection of important or redundant variables integrated in the method: therefore, when more traits than species are added, they will explain 100% of the variation, even if they are random. However, it is possible to perform variable selection to choose relevant traits (see for instance the approach of Blanchet et al., 2008).

Another point of attention concerns CA, which can struggle to recover the true traits with latent traits when trait matching is too strong, (see Appendix A). This can seem counter-intuitive, as we would expect a better performance of the method when the signal is stronger. But if we consider that CA has no external variable providing information on how to order species scores, when the gradient is not visible in the data (e.g. species are too specialized due to perfect trait matching), CA will not be able to recover the latent traits and will position species equidistant in the multivariate space, one specie on each axis. In the absence of additional information to determine species order, this actually makes sense, even if the true interactions obey to a unique trait gradient. In that case, the eigenvalues will be high for all axes, thus indicating a strong non-random structure. In this context, constrained approaches like CA and dc-CA can mitigate this problem by adding external information allowing to order species.

4 Define the interaction niche

The niche is a fundamental concept in ecology. Its broad definition is generally shared across all ecologists. However, specific aspects of the niche, such as optimum and breadth, have been defined and measured in numerous ways depending on the context (Ferry-Graham et al., 2002). The approach described here is no exception, and in this section I explicit the assumptions I made to define and measure the interaction niche.

In this framework, the niche is defined as an n-dimensional hypervolume, following Hutchinson (1957). First, I measured realized niches, as the analyses are based on observed ecological networks.

I also approximate the shape of the niche as an ellipse in dimension 2 (and a hyperellipsoid in n dimensions). This is a normal approximation of the niche space, which has no reason to be constrained to any particular sort of shape (Blonder et al., 2014). For this approximation to be meaningful, the niche of species has to be unimodal (unique optimum) and more or less symmetrical. The unimodal hypothesis is often assumed in trait matching studies (e.g. Benadi et al., 2022; Fründ et al., 2016; Albrecht et al., 2018) and posits that species have a unique trait preference: for example, we assume bird species have a preference for a single fruit size. This simplifying assumption is useful to draw general conclusions: however, I am not aware of any empirical verification in interaction networks. The symmetry assumption could also be discussed, because it is possible that the niche breadth extends more in one direction than the other (e.g. medium birds might be able to eat small fruits, but not large fruits; see for instance Peralta et al., 2024; Benadi et al., 2022). However, once the interactions are projected in the multivariate space, we could quantify the niche hypervolume using other measures than the hyperellipsoid approximation (for different volume measures, see Blonder et al., 2014).

Our framework also focuses on the part of the niche related to interactions. This is pertinent to study one specific aspect of the community functions: in the example here, seed dispersal. However, the sampled network might miss some species that are part of the seed dispersal network. Moreover, seed dispersal is not isolated of other processes at play in the community. For instance, some bird species complement their diet with insects as well, which could influence their fruit consumption. Multiplex networks, where more species types are taken into account, are increasingly recognized as relevant units to study community processes (Kéfi et al., 2016).

Finally, to define and quantify interaction niches, I drew strong parallels with species environmental (Grinnellian) niche. Indeed, there are some similarities, as the notion of hypervolume (Hutchinson, 1957) that can be applied to both niche types. However, there are also some fundamental differences. First, the interaction niche is a result of the coevolution of the interacting partners, whereas the environmental niche involves the evolution of the species only in an environment that is not subject to natural selection. This leads to fundamental differences in the modeling of these processes and in the interpretation of the scores measured by CA methods for consumer-resource

or species-by-sites matrices. Moreover, the environmental niche has more often been described as the conditions of persistence of a species, or "what a species *needs*", (Devictor et al., 2010), whereas the interaction (Eltonian) niche is more focused on the ecological processes in which the species takes part, "what a species *does*", (Devictor et al., 2010). In the end, the distinction between Eltonian and Grinnellian niche is dictated by the focus of the study. Indeed, we could argue that some interspecific interactions are essential for species persistence (e.g. trophic links), and that some species have functional effects through their environment (e.g. engineer species). Ultimately, the distinction between Eltonian and Grinnellian niches is not clear-cut and resides mostly in the eye of the observer (Devictor et al., 2010).

5 What structures networks and how much?

CA methods can be used to quantify the part of total variation in the interaction matrix explained by species traits. Variation partitioning thus allows to quantify the part of interactions explained by resource traits, consumer traits, or the matching between traits (see Figure 2.2.1). In the bird-fruit interaction network analyzed here, trait matching explained about 6% of the total variation, and more than 85% of the variation was not explained by any of the measured traits (see Table 2.2.1).

Quantifying the importance of trait matching on the structure of interaction networks allows us to measure the relative importance of this process. In our bird-fruit interaction network example, trait matching explains a minor part of the structure in the matrix. We can wonder how this result generalizes to other ecological communities, or network types. During exploratory analyses on other mutualistic networks (not presented in this thesis), I could see that trait matching was also a weak explanatory factor of network structure. The question is, then, what are the factors that structure these networks, and how much do these factor contribute to the observed pattern? Candidate factors include environmental variation, species abundance or phylogeny, but their relative importance remains to be quantified (Vázquez, Chacoff, & Cagnolo, 2009). Although CA could be constrained with phylogeny, there is no obvious way to include species abundances collected independently or environmental factors in the analysis.

6 Conclusion and perspectives

In this part, I used multivariate methods of the correspondence analysis family to quantify species interaction niches. These methods, some of which (CA and CCA) are well-known to ecologists, can be fruitfully re-employed and extended to study interaction networks.

An interesting development of these methods would be to quantify species niche overlap, which can contribute to the strength of competition between two species. This overlap could be quantified by computing the overlapping n-dimensional volume between the niche hyperellipsoids of species (with approaches similar to those used by Blonder et al., 2014 and Pappas and Stoermer, 1997).

An increasingly studied aspect of species niche is the individual aspect of the niche: "interactions, after all, occur between individuals" (Sexton et al., 2017). In our approach, the niche is defined at the species level, and related to species-averaged traits. It would be interesting to complement this approach with an individual-based one. It would allow us to see how the niche varies within species and between individuals (if it does at all) to better understand the species-specific patterns of niche measures (Sexton et al., 2017). For instance, we could imagine two species with the same species-level niche breadth, but in one of the species, all individuals have a wide niche and approximately the same niche optima and in the other species, all individuals have narrow niches but different optima.

Finally, I argue that the number of interacting partners of a species, or diversity indices based on this number, despite their popularity, are not necessarily good descriptors of the interaction niche breadth (Dehling & Stouffer, 2018). To understand this seemingly counter-intuitive statement, we can consider the parallel with the environmental niche of species. For a species to be considered an environmental generalist, merely being present in numerous locations is not enough: what matters is whether the occupied sites encompass diverse environmental conditions (e.g. elevation, temperature, precipitations). Similarly, I argue that to measure the interaction niche breadth of species, we should take into account the traits of their interacting partners, and how (dis)similar they are.

Bibliography

Albrecht, J., Classen, A., Vollstädt, M. G. R., Mayr, A., Mollel, N. P., Schellenberger Costa, D., Dulle, H. I., Fischer, M., Hemp, A., Howell, K. M., Kleyer, M., Nauss, T., Peters, M. K., Tschapka, M., Steffan-Dewenter, I., Böhning-Gaese, K., & Schleuning, M. (2018). Plant and animal functional diversity drive mutualistic network assembly across an elevational gradient. *Nature Communications*, *9*(1), 3177. https://doi.org/10.1038/s41467-018-05610-w

Araujo, A. I. L., de Almeida, A. M., Cardoso, M. Z., & Corso, G. (2010). Abundance and nestedness in interaction networks. *Ecological Complexity*, *7*(4), 494–499. https://doi.org/10.1016/j.ecocom.2010.02.004

Bascompte, J., & Jordano, P. (2007). Plant-Animal Mutualistic Networks: The Architecture of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 567–593. https://doi.org/10.1146/annurev.ecolsys.38.091206.095818

Bascompte, J., Jordano, P., Melián, C. J., & Olesen, J. M. (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, *100*(16), 9383–9387. https://doi.org/10.1073/pnas.1633576100

Bascompte, J., Melián, C. J., & Sala, E. (2005). Interaction strength combinations and the overfishing of a marine food web. *Proceedings of the National Academy of Sciences*, *102*(15), 5443–5447. https://doi.org/10.1073/pnas.0501562102

Bellwood, D., Wainwright, P., Fulton, C., & Hoey, A. (2005). Functional versatility supports coral reef biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1582), 101–107. https://doi.org/10.1098/rspb.2005.3276

Benadi, G., Dormann, C. F., Fründ, J., Stephan, R., & Vázquez, D. P. (2022). Quantitative Prediction of Interactions in Bipartite Networks Based on Traits, Abundances, and Phylogeny. *The American Naturalist*, *199*(6), 841–854. https://doi.org/10.1086/714420

Bender, I. M. A., Kissling, W. D., Blendinger, P. G., Böhning-Gaese, K., Hensen, I., Kühn, I., Muñoz, M. C., Neuschulz, E. L., Nowak, L., Quitián, M., Saavedra, F., Santillán, V., Töpfer, T., Wiegand, T., Dehling, D. M., & Schleuning, M. (2018). Morphological trait matching shapes plant–frugivore networks across the Andes. *Ecography*, *41*(11), 1910–1919. https://doi.org/10.1111/ecog.03396

Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward Selection of Explanatory Variables. *Ecology*, *89*(9), 2623–2632. https://doi.org/10.1890/07-0986.1

Blonder, B., Lamanna, C., Violle, C., & Enquist, B. J. (2014). The n-dimensional hypervolume. *Global Ecology and Biogeography*, *23*(5), 595–609. https://doi.org/10.1111/geb.12146

Blüthgen, N., Fründ, J., Vázquez, D. P., & Menzel, F. (2008). What do interaction network metrics tell us about specialization and biological traits. *Ecology*, *89*(12), 3387–3399. https://doi.org/10.1890/07-2121.1

Blüthgen, N., Menzel, F., & Blüthgen, N. (2006). Measuring specialization in species interaction networks. *BMC Ecology*, *6*(1), 1–12. https://doi.org/10.1186/1472-6785-6-9

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the Spatial Component of Ecological Variation. *Ecology*, *73*(3), 1045–1055. https://doi.org/10.2307/1940179

Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, *31*, 343–366. Retrieved April 8, 2024, from https://www.jstor.org/stable/221736

Dehling, D. M., Bender, I. M. A., Blendinger, P. G., Muñoz, M. C., Quitián, M., Saavedra, F., Santillán, V., Böhning-Gaese, K., Neuschulz, E.-L., & Schleuning, M. (2021). ANDEAN frugivory: Data on plant–bird interactions and functional traits of plant and bird species from montane forests along the Andes. https://doi.org/10.5061/DRYAD.WM37PVMN5

Dehling, D. M., Jordano, P., Schaefer, H. M., Böhning-Gaese, K., & Schleuning, M. (2016). Morphology predicts species' functional roles and their degree of specialization in plant–frugivore interactions. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1823), 20152444. https://doi.org/10.1098/rspb.2015.2444

Dehling, D. M., & Stouffer, D. B. (2018). Bringing the Eltonian niche into functional diversity. *Oikos*, *127*(12), 1711–1723. https://doi.org/10.1111/oik.05415

Dehling, D. M., Töpfer, T., Schaefer, H. M., Jordano, P., Böhning-Gaese, K., & Schleuning, M. (2014). Functional relationships beyond species richness patterns: Trait matching in plant–bird mutualisms across scales. *Global Ecology and Biogeography*, *23*(10), 1085–1093. https://doi.org/10.1111/geb.12193

Devictor, V., Clavel, J., Julliard, R., Lavergne, S., Mouillot, D., Thuiller, W., Venail, P., Villéger, S., & Mouquet, N. (2010). Defining and measuring ecological specialization. *Journal of Applied Ecology*, *47*(1), 15–25. https://doi.org/10.1111/j.1365-2664.2009.01744.x

Dolédec, S., Chessel, D., ter Braak, C. J. F., & Champely, S. (1996). Matching species traits to environmental variables: A new three-table ordination method. *Environmental and Ecological Statistics*, *3*(2), 143–166. https://doi.org/10.1007/BF02427859

Dolédec, S., Chessel, D., & Gimaret-Carpentier, C. (2000). Niche Separation in Community Analysis: A New Method. *Ecology*, *81*(10), 2914–2927. https://doi.org/10.1890/0012-9658(2000)081[2914:NSICAA]2.0.CO;2

Dormann, C. F., Fründ, J., & Schaefer, H. M. (2017). Identifying Causes of Patterns in Ecological Networks: Opportunities and Limitations. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 559–584. https://doi.org/10.1146/annurev-ecolsys-110316-022928

Dray, S., & Legendre, P. (2008). Testing the Species Traits–Environment Relationships: The Fourth-Corner Problem Revisited. *Ecology*, *89*(12), 3400–3412. https://doi.org/10.1890/08-0349.1

Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences*, *99*(20), 12917–12922. https://doi.org/10.1073/pnas.192407699

Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N. P., Dalsgaard, B., de Sassi, C., Galetti, M., Guimarães, P. R., Lomáscolo, S. B., Martín González, A. M., Pizo, M. A., Rader, R., Rodrigo, A., Tylianakis, J. M., Vázquez, D. P., & Allesina, S. (2013). The dimensionality of ecological networks. *Ecology Letters*, *16*(5), 577–583. https://doi.org/10.1111/ele.12081

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge University Press.

Elton, C. (1927). The animal community. In *Animal Ecology*. Sidgwick; Jackson.

Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, *6*(3-4), 290–297. https://doi.org/10.5486/PMD.1959.6.3-4.12

Erofeeva, E. A. (2021). Plant hormesis and Shelford's tolerance law curve. *Journal of Forestry Research*, *32*(5), 1789–1802. https://doi.org/10.1007/s11676-021-01312-0

Ferry-Graham, L. A., Bolnick, D. I., & Wainwright, P. C. (2002). Using Functional Morphology to Examine the Ecology and Evolution of Specialization. *Integrative and Comparative Biology*, *42*(2), 265–277. https://doi.org/10.1093/icb/42.2.265

Foster, M. S. (1987). Feeding Methods and Efficiencies of Selected Frugivorous Birds. *The Condor*, *89*(3), 566–580. https://doi.org/10.2307/1368645

Foster, M. S. (1990). Factors Influencing Bird Foraging Preferences among Conspecific Fruit Trees. *The Condor*, *92*(4), 844–854. https://doi.org/10.2307/1368720

Fründ, J., McCann, K. S., & Williams, N. M. (2016). Sampling bias is a challenge for quantifying specialization and network structure: Lessons from a quantitative niche model. *Oikos*, *125*(4), 502–513. https://doi.org/10.1111/oik.02256

Futuyma, D. J., & Moreno, G. (1988). The Evolution of Ecological Specialization. *Annual Review of Ecology and Systematics*, *19*, 207–233. Retrieved May 14, 2024, from https://www.jstor.org/stable/2097153

Gauch Jr., H. G., & Whittaker, R. H. (1972). Coenocline Simulation. *Ecology*, *53*(3), 446–451. https://doi.org/10.2307/1934231

Gimaret-Carpentier, C., Dray, S., & Pascal, J.-P. (2003). Broad-scale biodiversity pattern of the endemic tree flora of the Western Ghats (India) using canonical correlation analysis of herbarium records. *Ecography*, *26*(4), 429–444. https://doi.org/10.1034/j.1600-0587.2003.03356.x

Godoy, O., Bartomeus, I., Rohr, R. P., & Saavedra, S. (2018). Towards the Integration of Niche and Network Theories. *Trends in Ecology & Evolution*, *33*(4), 287–300. https://doi.org/10.1016/j.tree.2018.01.007

Green, R. H. (1971). A Multivariate Statistical Approach to the Hutchinsonian Niche: Bivalve Molluscs of Central Canada. *Ecology*, *52*(4), 543–556. https://doi.org/10.2307/1934142

Green, R. H. (1974). Multivariate Niche Analysis with Temporally Varying Environmental Factors. *Ecology*, *55*(1), 73–83. https://doi.org/10.2307/1934619

Greenacre, M. (2010). Correspondence analysis of raw data. *Ecology*, *91*(4), 958–963. https://doi.org/10.1890/09-0239.1

Greenacre, M. (2013). The contributions of rare objects in correspondence analysis. *Ecology*, *94*(1), 241–249. https://doi.org/10.1890/11-1730.1

Grinnell, J. (1924). Geography and Evolution. *Ecology*, *5*(3), 225–229. https://doi.org/10.2307/1929447

Hill, M. O. (1973). Reciprocal Averaging: An Eigenvector Method of Ordination. *The Journal of Ecology*, *61*(1), 237. https://doi.org/10.2307/2258931

Hill, M. O. (1974). Correspondence Analysis: A Neglected Multivariate Method. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *23*(3), 340–354. https://doi.org/10.2307/2347127

Hirschfeld, H. O. (1935). A Connection between Correlation and Contingency. *Mathematical Proceedings of the Cambridge Philosophical Society*, *31*(4), 520–524. https://doi.org/10.1017/S0305004100013517

Hotelling, H. (1936). Relations Between Two Sets of Variates. *Biometrika*, *28*(3/4), 321–377. https://doi.org/10.2307/2333955

Hutchinson, G. E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, *22*(0), 415–427. https://doi.org/10.1101/SQB.1957.022.01.039

Ings, T. C., Montoya, J. M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C. F., Edwards, F., Figueroa, D., Jacob, U., Jones, J. I., Lauridsen, R. B., Ledger, M. E., Lewis, H. M., Olesen, J. M., Van Veen, F. F., Warren, P. H., & Woodward, G. (2009). Ecological networks – beyond food webs. *Journal of Animal Ecology*, *78*(1), 253–269. https://doi.org/10.1111/j.1365-2656.2008.01460.x

Jordano, P., Bascompte, J., & Olesen, J. M. (2003). Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters*, *6*(1), 69–81. https://doi.org/10.1046/j.1461-0248.2003.00403.x

Kéfi, S., Miele, V., Wieters, E. A., Navarrete, S. A., & Berlow, E. L. (2016). How Structured Is the Entangled Bank? The Surprisingly Simple Organization of Multiplex Ecological Networks Leads to Increased Persistence and Resilience. *PLOS Biology*, *14*(8), e1002527. https://doi.org/10.1371/journal.pbio.1002527

Legendre, P., Galzin, R., & Harmelin-Vivien, M. L. (1997). Relating Behavior to Habitat: Solutions to The fourth-Corner Problem. *Ecology*, *78*(2), 547–562. https://doi.org/10.1890/0012-9658(1997)078[0547:RBTHST]2.0.CO;2

Legendre, P., & Legendre, L. (2012, August). *Numerical Ecology* [Google-Books-ID: DKlUIQcHhOsC]. Elsevier.

Lewinsohn, T. M., Inácio Prado, P., Jordano, P., Bascompte, J., & M. Olesen, J. (2006). Structure in plant–animal interaction assemblages. *Oikos*, *113*(1), 174–184. https://doi.org/10.1111/j.0030-1299.2006.14583.x

Maglianesi, M. A., Böhning-Gaese, K., & Schleuning, M. (2015). Different foraging preferences of hummingbirds on artificial and natural flowers reveal mechanisms structuring plant–pollinator interactions. *Journal of Animal Ecology*, *84*(3), 655–664. https://doi.org/10.1111/1365-2656.12319

McGill, B., Enquist, B., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, *21*(4), 178–185. https://doi.org/10.1016/j.tree.2006.02.002

Miele, V., Matias, C., Robin, S., & Dray, S. (2019). Nine quick tips for analyzing network data. *PLOS Computational Biology*, *15*(12), e1007434. https://doi.org/10.1371/journal.pcbi.1007434

Minchin, P. R. (1987). Simulation of multidimensional community patterns: Towards a comprehensive model. *Vegetatio*, *71*(3), 145–156. https://doi.org/10.1007/BF00039167

Moermond, T. C., & Denslow, J. S. (1985). Neotropical Avian Frugivores: Patterns of Behavior, Morphology, and Nutrition, with Consequences for Fruit Selection. *Ornithological Monographs*, (36), 865–897. https://doi.org/10.2307/40168322

Palacio, F., Valoy, M., Bernacki, F., Sánchez, M., Núñez-Montellano, M., Varela, O., & Ordano, M. (2017). Bird fruit consumption results from the interaction between fruit-handling behaviour and fruit crop size. *Ethology Ecology & Evolution*, *29*(1), 24–37. https://doi.org/10.1080/03949370.2015.1080195

Pappas, J. L., & Stoermer, E. F. (1997). Multivariate measure of niche overlap using canonical correspondence analysis. *Écoscience*, *4*(2), 240–245. https://doi.org/10.1080/11956860.1997.11682401

Pearson, D. L. (1971). Vertical Stratification of Birds in a Tropical Dry Forest. *The Condor*, *73*(1), 46–55. https://doi.org/10.2307/1366123

Pélissier, R., Dray, S., & Sabatier, D. (2002). Within-plot relationships between tree species occurrences and hydrological soil constraints: An example in French Guiana investigated through canonical correlation analysis. *Plant Ecology*, *162*(2), 143–156. https://doi.org/10.1023/A:1020399603500

Peng, F.-J., ter Braak, C. J. F., Rico, A., & Van den Brink, P. J. (2021). Double constrained ordination for assessing biological trait responses to multiple stressors: A case study with benthic macroinvertebrate communities. *Science of The Total Environment*, *754*, 142171. https://doi.org/10.1016/j.scitotenv.2020.142171

Peralta, G., CaraDonna, P. J., Rakosy, D., Fründ, J., Pascual Tudanca, M. P., Dormann, C. F., Burkle, L. A., Kaiser-Bunbury, C. N., Knight, T. M., Resasco, J., Winfree, R., Blüthgen, N., Castillo, W. J., & Vázquez, D. P. (2024). Predicting plant–pollinator interactions: Concepts, methods, and challenges. *Trends in Ecology & Evolution*, S0169534723003361. https://doi.org/10.1016/j.tree.2023.12.005

Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation Partitioning of Species Data Matrices: Estimation and Comparison of Fractions. *Ecology*, *87*(10), 2614–2625. https://doi.org/10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2

Phillips, R. D., Peakall, R., Van Der Niet, T., & Johnson, S. D. (2020). Niche Perspectives on Plant–Pollinator Interactions. *Trends in Plant Science*, *25*(8), 779–793. https://doi.org/10.1016/j.tplants.2020.03.009

Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: Why ecological interaction networks vary through space and time. *Oikos*, *124*(3), 243–251. https://doi.org/10.1111/oik.01719

R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Raffaelli, D. (2007). Food Webs, Body Size and the Curse of the Latin Binomial. In N. Rooney, K. S. McCann, & D. L. G. Noakes (Eds.), *From Energetics to Ecosystems: The Dynamics and Structure of Ecological Systems* (pp. 53–64). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5337-5_3

Schleuning, M., Blüthgen, N., Flörchinger, M., Braun, J., Schaefer, H. M., & Böhning-Gaese, K. (2011). Specialization and interaction strength in a tropical plant–frugivore network differ among forest strata. *Ecology*, *92*(1), 26–36. https://doi.org/10.1890/09-1842.1

Schoener, T. W. (1971). Theory of Feeding Strategies. *Annual Review of Ecology, Evolution and Systematics*, *2*(Volume 2, 1971), 369–404. https://doi.org/10.1146/annurev.es.02.110171.002101

Sexton, J. P., Montiel, J., Shay, J. E., Stephens, M. R., & Slatyer, R. A. (2017). Evolution of Ecological Niche Breadth. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 183–206. https://doi.org/10.1146/annurev-ecolsys-110316-023003

Shelford, V. E. (1931). Some Concepts of Bioecology. *Ecology*, *12*(3), 455–467. https://doi.org/10.2307/1928991

Sîrbu, I., Benedek, A. M., & Sîrbu, M. (2021). Variation partitioning in double-constrained multivariate analyses: Linking communities, environment, space, functional traits, and ecological niches. *Oecologia*, *197*(1), 43–59. https://doi.org/10.1007/s00442-021-05006-6

Smith, T. M., & Smith, R. L. (2015). *Elements of ecology* (Global ed., 9. ed). Pearson.

ter Braak, C. J. F. (1995). Ordination. In R. H. Jongman, C. J. F. ter Braak, & O. F. R. Van Tongeren (Eds.), *Data analysis in community and landscape ecology* (New ed., with corr). Cambridge University Press.

ter Braak, C. J. F. (1986). Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. *Ecology*, *67*(5), 1167–1179. https://doi.org/10.2307/1938672

ter Braak, C. J. F. (1987). The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, *69*(1), 69–77. https://doi.org/10.1007/BF00038688

ter Braak, C. J. F., Šmilauer, P., & Dray, S. (2018). Algorithms and biplots for double constrained correspondence analysis. *Environmental and Ecological Statistics*, *25*(2), 171–197. https://doi.org/10.1007/s10651-017-0395-x

ter Braak, C. J. F., & Verdonschot, P. F. M. (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Sciences*, *57*(3), 255–289. https://doi.org/10.1007/BF00877430

ter Braak, C. J., & Prentice, I. (2004). A Theory of Gradient Analysis. In *Advances in Ecological Research* (pp. 235–282, Vol. 34). Elsevier. https://doi.org/10.1016/S0065-2504(03)34003-6

Thiel, S., Willems, F., Farwig, N., Rehling, F., Schabo, D. G., Schleuning, M., Shahuano Tello, N., Töpfer, T., Tschapka, M., Heymann, E. W., & Heer, K. (2023). Vertically stratified frugivore community composition and interaction frequency in a liana fruiting across forest strata. *Biotropica*, *55*(3), 650–664. https://doi.org/10.1111/btp.13216

Thioulouse, J., & Chessel, D. (1992). A Method for Reciprocal Scaling of Species Tolerance and Sample Diversity. *Ecology*, *73*(2), 670–680. https://doi.org/10.2307/1940773

Thioulouse, J., Dray, S., Dufour, A.-B., Siberchicot, A., Jombart, T., & Pavoine, S. (2018). *Multivariate Analysis of Ecological Data with {ade4}*. Springer. doi.org/10.1007/978-1-4939-8850-1

Valdovinos, F. S. (2019). Mutualistic networks: Moving closer to a predictive theory. *Ecology Letters*, *22*(9), 1517–1534. https://doi.org/10.1111/ele.13279

Valdovinos, F. S., Brosi, B. J., Briggs, H. M., Moisset de Espanés, P., Ramos-Jiliberto, R., & Martinez, N. D. (2016). Niche partitioning due to adaptive foraging reverses effects of nestedness and connectance on pollination network stability. *Ecology Letters*, *19*(10), 1277–1286. https://doi.org/10.1111/ele.12664

Valiente-Banuet, A., Aizen, M. A., Alcántara, J. M., Arroyo, J., Cocucci, A., Galetti, M., García, M. B., García, D., Gómez, J. M., Jordano, P., Medel, R., Navarro, L., Obeso, J. R., Oviedo, R., Ramírez, N., Rey, P. J., Traveset, A., Verdú, M., & Zamora, R. (2015). Beyond species loss: The extinction of ecological interactions in a changing world. *Functional Ecology*, *29*(3), 299–307. https://doi.org/10.1111/1365-2435.12356

van Dam, A., Dekker, M., Morales-Castilla, I., Rodríguez, M.
bibinitperiod, Wichmann, D., & Baudena, M. (2021). Correspondence analysis, spectral clustering and graph embedding: Applications to ecology and economic complexity. *Scientific Reports*, *11*(1), 8926. https://doi.org/10.1038/s41598-021-87971-9

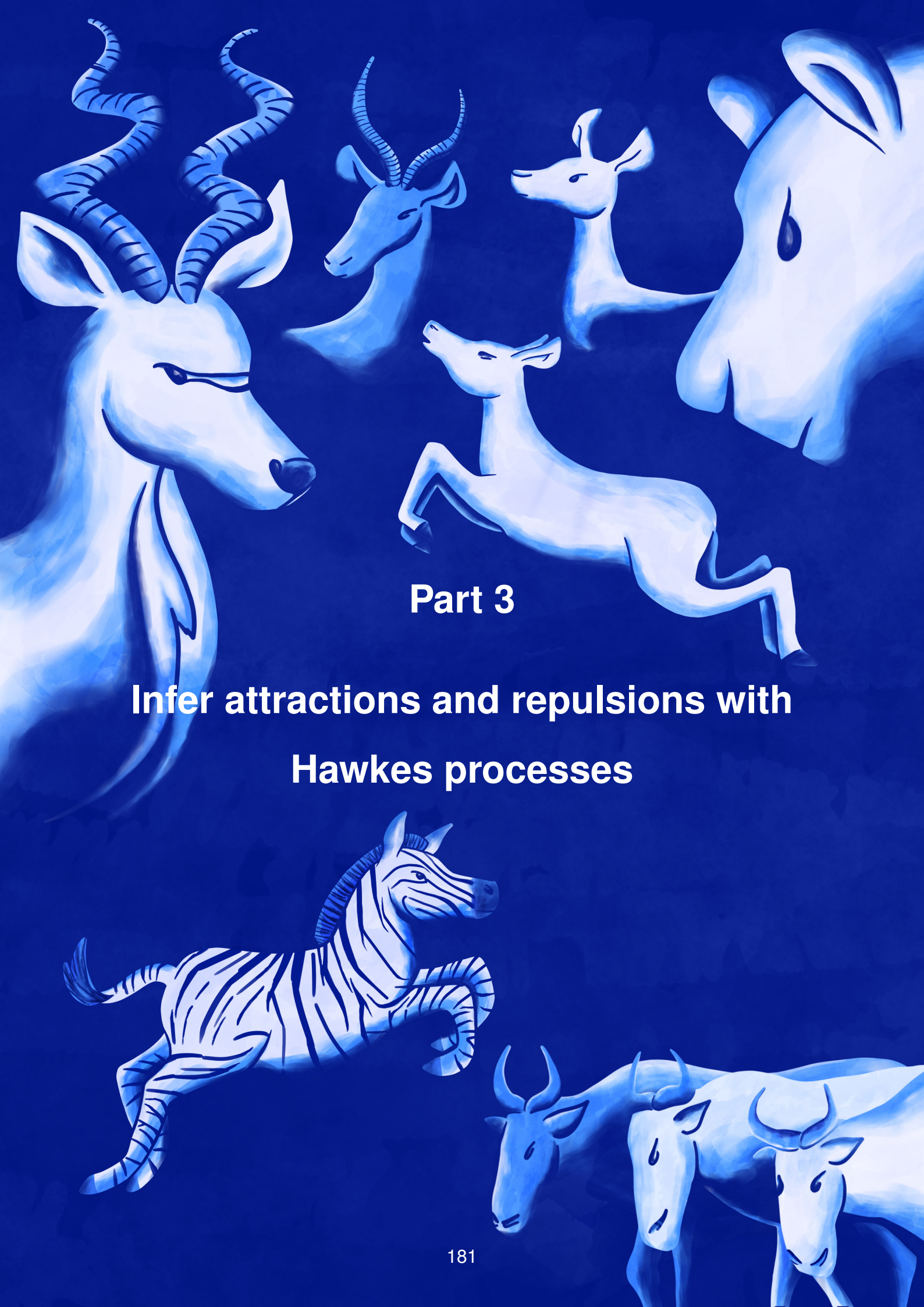
Vandermeer, J. H. (1972). Niche Theory. *Annual Review of Ecology and Systematics*, *3*, 107–132. Retrieved January 29, 2024, from https://www.jstor.org/stable/2096844

Vázquez, D. P., Blüthgen, N., Cagnolo, L., & Chacoff, N. P. (2009). Uniting pattern and process in plant–animal mutualistic networks: A review. *Annals of Botany*, *103*(9), 1445–1457. https://doi.org/10.1093/aob/mcp057

Vázquez, D. P., Chacoff, N. P., & Cagnolo, L. (2009). Evaluating multiple determinants of the structure of plant–animal mutualistic networks. *Ecology*, *90*(8), 2039–2046. https://doi.org/10.1890/08-1837.1

Wheelwright, N. T. (1985). Fruit-Size, Gape Width, and the Diets of Fruit-Eating Birds. *Ecology*, *66*(3), 808–818. https://doi.org/10.2307/1940542

Williams, R. J., & Martinez, N. D. (2000). Simple rules yield complex food webs. *Nature*, *404*(6774), 180–183. https://doi.org/10.1038/35004572



Part 3

Infer attractions and repulsions with Hawkes processes

Cover picture: the cyanotype technique consists in arranging objects on a surface coated with photosensitive compounds, which are then exposed to light and leave their imprint on the surface. This drawing represents the imprint of species occurrences on subsequent occurrences as captured with a cyanotype: from left to right and top to bottom, species are greater kudu, impalas, lion, zebra and wildebeest.

© Own drawing.

Introduction

1 Preamble: PhD context

My PhD was originally structured only around the work developed in this axis. However, at the end of my first year, we decided to extend the scope of the PhD, because inference on real data yielded only null interspecific interactions. Six months later, as I was tidying up the analyses to include this work in my thesis, I realized that there was an error in the code I had written to analyze data. This error randomized interspecific occurrence times and explained why I was not able to detect interactions with real data. I began working on this axis again, but even though the model fared better than expected, contrary to what we originally planned, the experimental design and the available data did not allow to robustly test ecological hypotheses using this model, so the PhD project was diversified with other axes.

During my PhD, I had the opportunity to spend 3 months in South Africa and to meet people working on the Snapshot Safari project, which collects the camera trap data used in this axis. This experience allowed me to have great exchanges with people who are part of the Snapshot Safari project, to better understand the biological system I analyzed and to be more familiar with the camera trap data protocol used for the Snapshot project.

This axis is the first one I worked on during my PhD, and holds a special place in my heart. It is the first subject I investigated in a research context, the axis I spend the most time on and I invested myself in the most and a stimulating and novel research question.

2 Introduction

Species distribution patterns are influenced by a variety of factors in interplay acting at different scales. First, the evolutionary history of the species influence their distribution. For instance, Malagasy Carnivora are endemic to the island of Madagascar because of the geological history of Madagascar, that has been isolated from other landmasses for more than 80 million years (Yoder et al., 2003). Environmental variables, like climate or habitat, also influence species distribution. For example, some plant species of alpine communities require cold temperatures to develop, as their germination is highly favored when the seed is exposed to cold temperatures under snow cover (Cavieres & Sierra-Almeida, 2018). Another important factor is habitat: for example, klipspringers (small antelopes of Eastern and Southern Africa) prefer rocky outcrops that their hooves are adapted to (Estes, 2012). Temporal variables like the season or the time of day can also affect species distributions, like for plankton communities which display a diel vertical migration following the day-night cycles (Haren & Compton, 2013). Finally, biotic interactions like competition or mutualism may influence species distribution as well. Thus, plants of the knotweed complex (*Faloppia* sp.) are able to limit the growth of other plants by reducing nutrient quality in the soil (Stoll et al., 2012; Aguilera et al., 2010). Conversely, some species tightly depend on mutualism, like figs of the genus *Ficus* and wasps of the *Agaonidae* family: figs depend on wasps for pollination, and wasps depend on figs to lay their eggs (Janzen, 1979; Weiblen, 2002).

These different processes act at various spatial and temporal scales: species evolutionary history is generally relevant to study broad spatial distribution pattern, while the effect of the time of day is more often studied at fine spatial scale. While large scales patterns are the focus of biogeography (Weiher et al., 2011), community ecology focuses on smaller scales (Hubbell, 1997). The ecological filters framework (Weiher et al., 2011) unites different processes at play at the community scale: a regional species pool, determined by species evolutionary history, is distributed into communities via environmental filtering and biotic interactions that determine the finer-scale patterns. This framework relies on the niche concept, which posits that species have different ecological preferences determining their range. But community assembly is affected by stochastic processes as well (Weiher et al., 2011; Gravel et al., 2006), a principle

developed by Hubbell's neutral theory Hubbell (2001).

We can also study species distribution in communities: indeed, several studies suggest that non-random fine-scale species distribution patterns occur inside communities. For instance, some species prefer specific habitats inside the community. In addition to the klipspringer example given above (Estes, 2012), we can also cite the example of water-dependent ungulates in arid savannas, like impala or giraffe, for which the habitat choice is strongly constrained by the distance to water (Valeix et al., 2009). Species distribution can also be influenced by fine scale climatic events, like precipitations or temperature: for example, to protect themselves from extreme temperatures, some species use unoccupied aardvarks burrows in South Africa (Whittington-Jones et al., 2011, document the use of these burrows by 27 other species). Other species inside the community can also affect species distribution. For instance, prey tend to adapt their distribution in response to the "landscape of fear" caused by predator presence (Say-Sallaz et al., 2019) at different spatio-temporal scales (Palmer et al., 2022).

In this part, I focus on the influence of other species from the community on species' spatio-temporal distribution at very short spatio-temporal scales. I focus on animal species to illustrate the method and questions, since animals (by opposition to plants) are able to change their distribution with short-term behavioral responses. I explore the extent to which camera trap data can be used to answer this question. Therefore, I use multivariate Hawkes processes that take advantage of the continuous-time detection process of camera traps. Hawkes processes are a family of self-exciting point processes, i.e. point processes modeling punctual occurrences in continuous time, where the probability of occurrences is affected by past occurrences. This framework can be applied to camera trap data, using species occurrences captured by in continuous time by camera traps. In chapter 3.1, we analyze camera trap data using a linear multivariate Hawkes process. We describe the model, evaluate its performance with a simulation approach and compare it to another method. We also analyze a real camera trap dataset and highlight attractions and repulsions between five mammal species from the South African savanna. This chapter has been published as a research article (Nicvert et al., 2024). In chapter 3.2, I enrich our analysis with a more exploratory chapter investigating the use of a non-linear multivariate Hawkes process with a temporal covariate. I perform a small simulation study to evaluate the model performance,

and I reanalyze the data from chapter 3.1 with a temporal covariate to take into account species circadian rhythms. Finally, I discuss the outcomes of these models in chapter 4.3.

2.1 Dataset

In this part, I illustrate the analyses on a camera trap dataset collected in semi-arid African savannas. I describe data in greater detail in chapter 3.1, but I give a brief overview here. This dataset was collected as part of the Snapshot Safari project (Pardo et al., 2021): here, I use a subset of data consisting in 5 species collected in 6 protected areas.

This dataset was collected over two years and gathers a large number of capture events (more than 70 000 overall for the five focal species), which is a lot compared to typical camera trap datasets (Burton et al., 2015). Figure 3.I.1 summarizes the capture counts for each species: impala is the most abundant and lion the rarest.

