

Parametric and nonparametric approaches to explain and predict nonlinear population dynamics in changing environments

by

Simone Cenci

B.S in Physics and Astrophysics, University of Rome "La Sapienza", 2012
M.S in Physics, University of Pisa, 2015

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Civil and Environmental Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

©Massachusetts Institute of Technology. All rights reserved

Signature redacted

Signature of Author

Department of Civil and Environmental Engineering

Signature redacted May 16, 2019

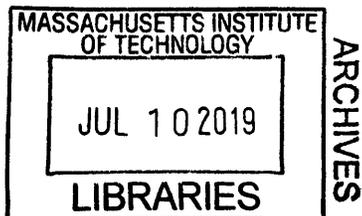
Certified by

Serguei Saavedra,
Assistant Professor, Department of Civil and Environmental Engineering
Thesis Supervisor,

Signature redacted

Accepted by

Heidi Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering,
Chair, Graduate Program Committee



Parametric and nonparametric approaches to explain and predict nonlinear population dynamics in changing environments

by

Simone Cenci

Submitted to the Department of Civil and Environmental Engineering On May 16, 2019, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil and Environmental Engineering

Abstract

Many aspects of human societies, from the sustenance of national economies to the control of population health, depend on the dynamics of biological populations within a given environment. Therefore, understanding and predicting the effects of changing environments on the dynamics of biological populations evolving in a continuously changing world is, nowadays, one of the most important challenges in biology.

In this thesis we have addressed this challenge using two different approaches. The first approach, called the structural approach, is deductive, i.e., the effects of changing environments on population dynamics are studied using parametric models under equilibrium assumptions. In this context, firstly we have shown that, while the approach was originally introduced to investigate the structural stability of the classic Lotka-Volterra dynamics; it can be applied to a much larger class of nonlinear models and to stochastic systems. Then, we used the approach to analyze empirical data to investigate how structure and dynamics of species interactions regulate species coexistence under fast environmental changes. The generalizability of this approach however, has been limited because equilibrium dynamics are seldom observed in nature and exact equations for population dynamics are rarely known.

Therefore, in the second part of the thesis we took an inductive approach. Specifically, we proposed a nonparametric framework to estimate the tolerance of non-equilibrium population dynamics to environmental variability. To apply the framework on empirical data we have improved/developed nonparametric computational methods to infer biotic interactions and their uncertainty from nonlinear time series data. Using our approach we were able to recover important ecological insights without the explicit formulation of parametric models. That is, we have shown that it is possible to build ecological theories inductively from observational data with minimal assumptions on the data-generating processes.

Overall, we believe that the increasing amount of biological data available nowadays paves the way for moving theoretical population biology from being a deductive, assumption driven science towards an inductive data-driven science. In this context, this study is a step forward towards the foundations of a nonparametric data-driven research to monitor and anticipate the response of populations to the increasing rate of environmental changes.

Thesis advisor: Serguei Saavedra

Title: Assistant Professor of Civil and Environmental Engineering

*Alla mia famiglia ora arricchita
dal piccolo Leonardo*

Acknowledgments

This is a fantastic occasion to say thank you to all those people that have been a great help and support during my studies and my life in the past few years. I want to start by thanking my parents. Their unconditional support of all my choices and my interests is the reason why I am having a very full, happy, fun and gratifying life. To them I owe all I have and there is nothing I can do and say to express my gratitude and love. I also want to thank Luca, my brother, for being there for me whenever I need it and for the uncountable experiences, conversations, and laugh we ever had. It is a luck to have a brother like him and a sister-in-law like Pina. I hope I can be the best uncle for their new little kid, Leonardo, and for all of those to come. Finally, I want to thank all my friends from Italy that are always there whenever I go back or are always there with me wherever I am. It is not easy to maintain friendships an ocean away but with them it come natural and I feel very lucky for this.

Let's know cross and ocean make landfall in the US. MIT is an amazing place that changed my way of looking at the world. Obviously, it is not the place that makes a difference but is the people I have found. Therefore, I want to thank all the people that I have met here and that become my very good friends. In particular I want to thank the Germans, Tim and Sandra, for all the fun we had together and the very special friendship we have built. Then I want to thank Chuliang, Lucas and Mohammad for the agreement and disagreement we always have on scientific problems. The discussion with them is always stimulating and educational. In particular I want to thank the oldest member of the group, Chuliang, whom has really contributed a lot to the thoughts that helped me writing this thesis. Finally, I want to thank Maria, Stephan, Diego, Marco, Enrico, Flavia, Fabiola and Justin for all the great fun. Last but surely not least I need to thank the MIT sailing club which made my summers here in Boston.

I want to thank the committee members starting with Mick Follows and Michael Neubert for their questions, suggestions and general support to my research. In particular, I want to thank George Sugihara for his enormous support to my research. Besides being an incredible person he is an unique scientist from which I have learned a completely new way of thinking about scientific questions. I was very lucky to have such a scientist as a mentor in my studies.

Finally, I am left with the two biggest acknowledgments. Firstly I want to start with my advisor, Serguei. To be honest I should really cut the acknowledgments here because if I really start enumerating everything he did for me in the last three years I will never stop writing. However, I want to try to summarize my acknowledgments with a thought experiment. If you sit with pencil and paper and creatively try to design a perfect advisor for a you as a young researcher this is what I bet you will probably create. A person who leaves you the freedom to ask your own questions but guides you whenever your inexperience is driving you towards dangerous paths. This ideal advisor should never put his/her interests before yours and should

let you prove him/her wrong or right time and time again. He/she should help you in your future career by actively supporting your choice and he/she should also be a good friend with whom you can laugh and talk about everything is in your mind. Well, this imaginary advisor was my advisor and I am incredibly lucky and grateful for this.

Last by again surely not least it is time to thank my beautiful wife. Here really I do not know where to start. She left her job and cross an ocean to come to marry me in California and to start a new life here in Boston. It was incredibly hard for her but she was unbelievably strong. We had so many adventures together and so much fun that I find hard to imagine anyone more lucky than I am because of her. I hope that I can always give her what she deserves and at least a fraction of what she gives to me. None of what I am or ever will be would be possible without her, so thank you Polly.

List of publications

[8] **S. Cenci**, G. Sugihara, and Serguei Saavedra *Predictability of nonlinear dynamics in changing environments* Submitted

[7] **S. Cenci** and S. Saavedra *Nonparametric estimation of the structural stability of non-equilibrium community dynamics* *Nature Ecol. Evol.* (2019)

[6] **S. Cenci**, G. Sugihara and S. Saavedra *Regularized S-map for inferring and forecasting with noisy ecological time series* *Methods Ecol. Evol.* **10**:650–660 (2019)

[5] **S. Cenci** and S. Saavedra *Uncertainty quantification of the effect of biotic interactions on community dynamics from nonlinear time series data* *J. Royal Soc. Interface*, **15**: 20180695 (2018)

[4] **S. Cenci***, C. Song* and S. Saavedra *Rethinking the importance of the structure of ecological networks under an environment-dependent framework* *Ecol.Evol.* **8**:6852–6859 (2018)

[3] **S. Cenci** and S. Saavedra *Structural Stability of Nonlinear Population Dynamics* *Phys. Rev. E* **97**, 012401 (2018)