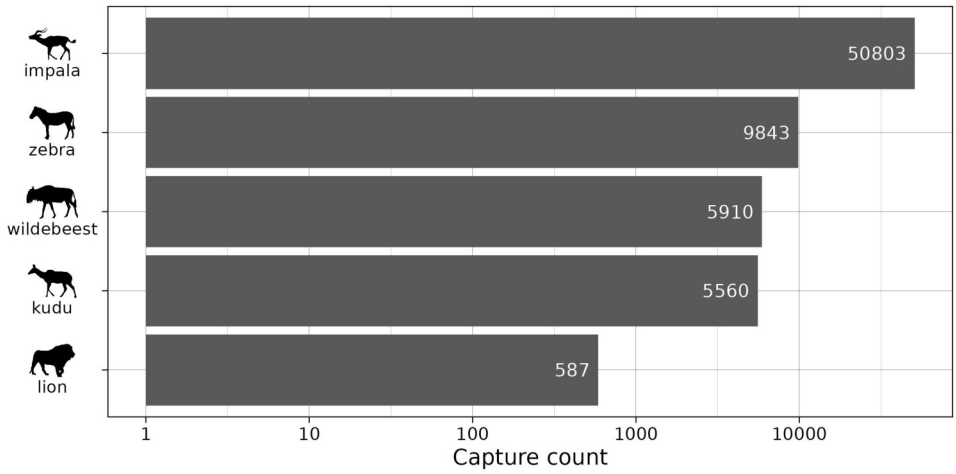


Figure 3.I.1: Capture count for each species. These counts take into account the number of times capture events without accounting for the number of individuals on the picture. The x-axis is in log-scale. Silhouette images from PhyloPic by Lukasiniho (wildebeest), Margot Michaud (lion), Robert Hering (kudu), Zimices (zebra) and an unknown author (impala).

I chose impala, greater kudu, lion, blue wildebeest and Burchell's zebra as focal species because they are thought to interact in diverse ways: lion predates on all species, and herbivores form mixed-species groups or follow a grazing succession. The ungulate species are also among the more common in savannas, which allowed to keep several sites in which they were all present and to have enough data to perform

a reliable inference. Figure 3.I.2 shows camera trap pictures used in the analysis for each species.



(a) Impala



(b) Greater kudu



(c) Lion



(d) Blue wildebeest



(e) Burchell's zebra

Figure 3.I.2: Camera trap pictures of focal species. Pictures from the Snapshot Safari project.

Chapter 3.1

Using the multivariate Hawkes process to study interactions between multiple species from camera trap data

The published version of the following article is available at https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237.

Check for updates

Received: 16 March 2023 | Revised: 13 October 2023 | Accepted: 10 November 2023

DOI: 10.1002/ecy.4237

ARTICLE

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

Using the multivariate Hawkes process to study interactions between multiple species from camera trap data

Lisa Nicvert[1] | **Sophie Donnet**[2] | **Mark Keith**[3] | **Mike Peel**[4,5,6] | **Michael J. Somers**[3] | **Lourens H. Swanepoel**[7] | **Jan Venter**[8,9] | **Hervé Fritz**[9,10] | **Stéphane Dray**[1]

[1]Universite Claude Bernard Lyon 1, LBBE, UMR 5558, CNRS, VAS, Villeurbanne, France

[2]Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay, Palaiseau, France

[3]Eugène Marais Chair of Wildlife Management, Mammal Research Institute, Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa

[4]Agricultural Research Council, Animal Production Institute, Rangeland Ecology, Pretoria, South Africa

[5]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa

[6]Applied Behavioural Ecology and Ecosystems Research Unit, University of South Africa, Florida, South Africa

[7]Department of Biological Sciences, Faculty of Science, Engineering and Agriculture, University of Venda, Thohoyandou, South Africa

[8]Department of Conservation Management, Faculty of Science, George Campus, Nelson Mandela University, George, South Africa

[9]REHABS, International Research Laboratory, CNRS-NMU-UCBL, Nelson Mandela University, George, South Africa

[10]Sustainability Research Unit, Nelson Mandela University, George, South Africa

Correspondence
Lisa Nicvert
Email: lisa.nicvert@univ-lyon1.fr

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR-18-CE02-0010

Handling Editor: Elise F. Zipkin

Abstract

Interspecific interactions can influence species' activity and movement patterns. In particular, species may avoid or attract each other through reactive responses in space and/or time. However, data and methods to study such reactive interactions have remained scarce and were generally limited to two interacting species. At this time, the deployment of camera traps opens new opportunities but adapted statistical techniques are still required to analyze interaction patterns with such data. We present the multivariate Hawkes process (MHP) and show how it can be used to analyze interactions between several species using camera trap data. Hawkes processes use flexible pairwise interaction functions, allowing us to consider asymmetries and variations over time when depicting reactive temporal interactions. After describing the theoretical foundations of the MHP, we outline how its framework can be used to study interspecific interactions with camera trap data. We design a simulation study to evaluate the performance of the MHP and of another existing method to infer interactions from camera trap-like data. We also use the MHP to infer reactive interactions from real camera trap data for five species from

Hervé Fritz and Stéphane Dray are co-last authors.

 © 2024 The Ecological Society of America.

South African savannas (impala *Aepyceros melampus*, greater kudu *Tragelaphus strepsiceros*, lion *Panthera leo*, blue wildebeest *Connochaetes taurinus* and Burchell's zebra *Equus quagga burchelli*). The simulation study shows that the MHP can be used as a tool to benchmark other methods of interspecific interaction inference and that this model can reliably infer interactions when enough data are considered. The analysis of real data highlights evidence of predator avoidance by prey and herbivore–herbivore attraction. Lastly, we present the advantages and limits of the MHP and discuss how it can be improved to infer attraction/avoidance patterns more reliably. As camera traps are increasingly used, the multivariate Hawkes process provides a promising framework to decipher the complexity of interactions structuring ecological communities.

KEYWORDS
African savanna, camera trap, interaction network, interspecific interactions, multivariate Hawkes process, reactive response, Snapshot Safari, spatio-temporal interactions

INTRODUCTION

Interspecific interactions affect many aspects of ecological communities. For instance, they influence ecosystem services (Valiente-Banuet et al., 2015), species assembly via biotic filtering (Ovaskainen et al., 2017) and the behavior of interacting species. In particular, interactions are one of the factors that structure the way in which animal species move in the landscape and adjust their habitat choices or activity times (Palmer et al., 2022). Mobile animals can respond to interactions by avoiding or seeking proximity with individuals of other species, depending on the positive or negative outcome of the interactions. For instance, prey can avoid their predators (Say-Sallaz et al., 2019), competing species can avoid each other (Cornhill et al., 2022; Searle et al., 2021), or herbivores can forage together to reduce predation risk or increase access to preferred foraging resources (Beaudrot et al., 2020). In this paper, we will use the term "interaction" to refer to the attraction or avoidance of a species by another one, even though "interaction" also refers to the underlying process of the attraction/avoidance pattern.

These interactions (as defined above) can occur in space and/or time, at different scales. Species can adjust their space use in response to the expected distribution of other species (proactive spatial interaction; Palmer et al., 2022). Species can also alter their daily activity patterns (e.g., Karanth et al., 2017) in response to other species (proactive temporal interaction). However, some species could also exhibit a reactive response to the presence of other species, that is, change their behavior in response to the actual presence of a species sometime before at a given location (e.g., Karanth et al., 2017; Parsons et al., 2016). This type of response could be mediated, for instance, by olfactory (Cornhill & Kerley, 2020; Kuijper et al., 2014) or auditory cues (Hettena et al., 2014). Investigating these reactive interactions is particularly promising as it allows us to identify fine-grained patterns that could be missed by approaches aggregating data in space or in time (Cusack et al., 2017; Frey et al., 2017; Parsons et al., 2022).

Investigating such fine-scale responses is very challenging, as it requires an intensive sampling effort to monitor multiple species in space and time. In this context, camera traps open new opportunities to study the spatial and temporal activities of multiple species (Caravaggi et al., 2017). Camera trap arrays allow the collection of multiple species occurrences and, therefore, the continuous monitoring of entire communities for large areas in time (Pardo et al., 2021). Hence, camera traps can produce massive amounts of data and offer new possibilities to study interactions between several species at multiple scales. Moreover, they are relatively cheap and easier to set up than classical fieldwork survey techniques (e.g., transects), especially for rare or elusive species or in remote areas. As camera traps become more affordable and automated species identification methods from pictures are being developed with deep learning, camera trap data (and other passive sensor data) will probably become more abundant in the future (Caravaggi et al., 2017).

With camera trap data, interspecific interactions are mostly studied at a broad spatial or temporal scale. To do this, data are often aggregated so that either the spatial or the temporal aspect is completely ignored. There are two main approaches for this purpose: comparing species' daily activities patterns (Ridout & Linkie, 2009) or spatial

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Université Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Universite Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

occupancy patterns (e.g., with the multispecies occupancy model of Rota et al., 2016). Such methods provide a measure of the proactive attraction or avoidance strategy, with species adapting their space or time use in anticipation of other species' presence or absence (Palmer et al., 2022). However, other approaches have combined spatial and temporal aspects to infer reactive attraction/avoidance strategies (frequently called spatio-temporal interactions in the literature; Karanth et al., 2017; Murphy et al., 2021; Niedballa et al., 2019; Prat-Guitart et al., 2020). Most methods using camera trap data quantify only the temporal aspect of reactive interactions; therefore, we will call the inferred patterns *reactive temporal interactions*. Most are based on the computation of time intervals between the detections of two species at a given place (e.g., Harmsen et al., 2009); here, we call this family of methods *interevent times methods*. The distribution of time intervals can then be contrasted according to the order of appearance of species (Parsons et al., 2016; Prat-Guitart et al., 2020) or summarized by a statistic that is compared with values obtained under a null model (usually data permutation; Cusack et al., 2017; Galindo-Aguilar et al., 2022; Karanth et al., 2017; Murphy et al., 2021). For a comparison of different approaches to infer reactive temporal avoidance with time interval measures, see Niedballa et al. (2019). Other more recent approaches use point processes, which allow us to integrate temporal dependence in a model-based framework (Kellner et al., 2022; Schliep et al., 2018).

Although all methods described above are useful to study reactive interactions, they usually focus on pairs of species and can therefore be unsuitable for studying complex interaction networks. For instance, these methods can identify spurious interactions between two species if other species are involved in the interaction network but not considered in the analysis. Moreover, they summarize the effect of a species on another one by a single value (e.g., the median of the time interval; Karanth et al., 2017), thus ignoring the multiscale and possibly time-dependent changes in the attraction/avoidance patterns (but see Cusack et al., 2017).

In this paper, we propose the multivariate Hawkes process (MHP) (Hawkes, 1971; Lambert et al., 2018) as a modeling framework to infer reactive interactions between multiple species from passive sensors such as camera traps. Hawkes processes belong to the family of point processes that allow the analysis of species capture events in continuous time, thus avoiding any data aggregation procedure. In Hawkes processes, species' interactions are modeled as pairwise interaction functions that depend on the time elapsed between species detections. The MHP used in this article is generative and offers the possibility to simulate occurrence data, given parameters specification. It also comes with an inference procedure that allows the adjustment of the pairwise interaction functions from observed data. It deals properly with indirect effects caused by species interaction chains, thus minimizing the risk of inferring spurious interactions. We believe that this model is a useful conceptual framework that is well suited for assessing reactive temporal interactions between species. We first present the Hawkes process and how it can be used to analyze camera trap data. Then, we describe the MHP used in this article, which was developed by Lambert et al. (2018) and implemented in the R package *UnitEvents* (Albert et al., 2021). We then show how this model can be used to simulate data to evaluate the performance of statistical methods or to infer interactions from camera trap data. We also apply the MHP on real camera trap data from the Snapshot Safari monitoring program (Pardo et al., 2021) to infer reactive temporal interactions between five mammal species. Finally, we discuss the usefulness of the MHP and the perspectives on how to develop this model further.

MATERIALS AND METHODS

All analyses were performed using R statistical software (v4.3.0; R Core Team, 2023) and the code and data (Nicvert et al., 2023) are available at https://doi.org/10.6084/m9.figshare.24552157.v5.

Model: The multivariate Hawkes process (MHP)

Hawkes processes are a family of point processes used to describe dependencies between punctual events. These processes belong to the class of self-exciting point processes for which the probability of occurrence at time t depends on the occurrences of the previous events. The first Hawkes process was introduced in 1971 by Alan G. Hawkes (Hawkes, 1971). Originally applied to model aftershocks following earthquakes (e.g., Ogata, 1988), Hawkes processes have been applied in various fields (Reinhart, 2018) for instance to model crime recurrence in cities (Mohler et al., 2018), the evolution of prices on the stock market (Hawkes, 2018) or the transmission of action potentials in a network of neurons (Reynaud-Bouret et al., 2013). The theoretical properties of Hawkes processes have also been thoroughly studied, and numerous extensions have been proposed.

Throughout this article, we define an occurrence as the detection of an individual at a camera at a given time,

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Universite Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and we do not take imperfect detection into account. To describe the model, we consider data on the occurrences of S species collected on C cameras. In our framework, the data collected on C cameras are seen as C independent realizations of the MHP. Let T_m^{li} denote the m-th instant of punctual occurrence for species i at camera l. Let N_i^l be the total number of occurrences for species i at camera l. We model the occurrence times $\left(T_m^{li}\right)_{m=1\ldots N_i^l, i=1\ldots S, l=1\ldots C}$ as C realizations of an MHP.

To model punctual occurrences, point processes use a latent intensity function, which is a measure of the rate at which events occur in time. When modeling species occurrences from camera trap data, the intensity for a given species represents the rate at which this species occurs at a camera. For species i, the intensity $\lambda_i^l(t)$ at camera l is formally defined as (Daley & Vere-Jones, 2003):

$$\lambda_i^l(t) = \lim_{\delta \to 0} \frac{P\left\{n^{\mathrm{li}}_{]t,t+\delta]} > 0\right\}}{\delta}, \tag{1}$$

where $n^{\mathrm{li}}_{]t,t+\delta]}$ is the number of points occurring between times t and $t+\delta$ for species i at camera l and δ is an infinitesimally small amount of time. Informally, the intensity of a point process multiplied by a small amount of time can be viewed as the probability that there will be at least one point occurring around time t.

In this work, we used the R package *UnitEvents* (Albert et al., 2021) available from https://sourcesup.renater.fr/frs/?group_id=3267, to simulate and infer MHPs. *UnitEvents* is only available on Linux and Mac OS. However, in the code and data repository for the article (Nicvert et al., 2023), we provide a Dockerfile allowing us to run the analyses from any operating system (including Windows).

UnitEvents implements the MHP described in Lambert et al. (2018). In this framework, the intensity of species i seen on camera l for a Hawkes process with S interacting species is written as:

$$\lambda_i^l(t) = \left(\nu_i + \sum_{j=1}^{S} \sum_{m|T_m^{lj}<t} f_{j\to i}\left(t - T_m^{lj}\right)\right)_+, \tag{2}$$

where $\lambda_i^l(t)$ represents the intensity for species i (as defined above) at camera l. ν_i is a positive parameter, the background rate: it represents the basal intensity of species i (in time^{-1}, e.g., day^{-1}) unrelated to previous occurrences. For instance, ν_i would be low for a rare species and higher for a common species.

$f_{j\to i}$ is the interaction function that represents the influence of an occurrence of species j on species i as a function of time delay: positive values of $f_{j\to i}$ represent an attraction of species i by species j, negative values represent a repulsion and null values independence. In the case $j=i$, the function $f_{i\to i}$ represents the interaction between individuals of the same species i. In that case, we will call $f_{i\to i}$ the auto-interaction function: it could reflect for instance the fact that some species are solitary or gregarious. $f_{j\to i}$ are defined as piecewise constant functions with K time bins of equal length δ:

$$f_{j\to i} = \sum_{k=1}^{K} \alpha_{j\to i}^k \mathbf{1}_{](k-1)\delta, k\delta]}, \tag{3}$$

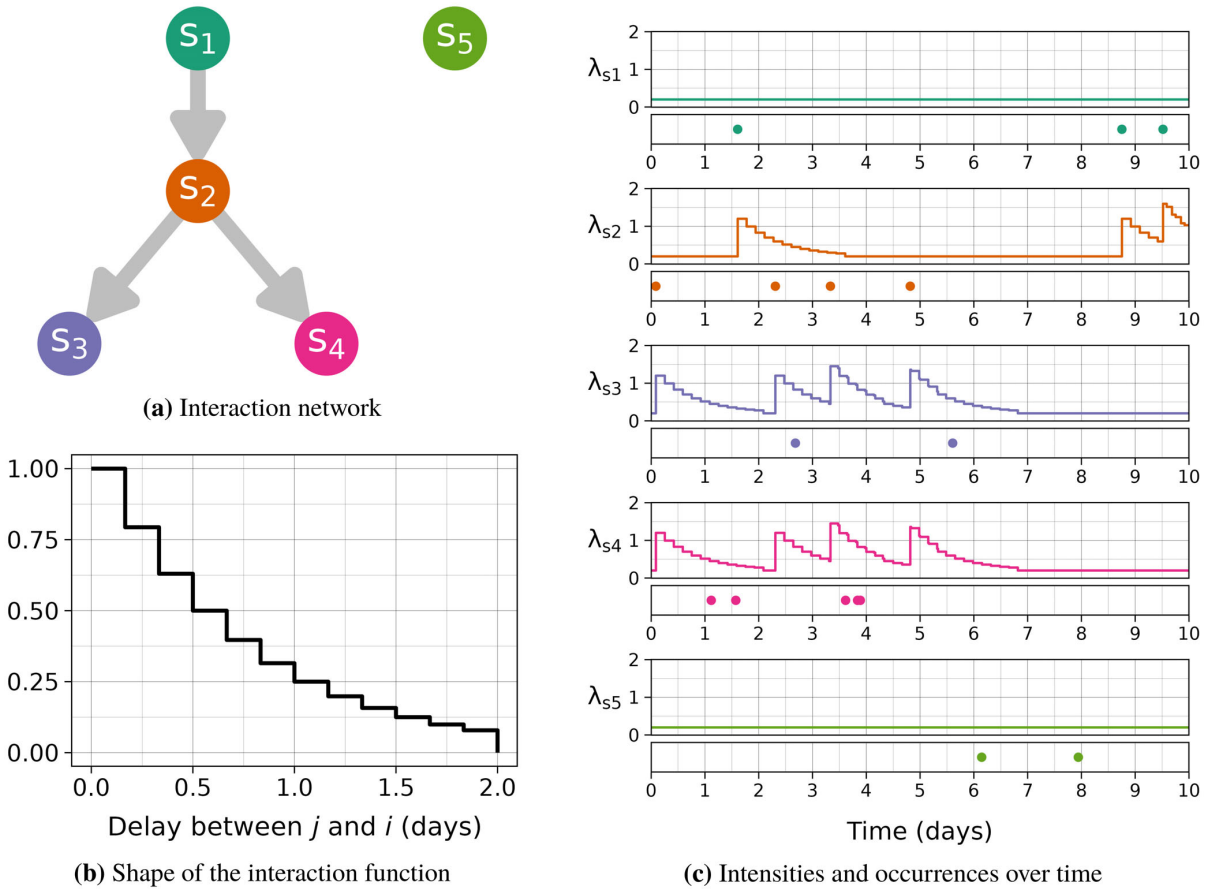
where $\mathbf{1}_{](k-1)\delta,k\delta]}$ denotes the indicator function between delays $(k-1)\delta$ and $k\delta$. The K coefficients $\alpha_{j\to i}^k$ represent the average number of occurrences of species i gained (if positive) or suppressed (if negative) by an occurrence of species j in the k-th interval after this occurrence of species j.

In this framework, $f_{j\to i}$ can take negative values, thus allowing modeling the repulsive effect of species j on species i. As the intensity λ_i^l must be positive by definition, Equation (2) includes a positive part $(\cdot)_+$. However, for mathematical reasons, in the following developments we will assume that the negative values of $f_{j\to i}$ are never too strong so that the intensity never becomes negative, and the positive part is not needed. To enforce this assumption, the repulsion terms can only be as strong as the other terms making up the total intensity.

Figure 1 illustrates a realization of an MHP with five species (measured at a single camera) simulated with *UnitEvents*. In this example, some species attract each other (see the interaction network in Figure 1a) with the same decreasing discrete exponential interaction function with $K=12$ time bins of width $\delta=4$ h (Figure 1b). The background rate is the same for all species and is fixed at 0.2 occurrences day^{-1}. The right panel in Figure 1c shows the simulated species occurrences and associated intensities over time. When nothing happens, the intensity is fixed at the background rate. When an attracting species occurs, the intensity of the attracted species peaks, making an occurrence more likely. For instance, each occurrence of species s_1 gives rise to a peak in the intensity of s_2. Moreover, when several attracting events occur, the interaction functions add up, which makes the occurrence of the target species even more likely.

Model inference

The inference procedure implemented in the *UnitEvents* package is a fast and scalable LASSO-penalized (least absolute shrinkage and selection operator) least-squares criterion. It allows the estimation of a single MHP from C



(a) Interaction network

(b) Shape of the interaction function

(c) Intensities and occurrences over time

FIGURE 1 Example of a realization of a multivariate Hawkes process. (a) Shows the interaction network between five species (each arrow represents a non-null interaction function). In this example, all autointeraction functions $f_{i\to i}$ are null. (b) Shows the shape of the interaction functions ($K = 12$ time bins of width $\delta = 4$ h) corresponding to arrows in the interaction network (a). (c) Shows a realization of the Hawkes process with the interaction network and the interaction functions shown in (a) and (b). For this simulation, the background rate was set to 0.2 occurrences day^{-1} for all species. For each species i, the above panel shows the intensity λ_i and the bottom panel shows the species occurrences. Each time an attracting species occurs, the intensity for the attracted species peaks and then decreases as dictated by the interaction function shape.

realizations (in our setting, this corresponds to C cameras).

Let $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_S)$ denote the parameters of interest for each species $i = 1, \ldots, S$. Each $\boldsymbol{\beta}_i$ is a vector of size $1 + SK$ containing the background rate of species i (ν_i) and the parameters of the interaction functions targeted to this species i for the S species and the K bins: $\boldsymbol{\beta}_i = \left(\nu_i, \left(\alpha^k_{j\to i}\right)_{j=1\ldots S, k=1\ldots K}\right)$. Each $\boldsymbol{\beta}_i$ is estimated as:

$$\widehat{\boldsymbol{\beta}}_i = \arg\min_{\boldsymbol{\beta}_i} \text{LASSO}(\boldsymbol{\beta}_i) \quad \text{where}$$

$$\text{LASSO}(\boldsymbol{\beta}_i) = \underbrace{-2\sum_{l=1}^{C} \boldsymbol{b}_i^{l\,T}\boldsymbol{\beta}_i + \boldsymbol{\beta}_i^T \sum_{l=1}^{C} \boldsymbol{G}^l \boldsymbol{\beta}_i}_{\text{least-squares}} + \underbrace{2\boldsymbol{d}_i^T |\boldsymbol{\beta}_i|}_{\text{penalization}}, \quad (4)$$

where ^{T} denotes transposition and $|\boldsymbol{\beta}_i|$ is the vector containing the absolute values of the coordinates of $\boldsymbol{\beta}_i$. $\boldsymbol{b}_i^l$ is an observable vector of size $1 + SK$. If camera l is active between times α_l and η_l, then:

$$\boldsymbol{b}_i^l = \left(N_i^l, \left(\int_{\alpha_l}^{\eta_l} n^{lj}_{[t-k\delta, t-(k-1)\delta[}\, \mathrm{d}n_t^{li}\right)_{j=1\ldots S, k=1\ldots K}\right). \quad (5)$$

Its first value is the total count of species i observed on camera l. The other values represent the total occurrence counts of the species j observed in the k-th bin before the occurrences of species i at camera l. $\boldsymbol{G}^l$ is also an observable matrix defined as:

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Universite Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\boldsymbol{G}^l = \int_{\alpha_l}^{\eta_l} \boldsymbol{c}_t^l \boldsymbol{c}_t^{l\,T} \mathrm{d}t, \tag{6}$$

where $\boldsymbol{c}_t^l$ is a vector of size $1+SK$ defined as $\boldsymbol{c}_t^l = (1, (n^{ij}_{\lfloor t-k\delta, t-(k-1)\delta \rfloor})_{j=1\ldots S, k=1\ldots K})$. Its first value is 1 and other values represent the occurrence counts of species j occurring on camera l in the k-th bin before time t.

The term $2\boldsymbol{d}_i^T |\boldsymbol{\beta}_i|$ of Equation (4) corresponds to the LASSO penalization: it can make some parameter values shrink to zero and thus avoid overparameterization. The strength of this LASSO penalization is controlled by the weights vectors $\boldsymbol{d}_i$, which are computed from the data and tuned by a unique user-chosen parameter γ (equation derived from Lambert et al. (2018) adapted from Hansen et al. (2015)):

$$\boldsymbol{d}_i = \sqrt{2\gamma \log\left(S+S^2K\right) \sum_{l=1}^{C} \int_{\alpha_l}^{\eta_l} \boldsymbol{c}_t^{l\,2} \mathrm{d}n_t^{li}} + \frac{\gamma \log\left(S+S^2K\right)}{3} \max_{l=1\ldots C} \left(\sup_{t \in [\alpha_l, \eta_l]} |\boldsymbol{c}_t^l| \right). \tag{7}$$

The choice of a suitable value for γ is crucial for model selection, because γ ensures that only relevant nonzero parameters are kept in the model. However, choosing a good value for γ is difficult: it has been evaluated by simulations in Lambert et al. (2018) and Hansen et al. (2015), and we proceeded similarly in this article.

In the current implementation of *UnitEvents*, three flavors of the LASSO penalization are available. We chose the "Bernstein Vanishing LASSO" (BVL), where the penalization in Equation (4) is first applied to discard weak interaction parameters. Then, the estimates of the remaining non-null parameters are obtained by minimizing the least-squares criterion. Lastly, an additional step is introduced to remove parameters smaller than a data-computed threshold (see Lambert et al., 2018, for details and justification).

In the implementation of *UnitEvents*, the bins width δ and the number of bins K for the interaction functions are fixed by the user, who also needs to choose a value of γ *a priori*. The other parameters (interaction functions coefficients $\alpha^k_{j\to i}$ and background rates ν_i) are fitted as described before.

Simulation study

We generated camera trap-like data under the MHP and used these simulated data to (1) evaluate the performance of a method and (2) tune the penalization parameter for inference on real data.

Simulation parameters

For these two objectives, we conducted two sets of simulations in the same conditions. We considered an interaction network with five species $s_{i=1\ldots5}$ where s_1 attracts s_2 and s_2 attracts s_3 and s_4 (network from Figure 1a). This network represents a difficult case as an inference method should detect direct interactions, but not spurious indirect interactions (e.g., $s_1 \to s_3$) and identify that species s_5 is not interacting with others. In this simulation, we define the true interactions by decreasing exponential functions to 2 days:

$$f(t) = \begin{cases} \alpha \exp\left(-\dfrac{\ln(2)}{0.5} t\right) & \text{if } t < 2 \\ 0 & \text{if } t \geq 2 \end{cases}, \tag{8}$$

where α is the interaction strength. The half-life of this function is the denominator of the decrease rate, so that this function will reach half of its initial value at $t = 0.5$ day. The interaction strength α for the true model varied from 0.01 to 1 day^{-1}. Here, the interaction strength represents the maximum intensity of the pairwise interaction function for $t = 0$. An analogous interaction function is shown in Figure 1b with $\alpha = 1$ and with discrete bins. The background rate was fixed at 0.1 day^{-1} for all species.

The simulated trapping length varied from 20 to 500 trapping days for each camera over 25 cameras (making up to 12,500 trapping days in total). For each condition, 30 different data sets were generated to evaluate the variability of the inference.

We evaluated the performance of the inference by computing the true positive and true negative rates. The true positive rate is the proportion of inferred nonzero interactions over the count of true nonzero interactions. The true negative rate is the proportion of inferred null interactions over the count of true null interactions.

Evaluating a method to infer reactive temporal interactions

We illustrated how synthetic data generated with the MHP can be used to evaluate the performance of a method to infer interspecific interactions, considering the interevent times method of Murphy et al. (2021). We applied the method described by Murphy et al. (2021) on simulated data (simulation settings are described in *Simulation parameters*). This method consists of computing the median time between directed pairwise species occurrences (excluding pairs from the same species) for observed and randomly permuted data (999 permutations). The permutation procedure involved randomly

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Universite Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

changing the cameras' labels of species occurrences (for details see Murphy et al., 2021). Finally, the statistical significance of interactions was estimated by comparing the median time for observed and permuted data. We used a significance threshold of 5% with a Holm correction for multiple testing.

Choice of the penalization parameter

To choose the best penalization parameter γ in the context of interaction inference, we used a simulation approach (simulation settings are described in *Simulation parameters*). We inferred MHPs with different values of γ (between 0.3 and 1) from the simulated datasets. For the inference parameters, we chose $K = 12$ bins of width $\delta = 4$ h (2 days in total, corresponding to the length of the simulated interaction functions). Then, we defined any inferred interaction function as null if all bins were zero over the function's support, and nonnull if at least one bin was not null.

Application: Analysis of interactions between five species in the African savanna

We used the MHP to infer interaction functions between five species of the southern African savanna: impala *Aepyceros melampus*, greater kudu *Tragelaphus strepsiceros*, lion *Panthera leo*, blue wildebeest *Connochaetes taurinus*, and Burchell's zebra *Equus quagga burchelli*.

Data collection

Camera trap data were collected as part of the long-term Snapshot Safari monitoring program (Pardo et al., 2021). Snapshot Safari is a network of camera trap grids set up in more than 30 locations in southern Africa. The camera trap design consists of grids of 5 km^2 in each location, in which cameras were fixed at ~50 cm high. Cameras were automatically triggered by motion or heat using passive infrared sensors. Each camera was programmed to take a series of three images within 1–5 s of each other by day, and only one image by night to minimize disturbance occasioned by white flash. For this analysis, we focused on six camera trap grids in the savanna biome in northern South Africa: the Associated Private Nature Reserves (around Kruger National Park), Kruger National Park, Madikwe Game Reserve, Pilanesberg National Park, Somkhanda Game Reserve, and Venetia Limpopo Nature Reserve (see Figure 2).

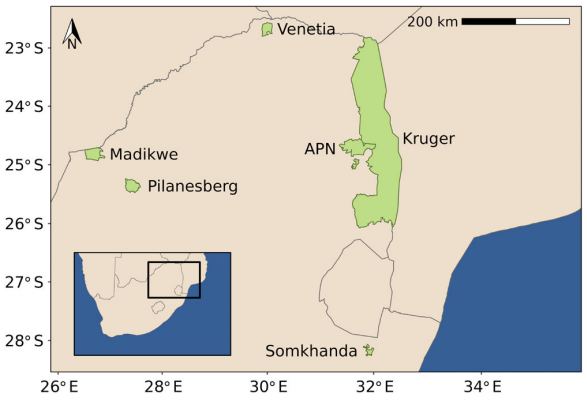


FIGURE 2 Study sites. Six protected areas were surveyed with camera traps for this study: the Associated Private Nature Reserves (APN), Kruger National Park, Madikwe Game Reserve, Pilanesberg National Park, Somkhanda Game Reserve, and Venetia Limpopo Nature Reserve. Data by OpenStreetMap contributors under ODbL license (https://www.openstreetmap.org/copyright).

Data preprocessing

Pictures were classified by citizen science using the Zooniverse platform (www.zooniverse.org), where pictures were available online and annotated by more than 150,000 volunteers (see Pardo et al., 2021, for more details).

For this analysis, we filtered out cameras where capture events were too rare (less than two pictures in total or fewer than one picture every 30 days on average). We did not filter for independence between occurrences of the same species. However, because the Hawkes model does not allow two capture events to occur simultaneously, if two or more individuals of different species were seen on the same capture event, their occurrence time was randomly shifted from 1 min in advance to 1 min later. For multiple individuals of the same species seen simultaneously, the occurrences of the individuals were counted as a single event (i.e., an occurrence corresponds to an individual or a group of individuals of a given species).

After the filtering procedure, 72,703 occurrence events (corresponding to 70,409 unique pictures) were collected on 179 cameras in total. Cameras were active during 503 ± 224 (SD) days on average (minimum: 19 days, maximum: 851 days), amounting to 90,176 trapping days on all cameras. All pictures were taken between June 2017 and November 2019.

Parameters inference

We inferred the parameters of a MHP using interaction functions defined by $K = 6$ bins of $\delta = 6$ h (36 h in total). This parametrization should allow us to capture the

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Universität Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

dynamics of reactive temporal interactions with enough granularity while keeping a relatively low number of parameters to estimate to allow reliable inference. Using the results of the simulation study (see *Choice of the penalization parameter*), we decided to set the value of the penalization parameter γ to 0.5.

RESULTS

Simulation study

Evaluating a method to infer reactive temporal interactions

We used data simulated under the MHP to evaluate the method of Murphy et al. (2021). As expected, the ability to detect interactions (true positive rate) increases with the strength of the interactions (Figure 3). Provided the interaction strength α is big enough (at least 0.1 day^{-1}), the ability to detect interactions increases with the number of trapping days, which indicates that a significant sampling effort is required to infer interactions from camera trap data (at least 300 trapping days for 25 cameras when the interaction strength is above 0.2 day^{-1}). More surprisingly, when the interaction strength is high (at least 0.5 day^{-1}), the true negative rate decreases with increasing sampling effort. This indicates that the method wrongly detects interactions between noninteracting species. Additional investigations (Appendix S1: Section S1) show that these errors mainly concern the detection of spurious indirect interactions between species involved in interaction chains (e.g., $s_1 \to s_3$).

Choice of the penalization parameter

The simulation study to find suitable values for the penalization parameter γ led to the results shown in Figure 4. Unsurprisingly, the ability to detect true interactions (true positive rate) increases with the number of trapping days and the strength of interactions. When the penalization is too low ($\gamma = 0.3$; top row), the model tends to identify interactions between noninteracting species (reducing the true negative rate) but this problem vanishes when the sampling effort increases. Conversely, a high penalization ($\gamma = 1$; bottom row), moderately improves the true negative rate, but more importantly dramatically hampers the ability to detect non-null interactions for small interaction strengths. A value of $\gamma = 0.5$ seems to be a good compromise allowing the efficient detection of true interactions when their strength is not too small (at least 0.1 day^{-1}) but avoiding the identification of false interactions. It gives good results especially when the sampling lasts more than 400 trapping days per camera. Hence, we decided to use a penalization parameter of $\gamma = 0.5$ to infer the parameters of an MHP from real data (see *Analysis of real data*). Lastly, supplementary analyses show that the spurious interactions are randomly distributed and not biased toward indirect interactions as with the interevent time method (Appendix S1: Section S1).

Analysis of real data

We fitted a MHP using the occurrence data of five species (impala, greater kudu, lion, blue wildebeest, and Burchell's

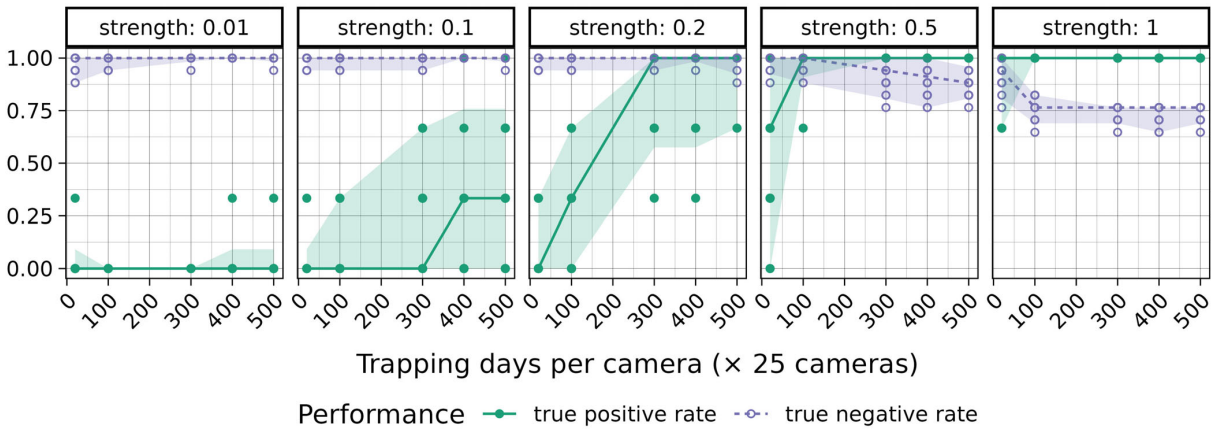


FIGURE 3 Evaluation of an interevent times method (Murphy et al., 2021). Panels represent different interaction strengths (maximum value of the interaction function). The *x*-axis represents the sampling length and the *y*-axis represents the performance: true positive rate (full dots, continuous line) or true negative rate (circles, dashed line). Points indicate values for the 30 repetitions, lines joins the medians, and the colored area represents the 2.5th and 97.5th percentiles.

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Université Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

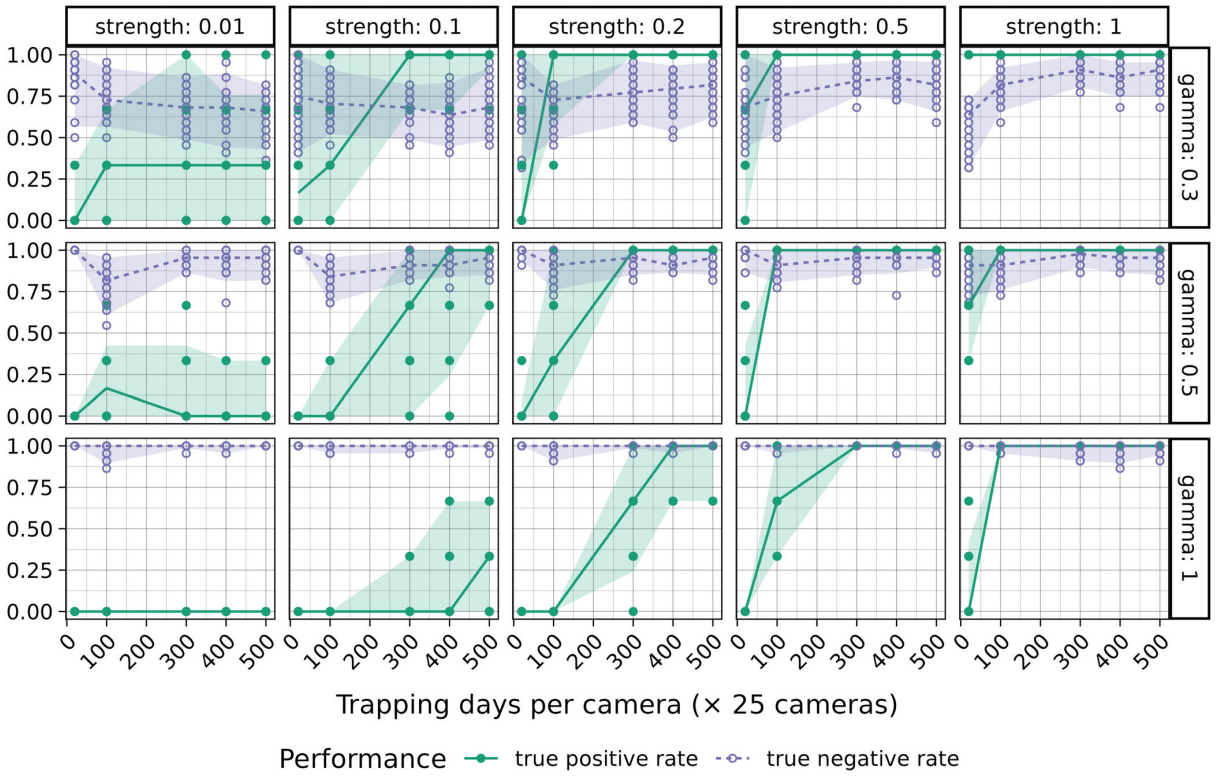


FIGURE 4 Performance of the inference with the multivariate Hawkes model. In columns, the interaction strength (maximum value of the interaction function). In rows, the different values of the penalization parameter γ. The *x*-axis represents the sampling length and the *y*-axis represents the performance (true positive rate or true negative rate). Lines, points and colors have the same meaning as in Figure 3.

zebra) collected with camera traps. Adjusting the model only took a few seconds on a personal computer. The resulting interaction functions are shown in Figure 5 and the inferred background rates are in Appendix S1: Section S2.

Background rates represent the basal intensity for each species, independently of the others. They vary greatly between species, with impala having a much higher background rate than other species (impala: 0.212 day^{-1}; zebra: 0.040 day^{-1}; kudu: 0.035 day^{-1}; wildebeest: 0.022 day^{-1}, and lion: 0.003 day^{-1}). As expected, they are strongly related to the total occurrence count of each species.

Regarding the interaction functions, the inferred parameters highlight a strong auto-attraction for the first bin (0–6 h), varying between 1.5 and 2.25 day^{-1} depending on the species. Regarding the cross-species interaction functions, many herbivores are attracted to each other. Impalas follow or avoid kudus (depending on the delay), wildebeests and zebras; zebras follow impalas, kudus, and wildebeests; wildebeests mainly follow zebras. Other interactions between herbivores are negligible. These herbivore–herbivore interactions are composed of a short-term attraction (during the first 6 h after an occurrence) and of a medium-term attraction (12 to 36 h after an occurrence) except impalas that are not attracted by zebras in the short term. Additionally, impalas seem to avoid kudus 6 to 12 and 30 to 36 h after an occurrence. We notice that these interactions are asymmetrical (impalas and zebras follow other species much more than they are followed). Regarding prey–predator interactions, lions do not follow or avoid any other species. Zebra and impala seem to avoid lions in the next 6 h following an occurrence of this predator. Finally, the inferred interactions are relatively robust to a change in bin width as we show in Appendix S1: Section S4, where we performed the inference on the same dataset with different bin widths (3 and 9 h).

DISCUSSION

It is now well established that identifying the signature of interspecific interactions from species occurrence data is generally difficult (Blanchet et al., 2020; Popovic et al., 2019). However, camera trap data provide additional

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Université Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

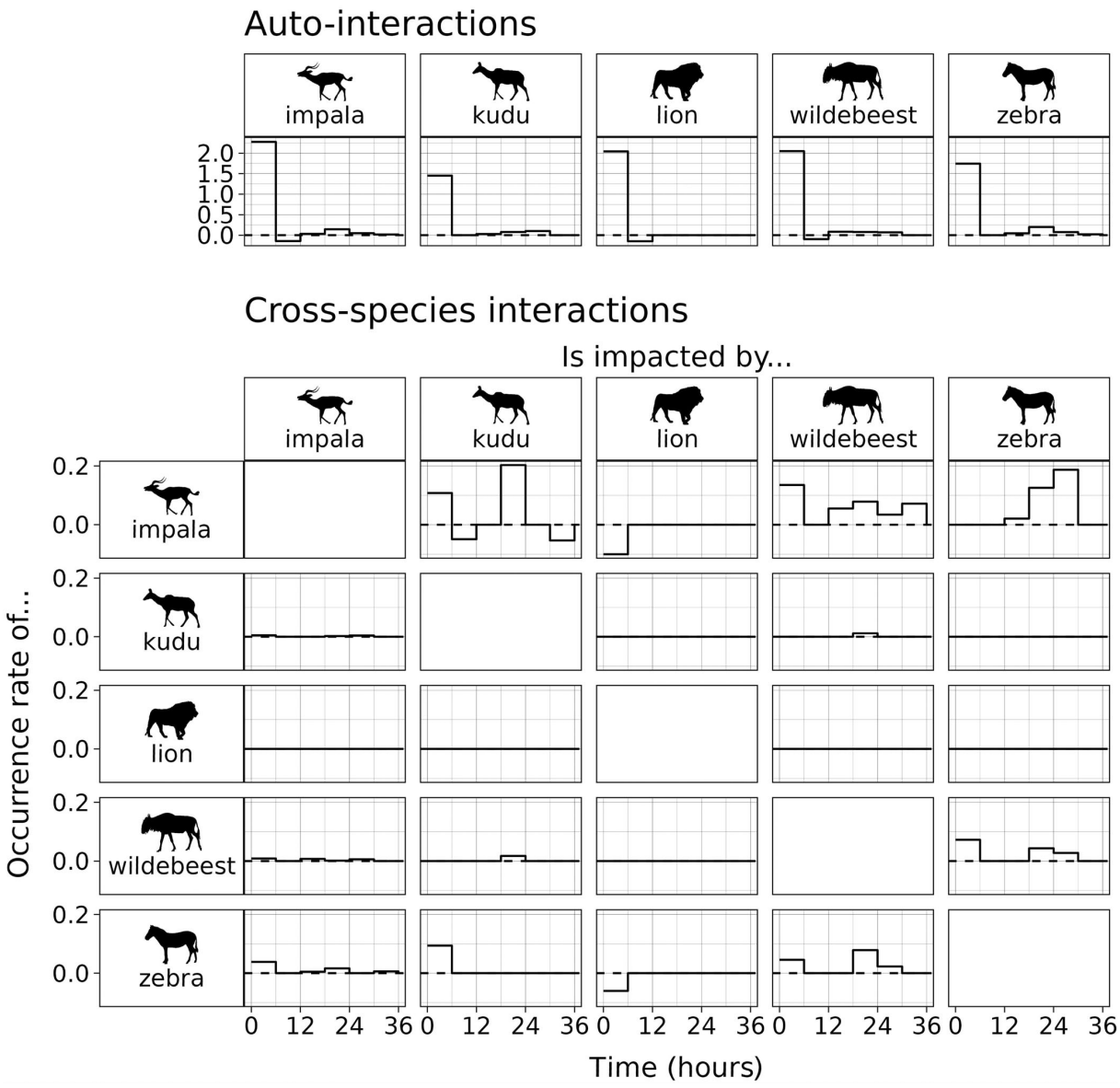


FIGURE 5 Inference of interactions from real data using the multivariate Hawkes model. The top plot shows the auto-interaction functions (between occurrences of the same species). The bottom plot shows cross-species interactions, where the intensity of species in rows is affected by species in columns. The horizontal dashed line represents zero. Note that the *y*-axis scale is different between autointeractions and cross-species interactions. Silhouette images from PhyloPic by Lukasiniho (wildebeest), Margot Michaud (lion), Robert Hering (kudu), Zimices (zebra), and an unknown author (impala).

information (time and order of occurrence of species) that can help to relate occurrence patterns to underlying interspecific interactions. In this context, the Hawkes model provides a new theoretical framework to analyze species occurrences sampled in continuous time using camera traps. This model aims to predict the probability of occurrence of a given species at a given time taking into account the previous occurrences for several species. By considering the exact time at which species occur, this model provides a detailed picture of species reactive temporal interactions under the form of interaction functions (here, the term "interaction" refers to the attraction/repulsion pattern). These functions allow a multiscale description of interactions as they characterize how the interaction strength varies with time, contrary to other methods that provide a single measure of attraction/

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Université Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Universite Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

avoidance. Moreover, these functions are directed: the inferred interactions can be asymmetrical, as expected for ecological interactions. The toolbox associated with this model offers the possibility to generate data and design simulation studies or to infer parameters from real data.

We used the MHP to generate camera trap-like datasets with different properties (sampling effort, strength of interactions) and showed how these simulated data can be used to evaluate the performance of a method or to tune inference parameters. Both simulation studies demonstrate that camera trap data can be used to detect reactive temporal interactions between species but this requires a substantial sampling effort, especially when the strengths of interactions are low. In our simulation setup (five species, background rates of 0.1 day^{-1} and only attractions), results suggest that at least 1 year of sampling with 25 cameras is required to obtain reliable inference, and this holds only if the interaction is strong enough (interactions of strength 0.01 day^{-1} are not reliably detected in our simulations). These requirements would probably be higher if more species were considered, especially for rare species (smaller background rate). Hence, we agree with Schliep et al. (2018) that more data are needed to estimate reliable reactive interaction patterns than to estimate species occupancy. In this context, the MHP provides a powerful simulation tool to design and assess the quality of sampling protocols in camera trap studies by adopting a virtual ecologist approach (Zurell et al., 2010). The simulation study also highlights the limits of methods focusing on pairs of species to analyze interactions between multiple species. By focusing only on two species at a time, these approaches are not able to disentangle direct interactions from indirect effects due to other species in interaction chains (Appendix S1: Section S1). Moreover, the correction for multiple testing we applied in our study was not sufficient to eliminate these spurious interactions, and we can assume that this issue is more important in the literature when no correction is considered. As a consequence, interevent times methods tend to overestimate the number of interactions, especially when their strength is high or the sampling effort increases. However, such spurious interactions were inferred only when we simulated quite strong interactions, and more investigations would be needed to estimate the range of interaction strengths we can expect in natural conditions. By contrast, the Hawkes process used here is multivariate by nature, so it works on all species simultaneously and thus allows the identification of interactions between two species conditionally to the other species, similar to graphical models in the context of co-occurrence analysis (Popovic et al., 2019). This modeling approach thus provides a better picture of the interaction network of the whole community.

The real dataset analysis shows how the MHP can be used to infer reactive interactions between five mammal species from the African savanna. In our example, because we defined an occurrence as the presence of an individual or a group of individuals, the values of the interaction functions represent the number of individuals or groups of individuals that are attracted/repulsed by other occurrences, and the typical group size to consider depends of the species. We identified strong auto-attractions for all species but also attractions between different herbivores and avoidance of lions by two herbivore species (impala and zebra). Whereas it could be tempting to interpret these results as behavioral responses of species to an underlying interaction (e.g., avoidance in response to predation), the Hawkes model only characterizes attraction/avoidance patterns and particular care should be taken when interpreting these results, especially because no covariates were included in this analysis. We discuss these different interpretations of the observed patterns in terms of ecological processes below and we make suggestions to improve the MHP to untangle the different hypotheses.

We identified auto-attractions for all species, indicating that the occurrence of a given species increases the probability of having another occurrence of the same species at the same place. This could be due to the same individual lingering in front of the camera, especially because no independence filter was applied (although cameras are configured to pause for 1 min between trigger events), or this could reflect sociality among individuals, as an individual or a group may attract other individuals for gregarious species. This could also stem from habitat selection processes, so that numerous subsequent occurrences could be observed at cameras located in species' preferred habitats. Lastly, circadian rhythms impose physiological constraints on the activity times of each species and thus could increase their occurrence rate at certain times of the day. When they are not taken into account, as is the case here, circadian rhythms could affect the interaction functions in the short term (0–6 h) and also induce a 24-h periodicity in the interaction functions. This issue is clearly illustrated using simulated data (see Appendix S1: Section S3) and could partly explain the short-term (0–6 h) auto-attraction, and probably most of the weak auto-attraction observed at ~24 h for impala, kudu, wildebeest and zebra (Figure 5).

Regarding the cross-species interactions, we observed attraction patterns between some herbivores, which could be explained by four mechanisms. First, temporal niche convergence could induce attraction between species when they are active at the same time of day and if they also share the same location. In our example, the four herbivore species are diurnal with crepuscular activity peaks. However, if the apparent attraction was due to shared

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Université Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

circadian rhythms, we would probably observe a symmetry of interaction functions (i.e., if $s_1 \to s_2$ is not null, $s_2 \to s_1$ is also not null) as the order of appearance of species at a camera during the activity time would be random. This is not always the case in the example depicted here, for instance between zebra and kudu. Second, species sharing the same kind of preferred environment might show apparent attraction. However, as for the temporal niche, this spatial niche should induce a symmetrical interaction pattern. Third, the apparent attraction between species could be due to the mixed-species grouping strategy, whereby some species forage together in mixed groups. Such groups are thought to mitigate predation risk and/or improve access to resources (Beaudrot et al., 2020). Moreover, when it comes to predation risk, in addition to the dilution effect (a simple number game), there is a possible benefit to being associated with more vulnerable species (Fitzgibbon, 1990). This implies a directionality in the choices of association, an asymmetry well captured by the MHP. In our analysis, some species are attracted by others in the first hours following an occurrence (impala follows kudu and wildebeest; zebra follows wildebeest, kudu and impala; and wildebeest follows zebra). Interestingly, these associations have been described in the literature (Meise et al., 2019; Pays et al., 2014; Schmitt et al., 2014). Finally, another mechanism that could explain interactions between herbivore species is grazing succession (Bell, 1971), which describes a strategy by which species sequentially use the same grazing area: less selective species come first (nonruminants and species with higher body mass), followed by more selective species (smaller ruminants). In our results, some herbivore species are attracted with a delay (impala following zebra, kudu, and wildebeest; wildebeest following zebra; zebra following wildebeest and impala). Impala following other (bigger) species and wildebeest following the nonruminant zebra are compatible with the grazing succession theory (Bell, 1971). However, the temporal scale of this potential grazing succession occurs at a temporal scale much shorter than the one classically described (McNaughton, 1976, 1985).

Regarding the apparent avoidance of lions by zebras and impalas, here again this could stem from temporal niche divergence (lion is a nocturnal species whereas impala and zebra are diurnal). This apparent repulsion could also reflect a strategy of impala and zebra to minimize predation risk by reactively avoiding lions, that is, responding to actual cues of lion presence (olfactory or auditory cues for instance) at a fine spatio-temporal scale, as documented for zebras (Courbin et al., 2016).

As discussed with the real dataset analysis, a major challenge remains linking attraction/repulsion patterns identified by the MHP to underlying ecological processes. To date, the implementation used in this paper cannot include covariates to model variations in species' background occurrence rates. This calls for two major improvements: first, we could include temporal covariates to account for the variation of species occurrence rate through the day according to their diel cycle. Second, we could include environmental covariates to account for species habitat preferences across the landscape. Works such as Fujita et al. (2018) for temporal covariates or Carstensen et al. (2010) for temporal and environmental covariates could be helpful in this perspective. Further developments include accounting for imperfect detection by camera traps, which is known to be an important issue (Burton et al., 2015). In this regard, Kellner et al. (2022) recently developed an occupancy model with a detection process occurring in continuous time with a Markov-modulated Poisson process, and a similar approach could be envisioned with the MHP.

Here, we inferred an MHP from camera trap data, but this modeling approach could be extended to other types of passive sensors collecting occurrence data in continuous time (e.g., microphones, hydrophones) that are increasingly used to monitor biodiversity. In particular, using a spatially explicit extension of the Hawkes process (first described by Ogata, 1998, in the context of earthquake occurrences) could be especially suited to include a spatial dependency between camera traps or to analyze GPS collar data and estimate interaction functions in time and space.

The Hawkes process could also be used for other applications than estimating interspecific interactions, for instance to study behavioral synchrony within a group (e.g., Pays et al., 2012) or to infer animal social networks from occurrence data (e.g., Jacoby et al., 2016).

Even if more developments are required to improve ecological inference, we contend that the MHP and other point-process methods offer an adapted theoretical framework for the analysis of time-continuous occurrence data while contributing to an explanation of interactions among herbivores and between herbivores and predators.

ACKNOWLEDGMENTS

We thank all the people involved in Snapshot Safari data collection and management, including the students and volunteer groups who contributed to maintaining the camera trap grids and the reserve owners and managers for opening their reserves. We also thank all Zooniverse volunteers who contributed to classifying pictures. We thank Gilles Scarella and Patricia Reynaud-Bouret, developers of the *UnitEvents* package, for their detailed

answers to our questions regarding the use of their package. We thank Mahendra Mariadassou, Vincent Miele, Sara Puijalon, and Stéphane Robin for their input on this work during discussions. We thank Stéphane Delmotte and Adil El Filali for their input on the code containerization. We are also grateful to four anonymous reviewers whose comments substantially improved the manuscript. This work was partially performed using the computing facilities of the CC LBBE/PRABI. This work was partially supported by the grant ANR-18-CE02-0010 of the French National Research Agency ANR (project EcoNet).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and code (Nicvert et al., 2023) are available on Figshare at https://doi.org/10.6084/m9.figshare.24552157.v5.

ORCID

Lisa Nicvert https://orcid.org/0009-0006-5763-0865
Sophie Donnet https://orcid.org/0000-0003-4370-7316
Mark Keith https://orcid.org/0000-0001-7179-9989
Mike Peel https://orcid.org/0000-0003-1284-3665
Michael J. Somers https://orcid.org/0000-0002-5836-8823
Lourens H. Swanepoel https://orcid.org/0000-0002-9955-8076
Jan Venter https://orcid.org/0000-0002-4548-2571
Hervé Fritz https://orcid.org/0000-0002-7106-3661
Stéphane Dray https://orcid.org/0000-0003-0153-1105

REFERENCES

Albert, M., Y. Bouret, J. Chevallier, M. Fromont, F. Grammont, T. Laloe, C. Mascart, et al. 2021. "UnitEvents: Unitary Events Method with Delayed Coincidence Count (MTGAUE or Permutation Method) and Bernstein Lasso Method for Hawkes Processes." https://sourcesup.renater.fr/frs/?group_id=3267.

Beaudrot, L., M. S. Palmer, T. M. Anderson, and C. Packer. 2020. "Mixed-Species Groups of Serengeti Grazers: A Test of the Stress Gradient Hypothesis." *Ecology* 101: e03163.

Bell, R. H. 1971. "A Grazing Ecosystem in the Serengeti." *Scientific American* 225: 86–93.

Blanchet, F. G., K. Cazelles, and D. Gravel. 2020. "Co-Occurrence Is Not Evidence of Ecological Interactions." *Ecology Letters* 23: 1050–63.

Burton, A. C., E. Neilson, D. Moreira, A. Ladle, R. Steenweg, J. T. Fisher, E. Bayne, and S. Boutin. 2015. "Wildlife Camera Trapping: A Review and Recommendations for Linking Surveys to Ecological Processes." *Journal of Applied Ecology* 52: 675–685.

Caravaggi, A., P. B. Banks, A. C. Burton, C. M. V. Finlay, P. M. Haswell, M. W. Hayward, M. J. Rowcliffe, and M. D. Wood. 2017. "A Review of Camera Trapping for Conservation Behaviour Research." *Remote Sensing in Ecology and Conservation* 3: 109–122.

Carstensen, L., A. Sandelin, O. Winther, and N. R. Hansen. 2010. "Multivariate Hawkes Process Models of the Occurrence of Regulatory Elements." *BMC Bioinformatics* 11: 456.

Cornhill, K. L., C. Kelly, and G. I. H. Kerley. 2022. "Lion Reintroduction Demonstrates that Resident Cheetah Have a Spatially Reactive Response to Lion." *African Journal of Ecology* 60: 1–12.

Cornhill, K. L., and G. I. H. Kerley. 2020. "Cheetah Communication at Scent-Marking Sites Can be Inhibited or Delayed by Predators." *Behavioral Ecology and Sociobiology* 74: 21.

Courbin, N., A. J. Loveridge, D. W. Macdonald, H. Fritz, M. Valeix, E. T. Makuwe, and S. Chamaillé-Jammes. 2016. "Reactive Responses of Zebras to Lion Encounters Shape their Predator-Prey Space Game at Large Scale." *Oikos* 125: 829–838.

Cusack, J. J., A. J. Dickman, M. Kalyahe, J. M. Rowcliffe, C. Carbone, D. W. MacDonald, and T. Coulson. 2017. "Revealing Kleptoparasitic and Predatory Tendencies in an African Mammal Community Using Camera Traps: A Comparison of Spatiotemporal Approaches." *Oikos* 126: 812–822.

Daley, D. J., and D. Vere-Jones. 2003. *An Introduction to the Theory of Point Processes*, 2nd ed. New York: Springer.

Fitzgibbon, C. D. 1990. "Mixed-Species Grouping in Thomson's and Grant's Gazelles: The Antipredator Benefits." *Animal Behaviour* 39: 1116–26.

Frey, S., J. T. Fisher, A. C. Burton, and J. P. Volpe. 2017. "Investigating Animal Activity Patterns and Temporal Niche Partitioning Using Camera-Trap Data: Challenges and Opportunities." *Remote Sensing in Ecology and Conservation* 3: 123–132.

Fujita, K., A. Medvedev, S. Koyama, R. Lambiotte, and S. Shinomoto. 2018. "Identifying Exogenous and Endogenous Activity in Social Media." *Physical Review E* 98: 052304.

Galindo-Aguilar, R. E., B. C. Luna-Olivera, M. Ramírez-Ibáñez, and M. C. Lavariega. 2022. "Spatiotemporal co-Occurrence of Predators and Prey in a Neotropical Mammal Community in Southern Mexico." *Journal of Tropical Ecology* 38: 285–294.

Hansen, N. R., P. Reynaud-Bouret, and V. Rivoirard. 2015. "Lasso and Probabilistic Inequalities for Multivariate Point Processes." *Bernoulli* 21: 83–143.

Harmsen, B. J., R. J. Foster, S. C. Silver, L. E. T. Ostro, and C. P. Doncaster. 2009. "Spatial and Temporal Interactions of Sympatric Jaguars (*Panthera onca*) and Pumas (*Puma concolor*) in a Neotropical Forest." *Journal of Mammalogy* 90: 612–620.

Hawkes, A. G. 1971. "Spectra of some Self-Exciting and Mutually Exciting Point Processes." *Biometrika* 58: 83–90.

Hawkes, A. G. 2018. "Hawkes Processes and their Applications to Finance: A Review." *Quantitative Finance* 18: 193–98.

Hettena, A. M., N. Munoz, and D. T. Blumstein. 2014. "Prey Responses to Predator's Sounds: A Review and Empirical Study." *Ethology* 120: 427–452.

Jacoby, D. M. P., Y. P. Papastamatiou, and R. Freeman. 2016. "Inferring Animal Social Networks and Leadership: Applications for Passive Monitoring Arrays." *Journal of the Royal Society Interface* 13: 20160676.

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Université Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Karanth, K. U., A. Srivathsa, D. Vasudev, M. Puri, R. Parameshwaran, and N. S. Kumar. 2017. "Spatio-Temporal Interactions Facilitate Large Carnivore Sympatry across a Resource Gradient." *Proceedings of the Royal Society B: Biological Sciences* 284: 20161860.

Kellner, K. F., A. W. Parsons, R. Kays, J. J. Millspaugh, and C. T. Rota. 2022. "A Two-Species Occupancy Model with a Continuous-Time Detection Process Reveals Spatial and Temporal Interactions." *Journal of Agricultural, Biological and Environmental Statistics* 27: 321–338.

Kuijper, D. P. J., M. Verwijmeren, M. Churski, A. Zbyryt, K. Schmidt, B. Jędrzejewska, and C. Smit. 2014. "What Cues Do Ungulates Use to Assess Predation Risk in Dense Temperate Forests?" *PLoS One* 9: e84607.

Lambert, R. C., C. Tuleau-Malot, T. Bessaih, V. Rivoirard, Y. Bouret, N. Leresche, and P. Reynaud-Bouret. 2018. "Reconstructing the Functional Connectivity of Multiple Spike Trains Using Hawkes Models." *Journal of Neuroscience Methods* 297: 9–21.

McNaughton, S. J. 1976. "Serengeti Migratory Wildebeest: Facilitation of Energy Flow by Grazing." *Science* 191: 92–94.

McNaughton, S. J. 1985. "Ecology of a Grazing Ecosystem: The Serengeti." *Ecological Monographs* 55: 259–294.

Meise, K., D. W. Franks, and J. Bro-Jørgensen. 2019. "Using Social Network Analysis of Mixed-Species Groups in African Savannah Herbivores to Assess how Community Structure Responds to Environmental Change." *Philosophical Transactions of the Royal Society B* 374: 20190009.

Mohler, G., J. Carter, and R. Raje. 2018. "Improving Social Harm Indices with a Modulated Hawkes Process." *International Journal of Forecasting* 34: 431–39.

Murphy, A., D. R. Diefenbach, M. Ternent, M. Lovallo, and D. Miller. 2021. "Threading the Needle: How Humans Influence Predator-Prey Spatiotemporal Interactions in a Multiple-Predator System." *Journal of Animal Ecology* 90: 2377–90.

Nicvert, L., S. Donnet, M. Keith, M. Peel, M. J. Somers, L. H. Swanepoel, and J. A. Venter. 2023. "Code and Data for: Using the Multivariate Hawkes Process to Study Interactions between Multiple Species from Camera Trap Data." https://doi.org/10.6084/m9.figshare.24552157.v5.

Niedballa, J., A. Wilting, R. Sollmann, H. Hofer, and A. Courtiol. 2019. "Assessing Analytical Methods for Detecting Spatiotemporal Interactions between Species from Camera Trapping Data." *Remote Sensing in Ecology and Conservation* 5: 272–285.

Ogata, Y. 1988. "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes." *Journal of the American Statistical Association* 83: 9–27.

Ogata, Y. 1998. "Space-Time Point-Process Models for Earthquake Occurrences." *Annals of the Institute of Statistical Mathematics* 50: 379–402.

Ovaskainen, O., G. Tikhonov, A. Norberg, F. Guillaume Blanchet, L. Duan, D. Dunson, T. Roslin, and N. Abrego. 2017. "How to Make More out of Community Data? A Conceptual Framework and its Implementation as Models and Software." *Ecology Letters* 20: 561–576.

Palmer, M. S., K. M. Gaynor, J. A. Becker, J. O. Abraham, M. A. Mumma, and R. M. Pringle. 2022. "Dynamic Landscapes of Fear: Understanding Spatiotemporal Risk." *Trends in Ecology & Evolution* 37: 911–925.

Pardo, L. E., S. P. Bombaci, S. Huebner, M. J. Somers, H. Fritz, C. Downs, A. Guthmann, et al. 2021. "Snapshot Safari: A Large-Scale Collaborative to Monitor Africa's Remarkable Biodiversity." *South African Journal of Science* 117: 1–4.

Parsons, A. W., C. Bland, T. Forrester, M. C. Baker-Whatton, S. G. Schuttler, W. J. McShea, R. Costello, and R. Kays. 2016. "The Ecological Impact of Humans and Dogs on Wildlife in Protected Areas in Eastern North America." *Biological Conservation* 203: 75–88.

Parsons, A. W., K. F. Kellner, C. T. Rota, S. G. Schuttler, J. J. Millspaugh, and R. W. Kays. 2022. "The Effect of Urbanization on Spatiotemporal Interactions between Gray Foxes and Coyotes." *Ecosphere* 13: e3993.

Pays, O., A. Ekori, and H. Fritz. 2014. "On the Advantages of Mixed-Species Groups: Impalas Adjust their Vigilance when Associated with Larger Prey Herbivores." *Ethology* 120: 1207–16.

Pays, O., E. Sirot, and H. Fritz. 2012. "Collective Vigilance in the Greater Kudu: Towards a Better Understanding of Synchronization Patterns." *Ethology* 118: 1–9.

Popovic, G. C., D. I. Warton, F. J. Thomson, F. K. Hui, and A. T. Moles. 2019. "Untangling Direct Species Associations from Indirect Mediator Species Effects with Graphical Models." *Methods in Ecology and Evolution* 10: 1571–83.

Prat-Guitart, M., D. P. Onorato, J. E. Hines, and M. K. Oli. 2020. "Spatiotemporal Pattern of Interactions between an Apex Predator and Sympatric Species." *Journal of Mammalogy* 101: 1279–88.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Reinhart, A. 2018. "A Review of Self-Exciting Spatio-Temporal Point Processes and their Applications." *Statistical Science* 33: 299–318.

Reynaud-Bouret, P., V. Rivoirard, and C. Tuleau-Malot. 2013. "Inference of Functional Connectivity in Neurosciences Via Hawkes Processes." In *2013 IEEE Global Conference on Signal and Information Processing* 317–320. Austin: IEEE.

Ridout, M. S., and M. Linkie. 2009. "Estimating Overlap of Daily Activity Patterns from Camera Trap Data." *Journal of Agricultural, Biological, and Environmental Statistics* 14: 322–337.

Rota, C. T., M. A. R. Ferreira, R. W. Kays, T. D. Forrester, E. L. Kalies, W. J. McShea, A. W. Parsons, and J. J. Millspaugh. 2016. "A Multispecies Occupancy Model for Two or More Interacting Species." *Methods in Ecology and Evolution* 7: 1164–73.

Say-Sallaz, E., S. Chamaillé-Jammes, H. Fritz, and M. Valeix. 2019. "Non-consumptive Effects of Predation in Large Terrestrial Mammals: Mapping our Knowledge and Revealing the Tip of the Iceberg." *Biological Conservation* 235: 36–52.

Schliep, E. M., A. E. Gelfand, J. S. Clark, and R. Kays. 2018. "Joint Temporal Point Pattern Models for Proximate Species Occurrence in a Fixed Area Using Camera Trap Data." *Journal of Agricultural, Biological and Environmental Statistics* 23: 334–357.

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Université Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Schmitt, M. H., K. Stears, C. C. Wilmers, and A. M. Shrader. 2014. "Determining the Relative Importance of Dilution and Detection for Zebra Foraging in Mixed-Species Herds." *Animal Behaviour* 96: 151–58.

Searle, C. E., J. B. Smit, J. J. Cusack, P. Strampelli, A. Grau, L. Mkuburo, D. W. Macdonald, A. J. Loveridge, and A. J. Dickman. 2021. "Temporal Partitioning and Spatiotemporal Avoidance among Large Carnivores in a Human-Impacted African Landscape." *PLoS One* 16: e0256876.

Valiente-Banuet, A., M. A. Aizen, J. M. Alcántara, J. Arroyo, A. Cocucci, M. Galetti, M. B. García, et al. 2015. "Beyond Species Loss: The Extinction of Ecological Interactions in a Changing World." *Functional Ecology* 29: 299–307.

Zurell, D., U. Berger, J. S. Cabral, F. Jeltsch, C. N. Meynard, T. Münkemüller, N. Nehrbass, et al. 2010. "The Virtual Ecologist Approach: Simulating Data and Observers." *Oikos* 119: 622–635.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

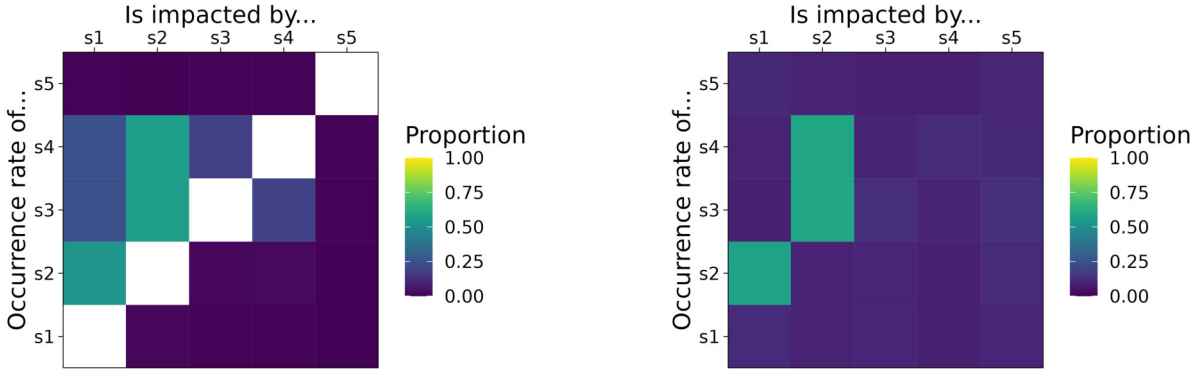
How to cite this article: Nicvert, Lisa, Sophie Donnet, Mark Keith, Mike Peel, Michael J. Somers, Lourens H. Swanepoel, Jan Venter, Hervé Fritz, and Stéphane Dray. 2024. "Using the Multivariate Hawkes Process to Study Interactions between Multiple Species from Camera Trap Data." *Ecology* 105(4): e4237. https://doi.org/10.1002/ecy.4237

19399170, 2024, 4, Downloaded from https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.4237 by Université Claude Bernard, Wiley Online Library on [02/04/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendices

S1 Bias in the inference methods

We are interested in knowing whether spurious interactions are more likely to be inferred between species which are part of an interaction chain. To investigate this question, we analyze the inferences performed with the inter-event times method of Murphy et al. (2021) and with the Hawkes method. The true interaction graph used for these simulations is depicted in Figure 1a in the main text (s_1 attracts s_2, and s_2 attracts s_3 and s_4). The total number of repetitions over all conditions were respectively $N = 750$ for the inter-event times method (Figure S1a) (5 durations $\times$ 5 strengths $\times$ 30 datasets) and $N = 2250$ for the Hawkes model (Figure S1b) (5 durations $\times$ 5 strengths $\times$ 30 datasets $\times$ 3 penalizations). With the inter-events times method, the model is biased towards inferring more often the indirect interactions $s_1 \to s_3$, $s_1 \to s_4$, $s_3 \to s_4$ and $s_4 \to s_3$. This is not the case with the Hawkes model.



(a) Inter-event times

(b) Hawkes model

Figure S1: **Bias in the inference.** These graphs show the proportion of times an interaction was inferred over all simulated data in different conditions for (a) the inter-event times method of Murphy et al. (2021) and (b) the Hawkes model. The true interactions are $s_1 \to s_2$, $s_2 \to s_3$ and $s_2 \to s_4$.

S2 Background rates inferred for real data

The figure below shows the background rates inferred for each species in the application example. These background rates are closely linked to the species abundances observed on camera traps, with impala being by far the most abundant species (50 803 occurrences), followed by zebra (9 843), then wildebeest and kudu (respectively 5910 and 5560) and lion being far more rare (587).

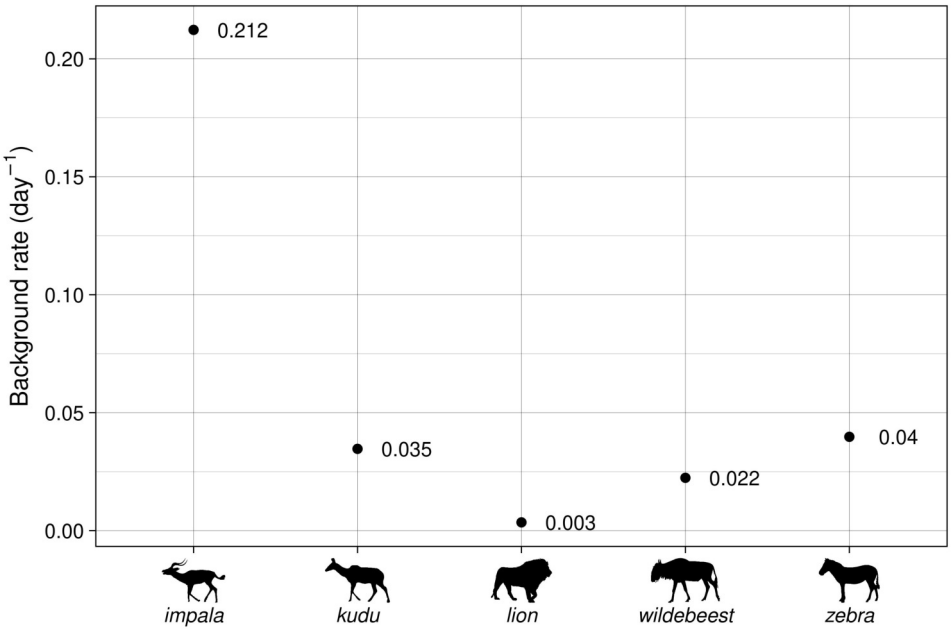
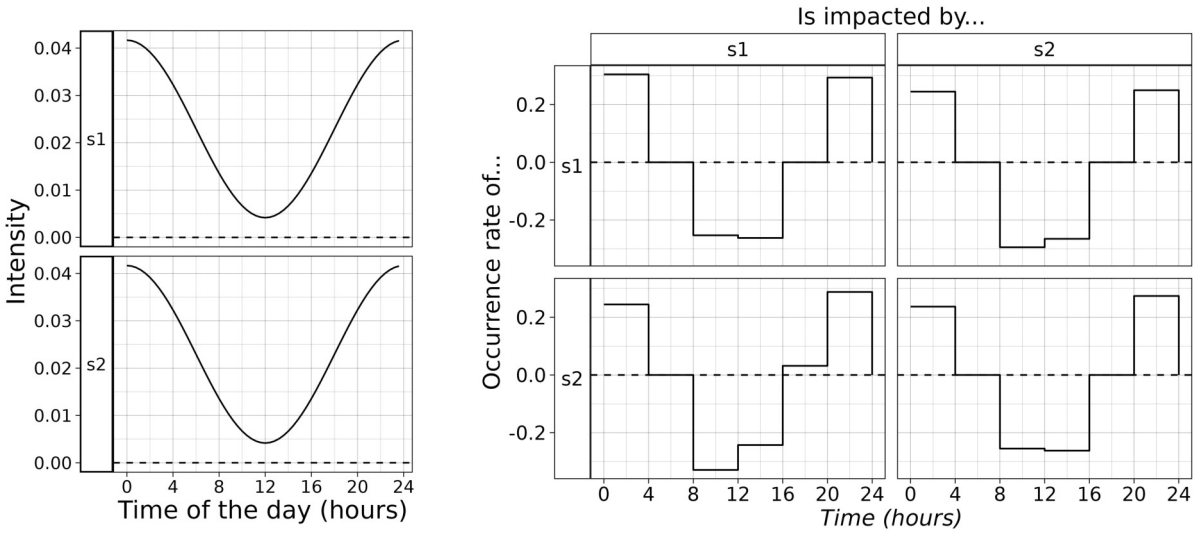


Figure S2: **Background rates inferred with the Hawkes process for each species.** The background rates values are written besides the points.
Silhouette images from PhyloPic by Lukasiniho (wildebeest), Margot Michaud (lion), Robert Hering (kudu), Zimices (zebra) and an unknown author (impala).