[2] **S. Cenci**, A. Montero-Castaño and S. Saavedra *Estimating the effect of the reorganization of the interactions on the adaptability of species to changing environments* *J.Theo.Bio* **437**: 115-125 (2018)

[1] S. Saavedra, **S. Cenci**, E. del-Val, K. Boege and R. P. Rohr *Reorganization of interaction networks modulates the persistence of species in late successional stages* *J. of Animal Ecology* **86**: 1136-1146, (2017)

CONTENTS

1	Introduction	13
1.1	The role of biological populations in human societies	13
1.2	Populations dynamics in changing environments	16
1.2.1	Determinism or stochasticity?	16
1.2.2	Biotic or abiotic drivers?	17
1.2.3	The need for models	17
1.2.3.1	Parametric approaches: assuming the model from theories	18
1.2.3.2	Nonparametric approaches: learning the model from data	21
1.3	Thesis statement	23
1.4	Summary	24
1.4.1	Structure of the thesis	24
I	Parametric approaches	27
2	Structural stability of nonlinear population dynamics	29
2.1	Introduction	30
2.2	Background: the structural approach	31
2.3	Generalization to nonlinear functional responses	34
2.4	Structural stability of stochastic dynamics	38
2.4.1	Derivation of the stochastic Lotka-Volterra model	38
2.4.2	The structural approach in finite populations	41
2.5	Conclusion	44
3	Species interactions in changing environments	47
3.1	Introduction	48
3.2	Linking network structure and persistence	49
3.3	Inconsistent conclusions about the importance of network structure	51
3.4	A plea for an environment-dependent framework	55
3.5	Final thoughts	58

4	Adaptability in changing environments	61
4.1	Introduction	62
4.2	Background	63
4.3	Theoretical Framework	64
4.3.1	Consumer-resource interaction networks	64
4.3.2	Feasibility analysis	65
4.3.3	Parameterization of the interaction matrix	66
4.3.4	Environmental changes	69
4.3.5	Measuring the size of the adaptation space of a group of species	69
4.4	Data	73
4.5	Results	75
4.5.1	Simulations	75
4.5.2	Data analysis	78
4.6	Discussion	81
4.7	Appendix	86
4.7.1	Appendix A. Stability of the fixed points	86
4.7.2	Appendix B. Discussions on the structural constraint	87
II	Nonparametric approaches	91
5	Statistical inference of Jacobian matrices from time-series data	95
5.1	Introduction	96
5.2	Materials and Methods	98
5.2.1	Noisy nonlinear time series	98
5.3	Results	109
5.3.1	Results on Synthetic data	109
5.4	Discussion	115
6	Uncertainty quantification of the Jacobian coefficients	119
6.1	Introduction	120
6.2	Methods	122
6.2.1	Background	122
6.2.2	Statistical inference of the effect of biotic interactions from multivariate time series	123
6.2.3	Identifiability issues of the regularized S-map	124
6.2.4	A model average algorithm to associate an uncertainty level with the effect of biotic interactions	125
6.2.5	Analysis of synthetic time-series data	128
6.2.6	Analysis of empirical time-series data	128
6.3	Results	130
6.3.1	Analysis of synthetic data	130
6.3.2	Analysis of empirical data	133
6.3.2.1	Bermuda Atlantic time series	133
6.3.2.2	Hawaii ocean time series	134
6.4	Conclusions	135

7	Nonparametric structural stability analysis	141
7.1	Introduction	141
7.2	Theoretical framework	143
7.2.1	Time-dependent response to changing environments	143
7.2.2	Structural stability of non-equilibrium dynamics	146
7.2.3	Validation on synthetic data	147
7.3	Structural stability of empirical populations	152
7.4	Conclusion	156
8	Structural stability and predictability of nonlinear time series in changing environments	159
8.1	Application of nonparametric structural stability analysis for predictive studies	159
8.1.1	Structural stability and predictability in changing environments	161
8.1.2	Predictability of empirical populations	162
8.2	Discussion	166
9	Conclusion	169
9.1	Conclusion and future research directions	169
9.1.1	What's next?	170
9.1.1.1	Artificial neural networks for model discovery	171
9.1.1.2	Causal inference	172
9.2	Concluding remarks	173
	Appendix	177
A	Supplementary Figures for chapter 4	177
B	Supplementary Figures for chapter 6	179
C	Supplementary Figures for chapter 7	183
D	Supplementary Figures for chapter 8	187
	Bibliography	195

CHAPTER 1

INTRODUCTION

This is an introductory chapter which is intended to provide the motivations for the work of this thesis. The chapter is divided in four sections. In the first section we will discuss, with an example, why there is growing interests in explaining and predicting the dynamics of biological populations in changing environments. Then, in the second section, we will review challenges and standard approaches proposed in the literature to tackle this issue. Finally, in the last two sections, we will state the goal of the thesis and outline its structure.

1.1 The role of biological populations in human societies

In 2004 (Thompson et al., 2004) have shown that microscopic plastic fragments and fibers (i.e., plastic debris of size smaller than $\sim 1\text{mm}$) are widespread in the oceans and, through years of plastic over consumption, have accumulated in the pelagic zone and sedimentary habitats. Since this seminal publication, extensive research has shown that long-term exposure to the physical presence of microplastics in the environment has measurable effects on marine organisms (Gallo et al., 2018). For example, it has been shown that microplastics can reduce reproductive output and fitness in marine species by altering their food consumption and energy allocation (Desforges et al., 2015; Sussarellu et al., 2016). Similarly, microplastics debris provide

habitats for the assemblage of communities different from the ones in the surrounding environments (Zettler et al., 2013) increasing the risk of emergence of competitive interactions between native and invasive species. Moreover, evidences show that microplastics can alter gene expression, have impact on cells and tissues and can alter how species develop and disperse (Rochman et al., 2016; Borrelle et al., 2017).

With laboratory or field experiments one can measure the effects of microplastics pollution on species' physiology and species-species interaction parameters. Yet, with such experiments alone we cannot provide predictions about how these effects will change the overall dynamics of populations (i.e., a community of interacting species). Importantly, this lack of knowledge of how impacts of microplastic on individual organisms might lead to ecological harm has been one of the main limiting factors for the development of stringent regulations to restrict the mass flow of plastic debris into the ocean (Galloway et al., 2017). To bridge the gap between our understanding of the effects of pollutant at organisms and population level we need a better understanding of population dynamics in changing environments

Notice that this is not only a problem of scientific interest. The example of microplastics pollution provide a good case to understand why predicting population dynamics in changing environments is relevant for different parties in human societies. Here we discuss three classic examples. Firstly, marine organisms are consumed as part of regular source of proteins worldwide and there is growing concern (not yet supported by enough quantitative results (Cauwenberghe & Janssen, 2014)) that the microplastics ingested by these organisms may have negative impact on human health. Secondly, the presence of microplastics in the ocean has an economic impact. The net export–import trade value of seafood products is estimated at US\$35 billion in 2012 (as term of comparison, this value is greater than the net trade income of the other agricultural commodities combined (Gallo et al., 2018)). Moreover, a large number (estimated at about 260 million) of jobs directly or indirectly depend on this trade (Teh & Sumaila, 2011; Lusher A, 2017). Hence, microplastics-driven decaying growth rates of marine species together with the possible risk associated with their

consumption jeopardize both the job availability and the economic development of many countries. Thirdly, marine organisms play a key role in climate regulation (Sun et al., 2016). For example, marine microbial species such as Diatoms and Cyanobacteria are important contributors to the ocean carbon cycle (Yool & Tyrrell, 2003; Rousk & Bengtson, 2014; Stuart et al., 2015). Thus, there is concern about the impact of marine biodiversity loss, possibly induced or accelerated by microplastics pollution, on the global climate.