S3 Influence of circadian rhythms

We simulated two species occurrences with cyclic rates in order to mimic circadian rhythms. We simulated occurrences over 1000 days on a single camera. For that, we used an inhomogeneous Poisson process for which species intensities varied periodically over 24 hours, with the Poisson non-homogeneous intensity defined with a cosine function with a period of 24 hours. Even though species occurrences were independent, the Hawkes process inferred a periodic attraction/repulsion pattern between and within species.



(a) Inhomogeneous Poisson intensity

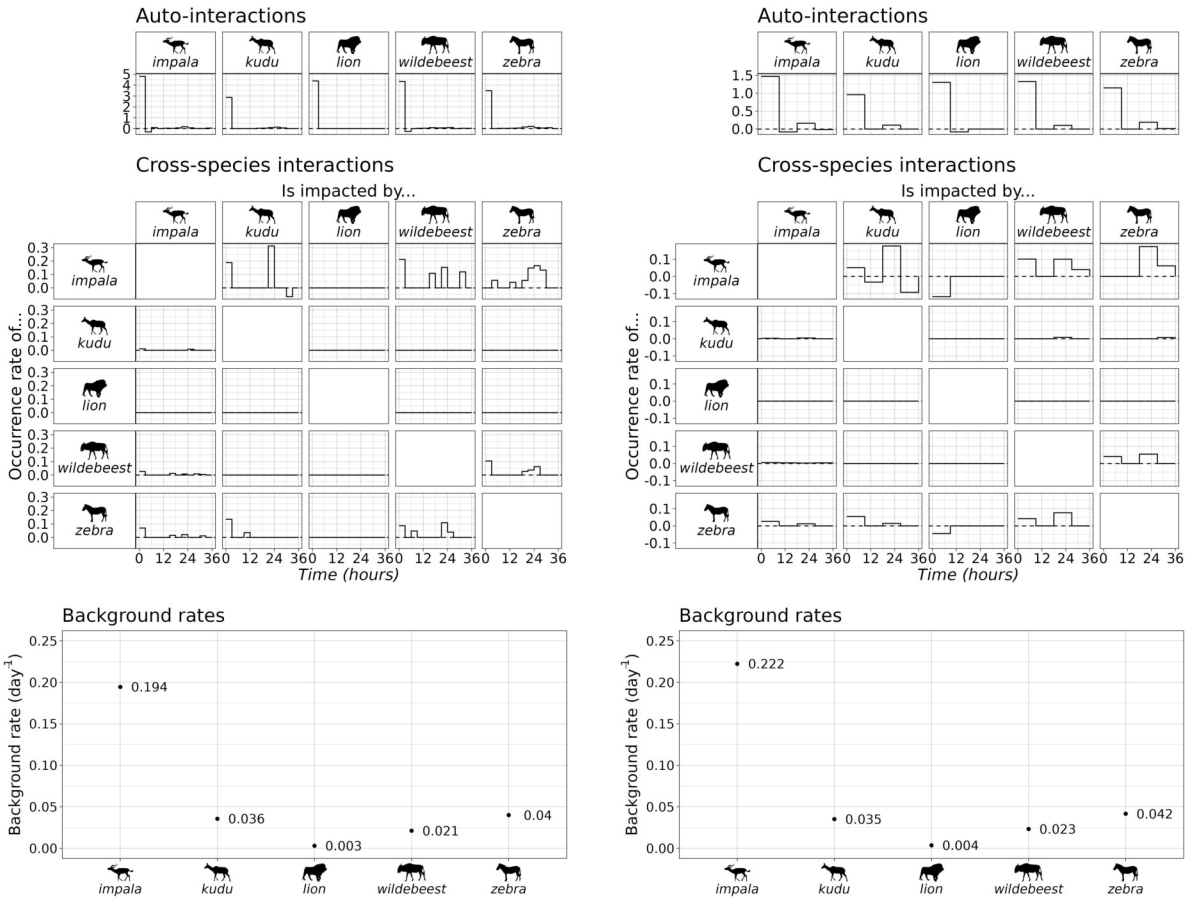
(b) Inferred interactions with the Hawkes process

Figure S3: **Inference of spurious interactions when circadian rhythms are present.** (a) shows the non-homogeneous Poisson intensities for the two species. (b) shows the interaction functions inferred with a multivariate Hawkes process from the data simulated with two non-homogeneous Poisson process for which the intensities are shown in (a).

S4 Influence of bin width

In order to check if bin width influences the inferred interaction functions, we changed the bin width used for the inference on the real data. Here, we inferred a Hawkes process with bins widths δ of 3 and 9 hours (compared to $\delta = 6$ hours in the main text).

Results are shown in Figure S4. Overall, the parameters are similar between 3, 6 and 9 hour bins. Regarding the 3 hours bins, the nonzero interactions are the same than with 6 hours bins and they have the same direction. Only the negative interactions (lion-impala and lion-zebra) are missed, possibly due to low power. The inferred background rates are similar to the rates inferred with 6 hours bins, except for impala which has a smaller background rate (0.194 compared to 0.212 occurrences day^{-1} in the main text).



(a) Bin width: 3 hours

(b) Bin width: 9 hours

Figure S4: **Inference of a multivariate Hawkes process with different bin widths.** (a) shows the parameters inferred with 3 hours bins and (b) the parameters inferred with 9 hours bins.
Silhouette images from PhyloPic by Lukasiniho (wildebeest), Margot Michaud (lion), Robert Hering (kudu), Zimices (zebra) and an unknown author (impala).

Regarding the 9 hours bins, all nonzero interactions are the same and they have

the same direction as in the 6 hours bins version. The inferred background rates are similar to those inferred with 6 hours bins as well.

Finally, we can note that the values of the interaction functions are smaller when the bins widths are larger. This is expected because the average number of occurrences gained or suppressed during a given interval is the integral of the function over this interval.

Chapter 3.2

Non-linear multivariate Hawkes process with a temporal covariate

1 Introduction and motivation

At the community scale, species distributions are influenced by a multitude of factors. One of these factors is the presence of other species of the community, mediated by interspecific interactions, an effect we have described and studied in the previous chapter. Indeed, the presence or absence of other animal species can have an effect on the distribution of species, in space and/or time and at different scales. But other important factors include environmental and temporal variables. Regarding environmental variables, for example, Gompper et al. (2016) showed that the type of habitat (like forests, roads or house densities) is a prominent factor shaping the distribution of five carnivore taxa in North America. Regarding temporal factors, several species have evolved activity patterns depending on external cues. Circadian clocks are widespread across the tree of life, and allow species to adapt their activity to the day-night cycle (Dibner et al., 2010). In this context, camera trap data allows to precisely quantify species daily activity patterns by collecting species records with precise date and time (Frey et al., 2017). Other temporal factors include seasonal variation: for instance, Fisher et al. (2020) showed that the distribution of white-tailed deer in a boreal forest shrinks in winter, reflecting the higher difficulty to move in a harsh environment.

In chapter 3.1, we used the multivariate Hawkes process (MHP) to infer the attraction and repulsion between species from camera trap data. We called the inferred pattern *reactive temporal interaction*, as it corresponds to a reactive response to another species presence in time. However, this model does not allow to include covariates, thus overlooking environmental and temporal factors affecting species distribution on top of species interactions. This hampers the interpretability of the model, as variables not taken into account might blur the signal or produce spurious reactive temporal interaction patterns. On particular, the results of the inference on real data in chapter 3.1 suggests a strong effect of circadian rhythms to explain the observed periodic patterns of attractions between species.

Recently, several Hawkes processes including covariates have been proposed (Truccolo et al., 2005; Mohler et al., 2018; Fallahi and Pourtaheri, 2024; Xu et al., 2018; see Reinhart, 2018 for a review). Covariates can be included in the model in different ways: for instance, `UnitEvents` allows the interaction functions to vary according to the time period (Albert et al., 2021). There are also spatio-temporal extensions of

the Hawkes processes allowing to model the interaction functions depending on spatial coordinates as well as time (Ogata, 1998; Reinhart, 2018). In our context, including covariates in the background rate of the Hawkes model is particularly interesting. Indeed, biologically, we expect the intrinsic probability of presence of a species to vary substantially according to spatial and temporal covariates such as habitat or circadian rhythms. Such models have been developed for instance by Truccolo et al. (2005) or Xu et al. (2018).

In this chapter, I test a non-linear multivariate Hawkes process (non-linear MHP) described by Carstensen et al. (2010) and implemented in the R package `ppstat` (Hansen, 2023). I first evaluate the inference using simulated data. Then, I infer interactions between species on real data using time of the day as a covariate, using the same dataset as the one used in chapter 3.1.

2 Material and methods

2.1 Model: the non-linear multivariate Hawkes process

In this chapter, I use a non-linear MHP to model species occurrences collected by camera trap data. This model is presented in Carstensen et al. (2010) and implemented in the R package `ppstat` (Hansen, 2023).

The main difference of this model with the linear MHP used in chapter 3.1 is the way interaction functions and background rates combine to define the species intensity λ: with the linear MHP, the effects of previous events on the focal species intensity are additive, whereas with the non-linear MHP they are multiplicative (see Equation (3.2.2)).

We consider S species observed on C cameras. For each camera l, let t_0^l be the time of the first event occurrence. The non-linear MHP used in `ppstat` is parameterized by a log-linear intensity, written as:

$$\ln(\lambda_i^l(t)) = \eta_i^l(t) + \sum_{j=1}^{S} \int_{[t_0^l, t]} g_{j \to i}(t-u) N_j^l(\mathrm{d}u) \tag{3.2.1}$$

where i denotes the focal species, j denotes the other species and l denotes the

camera. $N_j^l(t)$ is the number of occurrences of species j on camera l at time t. This log-linear parametrization allows to model repulsion between species, as $\lambda_i^l(t)$ is constrained to be positive with the logarithm.

Interpreting the model is easier when the equation directly describes intensity. By modifying Equation (3.2.1), we can rewrite the model as:

$$\lambda_i^l(t) = \underbrace{\nu_i^l(t)}_{\exp(\eta_i^l(t))} \times \prod_{j=1}^{S} \int_{[a,t]} \underbrace{f_{j\to i}(t-u)}_{\exp(g_{j\to i}(t-u))} N_j^l \mathrm{d}u \tag{3.2.2}$$

In this equation, $\nu_i^l(t)$ and $f_{j\to i}$ are respectively defined as the exponentials of $\eta_i^l(t)$ and $g_{j\to i}$ defined in Equation (3.2.2). $\nu_i^l(t)$ represents the background rate function, a function with positive values that represents the intensity of species i independent on previous occurrences. To model variations in the background rate, $\nu_i^l(t)$ can include covariates:

$$\nu_i^l(t) = \exp\left((\boldsymbol{\alpha}_i^l)^\top \mathbf{X}_i^l(t)\right) \tag{3.2.3}$$

where $\mathbf{X}_i^l(t)$ is a matrix of covariates that can depend on camera l, species i and time t, $\boldsymbol{\alpha}_i^l$ is a parameter vector of coefficients and $^\top$ denotes transposition. The first element of $\nu_i^l(t)$ corresponds to the baseline intensity for species i (the first element of $\boldsymbol{\alpha}_i^l$ is the logarithm of the baseline intensity and the corresponding elements of $\mathbf{X}_i^l(t)$ are 1). If no covariates are included, ν_i is a constant depending only on the species, as in the MHP presented in chapter 3.1.

The interactions functions $f_{j\to i}$ represent the multiplicative effect of occurrences of species j on the intensity of species i. They have values in $]0, +\infty[$: a value between zero and one indicates an avoidance of species j by species i, a value above one indicates an attraction and a value of one indicates no effect. In this framework, I choose to model these functions using cubic splines: the functions $g_{j\to i}$ are defined as cubic splines, which are piecewise polynomials of degree 3 that are jointed at specific points on the support of the function (the knots). The polynomial functions are defined using the B-spline basis, so that the functions $f_{j\to i}$ are written as:

$$f_{j\to i}(t-u) = \exp\left(\sum_{k=1}^{K+4} (\boldsymbol{\beta}_{j\to i}^k)^\top \mathbf{B}_k(t-u)\right) \tag{3.2.4}$$

where B_k are the B-spline basis functions, $\boldsymbol{\beta}_{i\to j}^k$ is the parameter vector corresponding to the k-th spline basis, and K is the number of equidistant knots. The number of knots together with the splines degree controls the degrees of freedom (here, it is equal to $K+4$ because I use cubic splines).

The spline parametrization (spline basis and knots) is chosen by the user a priori. A particular point of interest is the knots number and placement (Perperoglou et al., 2019). With the B-spline basis, I choose inner knots defining the support of the function, in addition to boundary knots which are outside the support and allow more flexibility. To choose the knots, I rely on the preliminary study in Carstensen et al. (2010) and choose similar parameters.

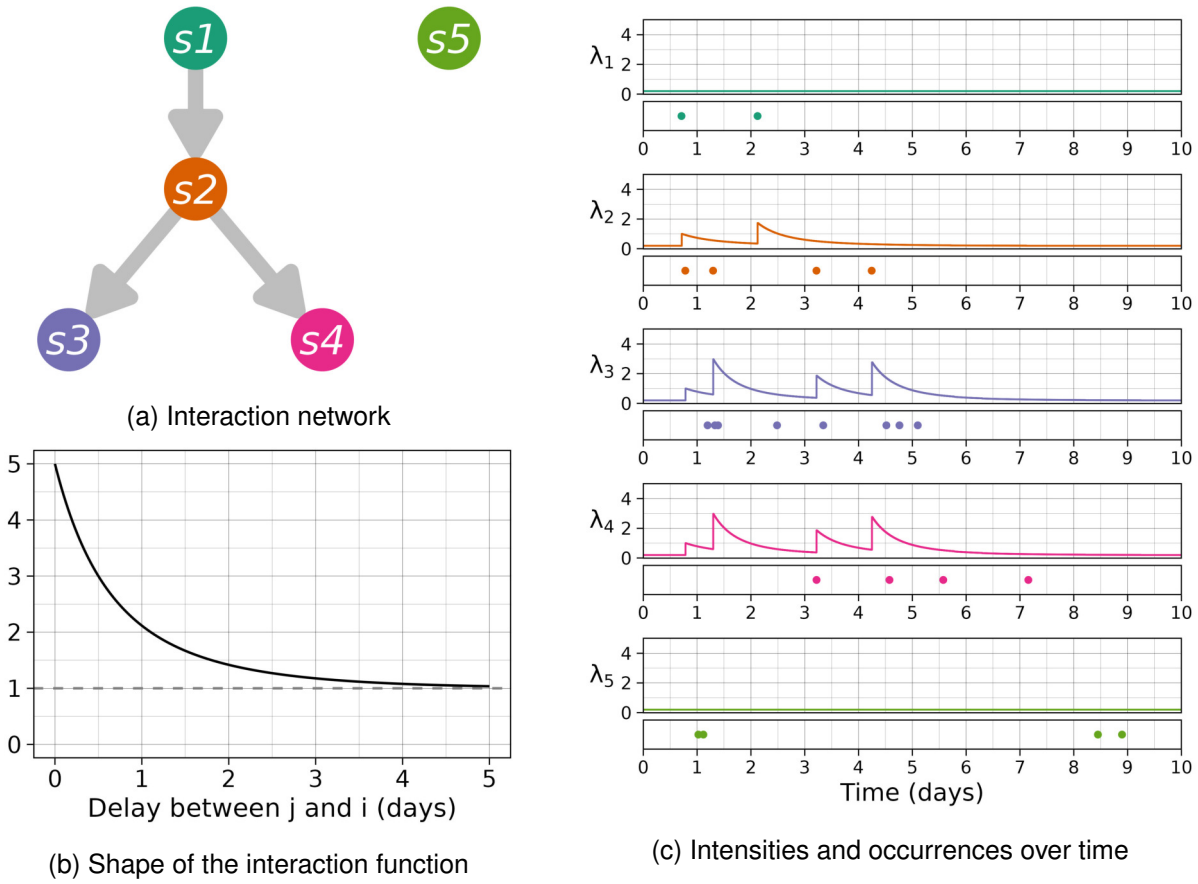
In order to illustrate this model, I simulate a simple realization of a non-linear MHP described with Equation (3.2.2) (Figure 3.2.1). In this example, five species interact following the graph in panel 3.2.1a, where the interaction functions $f_{j\to i}$ are shown in panel 3.2.1b. The right panel 3.2.1c shows the resulting occurrences and intensities. As with the linear MHP, an occurrence of an attracting species leads to a peak in the intensity of an attracted species (see for example the effect of an occurrence of s_1 on the intensity of s_2). But contrary to the linear MHP, the magnitude of the peak is not fixed, but instead depends on the intensity of the focal species at time t: it is very clear when comparing the magnitudes of the first and the second peak of s_3 and s_4.

2.2 Inference

To infer the parameters of a non-linear MHP from data, `ppstat` uses a maximum likelihood approach. The observed occurrence times are denoted T_m^{li} for the m-th instant of occurrence of species i on camera l. Let N_i^l be the total number of occurrences of species i at camera l. The log-likelihood to maximize is written as:

$$L(\boldsymbol{\Theta}) = \sum_{l=1}^{C}\left(\sum_{i=1}^{S}\left(\sum_{m=1}^{N_i^l}\log(\lambda_{\boldsymbol{\theta}_i}^l(T_m^l)) - \int_{t_0^l}^{t_{\max}^l}\lambda_{\boldsymbol{\theta}_i}^l(s)\mathrm{d}s\right)\right) \tag{3.2.5}$$

where $\boldsymbol{\theta}_i$ denotes the parameter vector for species i (covariates coefficients $\boldsymbol{\alpha}_i^l$ and spline coefficients and $\boldsymbol{\beta}_{j\to i}^k$) and $\boldsymbol{\Theta}$ is the parameter matrix for all species. $\lambda_{\boldsymbol{\theta}_i}^l$ is the intensity of species i on camera l computed with the parameters $\boldsymbol{\theta}_i$.



(a) Interaction network

(b) Shape of the interaction function

(c) Intensities and occurrences over time

Figure 3.2.1: Example of a realization of a non-linear multivariate Hawkes process. (a) shows the interaction network between five species (each arrow represents an interaction function $f_{i\to i}$ different from one). (b) shows the shape of the interaction functions $f_{i\to i}$, corresponding to arrows in the interaction network (a). The functions $f_{i\to j}$ are equal to $\exp(g_{i\to j})$, where $g_{i\to j}$ are defined as a decreasing exponentials. (c) shows a realization of the Hawkes process with the interaction network and the interaction functions shown in (a) and (b). For this simulation, the background rate ν_i^l was set to 0.2 occurrences day^{-1} for all species. For each species i, the above panel shows the intensity λ_i and the bottom panel shows the species occurrences. Each time an attracting species occurs, the intensity for the attracted species is multiplied by a factor dictated by the interaction function and then decreases following this function's shape.

The log-linear parametrization of the model guarantees to find the global maximum for this log-likelihood, if it exists (see Daley and Vere-Jones, 2003 (p 235), cited in Carstensen et al., 2010). This property allows `ppstat` to use a gradient descent procedure for the optimization (see the original publication by Carstensen et al. (2010) and its Appendix 11 for additional details).

2.3 Simulation

To evaluate the accuracy of the inference, I simulated data under a non-linear MHP and reinferred the model parameters from simulated data. For this pilot study, I chose to

simulate interactions between five interacting species (see interaction graph in Figure 3.2.1a) with a background rate ν fixed at 0.1 day^{-1} for all species. The interaction function is defined as

$$f_{j\to i}(t) = \begin{cases} \exp(g_{j\to i}) = \exp\left(a\exp\left(-bt\right)\right) & \text{if } t < 2 \\ 0 & \text{if } t \geq 2 \end{cases} \tag{3.2.6}$$

The shape of this function is constrained by the definition of $f_{j\to i}$ as the exponential of $g_{j\to i}$. Defining $f_{j\to i}$ as the exponential of a decreasing exponential still has good properties: indeed, this function converges to one (independent species) when time increases. Furthermore, we can define the parameters a and b to be interpretable. In this framework, I fixed $a = \ln(3)$. This parametrization makes the strength of the interaction at $t = 0$ equal to 3. This means that if an attracting species occurs at time t, the intensity will be multiplied by 3. b controls the decrease rate: I chose $b = -\frac{1}{0.5}\ln\left(\frac{1}{a}\ln\left(\frac{\exp(a)+1}{2}\right)\right)$. With this parametrization, we can show that half of the initial value of the function $f_{j\to i}$ is reached at $t = 0.5$ days.

These interactions were simulated on 25 cameras active for 300 days each, and the inference was repeated on 100 simulated datasets. The inference parameters are 200 B-spline bases of order 4, evenly spaced between -0.5 and 2 days with a step of 6 hours.

2.4 Data analysis

I reanalyzed the occurrences of $S = 5$ species collected on $l = 179$ camera traps of the Snapshot Safari network (same data as in chapter 3.1). The data collection and cleaning processes are described in chapter 3.1.

I fitted two models. The first model is a non-linear MHP with no covariates, which serves as a baseline for comparison with other models. It is written as:

$$\lambda_i^l(t) = \nu_i \times \prod_{j=1}^{S=5} \int_{[a=1.5,t]} f_{j\to i}(t-u) N_j^l \mathrm{d}u \tag{3.2.7}$$

Here, the background rate ν_i is a constant depending only on the species.

The second model is a non-linear MHP using the time of the day as a covariate,

which allows to model the background rate as a function of the time of day. It is written as:

$$\lambda_i^l(t) = \nu_i^l(t) \times \prod_{j=1}^{S=5} \int_{[a=1.5,t]} f_{j\to i}(t-u) N_j^l \mathrm{d}u \tag{3.2.8}$$

Here, the background rate $\nu_i^l(t)$ is a function of the time of the day. We note $\mathbf{x}^l(t)$ the vector of the time of day at time t on camera l. In this model, the time of day is represented as a fraction of a day (it has values in $[0, 1[$). $\nu_i^l(t)$ is defined with cubic splines:

$$\nu_i^l(t) = \exp\left((\boldsymbol{\alpha}_i^l)^\top \mathbf{B_1}(\mathbf{x}^l(t))\right) \tag{3.2.9}$$

Where $\mathbf{B_1}(\mathbf{x}^l(t))$ is a spline-transformation of the time of day. Its first row is only composed of ones and other columns represent the cubic spline transformations on the time of day. $\boldsymbol{\alpha}_i^l$ is the covariates vector (its first element corresponds to the logarithm of the baseline intensity and the other elements are the coefficients associated to the splines bases). I chose 15 equidistant knots between $-2/24$ and $1+2/24$ (1 knot every 2 hours, with two boundary knots), and I enforced symmetric spines bases around zero, so that the values loop around midnight. Because the dataset spans over 2 years, sunrise and sunset times change depending on the season. To correct that, I anchored the time of day to mean sunrise and sunset times using the `activity` package (Rowcliffe, 2023).

For both models, I consider only interactions occurring at most 36 hours after the occurrence ($a = 1.5$ days). The interaction functions $f_{j\to i}$ are modeled as exponentials of cubic splines, as described in Equation (3.2.4) (with $K = 200$ basis functions). I chose 9 equidistant knots between -0.5 and 1.5 days (inter-knot space is 6 hours).

3 Results

3.1 Simulation

I simulated data under the non-linear MHP, and used the simulated data to reinfer the parameters of the same model using the inference procedure of `ppstat`. Results are presented in Figure 3.2.2. Interactions are correctly recovered: the confidence intervals of the inferred functions exclude one when there are interactions and includes

one when there are no interactions. Together, these results suggest that the model is able to detect true interactions (no false negatives) and to avoid the spurious inference of interactions (no false positives).

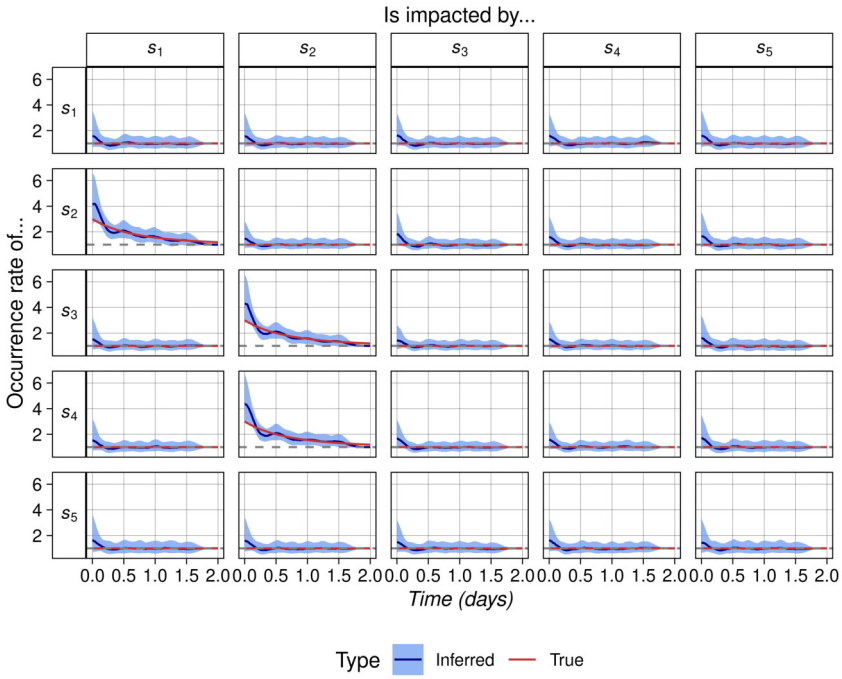


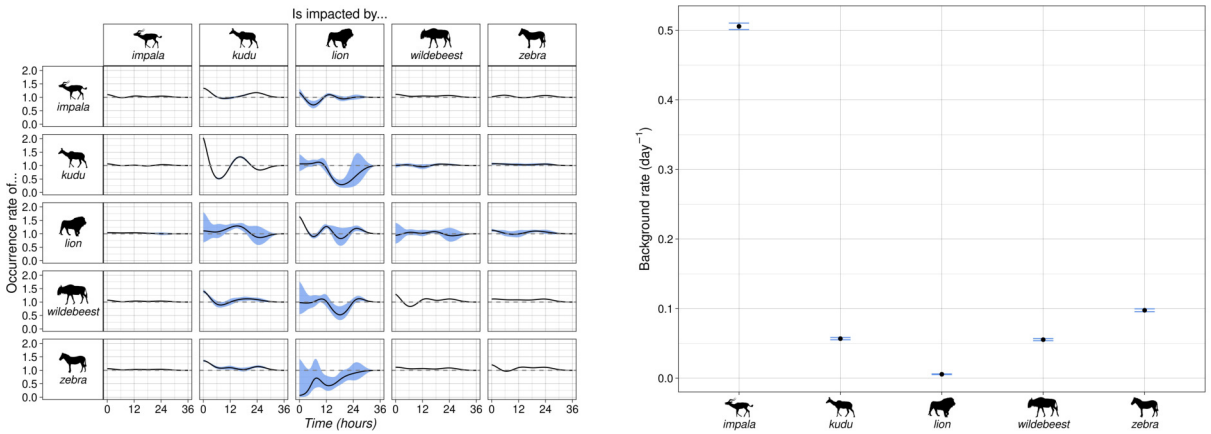
Figure 3.2.2: Performance of the inference with the non-linear multivariate Hawkes process. This graph shows the inferred function (blue, with confidence interval), compared to the true functions used in the simulation model (red) for each pair of species. The line for the inferred interactions functions represents the median of the function for the 100 repetitions and the confidence interval represents the medians of the 95% confidence interval computed by `ppstat`. The intensity of species in rows is affected by species in columns. The horizontal dashed line represents one. Interaction functions correspond to the $f_{j\to i}$ from Equation (3.2.2).

However, we observe some differences between the values of the inferred and the simulated interactions functions. In particular, a trend for spurious positive interaction at a small delay is observed for the majority of functions even when there are no simulated interactions. Similarly, the inferred values corresponding to a simulated interaction are also over-estimated at small delay. The inferred functions also tend to fluctuate around the true value: this is particularly visible for the non-one interaction functions between delays 0.25 and 1.

3.2 Data analysis

I analyzed species occurrences using two models: one model no covariates, and one model where the time of day is included as a covariate. Both models could be run on a

personal computer and the inference took around 5–10 minutes per model. As noted by Carstensen et al. (2010), fitting big models can be prohibitive because the inference involves computing a large model matrix.



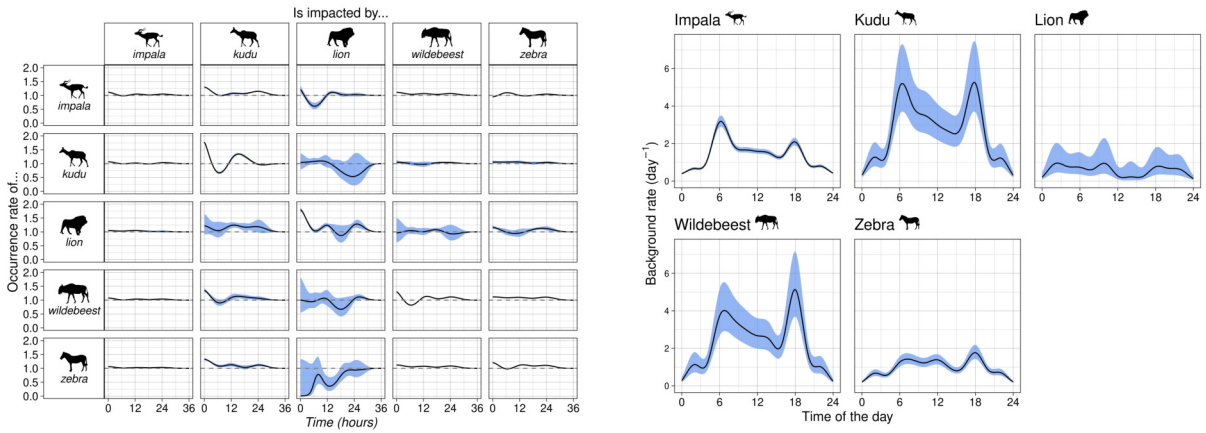
(a) Interaction functions

(b) Background rates

Figure 3.2.3: Model inferred without covariates. (a) Inferred interaction functions, where the intensity of species in rows is affected by species in columns. The horizontal dashed line represents one. Interaction functions correspond to the $f_{j\to i}$ from Equation (3.2.2). (b) Inferred background rate. With this model, background rates are a constant depending only on the species. They correspond to the ν_i^l defined in Equation (3.2.2). Silhouette images from PhyloPic by Lukasiniho (wildebeest), Margot Michaud (lion), Robert Hering (kudu), Zimices (zebra) and an unknown author (impala).

Regarding interaction functions, contrary to what I expected, the two models yielded very similar interaction functions (Figures 3.2.3a and 3.2.4a). Unless stated otherwise, the results below apply to both models. Regarding auto-attractions, kudu, lions, wildebeest and zebra (and to a lesser extent, impala) display a short-term attraction (up to 6 hours after an occurrence). For kudu and wildebeest, this short-term attraction is followed by a repulsion between 3 and 12 hours, and then an attraction again. For kudu, the auto-repulsion is attenuated with the model taking the time of day into account.

Regarding interspecific interactions, we observe weak herbivore-herbivore attractions. Impala follows kudu, and to a lesser extent wildebeest and zebra (surprisingly, the attraction with zebra does not occur in the immediate short term, but is initially one and increases, then decreases to be one again at 12 hours). Kudu does not display any clear following pattern with other herbivores. Wildebeest follows kudu mainly in the short term (< 6 hours), and zebra to a lesser extent and on a more extended timespan (up to 36 hours). Finally, zebra follows kudu (strong attraction in the short term, then decreases and extends to 36 hours) and to a lesser extent wildebeest (weak attraction up to 36 hours). The uncertainty around lion-related interactions are very large, so it is



(a) Interactions

(b) Background rates

Figure 3.2.4: Model inferred with a time of day covariate. (a) Inferred interaction functions, where the intensity of species in rows is affected by species in columns. The horizontal dashed line represents one. Interaction functions correspond to the $f_{j\to i}$ from Equation (3.2.2). (b) Inferred background rate. With this model, background rates are a function of the time of the day. They correspond to the ν_i^l defined in Equation (3.2.2). Silhouette images from PhyloPic by Lukasiniho (wildebeest), Margot Michaud (lion), Robert Hering (kudu), Zimices (zebra) and an unknown author (impala).

hard to draw solid conclusions. However, we can see a small attraction of lion by zebra between up to 6 hours and 12 to 24 hours after an occurrence and by kudu between 12 and 18 hours. Regarding the effect of lions on other herbivores, impala avoids lion between 3 and 9 hours, and this effect is accentuated when taking the time of day into account. Surprisingly, this avoidance is not immediately after a lion occurrence: the effect is initially one, then decreases and increases to reach one again after 12 hours. Kudu seems to avoid lion between 12 and 24 hours, an effect that vanishes when taking the time of day into account. Finally, zebra displays a tendency to avoid lion in the short-term and up to 24 hours, but the confidence interval indicates that this effect is very uncertain.

Background rates are estimated differently depending on the model. For the first model (Figure 3.2.3b), the occurrence rates are constants depending on the species. They follow the total occurrence counts of species (impala: 0.506; zebra: 0.098; kudu: 0.057; wildebeest: 0.055 and lion: 0.006 day^{-1}). For the second model (Figure 3.2.4b), the background rates are functions of the time of the day. They show that impala, kudu and wildebeest are diurnal with crepuscular activity peaks (particularly marked for impala), zebra is diurnal and lion is crepuscular. These function shapes are in accordance with the species circadian rhythms measured from data collected with these camera traps and also described in the literature. The baseline of the background rate,

which corresponds to the value of the function computed at midnight (intercept of the model), are proportional to species abundances collected on camera traps (impala: 0.321; zebra: 0.074; kudu: 0.026; wildebeest: 0.026; lion: 0.008 day^{-1}).

4 Discussion

In this chapter, I used a non-linear multivariate Hawkes process to analyze camera trap data. I included a temporal covariate to take circadian rhythms into account and measured patterns of reactive temporal interactions between species (more simply, "interactions" in this discussion).

4.1 Simulations

Simulations suggest that the model can accurately infer the presence or absence of interactions. However, the inferred shape of the interaction is not very precise. Notably, we notice a spurious trend to infer short-term attractions and the inferred interactions tend to fluctuate around the true value. Despite these issues on the estimates, the confidence interval around the estimated value encompasses the true interaction function shape.

These fluctuations are most probably due to the spline parametrization: as noted in Perperoglou et al. (2019), splines are sensitive to over-fitting. This problem could be overcome by fitting penalized splines, thought to be less sensitive to over-fitting (Eilers & Marx, 2010). The splines knots are also known to have a big influence on the estimate and are notoriously difficult to choose (Perperoglou et al., 2019). In this preliminary work, I chose the knots in a heuristic manner, but a solution could be to conduct a more thorough evaluation to choose optimal knots spacing and bounds. We could also use another function shape to specify the interaction functions or covariates response, for instance polynomial functions, as advised by Kroc and Olvera Astivia (2023): with the `ppstat` framework, the choice of the function is free, so this would be entirely possible.

The `ppstat` simulation algorithm does not allow to simulate data from a model including a covariate in the background rate. Therefore, I could not evaluate the accuracy of estimation procedure with covariates, in particular to check if the model is able to

disentangle the effect of covariates from the effect of other species. Finding a way to simulate data with species interactions and covariates effect would allow to better evaluate the inference performance.

4.2 Real data analysis

I analyzed real data with two models, one with no covariate, and one with the time of day included as a covariate. Contrary to what I expected, the inferred interactions remain similar whether or not the time of day is taken into account, which seems to indicate that the time of day has a negligible effect compared to interspecific interactions. This result is surprising, because species have very consistent and strong diel activity patterns in several systems, including in African savannas. Moreover, the inference with the second model suggests an important effect of the time of day on species background rates in these data (as shown in Figure 3.2.4b).

Regarding the inferred interactions, the biological interpretation is subject to caution because not all covariates were taken into account. In particular, these models do not take the environment into account, so the observed patterns could also stem from species' (dis)similarity in habitat preferences (as in chapter 3.1). However, the second model takes the time of day into account so it should remove the confounding effects of species temporal niche convergence/divergence.

Most intraspecific interactions are weaker than interspecific interactions. All species display auto-attractions: kudu has the strongest auto-attraction, then lion, zebra and finally impala. The weak intraspecific interaction for impala is surprising, but can be understood to a certain extent if we consider the high basal rate of impala and the multiplicative nature of the model: even a small increase in the interaction function can lead to many attracted individuals when the background rate is already high.

We observed several herbivore-herbivore interactions: impala follows kudu, wildebeest and zebra; wildebeest follows kudu and zebra and zebra follows kudu and wildebeest. These interactions can be explained by mixed-species grouping (Beaudrot et al., 2020; Fitzgibbon, 1990), especially for short-term attractions. They could also be due to grazing succession (Bell, 1971), especially for delayed attractions. Finally, similar habitat preferences might also influence these patters, in particular for the constant long-term attraction of zebra on wildebeest.

We observed four prey-predator interactions. Impala and zebra avoid lion, which could be the sign of an anti-predator response. These results are consistent with the literature for zebras (Courbin et al., 2016, 2019) and other ungulates (Valeix et al., 2009). Second, lion is weakly attracted by zebra and kudu, which could be explained by predator tracking behavior.

We can also note that many of the inferred interactions are either small variations around the one-line (no interaction), or short-term attractions. As the simulation study showed, the inferred models tend to infer spurious short-term attractions and fluctuations around the true value. In the simulation, these spurious patterns were not significant, contrary to what we see in the real data analysis: but real data might be more challenging for the model, so this calls for caution when interpreting the results.

4.3 Conclusions and perspectives

In this chapter, I described a non-linear MHP with covariates (Carstensen et al., 2010). I tested this model with a simulation approach and used it to analyze real data of species occurrences collected by camera traps.

In this model, covariates affect only the background rate. However, we could also imagine covariates that affect the shape of the interaction function, i.e. an interplay between environmental or temporal covariates and response to interactions. Such patterns have already been highlighted in the literature: for example, Valeix et al. (2009) showed a differentiated response of African herbivores to lion's presence depending on the distance to water and the type of habitat. Clare et al. (2023) also showed that white-tailed deer response to wolves and coyotes varies following the seasons, the avoidance of predators being reduced in winter, because deer cannot avoid predators as freely as they do at other times of the year due to resource scarcity. We could also imagine an interaction between several covariates, as observed for red deer from the Alps, that select sites of lower altitudes, south-facing and steeper slopes in winter, but not in summer (Zweifel-Schielly et al., 2009).