Microplastics pollution in the oceans is just one among numerous examples of globally relevant issues that, to be dealt with, require a better understanding of populations dynamics in changing environments. An incomplete lists of issues that can be better addressed with a full comprehension of the interactions between biological populations and their surrounding environments includes disease outbreaks (Cunningham et al., 2012), impact of gene-editing technologies on organisms and populations in the wild (Patrão Neves & Druml, 2017), spreading of antibiotic resistance (Hiltunen et al., 2017; Estrela & Brown, 2018), cancer dynamics (Korolev et al., 2014) and climate change (Worm et al., 2006; Scheffers et al., 2016).

What are the main challenges that prevent us to explain and predict population dynamics in changing environments? In the next section we will answer this question with a literature review. Before proceeding however, we would like to make a note to clarify some important terminology for the reader: throughout this chapter and the rest of this thesis we will use interchangeably the terms "community" and "biological populations". That is, a (biological) community is a set of biological populations co-occurring in the same location at the time of observation. Hence, following this nomenclature, we will also use interchangeably the terms "community dynamics" and "populations dynamics".

1.2 Populations dynamics in changing environments

The first step towards an understanding of how biological populations respond to changing environments is the determination of the driving forces of the dynamics. Intuitively, it is clear that the dynamics of living systems is driven by an interplay between stochastic (random) events and deterministic (predictable) processes (Chave et al., 2002; Mutshinda et al., 2009). Yet, which component (stochastic or deterministic) is dominant and which variable (biotic or abiotic) is the main driver of the dynamics has been a matter of debate for decades (Fargione et al., 2003; Houlahan et al., 2007; Mutshinda et al., 2009; Gamelon et al., 2017).

1.2.1 Determinism or stochasticity?

The debate about whether deterministic or stochastic forces are the main driver of the dynamics is originated from the difficulties associated with ruling out simple explanations of the data from empirical observations. The most striking example of this issue is the classic and challenging problem of distinguish noise from non-linearity in empirical time series. This problem originate from the fact that because populations dynamics is an intrinsic discrete process that take place in finite populations, fluctuations of species abundances emerge even without the influence of external drivers (changing environments) or nonlinear interactions (Bartlett, 1960). However, biological theories and experimental evidences suggest that demographic stochasticity cannot be the only driver of population dynamics. Other processes, such as species interactions and environmental factors, strongly contribute to the dynamics (Volterra, 1928; Hardin, 1960; Huisman & Weissing, 1999; Brown et al., 2001; Drake, 2005; Benincà et al., 2015). Conclusively distinguish between stochastic-driven and interaction-driven fluctuations required years of data collection (Bjørnstad & Grenfell, 2001), mathematical theories (Takens, 1981; Casdagli et al., 1991) and algorithm development (Sugihara, 1994).

1.2.2 Biotic or abiotic drivers?

It is now well established that, generally, demographic stochasticity play an important role, but fluctuations of species abundances are mainly driven by nonlinear (deterministic) biotic or abiotic interactions (Gamelon et al., 2017). However, establishing which of the two processes (biotic or abiotic interactions) is the main driver of the dynamics still remain a topic of active debate (Rosso et al., 2007; Freitas et al., 2009; Zunino et al., 2012; Ravetti et al., 2014). The issue is not trivial as, most likely, all forces (i.e., stochasticity, determinism, environments and species interactions) will act together nonlinearly (Mutshinda et al., 2009; Gamelon et al., 2017) and is therefore challenging to determine the relative importance of one process over another. Historically, the question has been addressed mainly by means of correlation studies (Mysterud et al., 2001; Rodó et al., 2002; Sugihara et al., 2012). Yet, nonlinear dynamics are characterized by ubiquitous mirage correlations that do not reflect true causations. Therefore, to properly disentangle the effects of different variables on the observed dynamics studies should abandon correlation analysis and focus on causality tests (Sugihara et al., 2012). Causal inference in nonlinear population dynamics is a growing and fascinating subject that we will touch upon in the last two chapter of these thesis. Moreover, in the conclusions we will discuss in further depth the role of causal inference for the development of population dynamics theories.

1.2.3 The need for models

A challenge strictly related to the topic of this thesis can be stated as the following question: provided we understand what are the relevant variables to include in our analysis, how do we model the effects of these variables on the dynamics of the populations? In other words, which laws do species obey when they interact within communities and with the surrounding environments? The problem raised by this question can be understood in terms of the example of the ocean microplastics discussed in section 1.1. As discussed above, extensive research has shown that microplastics is a variable to include in the analysis of marine organisms since it has

effects on their physiological properties such as their growth rates or fitness. Yet, how changes in the physiological properties (e.g. growth rate) will affect the overall dynamics of the community is a far more challenging issue to address. To predict these effects we need models. How do we build a population dynamics model? Broadly speaking we can identify two type of approaches: parametric and nonparametric.

1.2.3.1 Parametric approaches: assuming the model from theories

Parametric approaches have been the gold standard of population dynamics studies since the formulation of the first population models that date back to the first half of the nineteenth and twentieth century (Verhulst, 1838; Lotka, 1920; Volterra & Brelot, 1931). To formulate parametric models one needs to make assumptions about the mechanistic or phenomenological processes that generate the dynamics. Because the space of all possible assumptions is large (e.g., analytic form of the functional responses, including or ignoring spatial structure how to incorporate stochastic effects, ...) a large number of models have been proposed in the literature. These models are typically expressed in the form of discrete or continuous differential equations that depends on a number of parameters that have a clear biological interpretation. For example, probably the most notorious, parametric, model for population dynamics is the discrete Lotka-Volterra model:

$$\begin{aligned}x_{t+1} &= x_t[r_1 - \alpha_{11}x_t - \alpha_{12}y_t] \\y_{t+1} &= y_t[r_2 - \alpha_{21}x_t - \alpha_{22}y_t]\end{aligned}\tag{1.1}$$

In the equation above the parameters have a straightforward interpretation. Specifically, r_1, r_2 are interpreted as the intrinsic growth rate of a population (i.e., x, y , respectively); whereas the matrix α with entries α_{ij} is a matrix of interaction rates of individuals within a population (α_{ii}) and between populations (α_{ij}).

Eq. (1.1) is just an example of a parametric population dynamics model. Generally, when facing with the problem of constructing a parametric model we should attempt to achieve a parsimonious but robust characterization of the system under investigation (Wood & Thomas, 1999). That is, a model should be simple (i.e. it should depend on a small number of parameters) and robust to misspecification of either its parameters or the structure of the model itself. Simple and robust models such as the notorious Lotka-Volterra (Lotka, 1920; Volterra & Brelot, 1931), the replicator equation (Hofbauer & Sigmund, 1988) or the Leslie matrix model (Leslie, 1945) have contributed significantly to our understanding of the laws that regulate interactions between different species within populations. For example, parametric studies revealed the importance of intra- and inter-specific interactions of species in response to changing environments (May & Mac Arthur, 1972), the effect of different network structure on the stability properties of a system (May, 1972; McCann, 2000). Yet, while originally parsimony was the key driver for the formulation of population dynamics model, population biology literature is now characterized by the opposite trend: increasing computational power has lead to the formulation of models of ever-growing complexity (i.e., models that include many state variables and that depend on a large number of parameters) (Perretti et al., 2013b; Buckland et al., 2004).