To conclude, this chapter builds on the previous chapter by adding a temporal covariate to the model. However, it remains very exploratory: many conditions were not tested, the inference parameters were chosen empirically (knots spacing, time window), and spatial covariates were not taken into account. The simulation study could

be enriched by testing more conditions (as in the previous chapter). The inference could also be improved by using more stable functions specifications than splines, for instance polynomials or piecewise constant functions. These analyses still constitute a good starting point to refine this model, with the aim to draw more reliable and ecologically relevant inferences.

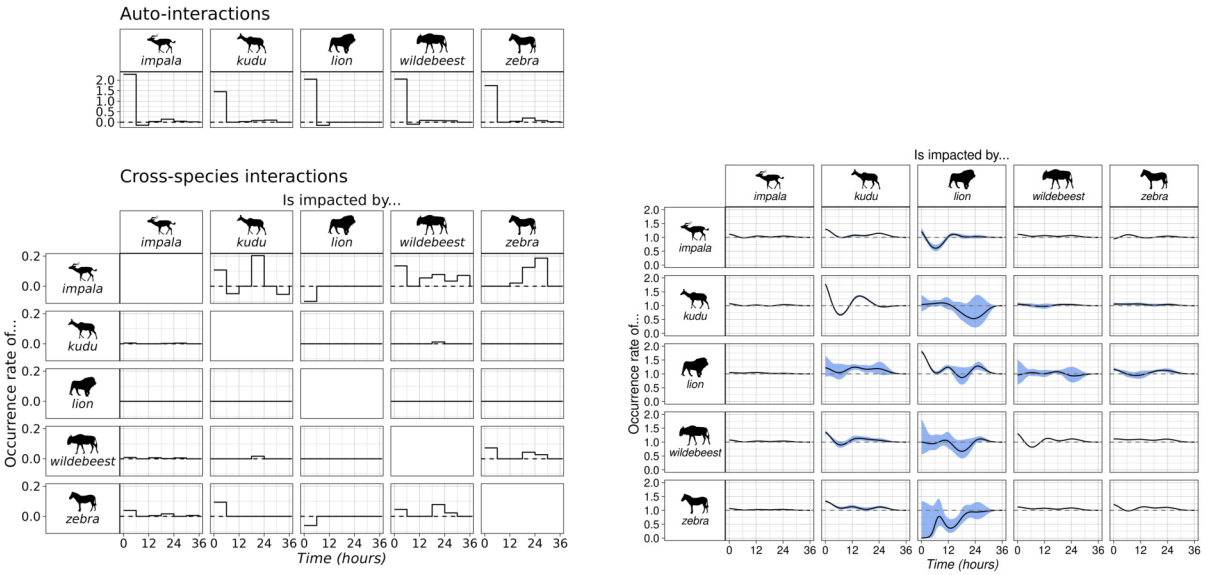
Discussion

In this part, I investigated the patterns of attraction/avoidance between species at short timescales (less than 2 days). I considered data collected with camera traps, and analyzed them with models from the Hawkes process family. In chapter 3.1, we used a linear multivariate Hawkes process (MHP). We first described this model, showed how its capacity to simulate data can be used to evaluate interaction inference methods, and exemplified it on real camera trap data. In chapter 3.2, I considered a non-linear MHP able to include covariates. I described the model, briefly evaluated its performance and illustrated it on the same camera trap dataset, comparing a model with no covariate to a model with a time of day covariate. I chose these two models notably because they had a ready-to-use implementation available (the `UnitEvents` and `ppstat` R packages), which was convenient to quickly evaluate the potential of Hawkes models to infer interactions. Here, I discuss what these two approaches can teach us about interspecific interactions and the ways forward.

1 Compare models outputs

In the two previous chapters, I used a linear and a non-linear MHP to model interactions between 5 species from the African savanna. With the non-linear MHP, I fitted two models, one of which included the time of day as a covariate: these two models yielded similar results, so in the following I only discuss the differences between linear and non-linear models. The figures depicting the interaction functions inferred with the linear and non-linear models (with a covariate) are reproduced in Figure 3.D.1.

The outputs of the inference on real data are rather similar for the linear and non-linear models. However, a striking difference is the lesser strength of auto-attractions in the non-linear MHP. The interaction functions for the two model families do not have the



(a) linear MHP

(b) non-linear MHP

Figure 3.D.1: Comparison of the linear and non-linear MHPs. (a) Linear MHP inferred in chapter 3.1. (b) Non-linear MHP with time of day as a covariate inferred in chapter 3.2. The intensity of species in rows is affected by the species in columns. Silhouette images from PhyloPic by Lukasiniho (wildebeest), Margot Michaud (lion), Robert Hering (kudu), Zimices (zebra) and an unknown author (impala).

same interpretation: for the linear MHP, the values of the interaction functions represent the average number of occurrences gained or suppressed by an occurrence. For the non-linear MHP, it represents the factor by which occurrence rate is affected by an occurrence. Even if we take this into account, the patterns are still different. For instance, if we consider the impala auto-attraction function, with the linear MHP, the strength in the first time step is 2 (so 2/6 = 0.3 day^{-1} for the instantaneous rate). It is about 10 times larger than impala's background rate, so we could expect the non-linear MHP to infer an auto-attraction of this magnitude, but the non-linear MHP instantaneous rate with no delay is 1.1.

If we compare the interspecific interactions, most patterns stay the same. Four additional interactions are inferred with the non-linear MHP: lion tends to follow kudu and zebra and wildebeest follows kudu and avoids lion (only for the model with no covariate). Moreover, some interaction shapes also vary: with the non-linear MHP, impala following interactions are weaker and the avoidance of lion by zebra is less clear.

Regarding background rates, they have a higher value with the non-linear MHP than with the linear MHP (here, we refer to the non-linear MHP without covariates, else the rates are much more difficult to compare). However, the relative values of the

background rates are consistent between models, with impala having the highest rate, followed by zebra, kudu, wildebeest, and lion.

2 From the data to the model

In this part, I aimed at inferring a reactive behavioral response of species to other species in their space use. However, the measure to consider (an effect of the interaction in time, what I called a *reactive temporal interaction*) was not based on biological considerations, but rather guided by the data at hand (camera trap data) and by the Hawkes model specification. We could imagine many other ways to measure the attraction/repulsion signal, and other models have been proposed in the literature (e.g. Karanth et al., 2017; Parsons et al., 2016; Murphy et al., 2021; Galindo-Aguilar et al., 2022).

I chose to model data with a multivariate Hawkes model. However, even within this framework, there is still room for different model specifications. Hawkes processes suppose that there is a cumulative effect of occurrence on intensity, but there are two ways these effects can add up: via an additive approach (linear MHP) or via a multiplicative approach (non-linear MHP) and we have no solid a priori on the best way to specify the model. The best way to check which model is the best would be to compare the fit of these models, but this is difficult due to different model optimization strategies. Here, I can make the cautious hypothesis that an additive effect (linear MHP) is more suited, because multiplicative models tend to produce highly clustered data and species occurrence data are not highly clustered.

The MHP models interactions as arising directly from punctual data, but species are not punctual entities: they move across the landscape in a continuous manner, and only the data collection process is punctual (via discrete camera traps triggers). Therefore, considering punctual data as the observation process underlying continuous species movement patterns, similarly to the multispecies occupancy model of Kellner et al. (2022), could be better suited. This would also make links between interspecific interactions and movement ecology by considering interspecific interactions as a dynamic continuous process, occurring in space and time.

In this part, I validated the MHPs with simulations but used the same model for

simulation and inference. This might be unrealistically easy for the model, and also confounds the ecological process we want to simulate (species interactions) with the sampling process (camera trap pictures) (Zurell et al., 2010). This choice was motivated by the fact that I could not find another model to simply generate punctual data with interactions between different species, except by designing an individual-based model, which was outside the scope of this thesis.

The modeling strategy used here is part of a larger movement of recent papers using continuous-time approaches to model camera trap data (Cusack et al., 2017; Karanth et al., 2017; Parsons et al., 2016; Galindo-Aguilar et al., 2022), some of them through point processes (Keim et al., 2019; Schliep et al., 2018; Kellner et al., 2022). But during the analysis, I had doubts on whether this was a good strategy. The relevance of continuous-time processes has recently been investigated in the literature for occupancy models by Pautrel et al. (2024) who concluded that continuous-time models provided little additional benefits. Point processes are complex models, requiring the fitting of more parameters than other currently used approaches like permutations (Murphy et al., 2021) or GLMs on delays (Parsons et al., 2016). Moreover, the Hawkes model used in chapter 3.1 uses a penalization parameter and not classical confidence intervals, which can make the results more difficult to interpret. I think that the real advantage of point process models such as the MHPs resides in the interpretability of the output and the formal model specification. Model parameters are easier to interpret in terms of effect sizes than outputs of simple permutation tests: for instance, the interaction functions can be interpreted as mean occurrence count gained or suppressed, whereas the mean time to event is difficult to interpreted biologically. Another advantage is the fact that MHPs are causal model that theoretically allow to explicitly model the effect of covariates and to disentangle the effect of multiple species.

3 Measuring interaction strength

Quantifying species interactions strength is difficult. It depends on how interactions are defined, but even if we consider that interactions are given by the effect of one species on the abundance of the other (typically measured by consumption rates in dynamic models), they are hard to quantify (Wootton & Emmerson, 2005). For finer processes

based on behavioral response, which are even harder to evaluate than population dynamics, we can think that it is even more difficult.

The strength of interspecific interactions as measured with the MHP, i.e. as a number of occurrences gained or suppressed by a species, is not often quantified in the literature. I had to proceed by trial and error to define model parameters for simulation and for inference, because I did not have a clear a priori idea of the magnitude of interactions or of potential confounding factors such as circadian rhythms or habitat preferences. In particular, it was unclear if confounding factors could completely obscure the signal of interactions and hamper reliable inferences of interspecies interactions. The camera trap literature about interspecific interactions is very focused on the spatial effect of interactions (via occupancy models) or on their temporal effect (via overlap measures), and even when some studies focus on finer scale spatio-temporal interactions, most of them do not provide an interpretable estimate of interaction strength and the outcomes of the different studies are also difficult to compare due to the variety of methods and study systems.

Overall, there are still few studies that investigate short-term interactions as the ones studied here. I think that this is not due to a nonexistent or uninteresting biological question, but that such fine-scale patterns are simply more difficult to highlight and to measure. Large and fine-scale datasets, collected for instance by camera trap data, acoustic monitoring devices or GPS collars, are relatively recent and still rare. In parallel, since these data have just emerged, the methods to analyze them are lacking or not widespread among the scientific community.

4 Including more covariates?

As noted before, species occurrence patterns are influenced by a multitude of environmental and temporal factors. In chapter 3.1, we only looked at the effect of other species without correcting for these confounding factors. In chapter 3.2, I used a model allowing to include covariates and fitted a model with a temporal covariate. Arguably, many factors are still missing from these analyses, like the seasonal variability or spatial heterogeneity due to roads, water points or vegetation.

Correcting for the time of day is a rather straightforward procedure, because circa-

dian rhythms are closely related to the day-night cycle. However, taking into account environmental variability would be much more challenging. Indeed, a multiplicity of features account for environmental variability, and taking them all into account would be difficult because of data collection, model power and interpretability.

When I started this analysis, I was convinced that refining the model with the more covariates possible was the way forward: now I am not so sure that this is the final end goal, and rather think that the most important is to select covariates that are important to take into account so that the signal-to-noise ratio is good enough (Levin, 1992).

Bibliography

Aguilera, A. G., Alpert, P., Dukes, J. S., & Harrington, R. (2010). Impacts of the invasive plant Fallopia japonica (Houtt.) on plant communities and ecosystem processes. *Biological Invasions*, *12*(5), 1243–1252. https://doi.org/10.1007/s10530-009-9543-z

Albert, M., Bouret, Y., Chevallier, J., Fromont, M., Grammont, F., Laloe, T., Mascart, C., Reynaud-Bouret, P., Rouis, A., Scarella, G., & Tuleau-Malot, C. (2021). UnitEvents: Unitary Events Method with Delayed Coincidence Count (MTGAUE or Permutation Method) and Bernstein Lasso method for Hawkes processes. https://sourcesup.renater.fr/frs/?group_id=3267

Beaudrot, L., Palmer, M. S., Anderson, T. M., & Packer, C. (2020). Mixed-species groups of Serengeti grazers: A test of the stress gradient hypothesis. *Ecology*, *101*(11), e03163. https://doi.org/10.1002/ecy.3163

Bell, R. H. (1971). A grazing ecosystem in the Serengeti. *Scientific American*, *225*(1), 86–93.

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Carstensen, L., Sandelin, A., Winther, O., & Hansen, N. R. (2010). Multivariate Hawkes process models of the occurrence of regulatory elements. *BMC Bioinformatics*, *11*(1), 456. https://doi.org/10.1186/1471-2105-11-456

Cavieres, L. A., & Sierra-Almeida, A. (2018). Assessing the importance of cold-stratification for seed germination in alpine plant species of the High-Andes of central Chile. *Perspectives in Plant Ecology, Evolution and Systematics*, *30*, 125–131. https://doi.org/10.1016/j.ppees.2017.09.005

Clare, J. D. J., Zuckerberg, B., Liu, N., Stenglein, J. L., Van Deelen, T. R., Pauli, J. N., & Townsend, P. A. (2023). A phenology of fear: Investigating scale and seasonality in predator–prey games between wolves and white-tailed deer. *Ecology*, *104*(5), e4019. https://doi.org/10.1002/ecy.4019

Courbin, N., Loveridge, A. J., Fritz, H., Macdonald, D. W., Patin, R., Valeix, M., & Chamaillé-Jammes, S. (2019). Zebra diel migrations reduce encounter risk with lions at night (J. Fryxell, Ed.). *Journal of Animal Ecology*, *88*(1), 92–101. https://doi.org/10.1111/1365-2656.12910

Courbin, N., Loveridge, A. J., Macdonald, D. W., Fritz, H., Valeix, M., Makuwe, E. T., & Chamaillé-Jammes, S. (2016). Reactive responses of zebras to lion encounters shape their predator-prey space game at large scale. *Oikos*, *125*(6), 829–838. https://doi.org/10.1111/oik.02555

Cusack, J. J., Dickman, A. J., Kalyahe, M., Rowcliffe, J. M., Carbone, C., MacDonald, D. W., & Coulson, T. (2017). Revealing kleptoparasitic and predatory tendencies in an African mammal community using camera traps: A comparison of spatiotemporal approaches. *Oikos*, *126*(6), 812–822. https://doi.org/https://doi.org/10.1111/oik.03403

Daley, D. J., & Vere-Jones, D. (2003). *An introduction to the Theory of Point Processes* (2nd, Vol. 1). Springer.

Dibner, C., Schibler, U., & Albrecht, U. (2010). The Mammalian Circadian Timing System: Organization and Coordination of Central and Peripheral Clocks. *Annual Review of Physiology*, *72*(1), 517–549. https://doi.org/10.1146/annurev-physiol-021909-135821

Eilers, P. H. C., & Marx, B. D. (2010). Splines, knots, and penalties. *WIREs Computational Statistics*, *2*(6), 637–653. https://doi.org/10.1002/wics.125

Estes, R. D. (2012). *The behavior guide to African mammals: Including hoofed mammals, carnivores, primates*. The University of California Press.

Fallahi, M., & Pourtaheri, R. (2024). Estimating high-dimensional Hawkes process with time-dependent covariates. *Communications in Statistics - Simulation and Computation*, *0*(0), 1–16. https://doi.org/10.1080/03610918.2024.2342916

Fisher, J. T., Burton, A. C., Nolan, L., & Roy, L. (2020). Influences of landscape change and winter severity on invasive ungulate persistence in the Nearctic boreal forest. *Scientific Reports*, *10*(1), 8742. https://doi.org/10.1038/s41598-020-65385-3

Fitzgibbon, C. D. (1990). Mixed-species grouping in Thomson's and Grant's gazelles: The antipredator benefits. *Animal Behaviour*, *39*(6), 1116–1126. https://doi.org/10.1016/S0003-3472(05)80784-5

Frey, S., Fisher, J. T., Burton, A. C., & Volpe, J. P. (2017). Investigating animal activity patterns and temporal niche partitioning using camera-trap data: Challenges and opportunities (M. Rowcliffe, Ed.). *Remote Sensing in Ecology and Conservation*, *3*(3), 123–132. https://doi.org/10.1002/rse2.60

Galindo-Aguilar, R. E., Luna-Olivera, B. C., Ramírez-Ibáñez, M., & Lavariega, M. C. (2022). Spatiotemporal co-occurrence of predators and prey in a neotropical mammal community in southern Mexico. *Journal of Tropical Ecology*, *38*(5), 285–294. https://doi.org/10.1017/S0266467422000189

Gompper, M. E., Lesmeister, D. B., Ray, J. C., Malcolm, J. R., & Kays, R. (2016). Differential Habitat Use or Intraguild Interactions: What Structures a Carnivore Community? (A. W. Reed, Ed.). *PLoS ONE*, *11*(1), e0146055. https://doi.org/10.1371/journal.pone.0146055

Gravel, D., Canham, C. D., Beaudet, M., & Messier, C. (2006). Reconciling niche and neutrality: The continuum hypothesis. *Ecology Letters*, *9*(4), 399–409. https://doi.org/10.1111/j.1461-0248.2006.00884.x

Hansen, N. R. (2023). Ppstat: Point Process Statistics. https://github.com/nielsrhansen/ppstat

Haren, H. v., & Compton, T. J. (2013). Diel Vertical Migration in Deep Sea Plankton Is Finely Tuned to Latitudinal and Seasonal Day Length. *PLOS ONE*, *8*(5), e64435. https://doi.org/10.1371/journal.pone.0064435

Hubbell, S. P. (1997). A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs*, *16*(1), S9–S21. https://doi.org/10.1007/s003380050237

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.

Janzen, D. H. (1979). How to be a fig. *Annual Review of Ecology and Systematics*, *10*, 13–51. https://doi.org/10.1146/annurev.es.10.110179.000305

Karanth, K. U., Srivathsa, A., Vasudev, D., Puri, M., Parameshwaran, R., & Kumar, N. S. (2017). Spatio-temporal interactions facilitate large carnivore sympatry across a resource gradient. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1848), 20161860. https://doi.org/10.1098/rspb.2016.1860

Keim, J. L., Lele, S. R., DeWitt, P. D., Fitzpatrick, J. J., & Jenni, N. S. (2019). Estimating the intensity of use by interacting predators and prey using camera traps. *Journal of Animal Ecology*, *88*(5), 690–701. https://doi.org/https://doi.org/10.1111/1365-2656.12960

Kellner, K. F., Parsons, A. W., Kays, R., Millspaugh, J. J., & Rota, C. T. (2022). A Two-Species Occupancy Model with a Continuous-Time Detection Process Reveals Spatial and Temporal Interactions. *Journal of Agricultural, Biological and Environmental Statistics*, *27*(2), 321–338. https://doi.org/10.1007/s13253-021-00482-y

Kroc, E., & Olvera Astivia, O. L. (2023). The case for the curve: Parametric regression with second- and third-order polynomial functions of predictors should be routine. *Psychological Methods*. https://doi.org/10.1037/met0000629

Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture. *Ecology*, *73*(6), 1943–1967. https://doi.org/10.2307/1941447

Mohler, G., Carter, J., & Raje, R. (2018). Improving social harm indices with a modulated Hawkes process. *International Journal of Forecasting*, *34*(3), 431–439. https://doi.org/10.1016/j.ijforecast.2018.01.006

Murphy, A., Diefenbach, D. R., Ternent, M., Lovallo, M., & Miller, D. (2021). Threading the needle: How humans influence predator–prey spatiotemporal interactions in a multiple-predator system. *Journal of Animal Ecology*, *90*(10), 2377–2390. https://doi.org/10.1111/1365-2656.13548

Nicvert, L., Donnet, S., Keith, M., Peel, M., Somers, M. J., Swanepoel, L. H., Venter, J., Fritz, H., & Dray, S. (2024). Using the multivariate Hawkes process to study interactions between multiple species from camera trap data. *Ecology*, *105*(4), e4237. https://doi.org/10.1002/ecy.4237

Ogata, Y. (1998). Space-Time Point-Process Models for Earthquake Occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402. https://doi.org/10.1023/A:1003403601725

Palmer, M. S., Gaynor, K. M., Becker, J. A., Abraham, J. O., Mumma, M. A., & Pringle, R. M. (2022). Dynamic landscapes of fear: Understanding spatiotemporal risk. *Trends in Ecology & Evolution*, *37*(10), 911–925. https://doi.org/10.1016/j.tree.2022.06.007

Pardo, L. E., Bombaci, S. P., Huebner, S., Somers, M. J., Fritz, H., Downs, C., Guthmann, A., Hetem, R. S., Keith, M., le Roux, A., Mgqatsa, N., Packer, C., Palmer, M. S., Parker, D. M., Peel, M., Slotow, R., Strauss, W. M., Swanepoel, L., Tambling, C., … Venter, J. A. (2021). Snapshot Safari: A large-scale collaborative to monitor Africa's remarkable biodiversity. *South African Journal of Science*, *117*(1/2), 1–4. https://doi.org/10.17159/sajs.2021/8134

Parsons, A. W., Bland, C., Forrester, T., Baker-Whatton, M. C., Schuttler, S. G., McShea, W. J., Costello, R., & Kays, R. (2016). The ecological impact of humans and dogs on wildlife in protected areas in eastern North America. *Biological Conservation*, *203*, 75–88. https://doi.org/10.1016/j.biocon.2016.09.001

Pautrel, L., Moulherat, S., Gimenez, O., & Etienne, M.-P. (2024). Analysing biodiversity observation data collected in continuous time: Should we use discrete- or continuous-time occupancy models? *Methods in Ecology and Evolution*, *15*(5). https://doi.org/10.1111/2041-210X.14314

Perperoglou, A., Sauerbrei, W., Abrahamowicz, M., & Schmid, M. (2019). A review of spline function procedures in R. *BMC Medical Research Methodology*, *19*(1), 46. https://doi.org/10.1186/s12874-019-0666-3

Reinhart, A. (2018). A Review of Self-Exciting Spatio-Temporal Point Processes and Their Applications. *Statistical Science*, *33*(3), 299–318. https://doi.org/10.1214/17-STS629

Rowcliffe, M. (2023). Activity: Animal Activity Statistics. https://CRAN.R-project.org/package=activity

Say-Sallaz, E., Chamaillé-Jammes, S., Fritz, H., & Valeix, M. (2019). Non-consumptive effects of predation in large terrestrial mammals: Mapping our knowledge and revealing the tip of the iceberg. *Biological Conservation*, *235*(October 2018), 36–52. https://doi.org/10.1016/j.biocon.2019.03.044

Schliep, E. M., Gelfand, A. E., Clark, J. S., & Kays, R. (2018). Joint Temporal Point Pattern Models for Proximate Species Occurrence in a Fixed Area Using Camera Trap Data. *Journal of Agricultural, Biological and Environmental Statistics*, *23*(3), 334–357. https://doi.org/10.1007/s13253-018-0327-8

Stoll, P., Gatzsch, K., Rusterholz, H.-P., & Baur, B. (2012). Response of plant and gastropod species to knotweed invasion. *Basic and Applied Ecology*, *13*(3), 232–240. https://doi.org/10.1016/j.baae.2012.03.004

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate Effects. *Journal of Neurophysiology*, *93*(2), 1074–1089. https://doi.org/10.1152/jn.00697.2004

Valeix, M., Loveridge, A. J., Chamaillé-Jammes, S., Davidson, Z., Murindagomo, F., Fritz, H., & Macdonald, D. W. (2009). Behavioral adjustments of African herbivores to predation risk by lions: Spatiotemporal variations influence habitat use. *Ecology*, *90*(1), 23–30. https://doi.org/10.1890/08-0606.1

Weiblen, G. D. (2002). How to be a Fig Wasp. *Annual Review of Entomology*, *47*(1), 299–330. https://doi.org/10.1146/annurev.ento.47.091201.145213

Weiher, E., Freund, D., Bunton, T., Stefanski, A., Lee, T., & Bentivenga, S. (2011). Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1576), 2403–2413. https://doi.org/10.1098/rstb.2011.0056

Whittington-Jones, G., Bernard, R., & Parker, D. (2011). Aardvark burrows: A potential resource for animals in arid and semi-arid environments. *African Zoology*, *46*(2), 362–370. https://doi.org/10.1080/15627020.2011.11407509

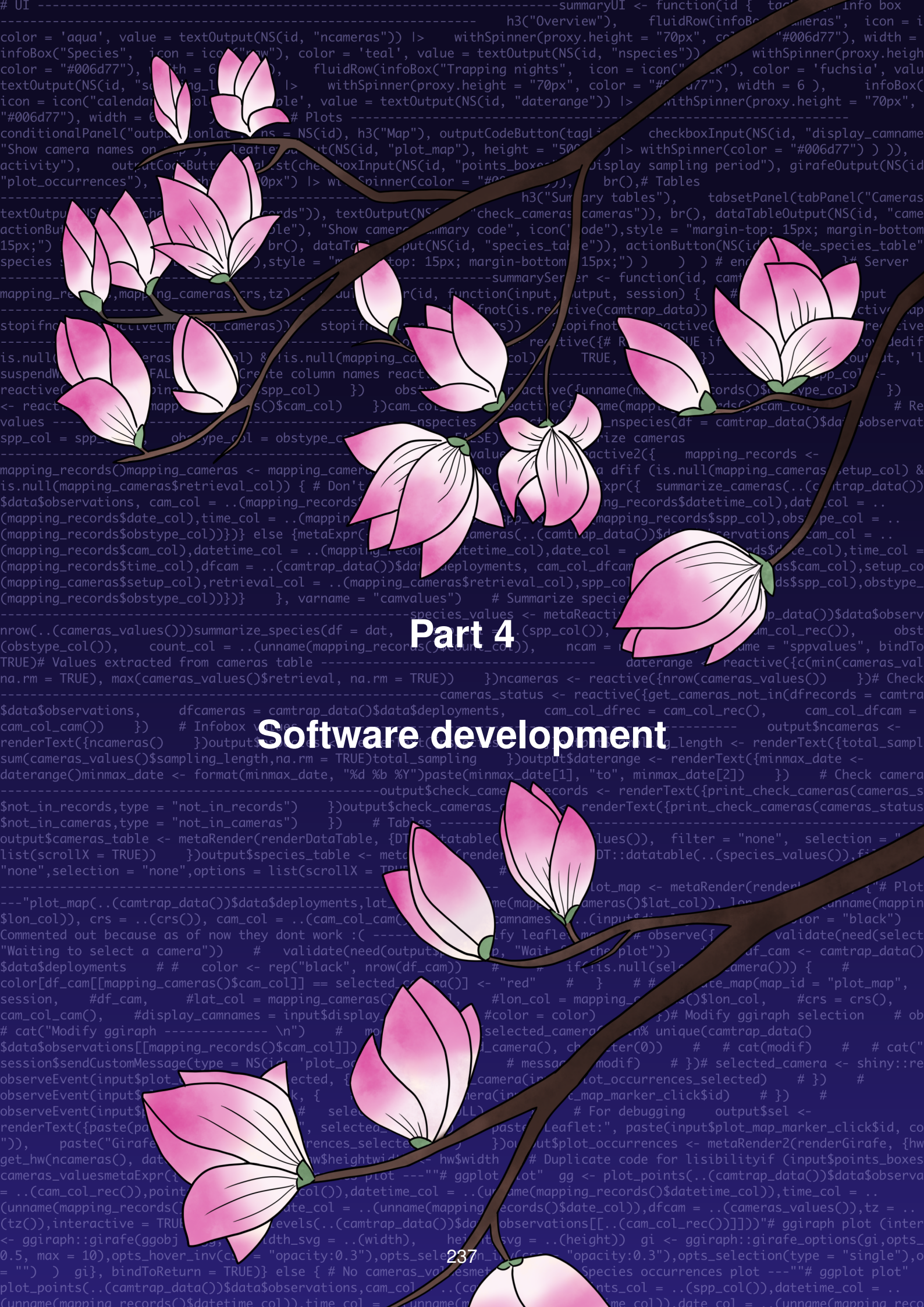
Wootton, J. T., & Emmerson, M. (2005). Measurement of Interaction Strength in Nature. *Annual Review of Ecology, Evolution, and Systematics*, *36*(1), 419–444. https://doi.org/10.1146/annurev.ecolsys.36.091704.175535

Xu, H., Luo, D., Chen, X., & Carin, L. (2018). Benefits from Superposed Hawkes Processes. *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, 623–631. Retrieved May 10, 2024, from https://proceedings.mlr.press/v84/xu18c.html

Yoder, A. D., Burns, M. M., Zehr, S., Delefosse, T., Veron, G., Goodman, S. M., & Flynn, J. J. (2003). Single origin of Malagasy Carnivora from an African ancestor. *Nature*, *421*(6924), 734–737. https://doi.org/10.1038/nature01303

Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., & Grimm, V. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, *119*(4), 622–635. https://doi.org/10.1111/j.1600-0706.2009.18284.x

Zweifel-Schielly, B., Kreuzer, M., Ewald, K. C., & Suter, W. (2009). Habitat selection by an Alpine ungulate: The significance of forage characteristics varies with scale and season. *Ecography*, *32*(1), 103–113. https://doi.org/10.1111/j.1600-0587.2008.05178.x



Part 4

Software development

Cover picture: this drawing represents magnolia flowers with R code from the `camtrapviz` application is written in the background.

© Own drawing.

Introduction

1 Preamble: PhD context

This last part is very important to me, because it embodies an aspect of science that I find crucial: harnessing scientific knowledge to act in the real world in a context of global biodiversity and climate crisis. Even though I am passionate about ecology and better understanding communities from a fundamental perspective, I am also aware of the emergency caused by global changes that impose rapid action. In this context, I feel that creating bridges between fundamental research and applied tools is very important.

We initially hesitated to include these results as a standalone of the manuscript because of its very applied nature. However, given the time needed to develop such tools and their scientific relevance, in my opinion they deserve to be included in the manuscript.

This research axis emerged from discussions with the team of people collecting and using camera trap data for the Snapshot Safari project. When I was in South Africa (in Summer 2022), I had the chance to take part in a workshop of the Snapshot Safari project. During discussions, we realized that there was a need for user-friendly tools to clean and visualize camera trap data, both for researchers and for conservationists and reserve managers. Moreover, I like coding and developing software tools, so I found it very enjoyable to work on this part.

2 Introduction

Camera traps are still relatively recent. Although the first camera setup with automatic trigger to detect wildlife dates back to the 1890s (O'Connell et al., 2011), it is only since the 2000s-2010s that computing facilities and decreasing cost made camera traps a popular and accessible survey tool (O'Connell et al., 2011). The only R package specifically developed for camera trap data that is widely used, `camtrapR` (Niedballa et al., 2016), is relatively recent, and no doubt other helping packages will follow (e.g. Bubnicki et al., 2024).

Camera trap data comes with specific methodological needs in the domains of sampling design, electronics and sensors and modeling for data analysis. In particular, data processing needs are relatively more important for camera traps than for traditional methods to survey communities. Indeed, the data collected with camera traps are images, which are large files, and demand specific processing. Specifically, camera trap data are often used to extract species occurrences, and identifying species from pictures is an important and time-consuming task. Machine learning models have been used to fulfill this task (e.g. Chen et al., 2014; Tabak et al., 2019; Norouzzadeh et al., 2018). Traceability and reproducibility are also important stakes with camera trap data, because the raw camera files need to be archived and properly linked to the final tabular data with species annotation. More generally, we need standardized and automated procedures to reliably analyze large amounts of data: in particular, for biodiversity data, we need reproducible workflows allowing to select and filter data, as pointed out by Ronquillo et al. (2024). Finally, there is also a need for data standardization for sharing camera trap data (Bubnicki et al., 2024).

Moreover, camera trap data have the potential to interest a wide range of researchers, applied scientists, reserve managers, and more largely anyone interested in collecting camera trap data. Therefore, there is a double challenge regarding, on the one hand, reproducibility of the data processing workflow, and on the other hand, accessibility of this workflow to allow people without a strong background in computer science to use these tools.

In this chapter, I develop two software tools to process camera trap data in R. In chapter 4.1, I develop a R package, `standardizeSnapshot` (https://github.com/SnapshotSafari/standard-merge), to process camera trap data, specifically designed for data

collected by the Snapshot Safari project. In chapter 4.2, I develop a more generalist R Shiny application, `camtrapviz` (https://lbbe-shiny.univ-lyon1.fr/camtrapviz/), to select, filter and visualize camera trap data, with a special attention to making the interactive analyses reproducible.

Chapter 4.1

standardizeSnapshot, a camera trap data cleaning tool

1 Introduction

Snapshot Safari is camera trap data collection project in Southern and Eastern Africa (see Figure 4.1.1) (Pardo et al., 2021). This camera trap data collection effort encompasses camera trap grids with a standardized design in more than 30 sites across 6 countries (Tanzania, Botswana, Kenya, Mozambique, South Africa, and Zimbabwe). Each camera trap grid comprises 8 to 245 camera traps, and most have been operating since 2018. Snapshot Safari is a long-term monitoring project designed to collect species occurrence data over a long period to gather information on biodiversity trends, and the cameras are intended to operate for around 10 years or longer.

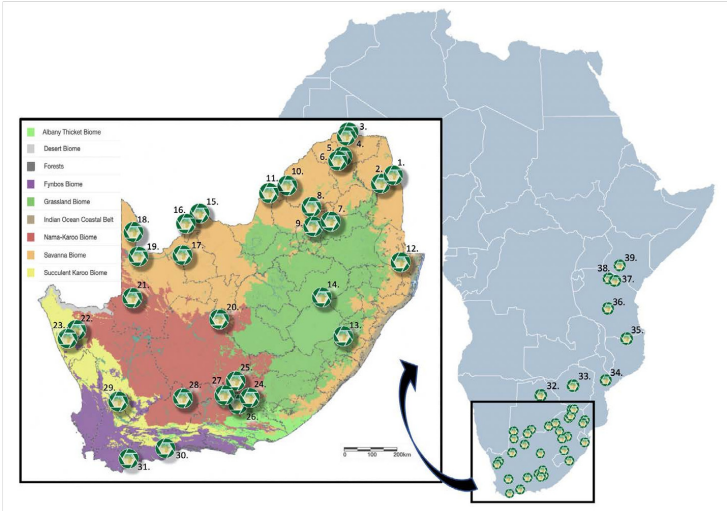


Figure 4.1.1: Location of the Snapshot Safari camera trap grids. Figure 1 from Pardo et al. (2021).

The amount of data collected with these cameras is very large and cannot be handled without automated data processing and cleaning procedures. To annotate species and/or behaviors from camera trap pictures, 3 methods have been used. Some pictures were processed using the Zooniverse platform (https://www.zooniverse.org/organizations/meredithspalmer/snapshot-safari), thanks to volunteer people from the public. Other pictures were annotated manually by researchers and graduate students using the Digikam software. Finally, an automated machine learning tool, TrapTagger, is increasingly used to annotate camera trap data collected with the Snapshot Safari project. These three tools produce different tabulated outputs, and do not always contain the same information. For instance, TrapTagger does not annotate species' behavior, so there are no associated columns, whereas Zooniverse and Digikam outputs contain behavior data (but not encoded in the same way).

Moreover, some of the encoded information follow to different standards, because the different projects were managed by different people. For instance, cameras identifiers sometimes differ between batches of data collection (e.g. MAD_A01 or A01); the same species could be identified with different codes (e.g., “batearedfox” or “fox-bateared” for the bat-eared fox), or the date and time encoded in different ways (e.g. 2020-01-13 and 01/13/2020).

Before designing this package, some code had already been written to clean and homogenize data, but it was not tested on all data and not distributed in an executable code bundle. Initially, there was no need for more, as people concentrated on analyses of small parts of the data that did not have discrepancies (or few discrepancies). But as more data was being collected, people began addressing questions requiring data from many sites: and using large parts of the dataset was hampered by these different data standards.

To overcome these issues, Lain Pardo, Sarah Huebner (postdocs working on the Snapshot Safari project) and me first defined a data standard to homogenize all formats. Then, I implemented a R package, `standardizeSnapshot`, with functions to automate data cleaning and conform data to the new data standard (available at: https://github.com/SnapshotSafari/standard-merge).

2 The data standard

To define the data standard, I exchanged with two postdocs working on the Snapshot Safari project, Lain Pardo and Sarah Huebner. Thanks to their knowledge on the data processing workflow and the specifics of naming conventions, together we could define rules to automate data standardization. To define the data standard, we used the recommendations from GBIF for camera traps column name (Reyserhove et al., 2023) and used a consistent camel case naming scheme for columns.

Standard	Nature of the data	Digikam	TrapTagger	Zooniverse
locationID	Camera trap grid site ID			
cameraID	Camera trap ID	Station	Cam.Site	site
season	Zooniverse batch			season
roll	ID for the camera trap period of activity between two services			roll
eventID	Capture event ID	Capture_ID		capture_id
snapshotName	Species	Species	capture_labels	question__species
eventDate	Capture event date	Date	date	capture_date_local
eventTime	Capture event time	Time	time	capture_time_local
filePath1	Path to picture file 1	file_path_1	capture_url	zooniverse_url_0
filePath2	Path to picture file 2			zooniverse_url_1
filePath3	Path to picture file 3			zooniverse_url_2
capture	Capture rank (by roll)			capture
captureID	Capture event ID	id		subject_id
countMax	Species maximum count			question__count_max
countMedian	Species count (or median count for Zooniverse)	metadata_Number	capture_sighting_count	question__count_median
countMin	Species minimum count			question__count_min
standing	Species behavior	Standing		question__standing
resting		Resting		question__resting
moving		Moving		question__moving
eating		Eating		question__eating
interacting		Interacting		question__interacting
drinking		Drinking		question__drinking
youngPresent	Young individuals	metadata_young_present		question__young_present
hornsVisible	Visible horns			question__horns_visible
consensusSpecies	Proportion of users that identified the species			p_users_identified_this_species
PielouEvennessIndex	Pielou Evenness index on species identification			pielou_evenness_index
classifier	Image classifier			

Table 4.1.1: Standard column names. The first column gives standard column names, the second column describes the nature of the data stored in the column and other columns give column names obtained with each data processing tool. Blank cells indicate that the column does not exist in the corresponding data format. Columns with no correspondence in the new standard are not shown.