Complex parametric models are problematic for two reasons: firstly they require unfeasibly large dataset to be validated. Secondly, even if we assume the existence of datasets over which to validate these type of models, which is a rarely justifiable assumption in population dynamics, the actual validation will be challenging. In fact, leaving over-fitting problems aside, large nonlinear models are rarely structurally identifiable (i.e., it is rarely possible to associate a unique value with the model parameters from empirical measurements) (Bellman & Aström, 1970; Villaverde et al., 2016). Hence, generally, complex parametric models are of limited use for explaining and predicting nonlinear population dynamics in changing environments (Perretti et al., 2013a; Ye et al., 2015).

Another fundamental drawback of complex models is known as "super-sensitivity

to model structure" (Wood & Thomas, 1999; Fussmann & Blasius, 2005). A model is super-sensitivity to its structure if small changes in the model formulation (i.e., model structure) can induce large changes in the overall dynamics (Wood & Thomas, 1999). This issue is particularly relevant for biological modeling but has not attracted much attention in the literature (Perretti et al., 2013a). In fact, in the physical science, a field from which biology is largely inspired, models are constructed based on well-understood mechanisms or first principles. Hence, there is strongly supported evidence that models (and most time their parameters) are well formulated. Conversely, in biology, and population dynamics in particular, it is rare to have such a clear understanding of the first principles that regulate the system and therefore models are prone to be misspecified (Mols et al., 2004; Gunawardena, 2014; Babbie et al., 2014). Moreover, even if a model was well specified (which, again, is a strong assumption in biology) both its true parameters and its true structure can change in time because species passively or actively respond to changing environments (e.g., the growth rate of a species can change in time as we saw discussed in the example in section 1.1 or species can adapt rapidly to changing environments (Bell & Gonzalez, 2011; Gonzalez & Bell, 2013)). Clearly, models that exhibit super-sensitivity to their structure are of limited use for the analysis of empirical data as small misspecification of errors in the estimation of their parameters will results in completely different conclusion about the system. Indeed, (Wood & Thomas, 1999) suggested that super-sensitivity to structure associated with a large uncertainty on model specification is the reason why many population dynamics model cannot be used for predictive studies.

Finally, a fundamental problem to bear in mind when a parametric model is fit to empirical data is that the conclusions drawn from the fitting and the analysis of the model are "*conclusions [...] about the model's mechanism, and not about nature's mechanism*" (Breiman, 2001). This statement is true regardless of the robustness or the simplicity of the proposed model. From this observation, it follows that, if the observed system does not actually follow the assumed model, then conclusions are necessary wrong. This is a key concept to bear in mind particularly in biology where, as discussed in the previous paragraph, model uncertainty is particularly problematic.

Overall, we believe that parametric approaches can be of great value to advance theories and even make predictions but only when the dynamics of the populations can be expressed in terms of tractable and robust equations. As we will see later, this issue is going to be one of the motivations for a key chapter of this thesis (chapter 2).

1.2.3.2 Nonparametric approaches: learning the model from data

A promising route to overcome the complexity associated with modeling the dynamics of biological populations in changing environments consist on learning the model directly from data (Perretti et al., 2013b,a; Ye et al., 2015; Martin et al., 2018). Statistical learning of model structure from observational data can be accomplished with different approaches. The most important and successful, at the time of writing this thesis, are: nonlinear state space reconstruction and symbolic regression.

Nonlinear state space reconstruction is a set of numerical techniques developed for constructing nonlinear predictive models from time series data (Takens, 1981; Casdagli, 1989; Kantz & Schreiber, 2004). The framework is based on a theorem by Floris Takens (Takens, 1981) which can be explained as follow: consider a scalar variable $x(t)$ which is observed (or sampled) at discrete (but constant) time intervals $t = n\tau$. We refer to $x(t)$ a scalar time series. Assume that the underlying data generating process of $x(t)$ is a nonlinear dynamical systems of dimension d . Takens theorem states that for a generic τ and $m \leq 2d + 1$ there exists a smooth map $\mathcal{S} : \mathbb{R}^m \rightarrow \mathbb{R}$ that for $1 \leq n < \infty$ satisfies:

$$f(x((n + m - 1)\tau), \dots, x(n\tau)) = x((n + m)\tau) \quad (1.2)$$

In Takens formalism m is called *embedding dimension* and τ is the *time lag* (Casdagli, 1989). These two parameters can be estimated directly from data following standard approaches (Kantz & Schreiber, 2004). In practice, Takens theorem provide a framework to reconstruct invariant manifolds $\mathcal{M} \in \mathbb{R}^m$ (i.e., a shadowed version of the true dynamics) from a scalar observation of $x(t)$ (i.e., the observed dynamics). From the reconstructed invariant manifold it is then possible to construct predictive models.

Moreover, it is possible to estimate important quantities of the dynamics such, as the largest Lyapunov exponent, that are preserved in the manifold reconstruction process. An important extension of Takens framework to reconstruct invariant manifold from multivariate time series data has been proposed in (Deyle & Sugihara, 2011).

Overall, nonlinear state space reconstruction is a powerful framework to build predictive models and, as we will see in the second half of this thesis, it will be a central focus of this work. Since its original formulation this framework has found numerous applications particularly in finance (Huang et al., 2017), climate science (Pérez-Muñuzuri, 2000) and physics (Descalzi & Rosso, 2018). The main limitation for the applicability of this approach in population dynamics is data availability. However, currently we are observing a worldwide data collection effort that is providing the scientific community with an unprecedented amount of information (Marx, 2013; Gharajeh, 2018). Hence we believe that nonlinear state space reconstruction methods will find continuously growing applications in population dynamics too.

An alternative approach to Takens theorem and its multidimensional generalizations for learning models from data is symbolic regression (Bongard & Lipson, 2007; Schmidt & Lipson, 2009; Brunton et al., 2016; Quade et al., 2016; Martin et al., 2018). Symbolic regression uses genetic algorithms to reconstruct the analytic form of the model itself (not just a shadow invariant manifold) from multidimensional time series data. Recovering the structure of a model from data would undoubtedly be the most promising route because, once exact equations are known, it is possible to both explain and predict the dynamics of biological populations in changing environments. However, while the method has been successfully applied to discover the governing equations of physical systems (Brunton et al., 2016; Quade et al., 2016), its applicability to biological dynamics is still far from reach. An interesting discussion about symbolic regression and example of its applicability to discover dynamic equations in biology can be found in (Martin et al., 2018).

1.3 Thesis statement

Going back to the original problem, the goal of this thesis is to develop analytical and numerical methodologies to estimate the tolerance of biological populations to changing environments. To this end, it is clear that no approach (i.e., parametric or nonparametric) can be superior to another under all circumstances. Parametric approaches fail if the model does not accurately describe the system under investigation. Similarly, nonparametric approaches fail if data are either scarce or not informative enough to properly implement the model learning algorithms. Generally, to deal with complex problems such as explaining and predicting population dynamics in changing environments we should tackle the issue from different perspective, i.e. we need to use different approaches and understand their advantages and disadvantages for each particular problem we are interested in. Moreover, to advance science there should always be a feedback of ideas and results from one world (parametric or nonparametric) to the other.