We defined this standard in the best way possible to not lose information in the final table. Because of that, some columns of the standard are not relevant to all data processing methods and may contain NAs. For instance, the new standard includes a column named `consensusSpecies` that indicates the proportion of users that identified this species, because on Zooniverse, a single picture was presented to several different users to be tagged. This column is relevant only for Zooniverse data, because for other processing tools, a single user or model identified the species. The correspondence between old and new column names is provided as an illustration in Table 4.1.1.

3 Main functionalities

The final R package provides functions to build a data standardization workflow, from reading raw files to writing files (Figure 4.1.2). The more important functions allow to standardize data: `standardize_snapshot_df` and `standardize_snapshot_list`, that allow to standardize either a dataframe object or that standardize a list of dataframes. The function `read_snapshot_files` allows to read files in a folder directly into a list of dataframes. This function works recursively, supports CSV and Excel files, and allows to specify files that should be ignored inside the folder. Finally, the functions `write_standardized_df` and `write_standardized_list` write the file(s) to a new location. These functions are essentially a wrapper around the R function `write.csv` to write dataframes objects to CSV files, but implements an automated naming scheme defined using the tables' column values, using the format "locationID_Sseason_Rroll.csv".

The standardization functions perform several cleaning operations at once: they homogenize dates and times format, species names and camera IDs, and standardize the column names to conform to the defined standard. These cleaning steps can be reproduced individually with package functions (that are combined inside the `standardize_snapshot*` functions) to allow a greater flexibility in the analyses.

In parallel, a function `create_logger` allows to create an optional logger object to redirect all R outputs, including functions messages, warnings and errors from the package functions to a file for traceability, thanks to the R package `log4r` (Myles White et al., 2022). The main functions of the package (described above and in Figure 4.1.2)

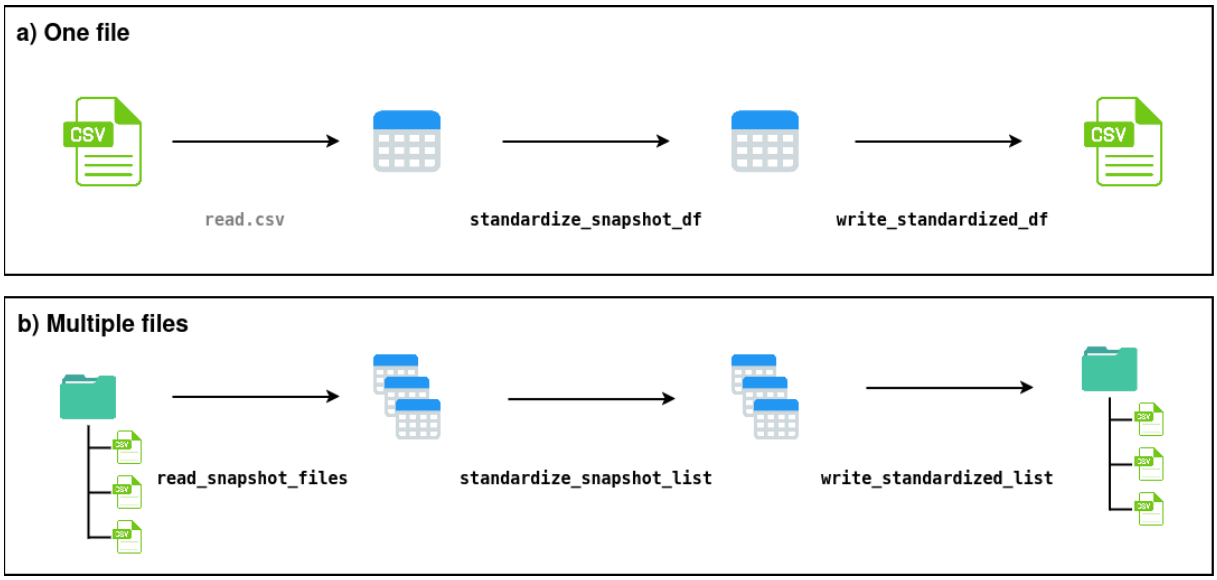


Figure 4.1.2: Workflow of `standardizeSnapshot`. a) Typical workflow for one file: the file is read into a dataframe (typically with the `read.csv` function). This dataframe is standardized and written to a CSV file. b) To handle multiple files at once, one can provide a path to a folder containing CSV files. The files are read to a list of dataframes, that are standardized and written to files. Icons: CSV by Awicon, folder by DinosoftlLabs and table by Pixel perfect from Flaticon.

have an optional `logger` argument allowing to specify a logger object used to redirect outputs to a file.

4 Software development

While developing the code, I made an effort to follow coding best practices, by testing and documenting the code, and using a version control system along the project development.

To validate the functions, I wrote unit tests for all functions developed in the package. Unit testing consists in formally testing individual components of a program to ensure the written code generates the expected results. Although all package development includes some degree of testing, these tests are usually informal, not exhaustive and deleted upon checking (Wickham & Bryan, 2023). By contrast, unit testing is a code development strategy that formally tests the written code and keeps the written tests available to re-run at any time, for example to ensure that adding a new functionality does not break existing code. To automate testing, the programmer provides code to test and the expected result and the result obtained by running the code is compared to the expected result. To write unit tests, I used the R package `testthat` (Wickham,

2011), which provides helper functions to test code and automate the process of running unit tests.

Unit testing has consequences on the code reliability and programming style. Because unit testing encourages to formally think about code testing, the tests tend to be more exhaustive, and edge cases can be tested to ensure the codes does not produce an unexpected behavior. Running the tests is automated, thus reducing the probability that new features or updated package dependency will break existing code. Finally, unit testing naturally encourages to write code such that broad functionalities are broken down in smaller steps to be tested separately. Since R is a functional programming oriented language (i.e. a language built around functions), the natural unit on which tests are performed is the function. Therefore, unit testing naturally encourages to write a function for each programming task. This encourages to follow a "DRY" (Do Not Repeat Yourself) coding strategy.

I wrote the code of these functions to make results as reliable as possible. Along with unit testing of functions, I also provided helper functions allowing to write a log file. This file stores the outputs of the functions printed to the R console, by redirecting these outputs to a text file where outputs can be kept and checked after code execution. Additionally, the main workflow function have a `verbose` argument allowing to control the amount of output displayed to the console.

I also documented the package by writing documentation for each function and exhaustive comments in the code. Along with the documentation, I provided examples and wrote two vignettes to demonstrate a typical workflow using the package. I also created data objects exported with the package and used in the examples and vignettes. The package installation procedure, functions documentation and vignettes are available on a website: https://snapshotsafari.github.io/standard-merge/ (created with `pkgdown` Wickham et al., 2024).

Along the project development, I used the version control system Git to keep a traceability of the package's advancement. I also unit tests and documentation writing along with the project development.

The package is available in open access and hosted on GitHub by the Snapshot Safari Organization (https://github.com/SnapshotSafari/standard-merge).

5 Conclusion and perspectives

I wrote a R package, `standardizeSnapshot`, to automate the standardization and data cleaning of the tabular data obtained after processing camera trap images. This package is available in open access, tested and documented.

Once the package was finished, I used it to standardize all available Snapshot Safari data. This package is now routinely used by researchers from the Snapshot Safari project to standardize camera trap data tables: at the moment, 15 active Snapshot subprojects (sites) using TrapTagger and Zooniverse as classifiers are using this package. The feedback so far regarding the package functionalities and usage has been positive.

The data standardization step addressed with this package is only one part of the global data collection and management workflow, from sampling design to data analysis by way of image processing. Within the Snapshot Safari project, many parts of this data management process constitute a logistic, programming or data storage challenge. Although I think that `standardizeSnapshot` constitutes a valuable tool inside the data management workflow, it could be better integrated with the other steps of data management, for instance by running the data standardization workflow automatically on processed data. Moreover, other parts of the data management process could be benefit from automation or standardization of some steps, but this is a real challenge considering the project scale.

Chapter 4.2

camtrapviz, a Shiny app to visualize camera trap data

1 Introduction

Camera trapping is an increasingly popular method to collect biodiversity data. It has many advantages, among which its automated nature, its non-invasiveness or its ability to record rare or elusive species (Burton et al., 2015). In the last 10 years, camera trapping has become a mainstream ecological data collection method (Delisle et al., 2021) and has prompted many methodological developments. Many models have been developed to analyze camera trap data, like the multispecies occupancy model of Rota et al. (2016) or time-continuous models like tomcat (Azzou et al., 2021); picture classification tools have also been proposed (e.g. Chen et al., 2014), and studies have evaluated sampling design (Kays et al., 2021; Fonteyn et al., 2021).

Despite these numerous developments, free and accessible tools to curate and explore camera trap data remain scarce. Yet camera traps can easily generate a large amount of data, which often includes spatial and temporal information collected for multiple species: this quantity and complexity calls for automated and transparent methods to handle data (Niedballa et al., 2016; MacFadyen et al., 2022). One of the main tools to process camera trap data is the R package `camtrapR`, dedicated to camera trap data management (Niedballa et al., 2016). This package is mainly focused on the conversion of annotated pictures to tabular data, with additional functionalities to format data or filter for temporal independence, and some built-in functions to visualize data. Another R package for camera trap data is `camtraptor` (Oldoni et al., 2024, still in development). This package allows to import and handle data in Camera Trap Data Package format (Bubnicki et al., 2024) in R.

One of the characteristics of camera trap data is that they can be useful for people with various backgrounds, in particular applied ecologists and reserve managers who generally have little to no training in computer science. Interactive data analysis tools are therefore particularly relevant for camera trap data analysis. More generally, quick and interactive exploration of camera trap data can be useful to anyone, even ecologists who are familiar with coding.

In this chapter, I present `camtrapviz`, a Shiny application and R package to analyze and visualize camera trap data. `camtrapviz` provides functions to filter, summarize and visualize camera trap data. It is also available as an interactive application allowing to perform these analyses interactively. This software is intended to visualize tabulated

camera trap data (once species have been annotated) primarily for exploratory purposes.

2 Preliminary work

Before developing this application, I exchanged with different actors from the camera trapping community to gather their opinion on the usefulness of this application and on useful features to include. I talked with people from the Snapshot Safari project, exchanged with PhD students that I knew who worked on camera trap data and contacted researchers from the camera trap community by email (Jürgen Niedballa, developer of the `camtrapR` package, and the development team of the Camtrap DP standard).

The responses of these people were positive and encouraging, so I began developing the application. To choose the analyses that the application should perform, I searched the literature, gathered the opinion of camera trap actors and used the knowledge of camera trap data developed during my PhD.

3 Camera trap data terminology

In this section, I briefly describe typical camera trap data and the associated terminology. For this application, I consider camera trap data collected on a grid of cameras setup in one area, which I refer to as a survey. Once camera traps have collected pictures, they are processed to identify species.

The resulting data are extracted to a table which compiles the species, the time of detection and the camera identifier, and possibly other information depending on the survey. Here, I call each row of the table a record, and call this table the records table. Often, a second table also compiles camera information, in particular coordinates, using cameras' identifier to link this data with records table. I call this second file the cameras information table.

More recently, another camera trap data format has been proposed by Bubnicki et al. (2024): the Camera Trap Data Package (or Camtrap DP), based on the Frictionless Data Package Standard. Camtrap DP comprises 3 CSV tables: Observations (corresponding to the records table), Deployments (corresponding to cameras informa-

tion) and Media (containing pictures information). A descriptor file in JSON format links these files and stores metadata like sampling design information or contributors names. This data format is intended to homogenize camera trap data to facilitate exchanges, and to enhance reproducibility by storing information in one data bundle (Bubnicki et al., 2024).

4 Shiny application

The `camtrapviz` interactive application was developed using the Shiny R package (Chang et al., 2023) for web applications development with R. `camtrapviz` allows to analyze camera trap data interactively, based on the functions developed in the associated R package. The application is available at https://lbbe-shiny.univ-lyon1.fr/camtrapviz/.

The interface of the Shiny application is organized as a dashboard with different tabs corresponding to different analyses steps, alternatively displayed in the main body (Figure 4.2.1).

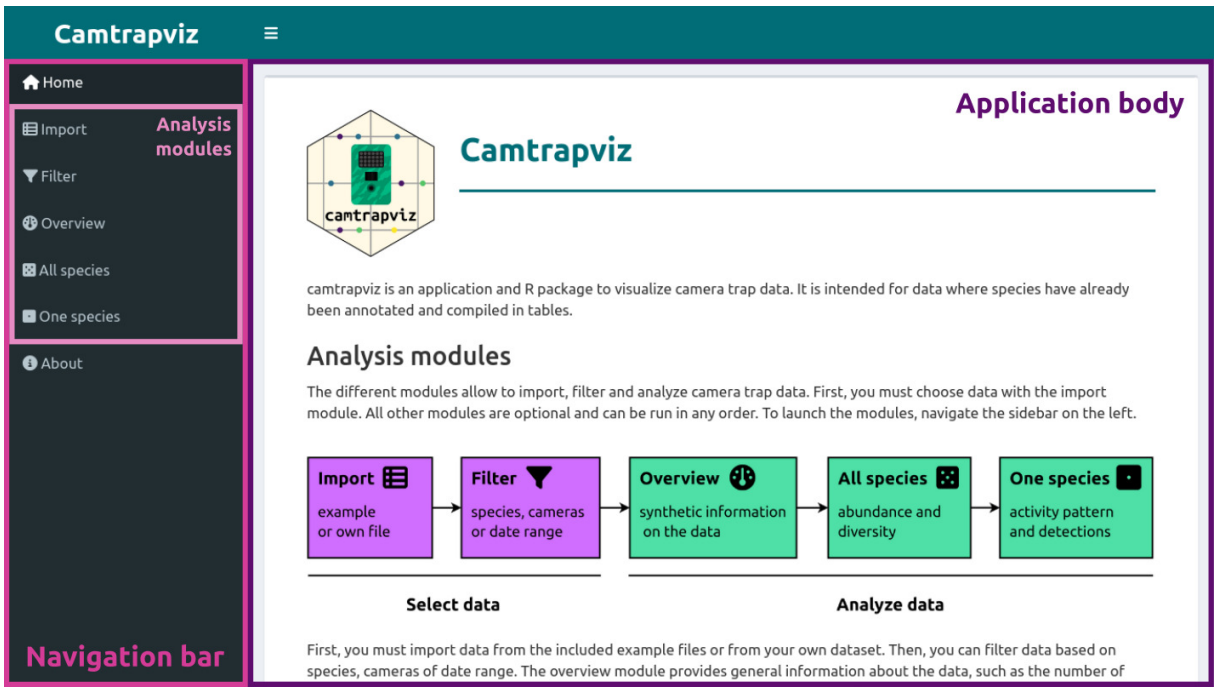


Figure 4.2.1: Overview of the `camtrapviz` application. The navigation bar on the left allows to switch between the 7 application tabs. The analysis modules (Figure 4.2.2) correspond to one tab each, and two additional "Home" and "About" tabs give more information about the application. The application body (right) displays the content of the selected tab (here, the home tab).

4.1 Modules

The application is organized in 5 modules, corresponding to different steps of the analysis (Figure 4.2.2). An overview of these modules is available in Appendices.

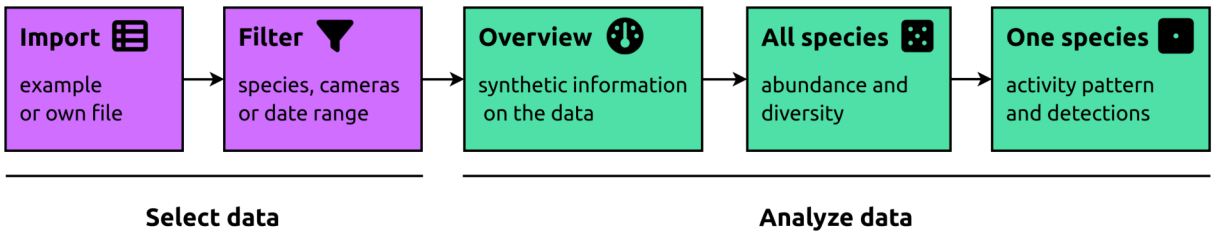


Figure 4.2.2: Analysis modules in the Shiny application. Two modules allow to select data, and three other modules allow to visualize and analyze data.

Import

First, users must choose a dataset in the "Import" module: they can choose one of the two example datasets or upload their own dataset. `camtrapviz` accepts 2 data formats: a single CSV file for records or 2 CSV files for records and cameras information. The application is built to be compatible with the camtrap DP format, meaning that all internal functions can handle camtrapDP data (although interactive import of camtrapDP data in the application is impossible to date).

Columns names of the imported data are free, and only 3 columns are mandatory: camera, species and date and time of detection (alternatively, the user can also provide separate date and time columns). If the user uploads their data, the code automatically attempts to match column names with the information they contain based on pattern matching: for example, if a column contains "species", it is assumed to contain species information.

Once data are imported, the dataset is formatted to be used by the application. I apply as few changes as possible: I only set data to the right data type (e.g. date instead of character) and when a single CSV was provided, I split data between records and cameras information.

Filter

The “Filter” module allows to select a subset of the data on which all subsequent analyses will be based. This module allows to filter data by species, cameras or pictures dates. These can be chosen manually, or based on another column in the dataset: for instance, if a user wants to select all carnivore species, they can select species manually from the species column, or if information is encoded in a “diet” column, they can select all species for which “diet” is “carnivore”. An interactive graph allows to see the filtered data in real time, by displaying all records and graying out the filtered out records. This module typically allows to remove empty pictures or cameras that malfunctioned, select species of interest or concentrate the analysis on a season.

Overview

The “Overview” module provides an overview of the survey with summary statistics for camera activity and detected species. A graph showing all records is displayed, with time on the x-axis and cameras on the y-axis. If the user provided cameras coordinates, a map of the camera trap grid is included. This module also displays summary tables: the camera summary table recapitulates information by camera, like the number of pictures, sampling length or setup and retrieval date. The species summary table includes species-specific information, for instance the number of pictures or the proportion of cameras on which the species was seen.

All species

The “All species” module displays more in-depth information about species in the community. This module aims at describing the community structure collected with camera trap data, by summarizing records counts by species (displayed as a barplot) and computing diversity indices by camera (displayed as a barplot or on a map if coordinates were provided).

One species

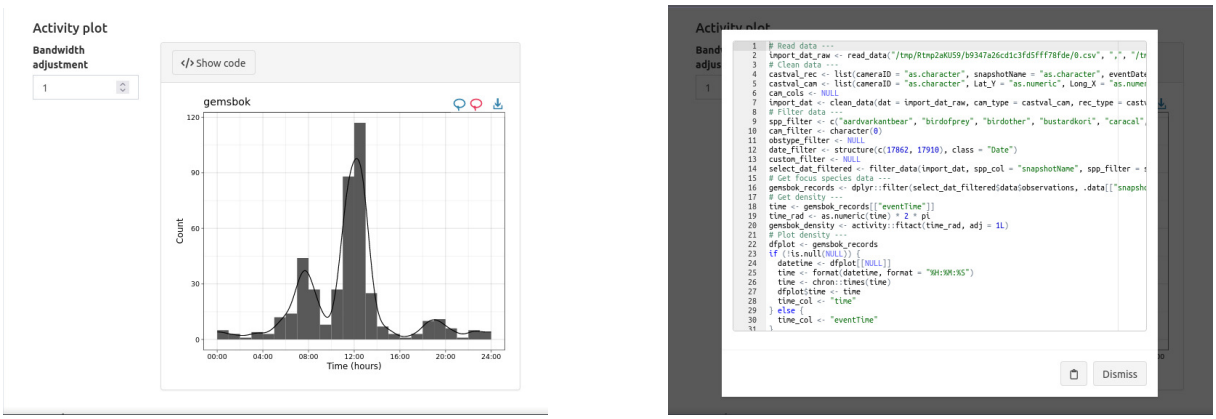
Finally, the “One species” module is designed to analyze more precisely data for user-chosen species, by examining their activity pattern and spatial distribution. For each

species chosen by the user, its activity pattern is displayed as an estimate of kernel density, computed with the R package `activity` (J. M. Rowcliffe et al., 2014; M. Rowcliffe, 2023), superimposed on a histogram of the observed records times. Descriptors of this species occurrences per camera are also displayed as a barplot or on a map (if coordinates were provided). The available descriptors are count, proportion or relative abundance index (number of sightings per time unit) of the species per camera.

4.2 Other features

Throughout the application, I paid particular attention to make analyses flexible. For instance, in the import module, users can import data with custom column names, dates and times formats or coordinates reference systems. To make the analysis workflow more free, the only mandatory module is the import module. In addition, although the underlying application logic performs the analyses in a sequential order (indicated in Figure 4.2.2), the user can navigate between modules at will because data are updated reactively in the background.

At each step of the analysis, users have the possibility to view the code allowing to reproduce the analysis in R (Figure 4.2.3). This is made possible by the `shinymeta` package (Cheng & Sievert, 2021), which provides functionalities to parse code used in the application to make it executable outside Shiny. The code of a Shiny application cannot be exported as is, because it uses special variables to manage user inputs that are only understandable in the context of the Shiny application: for instance, a Shiny numeric input from a user-filled field is encoded in the application as `input$mynumeric`. To export code using such a user input, `shinymeta` replaces this special variable as a hard-coded value corresponding to the user input (e.g. “42”). Users can also download code to reproduce the complete analysis workflow as a R Markdown file and its compiled HTML version, a functionality provided by `shinymeta`. Exporting R code from the interactive Shiny application improves code reproducibility, but also allows users to enhance the analyses conduced in the application by running and enriching them directly in R.



(a) Graph

(b) Associated R code

Figure 4.2.3: R code display in the Shiny application. (a) In the interface, the graph is associated with a “Show code” button (top left) allowing to display the associated R code (b).

5 Software development

As for the development of `standardizeSnapshot` package (chapter 4.1), I made an effort to follow best coding practices and make code reproducible. I used Git throughout project development, wrote unit tests and documentation. I performed unit tests on the functions of the `camtrapviz` package, with the same benefits described in chapter 4.1. I chose to write tests only for the package functions and did not write tests to check Shiny user interface or reactive server logic. This choice was made to gain time while focusing on testing the more critical parts of the project, i.e. ensuring the functions analyzed data without errors. I also wrote extensive documentation and 6 code vignettes, available at https://lisanicvert.github.io/camtrapviz/ (website created with `pkgdown` Wickham et al., 2024).

To enhance the readability and portability of the Shiny application, I followed some of the best practices of Shiny application development described in Wickham (2024a) by avoiding code duplication with functions, using Shiny modules and organizing the application as a R package.

I notably subdivided the Shiny application into modules. Modules in Shiny are akin to functions in base R: they are subdivisions of the full Shiny application that can take inputs and return values, but are isolated from the rest of the application (Wickham, 2024c). Each data treatment module from `camtrapviz` (Figure 4.2.2) is encapsulated in a Shiny module. Shiny modules can also be instantiated several times in the Shiny application, thus reducing code duplication: in this application, I used this property

for the “One species” module, where each instance of the analyses for one species corresponds to a Shiny module. Shiny modules enforce a better compartmentalization of the application functionalities and allow to separate the source code for the Shiny application into different R scripts, thus making the code easier to read and to maintain.

I also wrapped the application into a R package: in addition to the online version of the application, users can also install the `camtrapviz` package from https://github.com/LisaNicvert/camtrapviz and run the application locally. Developing a Shiny application as a R package has several advantages (Wickham, 2024b; Fay et al., 2022). First, this facilitates sharing the application by managing dependencies via package metadata (`DESCRIPTION` and `NAMESPACE` files). The package structure also provides features to document and test functions that can then be used in the server part of the application. In the case of `camtrapviz`, this double package/application nature is particularly relevant, since `camtrapviz` can also be seen as a standalone package to analyze camera trap data, thus allowing people to use functions directly in R.

6 Conclusion and perspectives

I developed `camtrapviz`, an open access Shiny application and R package to visualize and analyze camera trap data. This package is still in development, and many analyses can be added to make it more comprehensive. First, a module to filter for records independence would be essential. Indeed, it is common to set camera traps to take a series of pictures at each detection (usually 3 pictures) and filtering out these repeated pictures is important. The function `filterRecordTable` from the `camtrapR` package allows to filter for records independence and we could use this function directly in the Shiny application to filter for records independence. Second, including species occupancy modeling in the application would be very useful. Occupancy models are hierarchical models used to predict species’ latent presence or absence (“occupancy”), often using covariates to model the detection or occupancy process (MacKenzie et al., 2004). In the camera trapping context, these models are used to infer species’ environmental preferences on the camera trap grid. Occupancy models are one of the most popular analyses of camera trap data (Delisle et al., 2021), so including basic occupancy models in the application would be interesting.

Some existing analyses of the application could also be improved. First, the function to launch the Shiny application locally (from the `camtrapviz` package) could be improved to include an argument for users to provide their dataset directly, instead of interactively selecting a file in the import module. It would also be useful to make the camera column optional for users who only import data from a single camera (typically members from the public).

I think that `camtrapviz` could be useful for the scientific community if the developments listed above were made. The actors from the camera trapping community that I contacted were enthusiast regarding the idea of a Shiny application, especially since none of them knew any similar application. Recently, a Shiny application to analyze camera trap data interactively using functions from the `camtrapR` package was added as an extension of this package, which emphasizes the relevance of Shiny applications for camera trap data. Several recent analysis tools in ecology have also been deployed as Shiny applications, like `PhenoSpace` to study plant species in their trait space (Segrestin et al., 2021), `OCCUR` to design a workflow to select species occurrence records (Ronquillo et al., 2024), or `movedesign` to guide sampling design decisions in movement ecology (Silva et al., 2023).

The interactive application associated with `camtrapviz` could be especially useful for reserve managers or members of the general public interested in visualizing their camera trap data, notably since it is available online and users are not required to install anything on their computers. However, the online application is not dimensioned for intense traffic, so this solution is viable only if the number of users stays low.

To conclude, `camtrapviz` is a first step towards user-friendly analysis of camera trap data which combines the ease of use and interactivity of the Shiny application with the flexibility and reproducibility of coding.

Appendices

These appendices show a glimpse of the Shiny application for each analysis module.

A Import module

This module allows users to choose an example dataset or to import their own.

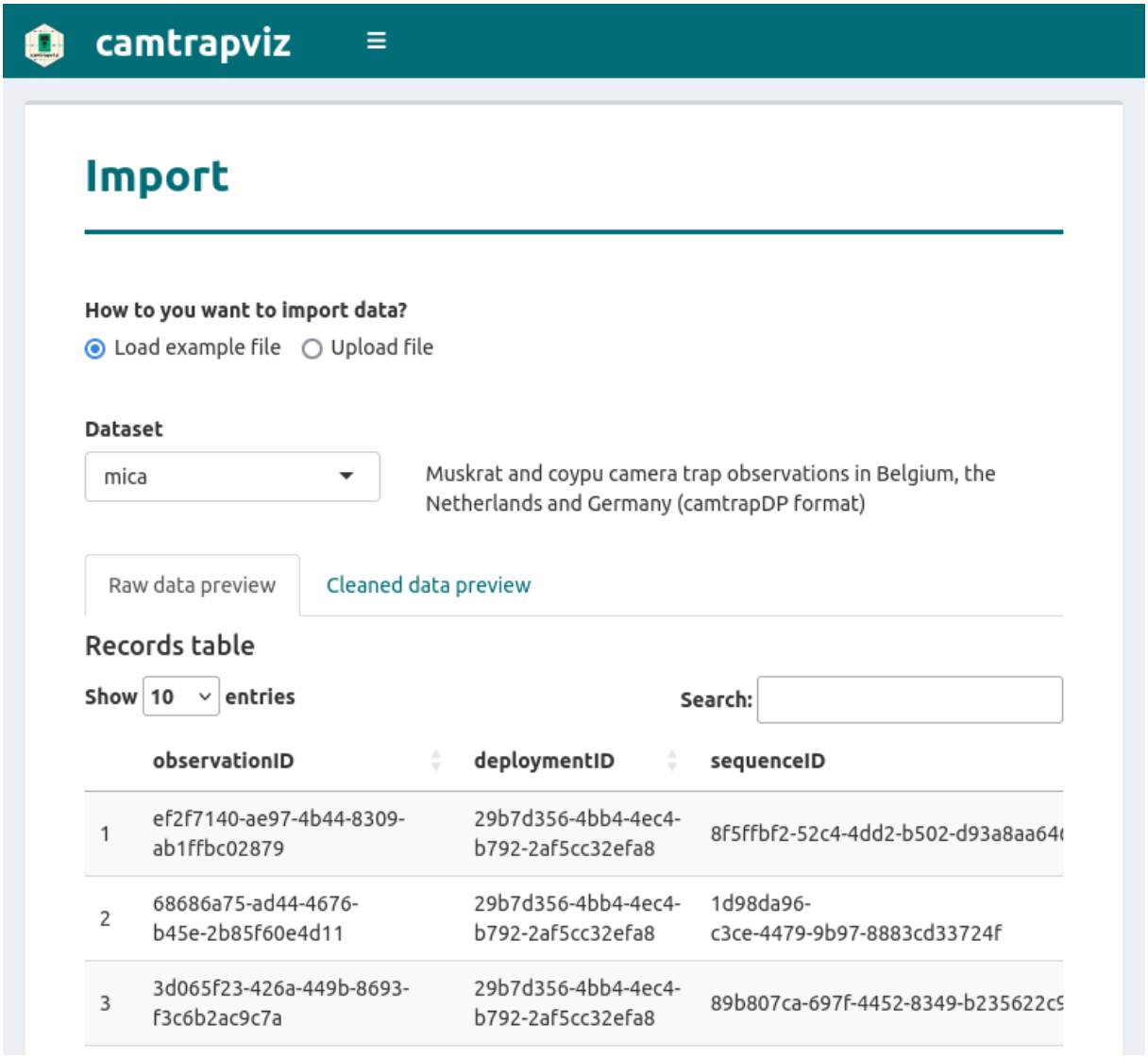


Figure A.1: Import module with the example datasets. Here, the “mica” dataset was chosen from the drop-down menu. An overview of the data is displayed below.

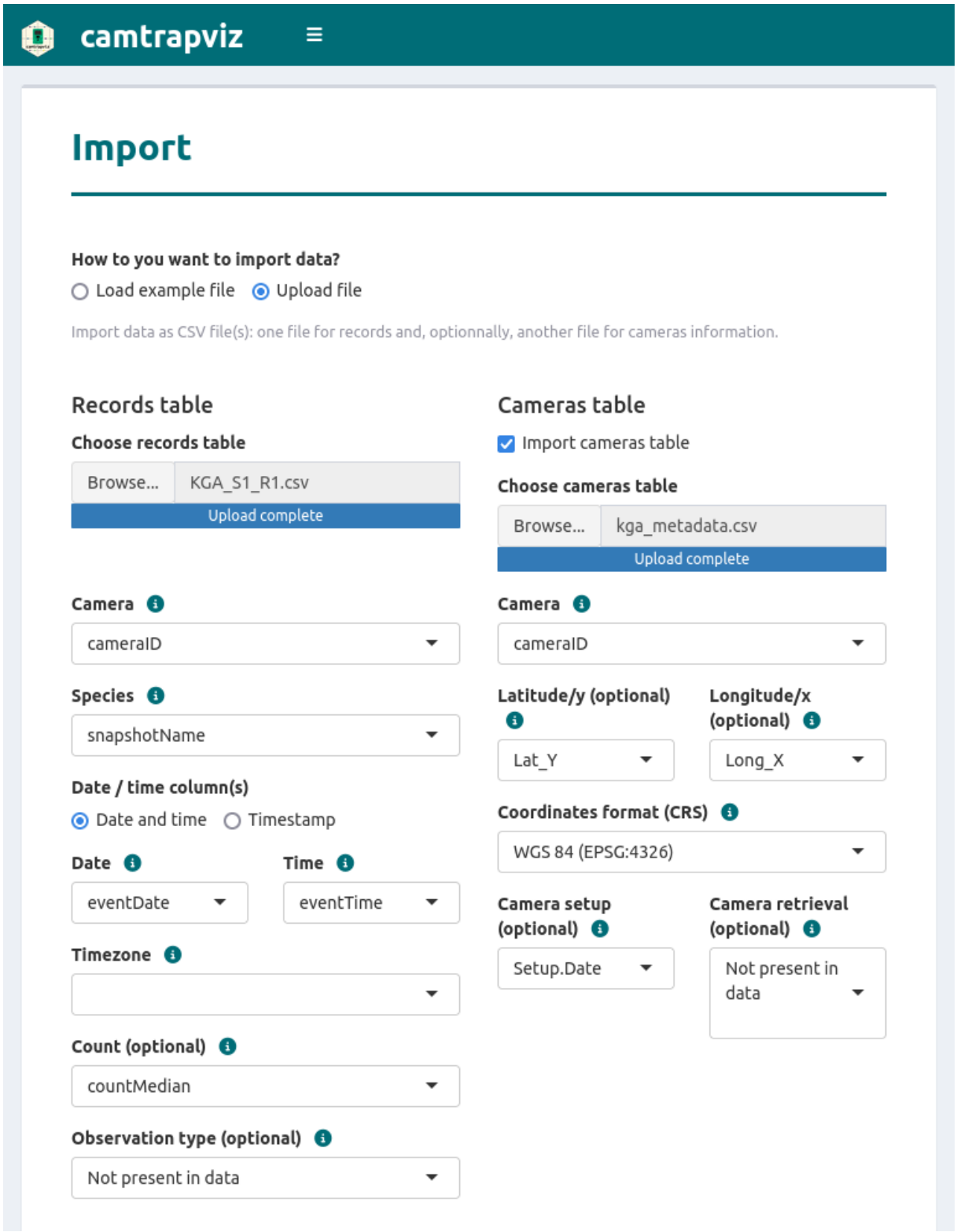


Figure A.2: Import module with a user-chosen dataset. Here, two CSV tables were imported (for the records and for the cameras). The drop-down menus allow to select the column names corresponding to the different variables.

B Filter module

This module allows users to select a subset of the data.

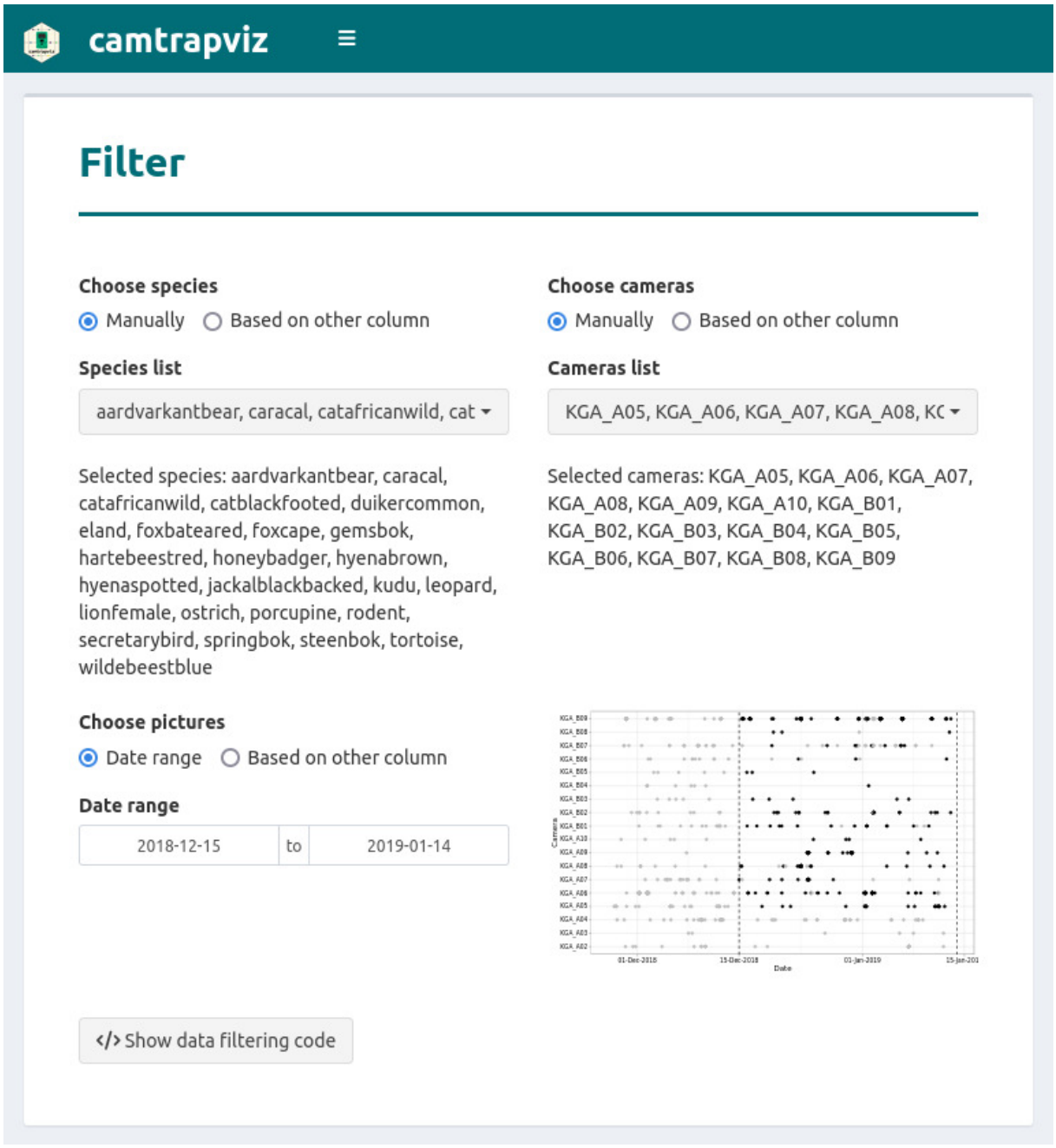


Figure B.1: Filter module. Three menus allow to select specific data subsets based on species, cameras or pictures date. A figure summarizing the filtering is displayed in the bottom-right corner.

C Overview module

This module provides general information about their data with summary statistics, graphs and summary tables.

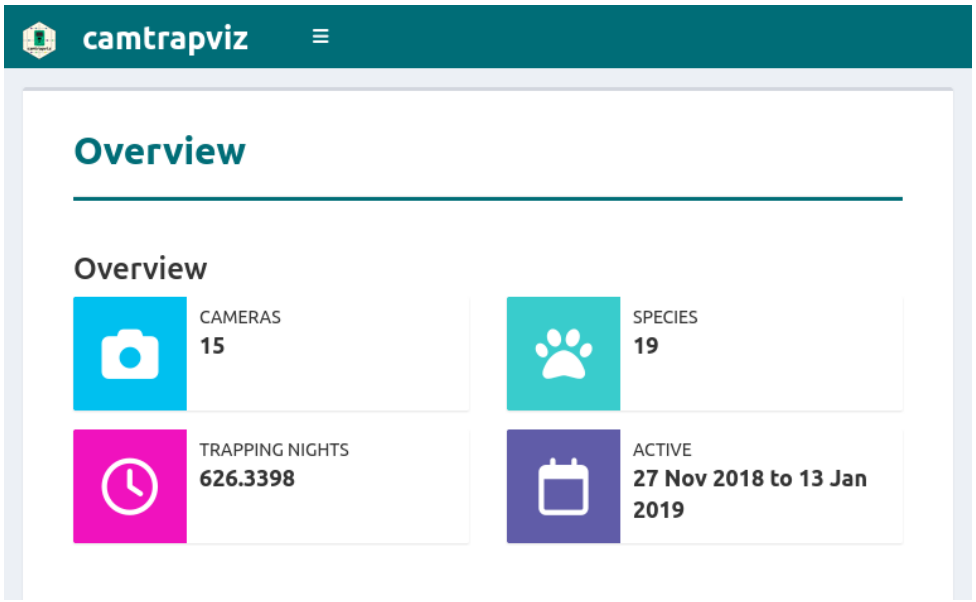


Figure C.1: Summary statistics showing the number of cameras and species, the total number of trapping nights over all cameras and the date range.

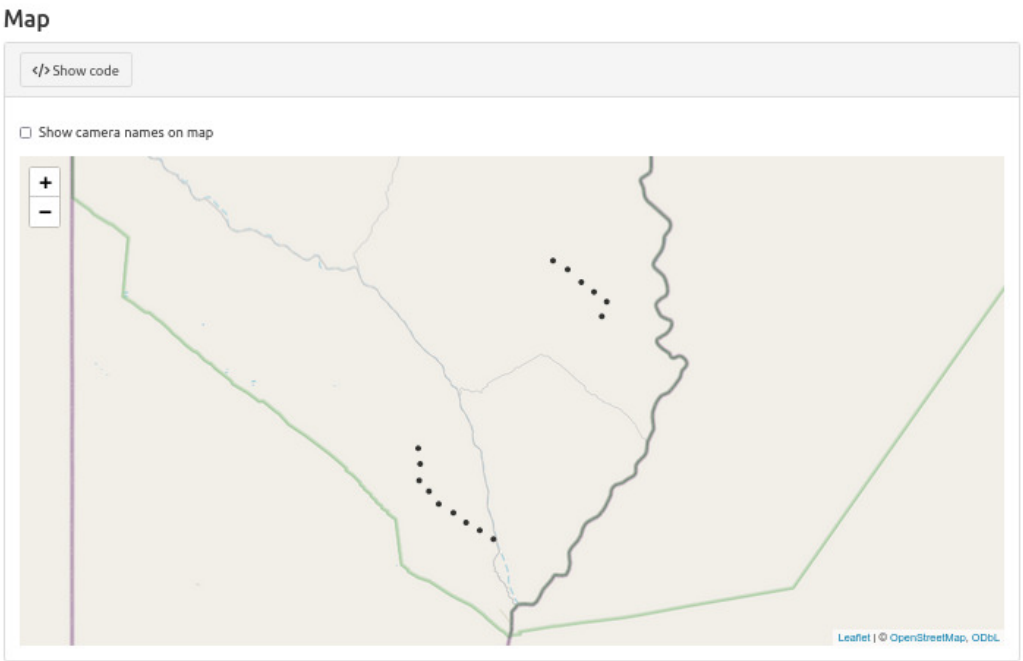
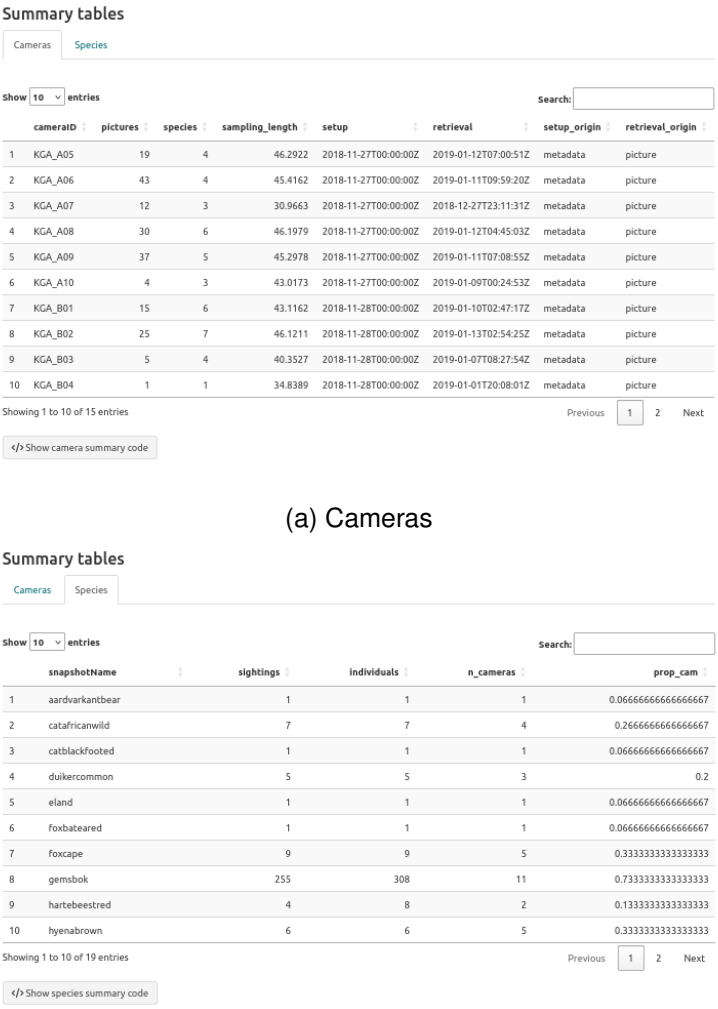


Figure C.2: Overview module map.



Figure C.3: Overview module records plot representing pictures taken by each camera in time.



Summary tables

Cameras | Species

Show 10 entries — Search:

	cameraID	pictures	species	sampling_length	setup	retrieval	setup_origin	retrieval_origin
1	KGA_A05	19	4	46.2922	2018-11-27T00:00:00Z	2019-01-12T07:00:51Z	metadata	picture
2	KGA_A06	43	4	45.4162	2018-11-27T00:00:00Z	2019-01-11T09:59:20Z	metadata	picture
3	KGA_A07	12	3	30.9663	2018-11-27T00:00:00Z	2018-12-27T23:11:31Z	metadata	picture
4	KGA_A08	30	6	46.1979	2018-11-27T00:00:00Z	2019-01-12T04:45:03Z	metadata	picture
5	KGA_A09	37	5	45.2978	2018-11-27T00:00:00Z	2019-01-11T07:08:55Z	metadata	picture
6	KGA_A10	4	3	43.0173	2018-11-27T00:00:00Z	2019-01-09T00:24:53Z	metadata	picture
7	KGA_B01	15	6	43.1162	2018-11-28T00:00:00Z	2019-01-10T02:47:17Z	metadata	picture
8	KGA_B02	25	7	46.1211	2018-11-28T00:00:00Z	2019-01-13T02:54:25Z	metadata	picture
9	KGA_B03	5	4	40.3527	2018-11-28T00:00:00Z	2019-01-07T08:27:54Z	metadata	picture
10	KGA_B04	1	1	34.8389	2018-11-28T00:00:00Z	2019-01-01T20:08:01Z	metadata	picture

Showing 1 to 10 of 15 entries — Previous 1 2 Next

Show camera summary code

(a) Cameras

Summary tables

Cameras | Species

Show 10 entries — Search:

	snapshotName	sightings	individuals	n_cameras	prop_cam
1	aardvarkantbear	1	1	1	0.06666666666666667
2	catafricanwild	7	7	4	0.2666666666666667
3	catblackfooted	1	1	1	0.06666666666666667
4	duikercommon	5	5	3	0.2
5	eland	1	1	1	0.06666666666666667
6	foxbateared	1	1	1	0.06666666666666667
7	foxcape	9	9	5	0.3333333333333333
8	gemsbok	255	308	11	0.7333333333333333
9	hartebeestred	4	8	2	0.1333333333333333
10	hyenabrown	6	6	5	0.3333333333333333

Showing 1 to 10 of 19 entries — Previous 1 2 Next

Show species summary code

(b) Species

Figure C.4: Overview module summary tables for cameras (a) and for species (b).

D All species module

This module provides graphical summaries related to all species captured on camera traps.

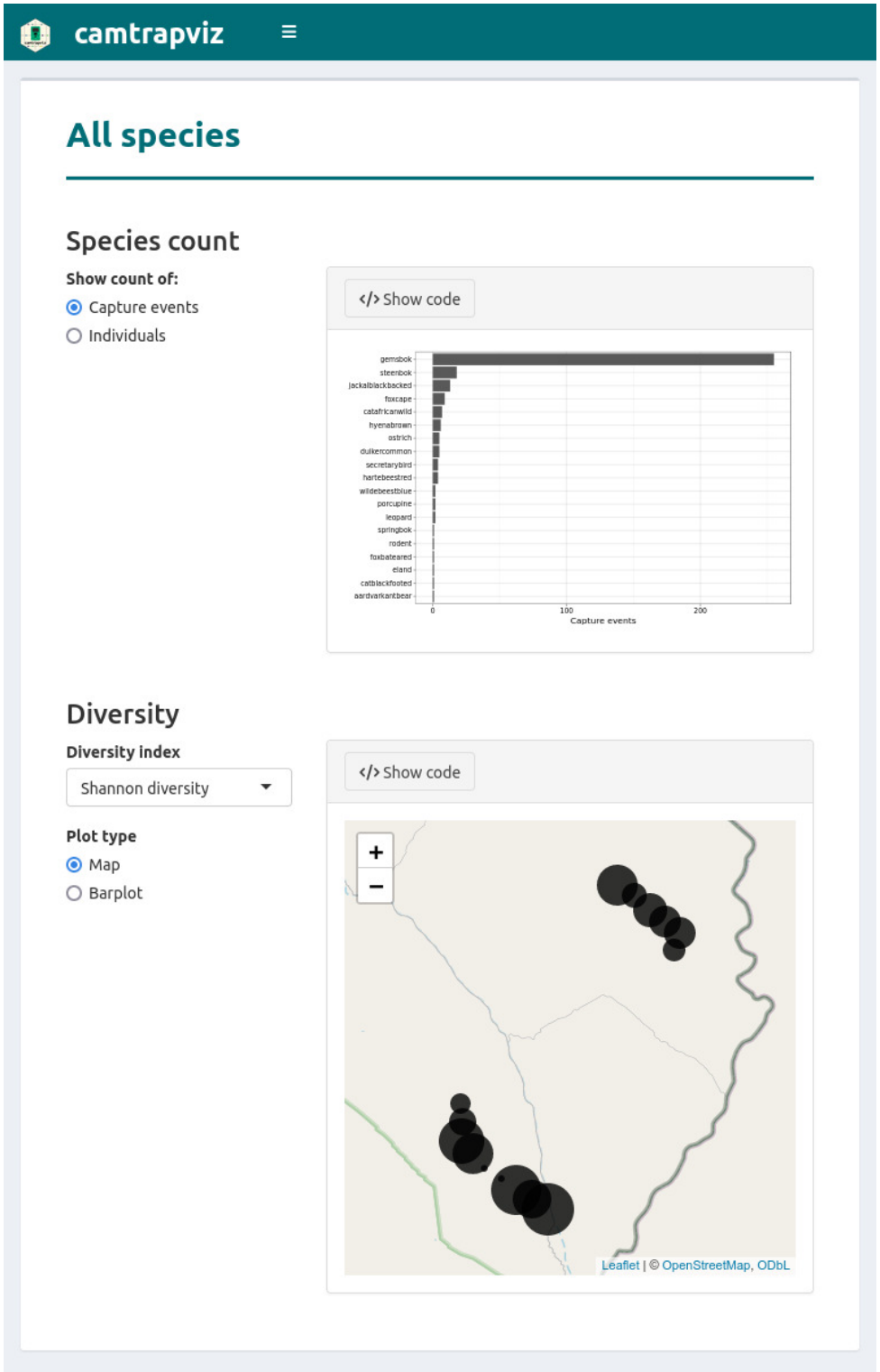


Figure D.1: All species module. The top graph shows capture counts by species and the bottom plot shows Shannon diversity by camera (circle size is proportional to the Shannon diversity).

E One species module

This module allows to investigate the activity rhythm and spatial distribution per species.

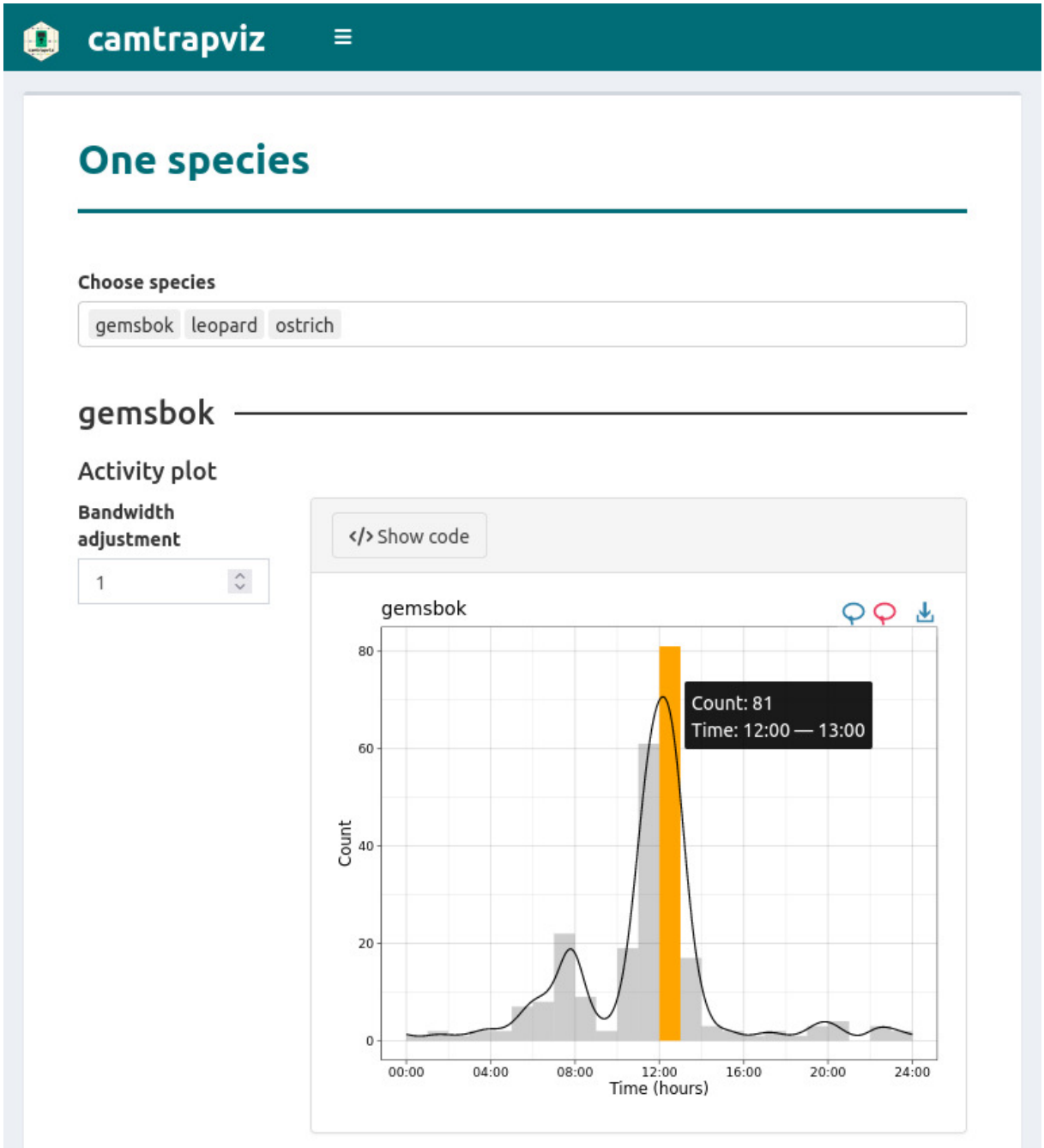


Figure E.1: One species module overview. The top menu allows to select species: here, gemsbok, leopard and ostrich are selected. Below, the activity and spatial distribution for these species are shown.

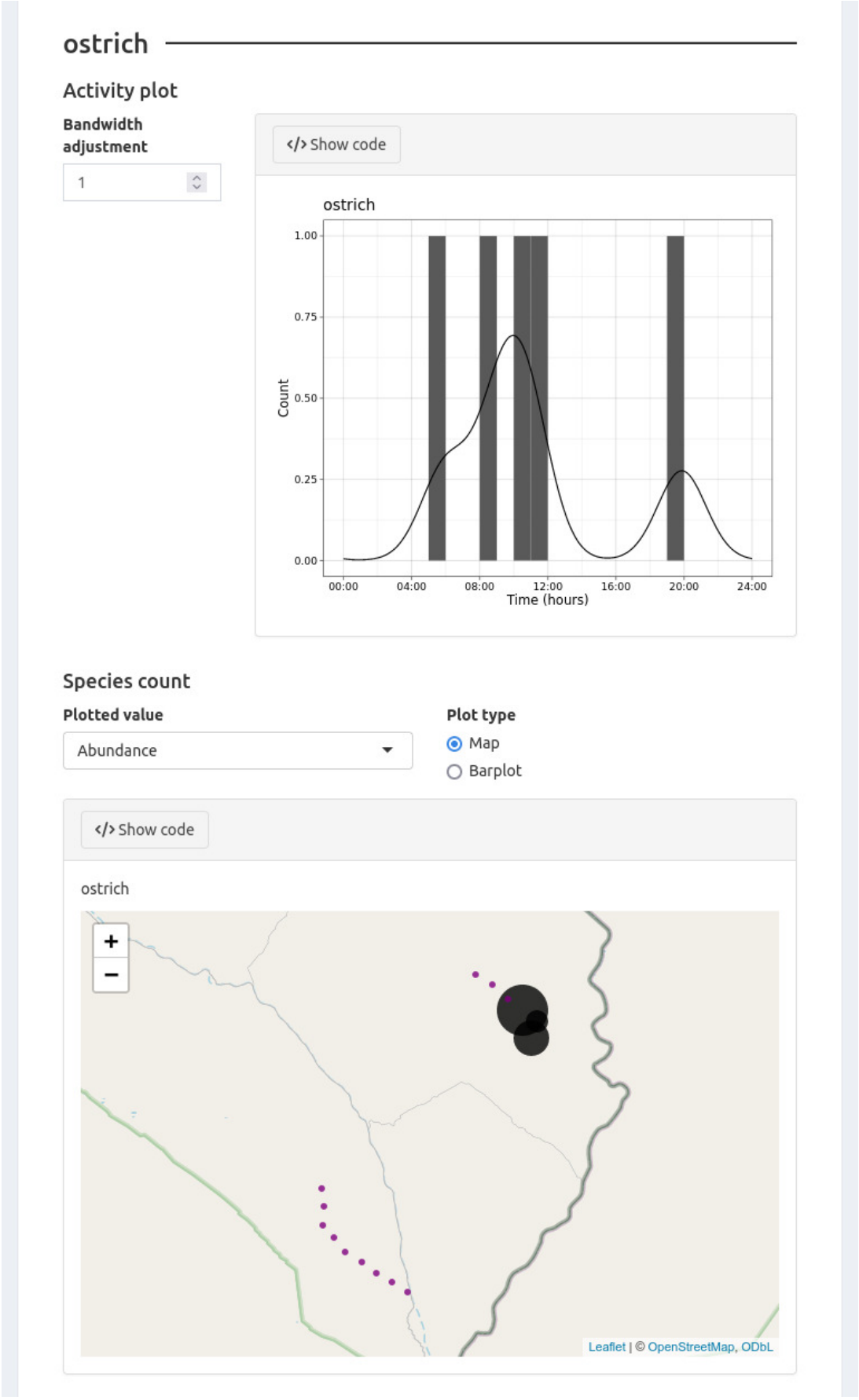


Figure E.2: One species module analyses showing the activity graph and presence map for ostrich.

Bibliography

Azzou, S. A. K., Singer, L., Aebischer, T., Caduff, M., Wolf, B., & Wegmann, D. (2021). A sparse observation model to quantify species distributions and their overlap in space and time. *Ecography*, *44*(6), 928–940. https://doi.org/10.1111/ecog.05411

Bubnicki, J. W., Norton, B., Baskauf, S. J., Bruce, T., Cagnacci, F., Casaer, J., Churski, M., Cromsigt, J. P. G. M., Farra, S. D., Fiderer, C., Forrester, T. D., Hendry, H., Heurich, M., Hofmeester, T. R., Jansen, P. A., Kays, R., Kuijper, D. P. J., Liefting, Y., Linnell, J. D. C., … Desmet, P. (2024). Camtrap DP: An open standard for the FAIR exchange and archiving of camera trap data. *Remote Sensing in Ecology and Conservation*, *10*(3). https://doi.org/10.1002/rse2.374

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2023). Shiny: Web Application Framework for R. https://CRAN.R-project.org/package=shiny

Chen, G., Han, T. X., He, Z., Kays, R., & Forrester, T. (2014). Deep convolutional neural network based species recognition for wild animal monitoring. *2014 IEEE International Conference on Image Processing (ICIP)*, 858–862. https://doi.org/10.1109/ICIP.2014.7025172

Cheng, J., & Sievert, C. (2021). Shinymeta: Export Domain Logic from Shiny using Meta-Programming. https://CRAN.R-project.org/package=shinymeta

Delisle, Z. J., Flaherty, E. A., Nobbe, M. R., Wzientek, C. M., & Swihart, R. K. (2021). Next-Generation Camera Trapping: Systematic Review of Historic Trends Suggests Keys to Expanded Research Applications in Ecology and Conservation. *Frontiers in Ecology and Evolution*, *9*. https://doi.org/10.3389/fevo.2021.617996

Fay, C., Rochette, S., Guyader, V., & Girard, C. (2022). Structuring Your Project. In *Engineering production-grade shiny apps* (First edition). CRC Press, Taylor & Francis Group. https://doi.org/10.1201/978-1-003-02987-8

Fonteyn, D., Vermeulen, C., Deflandre, N., Cornelis, D., Lhoest, S., Houngbégnon, F. G. A., Doucet, J.-L., & Fayolle, A. (2021). Wildlife trail or systematic? Camera trap placement has little effect on estimates of mammal diversity in a tropical forest in Gabon. *Remote Sensing in Ecology and Conservation*, *7*(2), 321–336. https://doi.org/10.1002/rse2.191

Kays, R., Hody, A., Jachowski, D. S., & Parsons, A. W. (2021). Empirical evaluation of the spatial scale and detection process of camera trap surveys. *Movement Ecology*, *9*(1), 41. https://doi.org/10.1186/s40462-021-00277-3

MacFadyen, S., Allsopp, N., Altwegg, R., Archibald, S., Botha, J., Bradshaw, K., Carruthers, J., De Klerk, H., de Vos, A., Distiller, G., Foord, S., Freitag-Ronaldson, S., Gibbs, R., Hamer, M., Landi, P., MacFadyen, D., Manuel, J., Midgley, G., Moncrieff, G., … Hui, C. (2022). Drowning in data,

thirsty for information and starved for understanding: A biodiversity information hub for cooperative environmental monitoring in South Africa. *Biological Conservation*, *274*, 109736. https://doi.org/10.1016/j.biocon.2022.109736

MacKenzie, D. I., Bailey, L. L., & Nichols, J. D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, *73*(3), 546–555. https://doi.org/https://doi.org/10.1111/j.0021-8790.2004.00828.x

Myles White, J., White, K., Müller, K., & Jacobs, A. (2022). Log4r: A Fast and Lightweight Logging System for R, Based on 'log4j'. https://CRAN.R-project.org/package=log4r

Niedballa, J., Sollmann, R., Courtiol, A., & Wilting, A. (2016). camtrapR: An R package for efficient camera trap data management. *Methods in Ecology and Evolution*, *7*(12), 1457–1462. https://doi.org/10.1111/2041-210X.12600

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, *115*(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

O'Connell, A. F., Nichols, J. D., & Karanth, K. U. (Eds.). (2011). *Camera Traps in Animal Ecology*. Springer Japan. https://doi.org/10.1007/978-4-431-99495-4

Oldoni, D., Desmet, P., & Huybrechts, P. (2024). Camtraptor: Read, Explore and Visualize Camera Trap Data Packages.

Pardo, L. E., Bombaci, S. P., Huebner, S., Somers, M. J., Fritz, H., Downs, C., Guthmann, A., Hetem, R. S., Keith, M., le Roux, A., Mgqatsa, N., Packer, C., Palmer, M. S., Parker, D. M., Peel, M., Slotow, R., Strauss, W. M., Swanepoel, L., Tambling, C., … Venter, J. A. (2021). Snapshot Safari: A large-scale collaborative to monitor Africa's remarkable biodiversity. *South African Journal of Science*, *117*(1/2), 1–4. https://doi.org/10.17159/sajs.2021/8134

Reyserhove, L., Norton, B., & Desmet, P. (2023). Best Practices for Managing and Publishing Camera Trap Data. Retrieved April 10, 2024, from https://docs.gbif.org/camera-trap-guide/en/

Ronquillo, C., Stropp, J., & Hortal, J. (2024). OCCUR Shiny application: A user-friendly guide for curating species occurrence records. *Methods in Ecology and Evolution*, *15*(5). https://doi.org/10.1111/2041-210X.14271

Rota, C. T., Ferreira, M. A. R., Kays, R. W., Forrester, T. D., Kalies, E. L., McShea, W. J., Parsons, A. W., & Millspaugh, J. J. (2016). A multispecies occupancy model for two or more interacting species (D. Warton, Ed.). *Methods in Ecology and Evolution*, *7*(10), 1164–1173. Retrieved June 11, 2020, from https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12587

Rowcliffe, J. M., Kays, R., Kranstauber, B., Carbone, C., & Jansen, P. A. (2014). Quantifying levels of animal activity using camera trap data (D. Fisher, Ed.). *Methods in Ecology and Evolution*, *5*(11), 1170–1179. https://doi.org/10.1111/2041-210X.12278

Rowcliffe, M. (2023). Activity: Animal Activity Statistics. https://CRAN.R-project.org/package=activity

Segrestin, J., Sartori, K., Navas, M.-L., Kattge, J., Díaz, S., & Garnier, E. (2021). PhenoSpace: A Shiny application to visualize trait data in the phenotypic space of the global spectrum of plant form and function. *Ecology and Evolution*, *11*(4), 1526–1534. https://doi.org/10.1002/ece3.6928

Silva, I., Fleming, C. H., Noonan, M. J., Fagan, W. F., & Calabrese, J. M. (2023). Movedesign: Shiny R app to evaluate sampling design for animal movement studies. *Methods in Ecology and Evolution*, *14*(9), 2216–2225. https://doi.org/10.1111/2041-210X.14153

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B., Beasley, J. C., Schlichting, P. E., Boughton, R. K., Wight, B., Newkirk, E. S., Ivan, J. S., Odell, E. A., Brook, R. K., … Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, *10*(4), 585–590. https://doi.org/10.1111/2041-210X.13120

Wickham, H. (2011). Testthat: Get Started with Testing. *The R Journal*, *3*, 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf

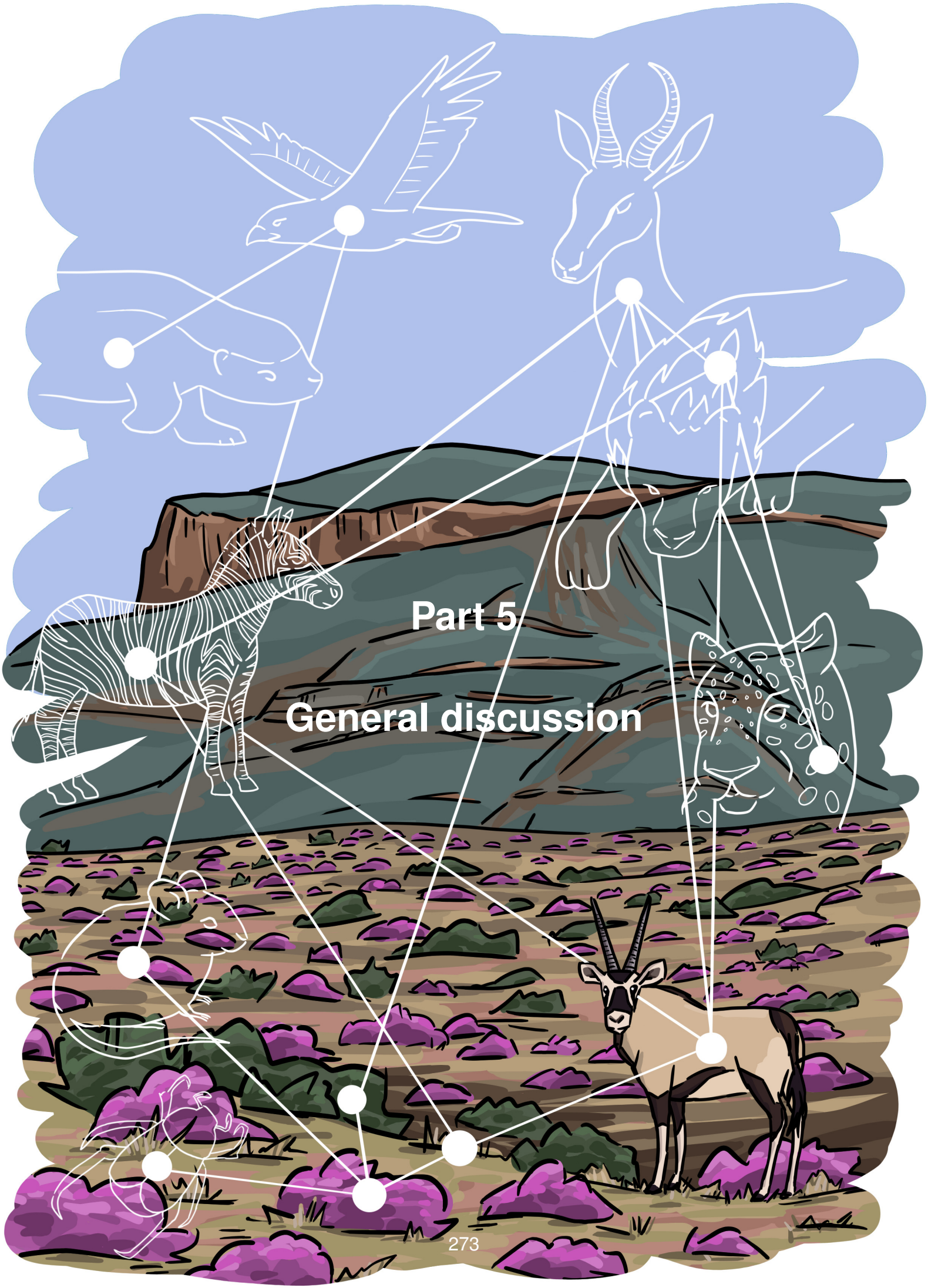
Wickham, H. (2024a). *Mastering Shiny*. O'Reilly Media, Inc. Retrieved April 17, 2024, from https://mastering-shiny.org/index.html

Wickham, H. (2024b). Packages. In *Mastering Shiny*. O'Reilly Media, Inc. Retrieved April 16, 2024, from https://mastering-shiny.org/scaling-packaging.html?q=package#package-workflow

Wickham, H. (2024c). Shiny modules. In *Mastering Shiny*. O'Reilly Media, Inc. Retrieved April 16, 2024, from https://mastering-shiny.org/scaling-modules.html

Wickham, H., & Bryan, J. (2023). Testing basics. In *R Packages: Organize, Test, Document, and Share Your Code* (2nd ed.). O'Reilly Media, Inc. Retrieved April 12, 2024, from https://r-pkgs.org/testing-basics.html

Wickham, H., Hesselberth, J., & Salmon, M. (2024). Pkgdown: Make Static HTML Documentation for a Package. https://CRAN.R-project.org/package=pkgdown



Part 5

General discussion

Cover picture: this drawing goes back over the cover for the general introduction. Some time has passed, and as we observed this system we realized that species are linked with a network of interactions. The species (nodes of the network) are, from top to bottom and left to right: pale chanting goshawk, impala, honeybadger, lion, zebra, leopard, Karoo bush rat, gemsbok, a bush species, a grass species, a monkey beetle (*Hopliini*) and the flowering succulent *Drosanthemum eburneum*. These species engage in a complex interaction network, and edges represent various interactions such as competition, predation, herbivory, commensalism or pollination.

© Own drawing inspired from a picture taken in the Karoo National park.

Discussion

This thesis was the opportunity to explore different aspects of ecological networks and communities, from a methodological and an ecological point of view. In this general discussion, I first synthesize the main results from the different chapters of this thesis. Then I reflect on open science, especially on the contribution of clean and open code for research. After that, I discuss model validation approaches and philosophical implications of ecological data modeling, and conclude with more personal thoughts on generalist versus specialist approaches in this thesis and more broadly in ecology.

1 Synthesis of the main results

In this thesis, I explored different facets of interspecific interactions and ecological communities. I investigated trait matching in ecological networks (part 2), attraction/repulsion patterns between species (part 3) and provided tools for processing large amounts of multi-species occurrence data, which are often necessary to study ecological networks (part 4). Table 5.1 synthesizes the main methodological and data analysis results for each chapter.

1.1 Investigate trait matching in interaction networks

In part 2, I investigated causes of structure in ecological networks, and more precisely the importance of trait matching on interactions in ecological communities, using the notion of interaction niche. To do so, I extended, developed and evaluated multivariate methods of the correspondence analysis (CA) family. I illustrated these methods on a mutualistic bird-fruit interaction network (data from Dehling et al., 2021). In this part, an interaction is defined as a physical encounter between species, here birds eating

Part	Chapter	Methodological results	Data analysis results
2	2.1	Evaluation and interpretation of CA and reciprocal scaling in ecological networks	Bird and plants niche breadths are related to latent traits
	2.2	Interpretation of CCA and dc-CA in ecological networks Definition of variation partitioning in ecological networks	Trait matching explains 6% of variation in the network
	2.3	Extension of reciprocal scaling to CCA and dc-CA	Visualize birds and plants niches constrained with traits
3	3.1 & 3.2	Evaluation and interpretation of MHPs to infer interactions from camera trap data	Zebra and impala follow other herbivores and avoid lion
	3.2	Specification and interpretation of a MHP with a temporal covariate	Circadian rhythms have little effect on interactions
4	4.1	Implementation of a R package to process Snapshot Safari data	Standardize Snapshot Safari database
	4.2	Implementation of a R package and Shiny application to analyze and visualize camera trap data	

Table 5.1: Synthesis of the main results of the thesis. In part 2, I analyzed a bird-fruit interaction network; In part 3, I analyzed mammal occurrence data collected with camera traps and in chapter 4.1, I standardized data collected by the Snapshot Safari program.

plant fruits.

In chapter 2.1, I used CA and reciprocal scaling to analyze the interaction niche of species by applying methods initially designed for environmental niches to study interaction niches, thus providing a new ecological interpretation of these methods. Evaluation on simulated data suggests that the model can accurately infer interaction niche breadths and optima. The bird-fruit network analysis also shows how these methods can be used to answer ecological questions about the drivers of species niches breadths and optima in relation to their traits. In particular, intermediate size bird and plant species tend to have wider niches than small or large birds and plants.

In chapter 2.2, I employed constrained multivariate methods, canonical and double-constrained correspondence analyses (respectively CCA and dc-CA), to quantify more precisely the part of the network structure due to species traits. I illustrated these

methods by performing variation decomposition on the bird-fruit interaction network, showing that matching of the available traits accounts, at best, for a small part of the total variability (less than 6%).

Finally, in chapter 2.3, I developed a method allowing to measure species niche breadth from constrained analyses, by extending reciprocal scaling to CCA and dc-CA. I exemplified the model by measuring interaction niche breadths constrained with traits on the bird-fruit network.

To summarize, in this part I studied networks through the lens of the interaction niche by using statistical methods and ecological concepts developed with environmental niches in mind to study interaction niches. I also proposed to measure interaction niche indices by taking species traits into account.

1.2 Infer attractions and repulsions

In part 3, I investigated the consequences of interspecific interactions in a network of interacting species. I explored how the interactions between different species can affect other species' spatio-temporal distribution patterns: in this part, an interaction is defined as an effect of species on other species distribution patterns. I used multivariate point processes in the Hawkes process family to analyze data collected in continuous time with camera traps data. The goal was to use the temporal information provided by camera traps, which is often overlooked, to inform the inference of interspecific interactions. I illustrated Hawkes processes on a dataset of five species occurrences in the South African savanna collected with camera traps (data available at Nicvert et al., 2023).

In chapter 3.1, I used a linear multivariate Hawkes process (linear MHP) to analyze camera trap data. Simulations performed to evaluate the method suggest that it can successfully infer interactions when using enough data and when interactions are strong enough. Real camera trap data analysis highlighted ecologically interpretable interactions between species: attractions between individuals of the same species, attractions between herbivore species (in particular zebra and impala following other herbivores) and avoidance of lion by impala and zebra. These patterns can be interpreted using known behavioral interactions between species such as predation, grazing succession or mixed-species grouping, but may also be influenced by environmental and

temporal factors.

In chapter 3.2, I used a non-linear MHP with a temporal covariate to reanalyze this dataset and tease apart species' occurrence patterns due to circadian rhythms from patterns due to interspecific interactions. Results suggest that circadian rhythms have little effect on the inferred interactions, and are overall consistent with results from the linear MHP.

To sum up, I showed that multivariate Hawkes processes can successfully analyze ecological data. Compared to other existing methods to infer interactions from camera trap data, this model has more interpretable outputs, but this comes at the cost of more statistical complexity.

1.3 Software development

In part 4, I took a more applied stance and developed two software tools to process, clean and visualize multi-species data collected with camera traps. This part is in line with the growing need for automated analysis tools to process increasing amounts of data in ecology.

In chapter 4.1, I developed a R package, `standardizeSnapshot`, to automate the internal data standardization and cleaning process of camera trap data collected with the Snapshot Safari program. I used this package to standardize all available Snapshot Safari data collected to date. `standardizeSnapshot` is tested, documented and distributed on GitHub (https://github.com/SnapshotSafari/standard-merge), and now routinely used to process Snapshot Safari data.

In chapter 4.2, I developed a Shiny application and R package, `camtrapviz`, to summarize and visualize multi-species data collected with an array of camera traps (https://github.com/LisaNicvert/camtrapviz). The package provides functions to analyze and summarize camera trap data with R, and the application allows to interactively analyze data and to export code to reproduce the analyses outside the application. `camtrapviz` is still in development and I plan to enrich it with new analyses.