That being said, it is important to be aware of the fundamental limitations of the chosen approach. The major drawback of parametric approaches is the difficulty of translating unknown biological mechanisms into models that are at the same time tractable (depends on few parameters) and robust (can be misspecified and still produce consistent results). This is a major drawback that cannot be overcome with neither more experiments nor more theory, i.e., if the true data generating process depends nonlinearly on many parameters then no tractable model can be derived. Moreover, many results derived from parametric approaches hold true only under the restrictive assumptions of the model, making the derivation of general theories difficult to justify. In this context, an important goal of the first half of this thesis will be to establish the range of validity of a novel parametric framework that has been recently used to validate important biological theories. We will show that indeed the framework is successful because it does not have to rely on strict assumptions about the data generating process.

Issues with nonparametric approaches are of different nature. Their major drawbacks are (1) strong dependence on data quality and availability and (2) the diffi-

culties associated with using predictive models for explanatory purposes. The first issue can be overcome with the advancement of data collection technologies and more funding to experimental or observational studies. Hence is not a fundamental limitation. Overcoming the second issue, which is more fundamental, will be the ultimate goal of this thesis. Indeed, chapter seven, which is the major contribution of this thesis is devoted to this goal.

Overall, we believe that the increasing amount of biological data available nowadays paves the way for moving theoretical population biology from being an assumption-driven towards a data-driven science. Hence we foresee that nonparametric approaches will find continuously growing applications.

1.4 Summary

In section 1.1 we have used a classic example to illustrate the reasons why it is important to explain and predict the dynamics of biological population in changing environments. Then, in section 1.2 we have reviewed what are the main challenges that we face when we try to address this issue. The goal of the section was to illustrate advantages and disadvantages of both parametric and nonparametric approaches to modeling. The rest of this thesis is devoted to develop a systematic framework to understand the effect of changing environments on the dynamics of biological populations and to predict how populations may respond to these changes.

1.4.1 Structure of the thesis

Each chapter of the thesis has an abstract that link the chapter with the previous discussions and summarize its main contributions. The thesis is organized as follow:

Part I: parametric approaches

- In chapter two we introduce the basic approach (known as the structural approach) upon which we construct our parametric framework. Then we show

that this approach does not suffer of the problems discussed in section 1.2.3.1. Specifically, it is robust to perturbations in model structure. Hence, conclusions drawn based on this framework are not based on the assumption of a specific parametric form of a model but on the assumption that the observed dynamics is generated by a model that belong to a large class of possible parametric structures. The chapter is based on reference (Cenci & Saavedra, 2018a)

- In chapter three we use the structural approach to challenge the view that the structure of species interaction networks is a good predictor of species's persistence in changing environments. The chapter is based on reference (Cenci et al., 2018b)
- In chapter four we develop a framework based on the structural approach to predict species' likelihood to adapt to fast changing environments. The chapter is based on reference (Cenci et al., 2018a)

Part II: nonparametric approaches

The first two chapter of this section are devoted to develop a methodology to infer biotic interactions and their associated uncertainty from noisy nonlinear time series data. The results of these two chapters will then be used in the chapter seven to explain and predict, nonparametrically, biological population dynamics in changing environments

- In chapter five we propose a methodology to improve a state of the art inference algorithm to infer Jacobian coefficients from noisy nonlinear time series data. The chapter is based on reference (Cenci et al., 2019).
- In chapter six we proposed a methodology to quantify the uncertainty associated with Jacobian coefficients estimated from empirical data with the algorithm proposed in the previous chapter. The chapter is based on reference (Cenci & Saavedra, 2018b)
- Chapter seven is the most important chapter of the thesis. Here we propose a novel nonparametric approach to estimate the effects of changing environ-

ments on the dynamics of biological populations. The chapter is based on reference (Cenci & Saavedra, In press)

- In chapter eight, we present an application of the theoretical framework derived in the previous chapter.
- Finally, in chapter nine we summarize the main finding of the thesis and we discuss future directions of research.

Part I

Parametric approaches

CHAPTER 2

STRUCTURAL STABILITY OF NONLINEAR POPULATION DYNAMICS

In this chapter we introduce an established framework, known as the structural approach, developed to systematically investigate the effect of changing environments on the dynamics of biological populations. This parametric approach is build upon the concept of structural stability, i.e., the stability of a dynamical system against perturbation of the model structure or its parameters. Originally, this approach was introduced to explain what type of structure of species interactions maximize the capacity of a system to withstand environmental perturbations. Yet, the approach could only be applied to communities that follow the standard Lotka-Volterra dynamics. In contrast, here we show that the validity of the approach extend beyond the standard Lotka-Volterra dynamics to a much larger class of nonlinear models. Importantly, this class covers a large number of nonlinear functional responses that have been intensively investigated both theoretically and experimentally. We also investigate the applicability of the structural approach to stochastic dynamical systems and we provide a measure of structural stability for finite populations. Overall, we show that the structural approach can provide reliable and tractable information about the qualitative behavior of many nonlinear dynamical systems.

2.1 Introduction

Quantifying the tolerance of interacting populations to environmental perturbations has been the center of theoretical population-biology research for many decades (Levins, 1968; Thom & Fowler, 1975; Ye et al., 2015). As already discussed in chapter 1, the big challenges behind this task are: the difficulty in producing a large number of exact experimental replicates on which to test theories, the analytical intractability of complex mechanistic models, and the possibility that the functional form of the model itself can change throughout time (Vandermeer, 1969; Ye et al., 2015). In order to deal with these challenges, a typical approach is to focus on simple, yet, insightful models derived from first principles (Logofet, 1993). However, while simple models of population dynamics can be tractable, they are often criticized for providing a non-realistic deterministic version of the factors shaping the time evolution of population abundances (Tilman, 1987). The classic example of a simple deterministic model that has attracted much of attention, but that has been strongly criticized, is the Lotka-Volterra (LV) model (Lotka, 1920; Volterra & Brelot, 1931). While this model has been derived from thermodynamics principles (Michaelian, 2005), principles of conservation of mass and energy (Logofet, 1993), from chemical kinetics in large populations (Täuber, 2011), and can exhibit a rich behavior such as chaotic dynamics (Vano et al., 2006) and limit cycles (Hofbauer & So, 1994); it seems unreasonable to believe that such a simple deterministic model could describe the time evolution of population abundances that we observe in nature (Tilman, 1987).

Certainly, despite its mechanistic foundation, the classic LV dynamics does not take into account many important biological and environmental processes ranging from fluctuations in birth and death processes to saturating effects of species growth (Logofet, 1993). In order to deal with some of these limitations, recent theoretical work has developed structural approaches that are based on the concept of structural stability (Thom & Fowler, 1975; Lewontin, 1969). That is, stability of the qualitative behavior of a dynamical system against fluctuations of its parameters (Rohr et al., 2014; Saavedra et al., 2014, 2017b). Following a structural approach, studies have been able to reconcile foundational hypothesis in ecology with observational data

(Saavedra et al., 2014, 2016b, 2017a; Song et al., 2017). Importantly, here we show that this approach can be directly applied to a larger class of population dynamics models with nonlinear functional responses. The main implication of our findings is that the applicability of the structural approach does not depend exclusively on the assumption of pairwise interactions or that dynamics are generated by one single deterministic model whose functional form remains fixed in time. We also show how to extend the approach to investigate the structural stability of stochastic dynamics.