This part exemplifies how software and coding tools can be used to automate the data cleaning and data exploration processes. More generally, these developments are part of the open science movement to make scientific results available for more people by releasing the code freely. They also aim at making scientific analyses more reliable

by putting a greater effort on writing clean, documented and tested code.

2 Open science and reproducibility

To be reliable, research should be carried out with appropriate methods, conducted honestly (i.e. without fraud) and reported clearly. Owing to the scale and temporal dynamics of ecological systems, ecology faces specific challenges regarding data collection (Filazzola & Lortie, 2022) and causal inference (Arif & MacNeil, 2022; Tredennick et al., 2021). While these broad issues are important, in this section, I focus more specifically on how clean and open code and data can improve the reliability of ecological research. I begin by defining replicability and reproducibility, outline how clean and open code and data can help mitigate reproducibility issues, and show how reproducible and open code has been implemented in this thesis.

2.1 Replicability and reproducibility

Replicability is the ability to qualitatively confirm the conclusions of research on a different dataset (Popovic et al., 2024). The successful replication of a scientific result is essential to increase the confidence we have in this result and/or generalize findings (Ioannidis, 2005).

Reproducibility is the ability to reproduce results using the same analyses on the same dataset, and can be seen as a first step towards replicability (see Figure 5.1). In the context of growing use of code in ecology, reproducibility includes making code and data available (at least on demand), ensuring that the code can run (in particular, checking that the needed libraries can be installed and specifying their version) and that running the code produces consistent and desired results (Ivimey-Cook et al., 2023).

Replicability is so fundamental that it has been described as a distinction between science and non-science (Filazzola & Cahill Jr, 2021). However, the replicability of scientific research has recently been challenged, and research has been described as undergoing a “replication crisis” (Filazzola & Cahill Jr, 2021; Shrout & Rodgers, 2018). In psychology, a systematic replication experiment across many labs estimated the rate of replication success for published studies at only 36% (Open Science Collaboration, 2015). To my knowledge, no similar estimation exists in ecology, but the extremely

low rate of replication studies, estimated to 0.023% (Kelly, 2019), makes it difficult to evaluate the replicability rate. This replication crisis can be due to intentional fraud, thought to be very rare, (Shrout & Rodgers, 2018), or more commonly to unintentional factors, like publication bias for significant results (Ioannidis, 2005) or questionable use of statistics (Popovic et al., 2024).

To improve reproducibility and replicability of scientific studies, several avenues have been proposed, including encouraging the publication of negative results (Ioannidis, 2005), using better statistical practices (Popovic et al., 2024) or developing studies preregistration to avoid HARKing (Hypothesizing After Results are Known) (Kerr, 1998; Shrout & Rodgers, 2018). Below, I focus on solutions involving better coding practices and open science, which have the potential to greatly improve the reproducibility of scientific studies at a relatively low cost. Indeed, code has become a cornerstone of ecological research (Filazzola & Lortie, 2022; Markowetz, 2017) but is still surprisingly little scrutinized today. In ecology, although more than 90% of published papers use code to produce results, only 27% provide the code to reproduce the analyses (Culina et al., 2020). Additionally, this code is usually not reviewed (Ivimey-Cook et al., 2023). This is an issue, because not disclosing code or disclosing code with errors can hamper reproducibility and replicability.

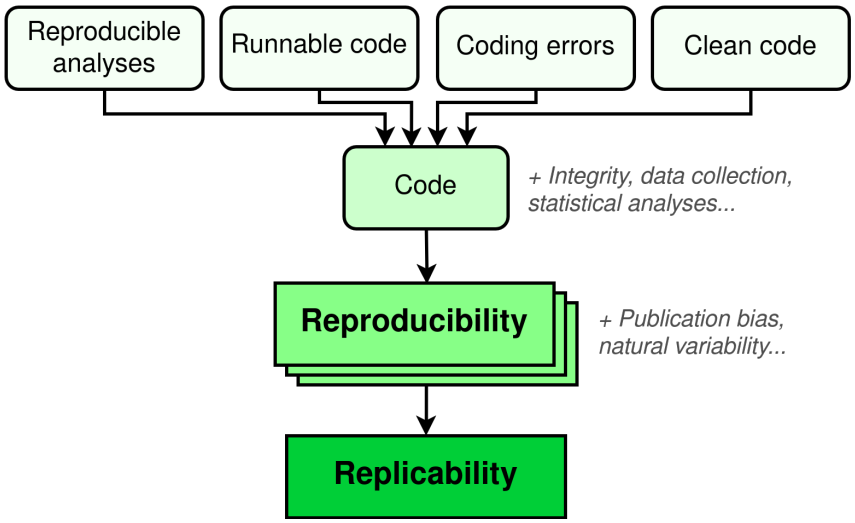


Figure 5.1: Link between coding practices, reproducibility and replicability. Code influences replicability, along with other factors such as scientific integrity or statistical analyses. Many reproducible studies yielding the same result allow to reach replicability, but it is also affected by sampling bias or the natural variability of the studied systems.

2.2 Coding practices to improve reproducibility

The coding process provides several opportunities to make data analyses as reproducible as possible (see Figure 5.1) (Gomes et al., 2022; Ivimey-Cook et al., 2023; Filazzola & Lortie, 2022). The first element is to conduct as many analysis steps as possible in a reproducible manner: non-reproducible analyses include using a proprietary software to process data or processing data manually (Gomes et al., 2022). If code automation is not possible, or too costly to implement, analyses steps conduced in a non-reproducible manner can be thoroughly described and reported (Gomes et al., 2022).

Another way to increase reproducibility is by verifying that the code runs with no errors (Ivimey-Cook et al., 2023). These errors include syntax errors (e.g. a missing parenthesis), programming errors (e.g. inverted row and column coordinates), or missing or wrong versions of libraries.

Third, reproducibility includes ensuring that running the code yields the conclusions supported by the publication (Ivimey-Cook et al., 2023), meaning that data analysis code should fulfill its intended purpose with no errors (Popovic et al., 2024; Ivimey-Cook et al., 2023). The increasing use of code to analyze data has the potential to advance science, but can also lead to new sources of errors that need to be accounted for (Ivimey-Cook et al., 2023). An example of coding error in the scientific literature is related to misspelled genes names due to automatic type conversion when entering these names in Microsoft Excel (e.g. gene *MARCH1* being converted to the date "Mar-1"). A study by Ziemann et al. (2016) examining the supplementary material of published articles in genomics found that about 20% of articles included gene names errors attributable to Microsoft Excel. This high prevalence of errors prompted the Human Gene Name Consortium to rename some of the offending genes in 2020: for example, *MARCH1* became *MARCHF1* (Vincent, 2020).

Finally, writing clean code, i.e. code that is easily readable and understandable by humans, is also useful to avoid coding errors (Filazzola & Lortie, 2022). Writing clean code includes choosing relevant variable names, organizing analyses in a clear way, commenting and documenting code and avoiding code duplication (Filazzola & Lortie, 2022). Writing clean code is not only beneficial for future code reviewers or readers, but also for the author of the code themselves if they need to reuse or modify their

code some time after writing it. The process of cleaning and testing code can also lead to the detection of errors in the analyses that would not have been spotted otherwise (Filazzola & Lortie, 2022). For example, during this thesis, I realized I had made a mistake in the code to analyze data with the linear MHP (chapter 3.1) while cleaning code.

Using code to analyze results makes it an integral part of the research process, so best coding practices contribute to the quality of scientific research. We can also view code as a way to communicate scientific results in its own right (Filazzola & Lortie, 2022). Therefore, code can be included in the material involved in open science alongside data and publications.

2.3 Open science, code and data

Open science is the movement advocating for free and open access to scientific results (Racimo et al., 2022). It is increasingly recognized as an efficient means to advance scientific knowledge (Gomes et al., 2022): open science benefits researchers, who can access research findings freely and rapidly and avoid to "reinvent the wheel" (Hampton et al., 2015; Racimo et al., 2022), but it also benefits society at large, as anyone from the general public that might be interested in scientific results can access them.

Opening code increases the likelihood to spot analysis and coding errors (Ivimey-Cook et al., 2023). In research articles, it is commonly accepted that the material and methods should be reported in sufficient details in order to allow anyone to critically examine the research protocol, and to reproduce or replicate results. In the context of increasing use of code, this includes accurately describing the coding procedure, and even better, including the code as supplementary material in publications. Opening code for scrutiny has the potential to improve the reliability of research results. The experience related by Andrew Anderson is eloquent in this respect: in this story published on the SORTEE (Society for Open, Reliable, and Transparent Ecology and Evolutionary biology) Blog (Anderson, 2023), he relates how he fixed a coding error in a published article. In 2022, he published an article about sexual selection in primates, including open code and data. One year after publication, a reader reached out by email pointing out an error they had spotted in the data analysis code that substantially affected the study conclusions. The article was then retracted, and replaced with

another article with correct analyses.

Open code and data also have the potential to accelerate the spread of research results and ideas, which can be essential to act fast in response to crises (Racimo et al., 2022). For example, during the COVID-19 crisis, preprint servers have been increasingly used to disseminate scientific results and accelerate research in the face of the pandemic (Fraser et al., 2021; Gomes et al., 2022).

2.4 In this thesis

For all reasons above, in this thesis I made an effort to write reproducible code and to open and document the code and the underlying data (see List of online resources). All code for part 2 is stored in a GitHub repository (which is private at the moment). The published article about the linear MHP from chapter 3.1 was published with open code and data, and the complete code for the analyses performed in part 3 (updated with code from chapter 3.2 to analyze non-linear MHPs) is available on GitHub. The source code of software applications developed in chapters 4.1 and 4.2 are available on GitHub, and the Shiny application is also available freely on a server.

To make code readable, I distributed the analyses in computational documents created using Quarto. I also made an effort on documentation, using a global README to document all code repositories and additionally using documentation websites for chapters 4.1 and 4.2. All code repositories are distributed as R packages, which facilitate dependency management and function documentation. For software tools developed in chapters 4.1 and 4.2, I also used unit tests to check results.

2.5 Additional remarks

Opening code and data entails many benefits for the scientific community. Sometimes, sharing code and data may not be possible (e.g. proprietary software) or desirable (e.g. sensitive biodiversity data on poached species) (Jenkins et al., 2023). However, these exemptions do not prevent to archive code privately, and code can still be cleaned and documented.

Here, I also want to stress that, as Gomes et al. (2022) say it, “there is no such thing as ‘perfect code”’. Writing clean code is a continuous process, and the code can

always be improved. In the end, clean coding conventions rest on a trade-off balancing, among other things, readability, efficiency, longevity and invested cleaning time, and also depends on arbitrary preferences between coding standards. It is particularly important to recognize this last point, because one of the barriers to sharing code and data is a fear of being judged because one codes "poorly" or does not use sophisticated coding tools (Gomes et al., 2022; Hampton et al., 2015). Another factor is the fear of exposing coding errors (Hampton et al., 2015). However, even though exposing errors can be hard (Anderson, 2023), it is in the interest of scientific research: therefore, I think it is essential to develop a culture of sharing code and data in ecology, as well as fostering a positive and kind research environment to mitigate the fear of being judged (Gomes et al., 2022).

3 Model validation: with simulations or real data?

For the statistical models developed in this thesis (models of the CA family in part 2 and Hawkes processes in part 3), I validated models using both simulation approaches and real data analyses. Here, I define model validation as the testing procedures conduced to determine how well the model performs, and whether it is suitable to estimate the parameters of interest for real data (after Rykiel, 1996). In this section, I reflect on the pros and cons of simulation and real data analysis to evaluate models.

3.1 Validation with a simulation approach

The simulation approach consists in simulating the ecological process and the data collection, fitting the model, and comparing the model parameters to the simulated ground truth: this is the "virtual ecologist approach" described by Zurell et al. (2010).

Validating models with this approach has several advantages. First, simulating data allows to know the ground truth for the parameters values, which is not possible in general for real data. For instance, in chapter 2.1, we simulated a probability of interaction based on trait matching between species: this probability is unknown in the real dataset. Second, simulations allow to test the model performance under different conditions. For example, the simulation of the linear MHP in chapter 3.1 showed how the model's performance changes with different sample sizes and interaction strength

and allowed to determine conditions for which we can expect the model to produce a reliable inference. Finally, simulations can also be used to choose the model's hyperparameters. For instance, I used simulations for the linear MHP in chapter 3.1 to choose a suitable value of the penalization parameter γ: without this simulation approach, the value of γ would have been a random guess.

Although simulations can be very useful, they also have limits. First, defining realistic simulation parameters can be challenging (Gotelli & Ulrich, 2012). For instance, to simulate data for the MHPs in part 3, I had to proceed by trial and error to determine the ecologically realistic range of interaction strengths. Second, the simulation model can also be computationally intensive or challenging to implement. For example, again regarding the validation of the MHPs in part 3, I used the same model for simulation and inference because implementing an individual-based model would have been too time-consuming. Another limit of simulations is that they only inform us on the performance of the model we can expect under ideal conditions, i.e. with no factors other than the ones used in the simulation model. Therefore, evaluating models using simulations provides an upper bound for the performance we can expect on real, noisy, data. For instance, the simulation of species interactions in chapter 2.1 assumes a normal interaction niche, which is not necessarily true for real species.

3.2 Validation using real data

Considering the limits of simulation approaches, it is tempting to use real data to validate the model. A major limit of real data is that the ground truth is unknown: but we can use different criteria to validate the model. First, considering a statistical model that is fitted on data, we can verify that it converges on real data, i.e. that the optimization algorithm finds an optimal solution. In particular, the model might fail to converge when fitted on too sparse data (this is the case for example for the non-linear MHP presented in chapter 3.2). We can also check the model outputs for unexpected patterns, when a priori knowledge on the system allows it. Indeed, some factors that are ignored in the model can affect the patterns observed in real data: spotting these unexpected effects can hint to an unmodeled effect and guide model specification and interpretation. For instance, in chapter 3.1, the inference of a MHP on real data yielded a cyclic interaction between some species. This led us to suspect an effect of circadian rhythms, and

prompted the design of a simulation experiment to check if circadian rhythms could produce such observed patterns (and they could, see Appendix S3). I think that testing models on real data is essential: after all, the aim of ecological models, such as the ones developed in this thesis, is to understand and analyze real data.

3.3 Two complementary approaches and other perspectives

I think that evaluating models on simulated and real data are highly complementary approaches in the context of ecological data analysis. On one hand, simulation confers more control on the data and allows to check the model performance in ideal conditions. However, good model performances under simulation do not guarantee that the model is flawless. Indeed, the model might perform well with simulated data, but fail to account for a critical ecological process, or perform well only for ecologically unrealistic parameter values. On the other hand, an application on real data is essential to test whether the model is able to infer the patterns of interest despite the noise owing to the multiple factors at play in ecological systems. However, with real data, we do not control factors affecting the observed patterns, and some model parameters might be wrongly estimated because of noise introduced by unmodeled variables. For instance, the avoidance of lion by zebra detected in chapter 3.1 might be due to the time of day, which is not taken into account in the model. Therefore, simulation and real data analysis are both useful to validate a model.

Another important characteristic of a model is its conceptual validity (sensu Rykiel, 1996), i.e. whether the model is an accurate description of real world mechanisms. In this thesis, I did not explicitly assess the conceptual validity of models: I did not test the assumption that species traits generate structure in interaction networks (part 2) or that interspecific attraction/avoidance generate non-random patterns in occurrence data (part 3). To evaluate models conceptual validity, it could be useful to compare them to other models. In particular, I could have used null or neutral models: null models are based on randomization of real data to produce a pattern that is expected in the absence of a given mechanism (Gotelli & McGill, 2006), and neutral models allow to simulate data without taking a given mechanism into account (Gotelli & McGill, 2006). These models allow to evaluate the effect of a mechanism, by comparing real data to data generated with models where the mechanism is absent. However, null and

neutral models are difficult to design, because it is difficult to account for all processes affecting the observed patterns besides the mechanism we are interested in. Indeed, if some structuring factors are omitted from the null or neutral model, real data might hold more structure than simulated/randomized data, even if the mechanism we want to test has no importance (Gotelli & Ulrich, 2012).

More broadly, we can also ask ourselves whether scientific models can be good descriptors of the real world, and how accurate this description is. Below, I briefly discuss this issue and its implications in the frame of this thesis.

4 Modeling ecological data

Models are an important component of scientific process. Scientific models are extremely diverse (Frigg & Hartmann, 2020; Hughes, 1997; Varenne, 2013): a few examples of models include the model of the atom in physics, the Lotka-Volterra model of prey-predator dynamics in ecology or the plant *Arabidopsis thaliana* as a model organism in biology. Owing to this diversity, giving a general definition of scientific models is difficult, and some have even argued it is not relevant (Callender & Cohen, 2005). One of the common determinants of models is that they are tools to understand the real world, often (but not always) by means of simplified representation of a phenomenon (Varenne, 2013; Legay, 1997).

In this thesis, I developed and studied two main models: a model to infer species interaction niches (part 2) and a model to infer and simulate attraction and avoidance between species (part 3). Both models rely on mathematical equations to formalize the study system, and both models can be fitted to data to infer information about real systems. This section focuses on the nature of such models, and on what they can and cannot teach us on natural systems.

An important question regards the relationship of models with the real world and can be put as follows: what kind of representation of the real world do models provide, and how accurate is it? This question has spurred much debate, and there is no philosophical consensus on the answer to these questions today (Frigg & Nguyen, 2016). Different authors have proposed different conceptions of the link between scientific models and reality: two views that are particularly relevant for ecology are the

axiomatic and the pragmatic view (see Figure 5.2) (Travassos-Britto et al., 2021).

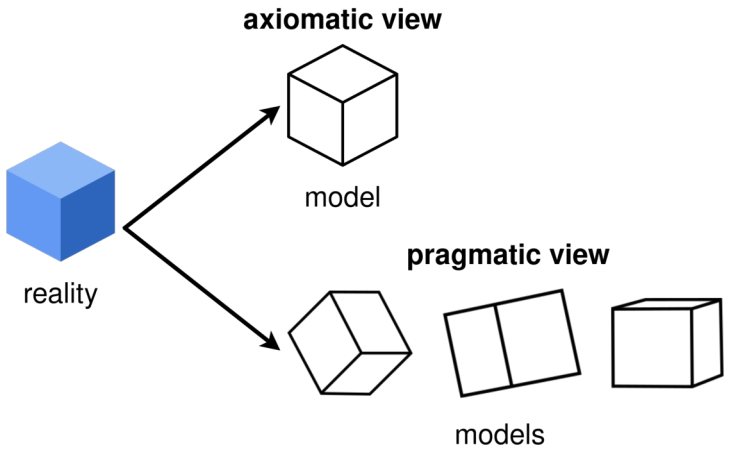


Figure 5.2: Axiomatic and pragmatic view of models. The axiomatic view postulates that the model is an approximation of the reality, while the pragmatic view adds that models also represent a point of view of a given phenomenon occurring in reality. Cube for reality and axiomatic view by kmg design. Other cubes are own designs.

The axiomatic view posits that models can be formulated with a set of axioms, i.e. statements about a phenomenon, and expressed with a model that will then be tested against the real phenomenon (Travassos-Britto et al., 2021). To exemplify how this reasoning can apply, I take the example of the Hawkes model used in part 3. We can formulate the axiom : "some species attract or avoid each other". This axiom can be translated into the equation of the Hawkes process, modeling the occurrence rate of a species $\lambda_i(t)$ as a function of the occurrences of other species: for instance, with the linear MHP, the model is described with Equation (2) of chapter 3.1: $\lambda_i^l(t) = \left(\nu_i + \sum_{j=1}^{S} \sum_{m \mid T_m^{lj} < t} f_{j \to i}(t - T_m^{lj})\right)_+$. Then, this model can be fitted to real data of species occurrences and we can see if, and which species attract or avoid each other.

There is a fundamental philosophical issue with the axiomatic view applied in this context. Indeed, axioms tend to simplify the system by stressing one particular aspect of the system functioning (Travassos-Britto et al., 2021). For instance, here, considering species simply in terms of occurrences dictated by interactions with other species neglects other aspects of the ecological system like the environment or the personality of each individual, which can both affect species occurrences. Is it to say that this model is useless because it does not depict the real world accurately? I think that the answer is no, and from a philosophical perspective this issue is resolved by adopting a pragmatic view.

The pragmatic view shifts the emphasis of the model as a representation of reality

to a representation of a given phenomenon about reality: Travassos-Britto et al. (2021) summarize this difference between the axiomatic and the pragmatic view by stating that "[the axiomatic view defines] the notion of a 'model of something', stressing how models represent the world, [and the pragmatic view defines] the notion of 'model for something', laying emphasis on how models are used to learn about the world". Taking the example of the Hawkes model, it is designed specifically to study the effect of previous species occurrences on the occurrences of a given species.

I think that the models developed in this thesis fit the pragmatic view. In part 2, the objective was to quantify measures of the interaction niche in networks measured as the diversity of (latent) traits of their interacting partners. Other phenomenons influence the probability of interactions (e.g. phylogeny or seasonality), and they are not explicitly modeled or taken into account in this part. This model also constrains the shape of species niches to be hyperellipsoids. In part 3, the objective was to quantify the effect of previous species occurrences on the occurrence of a given species. Other phenomenons were neglected as well (e.g. environment, personality, spatial aspect), and the mathematical model constraints the shape of the effect of the interaction (e.g. additive effect in chapter 3.1 or multiplicative effect in chapter 3.2).

Under the pragmatic view, models are not necessarily isomorphic representations of the real world, i.e. models structure does not necessarily represent the causal relationships occurring in real systems. Therefore, the question of what exactly these models can and cannot tell us about the causal relationships occurring in the real world is crucial. In particular, two questions are of interest: evaluate the agreement of the model with data, and quantify how the model's results might generalize. Evaluating the agreement of a model with data is the domain of statistics (Romeijn, 2014). Several statistical methods exist to do so, for instance frequentist hypothesis testing, Bayesian inference, and models comparison (Romeijn, 2014; Tredennick et al., 2021): in this thesis, I evaluated models with simulations and inference on real data (see section 3). Another open question is the generalization of the model to other study systems or settings. In this thesis, I inferred interaction niches of bird and plant species using data collected in the Peruvian montane forest between 2009 and 2010 (part 2), and the attraction and avoidance between 5 mammal species from 6 protected areas in South African savannas collected between 2017 and 2019 (part 3). Inferences were con-

duced on these particular datasets, and their results and statistical validation applies to these instances only. However, the results of these models would be more useful if they could be generalized: for example, in chapter 2.2, I estimated the inertia explained by trait matching in the Peruvian network to approximately 6%, and an open question is to determine if this low percentage is common across all bird-fruit mutualistic networks, or even across all ecological networks. Similarly, in chapter 3.1, we inferred an avoidance of lion by zebra and impala in the studied sites, and we can wonder if this results holds true for other locations. To generalize these conclusions, we would need to replicate the study by applying the models to other data (see section 2 for a definition and discussion of scientific replication).

Adopting a pragmatic view of models has practical consequences for the formalization of scientific models. At the beginning of my thesis, I wanted to build a "perfect" model to represent species occurrences at camera traps. So when I realized that the Hawkes model lacked some explanatory factors affecting species occurrences, I was disappointed and thought that this model was useless to draw reliable inferences on ecological data. But by adopting the pragmatic view, I realized that, even if some aspects of the system are omitted, a model can still inform us about the phenomenon we want to study. Under the pragmatic view, a more complete knowledge of the system can then be obtained by confronting models with different assumptions to data (Travassos-Britto et al., 2021). Therefore, a single model (and a single modeler) does not necessarily need to account for all factors in a system, nor do they need to be a specialist in every aspect of the functioning of a given system.

5 Generalism and specialization in science

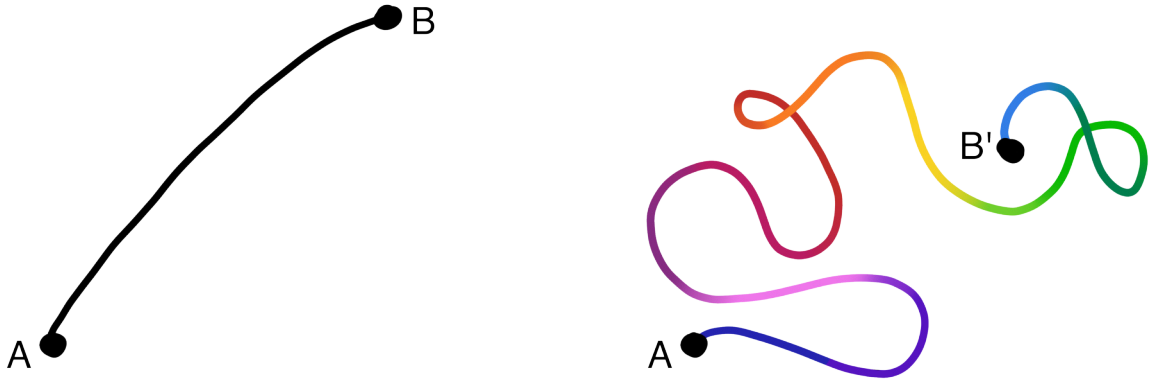
This thesis is focused on developing tools to improve ecological data analysis, either by using statistical models or software tools. All axes study complex systems, and in particular ecological communities and their multiple species (parts 2, 3 and 4) and multiple spatial and temporal sampling units (3 and 4). However, this thesis also gathers three diverse parts. They differ in the scientific question addressed: how to explain the structure of networks (part 2), what effects do they have (part 3) and how to improve data processing for community data (part 4). Study systems are also varied: Peruvian

bird-fruit network (part 2), African mammals community (part 3) and any animal community investigated with camera traps (part 4). Finally, the methods that are used and developed in each part are also distinct: multivariate methods in part 2, point process models in part 3 and software tools in part 4.

This multiplicity of approaches was not a choice: rather, it was driven by the contingencies of the thesis project and unexpected developments compared to the initially planned project. It was also driven by our skills and interests, as part 2 was suggested by one of my supervisors and part 4 arose from my will to conduct more applied research.

Deviating from a planned research project (see Figure 5.3) seems to be very common, if I trust the numerous discussions I had with other scientists (see also the personal introduction in Barabási and Pósfai, 2016). However, this reality is rarely apparent in the way that science is shared, with the most prominent form of research dissemination, research articles, sharing a common and rigid structure (introduction, material and methods, results and discussion). I understand the need to present research linearly and to discard some of the explored paths for clarity. But at the same time, this homogenization of research findings and the current publishing system combine to yield unwanted effects: notably, because the final research product tends to focus on significant results (by performing HARKing, or Hypothesizing After the Results are Known Kerr, 1998), because it tends to overemphasize results significance (Corneille et al., 2023), and more largely because the way science is presented in articles largely distorts the way science is practiced.

This thesis is no exception, and while I presented results linearly, here I acknowledge that this is a simplification of the unfolding of this PhD. The planned research project was organized to proceed linearly from point A to point B (Figure 5.3a), but it is very different from the work presented in this manuscript, that is the result of trial and error, explored and abandoned research axes and unexpected results and developments. This journey leads to results B' that are different from the planned results (Figure 5.3b). Moreover, my experience of pursuing a research project was not neutral: even though as a scientist, my research is based on logic and reason, there was also a component of personal and emotional investment in this research. Succeeding to write a proof, or obtaining an ecologically meaningful result brought me a lot of joy,



(a) Expected research trajectory

(b) Real research trajectory

Figure 5.3: Real and expected research trajectories. At the beginning of this PhD, I was expecting to pursue a linear research trajectory (a), but the trajectory I followed in the end was much more convoluted and colorful due to personal and emotional investment in research results (b).

while being stuck on an equation or a bug in the code was extremely frustrating.

This convoluted progression, as well as the position of my work at the interface between ecology and statistics, was initially the source of much doubt and discomfort. This was in contradiction with the idea of research that I had in mind, which was following a linear path towards increasingly specialization in a given subject, its literature and associated methods. In the contrary, as I progressed in my PhD, I was continuously taking on new methods and study systems.

Now that I approach the end of this PhD, I reflect on this experience much more positively. I think that a diversity of approaches and methods can increase one's understanding of a subject (Legay, 1986), and that generalist scientific culture is important to link seemingly unrelated subjects. I also see my position at the interface of statistics and ecology more positively. Indeed, interdisciplinary research in statistics and ecology can be very fruitful (King, 2014; Gimenez et al., 2014; Warton, 2015), but to be valuable, statisticians and ecologists must understand one another, by using the same language and understand biological or statistical constraints. Sometimes, people knowing concepts between two different fields can help bridge the gap between disciplines. For instance, causal models, which are increasingly advocated in ecology, require not only statistical skills, but also ecological knowledge of the system (Arif & MacNeil, 2022; McGowan et al., 2023). This came with the growing realization that science is not necessarily an individual endeavor, but can be conduced in teams: so it is not necessary to know everything to make good science.

Bibliography

Anderson, A. (2023, December). Setting the record straight: How data and code transparency caught an error and how I fixed it [Section: blog]. Retrieved December 6, 2023, from https://www.sorte e.org/blog/2023/12/04/2023_setting_record_straight/

Arif, S., & MacNeil, A. (2022). Predictive models aren't for causal inference. *Ecology Letters*, *25*(8), 1741–1745. https://doi.org/10.1111/ele.14033

Barabási, A.-L., & Pósfai, M. (2016). *Network science* [OCLC: ocn910772793]. Cambridge University Press. http://networksciencebook.com/

Callender, C., & Cohen, J. (2005). There Is No Special Problem About Scientific Representation. *Theoria*, *55*(1), 67–85. Retrieved May 6, 2024, from http://philsci-archive.pitt.edu/2177/

Corneille, O., Havemann, J., Henderson, E. L., IJzerman, H., Hussey, I., Orban de Xivry, J.-J., Jussim, L., Holmes, N. P., Pilacinski, A., Beffara, B., Carroll, H., Outa, N. O., Lush, P., & Lotter, L. D. (2023). Beware 'persuasive communication devices' when writing and reading scientific articles. *eLife*, *12*, e88654. https://doi.org/10.7554/eLife.88654

Culina, A., Berg, I. v. d., Evans, S., & Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. *PLOS Biology*, *18*(7), e3000763. https://doi.org/10.1371/journal.pbio.3000763

Dehling, D. M., Bender, I. M. A., Blendinger, P. G., Muñoz, M. C., Quitián, M., Saavedra, F., Santillán, V., Böhning-Gaese, K., Neuschulz, E.-L., & Schleuning, M. (2021). ANDEAN frugivory: Data on plant–bird interactions and functional traits of plant and bird species from montane forests along the Andes. https://doi.org/10.5061/DRYAD.WM37PVMN5

Filazzola, A., & Cahill Jr, J. F. (2021). Replication in field ecology: Identifying challenges and proposing solutions. *Methods in Ecology and Evolution*, *12*(10), 1780–1792. https://doi.org/10.1111/2041-210X.13657

Filazzola, A., & Lortie, C. (2022). A call for clean code to effectively communicate science. *Methods in Ecology and Evolution*, *13*(10), 2119–2128. https://doi.org/10.1111/2041-210X.13961

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., & Coates, J. A. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology*, *19*(4), e3000959. https://doi.org/10.1371/journal.pbio.3000959

Frigg, R., & Hartmann, S. (2020). Models in Science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020). Metaphysics Research Lab, Stanford University. Retrieved March 23, 2022, from https://plato.stanford.edu/archives/spr2020/entries/models-science/

Frigg, R., & Nguyen, J. (2016). Scientific Representation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. Retrieved May 6, 2024, from https://plato.stanford.edu/archives/win2021/entries/scientific-representation/

Gimenez, O., Buckland, S. T., Morgan, B. J. T., Bez, N., Bertrand, S., Choquet, R., Dray, S., Etienne, M.-P., Fewster, R., Gosselin, F., Mérigot, B., Monestiez, P., Morales, J. M., Mortier, F., Munoz, F.,

Ovaskainen, O., Pavoine, S., Pradel, R., Schurr, F. M., . . . Rexstad, E. (2014). Statistical ecology comes of age. *Biology Letters*, *10*(12), 20140698. https://doi.org/10.1098/rsbl.2014.0698

Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martinez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., & Gaynor, K. M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1987), 20221113. https://doi.org/10.1098/rspb.2022.1113

Gotelli, N. J., & McGill, B. J. (2006). Null Versus Neutral Models: What's The Difference? *Ecography*, *29*(5), 793–800. https://doi.org/10.1111/j.2006.0906-7590.04714.x

Gotelli, N. J., & Ulrich, W. (2012). Statistical challenges in null model analysis. *Oikos*, *121*(2), 171–180. https://doi.org/10.1111/j.1600-0706.2011.20301.x

Hampton, S. E., Anderson, S. S., Bagby, S. C., Gries, C., Han, X., Hart, E. M., Jones, M. B., Lenhardt, W. C., MacDonald, A., Michener, W. K., Mudge, J., Pourmokhtarian, A., Schildhauer, M. P., Woo, K. H., & Zimmerman, N. (2015). The Tao of open science for ecology. *Ecosphere*, *6*(7), art120. https://doi.org/10.1890/ES14-00402.1

Hughes, R. I. G. (1997). Models and Representation. *Philosophy of Science*, *64*(S4), S325–S336. https://doi.org/10.1086/392611

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ivimey-Cook, E. R., Pick, J. L., Bairos-Novak, K. R., Culina, A., Gould, E., Grainger, M., Marshall, B. M., Moreau, D., Paquet, M., Royauté, R., Sánchez-Tójar, A., Silva, I., & Windecker, S. M. (2023). Implementing code review in the scientific workflow: Insights from ecology and evolutionary biology. *Journal of Evolutionary Biology*, *36*(10), 1347–1356. https://doi.org/10.1111/jeb.14230

Jenkins, G. B., Beckerman, A. P., Bellard, C., Benítez-López, A., Ellison, A. M., Foote, C. G., Hufton, A. L., Lashley, M. A., Lortie, C. J., Ma, Z., Moore, A. J., Narum, S. R., Nilsson, J., O'Boyle, B., Provete, D. B., Razgour, O., Rieseberg, L., Riginos, C., Santini, L., . . . Peres-Neto, P. R. (2023). Reproducibility in ecology and evolution: Minimum standards for data and code. *Ecology and Evolution*, *13*(5), e9961. https://doi.org/10.1002/ece3.9961

Kelly, C. D. (2019). Rate and success of study replication in ecology and evolution. *PeerJ*, *7*, e7654. https://doi.org/10.7717/peerj.7654

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

King, R. (2014). Statistical Ecology. *Annual Review of Statistics and Its Application*, *1*(Volume 1, 2014), 401–426. https://doi.org/10.1146/annurev-statistics-022513-115633

Legay, J.-M. (1986). Quelques réflexions à propos d'écologie : Défense de l'indisciplinarité. *Acta oecologica. Série Oecologia generalis*, *7*(4), 391–398.

Legay, J.-M. (1997). *L'Expérience et le modèle* [Bibliographie_available: 0 Cairndomain: www.cairn.info Cite Par_available: 0 Publisher: Éditions Quæ]. INRA. Retrieved March 23, 2022, from http://www.cairn.info/l-experience-et-le-modele--9782738007780-page-9.htm

Markowetz, F. (2017). All biology is computational biology. *PLOS Biology*, *15*(3), e2002050. https://doi.org/10.1371/journal.pbio.2002050

McGowan, L. D., Gerke, T., & Barrett, M. (2023). Causal inference is not just a statistics problem. *Journal of Statistics and Data Science Education*, *0*(ja), 1–9. https://doi.org/10.1080/26939169.2023.2276446

Nicvert, L., Donnet, S., Keith, M., Peel, M., Somers, M. J., Swanepoel, L. H., & Venter, J. A. (2023). Code and data for: Using the multivariate Hawkes process to study interactions between multiple species from camera trap data [Publisher: figshare]. Retrieved November 13, 2023, from https://doi.org/10.6084/m9.figshare.24552157.v3

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Popovic, G., Mason, T. J., Drobniak, S. M., Marques, T. A., Potts, J., Joo, R., Altwegg, R., Burns, C. C. I., McCarthy, M. A., Johnston, A., Nakagawa, S., McMillan, L., Devarajan, K., Taggart, P. L., Wunderlich, A., Mair, M. M., Martínez-Lanfranco, J. A., Lagisz, M., & Pottier, P. (2024). Four principles for improved statistical ecology. *Methods in Ecology and Evolution*, *15*(2). https://doi.org/10.1111/2041-210X.14270

Racimo, F., Galtier, N., De Herde, V., Bonn, N. A., Phillips, B., Guillemaud, T., & Bourget, D. (2022). Ethical Publishing: How Do We Get There? *Philosophy, Theory, and Practice in Biology*, *14*(0). https://doi.org/10.3998/ptpbio.3363

Romeijn, J.-W. (2014). Philosophy of Statistics. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2020). Metaphysics Research Lab, Stanford University. Retrieved May 6, 2024, from https://plato.stanford.edu/archives/spr2020/entries/statistics/

Rykiel, E. J. (1996). Testing ecological models: The meaning of validation. *Ecological Modelling*, *90*(3), 229–244. https://doi.org/10.1016/0304-3800(95)00152-2

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology*, *69*(1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Travassos-Britto, B., Pardini, R., El-Hani, C. N., & Prado, P. I. (2021). Towards a pragmatic view of theories in ecology. *Oikos*, *130*(6), 821–830. https://doi.org/https://doi.org/10.1111/oik.07314

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, *102*(6), e03336. https://doi.org/10.1002/ecy.3336

Varenne, F. (2013). Modèles et simulations dans l'enquête scientifique : Variétés traditionnelles et mutations contemporaines [Pages: pp. 11]. In *Modéliser & Simuler. Epistémologies et pratiques de la modélisation et de la simulation* (Vol. 1). Editions Matériologiques. Retrieved March 23, 2022, from https://halshs.archives-ouvertes.fr/halshs-03080674

Vincent, J. (2020). Scientists rename human genes to stop Microsoft Excel from misreading them as dates. *The Verge*. Retrieved May 1, 2024, from https://www.theverge.com/2020/8/6/21355674/human-genes-rename-microsoft-excel-misreading-dates

Warton, D. I. (2015). New opportunities at the interface between ecology and statistics. *Methods in Ecology and Evolution*, *6*(4), 363–365. https://doi.org/10.1111/2041-210X.12345

Ziemann, M., Eren, Y., & El-Osta, A. (2016). Gene name errors are widespread in the scientific literature. *Genome Biology*, *17*(1), 177. https://doi.org/10.1186/s13059-016-1044-7

Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B., & Grimm, V. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, *119*(4), 622–635. https://doi.org/10.1111/j.1600-0706.2009.18284.x