This chapter is organized as follows. In the second section (2.2), we briefly review the details of the structural approach focusing on its geometrical interpretation. In the third section (2.3), we derive the conditions under which we can apply the structural approach to study the structural stability of nonlinear population dynamics. In the fourth section (2.4), we examine the applicability of this approach to stochastic dynamical systems focusing on the stochastic LV model. Finally, in the last section (2.5), we discuss the implications of our findings.

2.2 Background: the structural approach

The structural approach is a geometric methodology recently introduced (Rohr et al., 2014; Saavedra et al., 2014, 2017b) to provide a quantitative measure for the structural stability of the classic LV dynamics:

$$\dot{\vec{x}} = \vec{x}(\vec{r} - \mathcal{A}\vec{x}), \quad (2.1)$$

where \vec{x} is the vector of population abundances (i.e., the state variables), \vec{r} is the vector of intrinsic growth rates (i.e., difference between death and birth rate), \mathcal{A} is the $d \times d$ interaction matrix, and d is the dimension of the state space (i.e., the number of species). The classic LV dynamics is a deterministic model derived from the infinite population limit of a multispecies stochastic process with mass action assumption on the transition rates (Constable & McKane, 2015, 2017).

A dynamical system is said to be structurally stable if smooth variations of the model (e.g., its parameters) leave some properties of the system unchanged (Mayo-Wilson, 2015). For example, hyperbolic dynamical systems are a typical case of structurally stable models because smooth changes of their parameters will leave the sign of the largest eigenvalue, and therefore the stability property of the fixed points, unchanged (Hofbauer & Sigmund, 1988). Interestingly, structural stability has a geometrical interpretation. That is, the parameter space of structurally stable dynamical systems can be partitioned into structurally stable domains separated by regions that are structurally unstable (Kauffman, 1993). Within the structurally stable domains, smooth changes of the parameters (caused for example by environmental perturbations) do not change the nature of the fixed point. For instance, if a fixed point of a dynamical system exists and is feasible (i.e., all species have positive abundances on the attractor), perturbations that keep the parameters within the structurally stable domains preserve its feasibility.

The structural approach uses the size of the structurally stable domains of feasible fixed points as a measure for the structural stability of the population dynamics model (Rohr et al., 2014; Saavedra et al., 2014, 2017b). This domain is known as the feasibility domain (Logofet, 1993; Saavedra et al., 2014), i.e., a convex region within the parameter space of a dynamical system from which it is possible to sample parameters that generate feasible solutions. The larger the feasibility domain, the larger the structural stability of a feasible fixed point (see Figure 2.1). Formally, for classic LV dynamics, because of the linearity of the functional response, the feasibility domain is a convex hull in the parameter space of the intrinsic growth rates \vec{r} spanned by the columns of the interaction matrix \mathcal{A} :

$$D_F(\mathcal{A}) = \{\vec{r} \in \mathbb{R}^d \mid \exists \lambda_1, \dots, \lambda_d > 0, r_j = \sum_i \lambda_i a_{ij}\} \quad (2.2)$$

A fixed point is feasible if the vector of intrinsic growth rates lies within the feasibility domain (Logofet, 1993) (see Figure 2.1 for a graphical example). In fact, by considering a non-singular interaction matrix \mathcal{A} , the vector of state variables can be

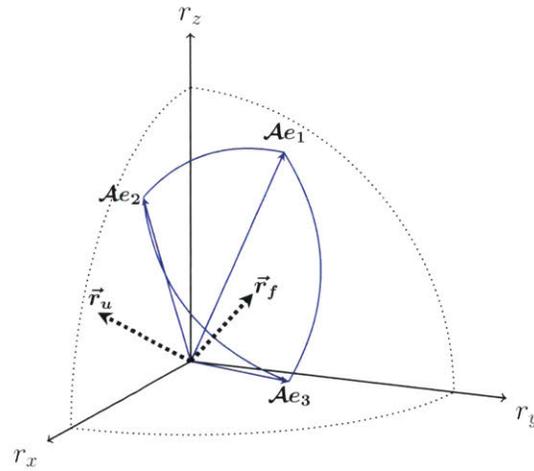


Figure 2.1: **Structural approach to infer the structural stability of feasible fixed points in the classic Lotka-Volterra dynamics.** The figure shows the feasibility domain in the parameter space of the intrinsic growth rates. A fixed point of the classic LV dynamics is feasible if the vector of intrinsic growth rates falls within the feasibility domain spanned by the columns of the interaction matrix \mathcal{A} , such as the vector \vec{r}_f illustrated in the figure. Otherwise, a vector \vec{r}_u would generate an unfeasible solution, i.e., one or more species would be extinct at the fixed point. The structural approach uses the size of the feasibility domain as a measure for the structural stability of the feasible fixed point.

rewritten as a combination of a standard basis of \mathbb{R}^d , i.e., $\vec{x} = x_1 e_1 + \dots + x_n e_n$ and $\mathcal{A}\vec{x} = \sum_j x_j \mathcal{A}e_j$. Therefore the positive orthant (i.e., the state space of LV) is contracted to a n -hedral angle with generatrices formed by the columns of the interaction matrix. Hence a feasible (positive) solution of $\mathcal{A}\vec{x} = \vec{r}$ exists if $\vec{r} \in D_F(\mathcal{A})$ (Logofet, 1993; Saavedra et al., 2014). This is precisely the non-zero fixed point of Eq. (6.5).

Following the structural approach (Saavedra et al., 2014, 2017b), the structural stability of the classic LV dynamics can be quantified by measuring the volume of the feasibility domain in the L_p norm:

$$\Omega = \frac{|\det \mathcal{A}|}{\|\mathcal{A}e_1\|_{L_p} \dots \|\mathcal{A}e_d\|_{L_p}} \quad (2.3)$$

where d is the dimension of the system. The choice of the norm or the normalization factors of the interaction matrix do not change the computation of the volume

(Rohr et al., 2014; Logofet, 1993). Particularly, the relevance of the feasibility domain relies on its quantitative characterization of structural stability and its conceptual interpretation as the tolerance of a dynamical system to environmental perturbations (Saavedra et al., 2017b,a; Cenci et al., 2018a).

2.3 Generalization to nonlinear functional responses

One important question that remains to be answered is to which other models the structural approach can be extended. To answer this question, let us consider a more general form for the classic LV dynamics:

$$\dot{\vec{x}} = \vec{x}(\vec{r} - \mathcal{A}f(\vec{x})), \quad (2.4)$$

where $f(\vec{x})$ denotes a general functional response. This type of models have been typically called generalized Lotka-Volterra (GLV), and their association with the classic LV dynamics has been extensively studied (Takeuchi, 1996). For example, under classic LV dynamics, the non-negative equilibrium point is the solution to the linear complementary problem (LCP) for \mathcal{A} and \vec{r} (i.e., $\text{LCP}(\mathcal{A}, -\vec{r})$). Similarly, the non-negative equilibrium point of GLV is the solution to the nonlinear complementary problem (NCP) for $\mathcal{A}f(\vec{x})$ and \vec{r} (i.e., $\text{NCP}(\mathcal{A}f(\vec{x}), -\vec{r})$). It is also known that if $f(\vec{x})$ is continuous and monotone in \mathbb{R}^d , then there exists a non-negative equilibrium of Eq. (2.4) for some $\vec{r} \in \mathbb{R}^d$. Importantly, these associations allow us to be more concrete and redefine the question above to whether it is possible to apply the structural approach to study the structural stability of the feasible fixed point of Eq. (2.4).

Here, we derive the sufficient conditions for the applicability of the structural approach to Eq. (2.4). We do this by mapping the nonlinear functional response into a linear response and finding the conditions for $f(\vec{x})$ under which this change of variable leaves the state space of Eq. (2.4) unchanged. Then, we ensure that the system in the new variable is topologically equivalent to LV. Formally, let us call $\vec{z} = f(\vec{x})$. Then, $\vec{x} = f^{-1}(\vec{z})$:

$$\dot{\vec{x}} = \vec{x}(\vec{r} - \mathcal{A}f(\vec{x})) \Rightarrow \frac{d(f^{-1}(\vec{z}))}{dt} = f^{-1}(\vec{z})(\vec{r} - \mathcal{A}\vec{z}), \quad (2.5)$$

which can be rewritten as

$$\frac{d\vec{z}}{dt} = \frac{1}{\nabla(f^{-1}(\vec{z}))} f^{-1}(\vec{z})(\vec{r} - \mathcal{A}\vec{z}). \quad (2.6)$$

Because we are interested in a feasible fixed point, we need to guarantee that the inverse of the transformation must map points in the state space \mathcal{S} of Eq. (2.6) to points in the positive orthant of the original dynamical system of Eq. (2.4) (i.e., $f^{-1} : \mathcal{S} \subseteq \mathbb{R}_{>0}^d \rightarrow \mathbb{R}_{\geq 0}^d$). The topological equivalence to the classic LV dynamics also implies that Eq. (2.4) and Eq. (2.6) must have only two fixed points in \mathcal{S} : one feasible stable point and one feasible (or unfeasible) unstable point. Thus, the change of variable also must map the unstable fixed point of Eq. (2.4) onto the unstable fixed point of Eq. (2.6) and the same for the stable one. Specifically, this translates into requiring the condition that $\frac{f^{-1}(\vec{z})}{\nabla(f^{-1}(\vec{z}))}$ is well defined in \mathcal{S} (i.e., $\nabla(f^{-1}(\vec{z})) \neq 0$) and that its (only) root is an unstable fixed point of Eq. (2.6). Formally, the conditions under which we can consider Eq. (2.4) and Eq. (2.6) to be topologically equivalent to the classic LV can be written as:

$$\exists! \vec{x} \in \mathbb{R}_{>0}^d \mid \vec{r} - \mathcal{A}f(\vec{x}) = 0, \quad f^{-1} : \mathcal{S} \subseteq \mathbb{R}_{>0}^d \rightarrow \mathbb{R}_{\geq 0}^d, \quad \exists! \vec{z}^* \in \mathbb{R}_{\geq 0}^d \mid \frac{f^{-1}(\vec{z}^*)}{\nabla(f^{-1}(\vec{z}^*))} = 0, \quad (2.7)$$

where \vec{z}^* is an unstable fixed point of Eq. (2.6). Note that the first two conditions derive directly from previous results on GLV dynamics (Takeuchi, 1996). To illustrate the conditions of Eq. (2.7), let us consider a practical example. If we take $f(\vec{x}) = \vec{x}^\beta \Rightarrow \vec{x} = \vec{z}^{\frac{1}{\beta}}$:

$$\dot{\vec{x}} = \vec{x}(\vec{r} - \mathcal{A}\vec{x}^\beta) \quad (2.8) \quad \dot{\vec{z}} = \beta\vec{z}(\vec{r} - \mathcal{A}\vec{z}) \quad (2.9)$$

then a feasible solution of Eq. (2.9) guarantees the feasibility of Eq. (2.8) given that the inverse function f^{-1} maps each point onto the positive orthant of Eq. (2.9) to

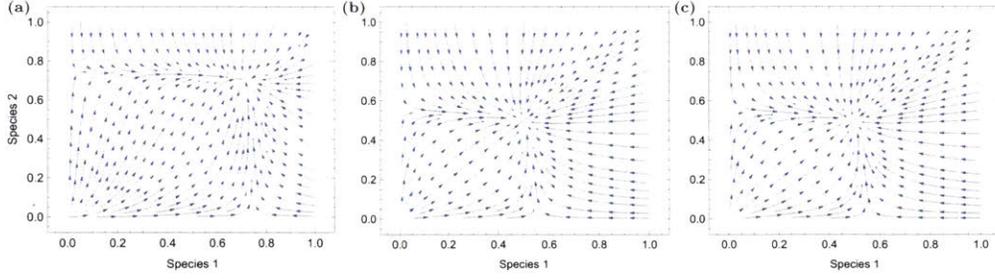


Figure 2.2: **Topological equivalence between classic and generalized Lotka-Volterra dynamics for a polynomial response function.** Panel (a) shows the direction field of Eq. (2.4) with a polynomial response function, i.e., $f(\vec{x}) = \vec{x}^2$. Panel (b) shows the direction field of the linear transformation of GLV dynamics, as given by Eq. (2.9) with $\beta = 2$. As a reference, panel (c) shows the direction field of the classic LV dynamics, i.e., Eq. (6.5). The figure shows that the state spaces are equivalent. Note that changing the exponent of the polynomial would just modify the qualitative picture of the state space but would leave the structure of the state space unchanged. The polynomial response function satisfies the conditions in Eq. (2.7). Parameters: $r_i = 0.5$, $\mathcal{A}_{ii} = 0.9$ and $\mathcal{A}_{ij} = 0.1$

the positive orthant of the original system. Furthermore, because the system \vec{z} has a linear stationary state (see Figure 2.2), we can apply the structural approach to Eq. (2.9) to measure the structural stability of the fixed point of Eq. (2.8).

Following the rationale above, we can generalize our results beyond GLV dynamics. To illustrate this, let us now consider the following general population dynamics model:

$$\dot{\vec{x}} = g(\vec{x})[\vec{r} - \mathcal{A}f(\vec{x})]. \quad (2.10)$$

Using the change of variable $\vec{z} = f(\vec{x})$, we can rewrite Eq. (2.10) as

$$\frac{d\vec{z}}{dt} = \frac{1}{\nabla(f^{-1}(\vec{z}))} g(f^{-1}(\vec{z}))(\vec{r} - \mathcal{A}\vec{z}). \quad (2.11)$$

We call \mathcal{S}_1 and \mathcal{S}_2 the state spaces of Eq. (2.10) and Eq. (2.11), respectively. Note that \mathcal{S}_1 is not necessarily the entire positive orthant, but we continue to assume that the function $g(\vec{x})$ has at most one zero in \mathcal{S}_1 and it is an unstable fixed point of

Eq. (2.10). It is now straightforward to see that the conditions leading to topological equivalence are: (1) the system $\vec{r} - \mathcal{A}f(\vec{x}) = 0$ has only one solution in the positive orthant, (2) $f^{-1} : \mathcal{S}_2 \subseteq \mathbb{R}_{>0}^d \rightarrow \mathcal{S}_1$, and (3) $\frac{g(f^{-1}(\vec{z}))}{\nabla(f^{-1}(\vec{z}))}$ has at most one root in \mathcal{S}_2 . This last condition needs to be satisfied given that we have imposed that $g(\vec{x})$ has only one root.

In sum, in this section, we have provided the sufficient conditions that need to be assumed on a response function in order to study the structural stability of nonlinear models under the structural approach derived for linear functional responses. As a final note, we want to stress that the conditions in Eq. (2.7) are sufficient but not necessary. To illustrate this point, let us consider the case of a functional response of type II, that is $f_{\text{II}}(\vec{x}) = \mathcal{M}\vec{x}$, where \mathcal{M} is a diagonal matrix with diagonal entries $\mathcal{M}_{ii} = \frac{1}{1+x_i}$. For this type of functional response the conditions in Eq. (2.7) are not satisfied. Specifically, because $\vec{z} \in \mathbb{R}_{\geq 0}^d$, then $f^{-1} : \mathbb{R}_{\geq 0}^d \rightarrow \mathbb{R}^d$ and there also exists an additional feasible fixed point which is not globally (and locally) stable. Nevertheless, it has already been shown that the structural approach can still be applied to these type of functional responses (Saavedra et al., 2017b). As illustrated in Figure 2.3, the basin of attraction of the only stable fixed point is the only subregion of \mathcal{S}_2 that can be mapped back into the positive orthant of the original system. The basin of attraction of this feasible stable fixed point (plus the origin) is, indeed, topologically equivalent to the classic LV dynamics. This example suggests that (1) if there exists a basin of attraction \mathcal{B} of the stable fixed point of Eq. (2.11) topologically equivalent to LV, and (2) this is the only subregion of \mathcal{S}_2 that is mapped back by f^{-1} to \mathcal{S}_1 , then we can still apply the structural approach to the restriction of the state space on \mathcal{B} . These results imply that we can relax the conditions in Eq. (2.7), and extend the validity of the structural approach to a much larger class of functional responses if there exists a unique subregion of the state space of Eq. (2.11) for which the conditions in Eq. (2.7) hold.

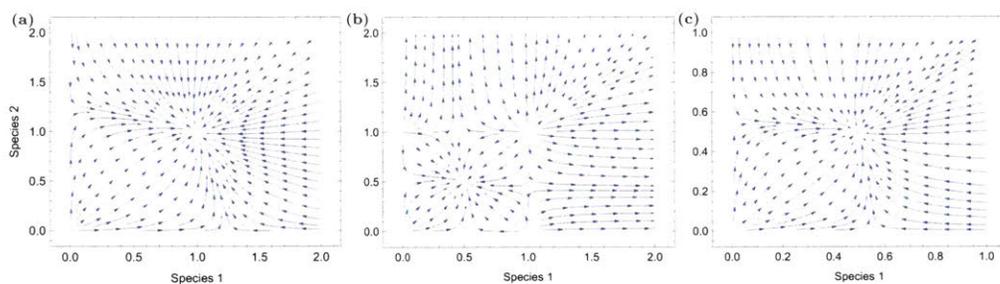


Figure 2.3: **Direction field of Lotka-Volterra dynamics with a functional response of type II.** While the state space of Eq. (2.10) with a functional response of type II (Panel (a)) is mapped into a state space that is not topologically equivalent to LV (Panel (b)); the restriction of this state space on the basin of attraction $\mathcal{B} = [0, 1) \times [0, 1)$ of the stable fixed point it is. As a reference, Panel (c) shows the direction field of the classic LV dynamics. Parameters: $r_i = 0.5$, $\mathcal{A}_{ii} = 0.9$ and $\mathcal{A}_{ij} = 0.1$

2.4 Structural stability of stochastic dynamics

The analysis in the previous section was limited to deterministic dynamical systems. This is a good approximation in the limit of infinitely large populations. However, when this limit cannot be assumed, noise in the state variables (i.e., demographic stochasticity) cannot be neglected. That is, fluctuations of the state variables need to be considered explicitly. In this section, we show under which conditions we can study the structural stability of stochastic LV dynamics (SLV) under the structural approach derived for linear functional responses. Then, we introduce an extension of the structural approach to take into account the extinction probability in finite populations.

The SLV can be derived from the linear noise approximation (LNA) of the chemical master equation (Van Kampen, 1992; Gardiner, 2004) for the probability distribution of the stochastic process that generate the classic LV dynamics.

2.4.1 Derivation of the stochastic Lotka-Volterra model

To derive the stochastic dynamics we write down the master equation, which in its most general form is written as:

$$\frac{\partial p(\vec{n}, t)}{\partial t} = \sum_{l=1}^{\text{num of reactions}} \left[\underbrace{T_l(\vec{n}|\vec{n} - \vec{\mu}_l)}_{\text{gain}} p(\vec{n} - \vec{\mu}_l, t) - \underbrace{T_l(\vec{n} + \vec{\mu}_l|\vec{n})}_{\text{loss}} p(\vec{n}, t) \right], \quad (2.12)$$

where $\vec{n} = (n_1, \dots, n_d)$ is the state vector, $p(\vec{n}, t)$ is the probability of observing a system in state \vec{n} and $\vec{\mu}_1 = (a, b, c), \dots, \vec{\mu}_{\text{num of reactions}}$ are stoichiometric coefficients (Constable & McKane, 2015). The stoichiometric coefficients are introduced for a notational convenience that will be clear in Eq. (2.16).

The solution of the master equation provides us with the probability distribution of observing a state \vec{n} at time t . The reaction rate can be written easily in terms of chemical reactions. As an example take the reaction $A + B \rightarrow A + A$ that takes place at rate \mathcal{A}_{AB} . Then, if we call n_A and n_B the number of reactant A and B, respectively, the reaction term reads:

$$T_\alpha(n_A + 1, n_B - 1|n_A, n_B) = \mathcal{A}_{AB} \frac{n_A n_B}{V^2}, \quad (2.13)$$

where we have assumed mass action on the reaction rates and V is a measure of system size. Obviously, the solution of the master equation in a closed analytical form is extremely challenging to find. Therefore, following standard methods of stochastic dynamics, we take the Van Kampen system size expansion of the master equation to recover the Fokker-Plank equation (Van Kampen, 1992; Gardiner, 2004):

$$\frac{\partial p(\vec{x}, \tau)}{\partial \tau} = - \sum_{i=1}^d \frac{\partial}{\partial x_i} D_i(\vec{x}) p(\vec{x}, \tau) + \frac{1}{2V} \sum_{i,j=1}^d \frac{\partial}{\partial x_i \partial x_j} \mathcal{T}_{ij}(\vec{x}) p(\vec{x}, \tau), \quad (2.14)$$

where we have used the change of variable: $\vec{x} = (n_1/V, \dots, n_d/V)$ and $\tau = t/V$.

The drift and diffusion term of the Fokker-Plank equation (i.e., first and second term on the right hand side of Eq. (2.14)) can be written explicitly following (Constable & McKane, 2015):

$$D_i(\vec{x}) = \sum_{k=1}^R \mu_{i,k} T_k(V\vec{x} + \mu_k | V\vec{x}) \quad (2.15)$$

$$\mathcal{T}_{ij}(\vec{x}) = \sum_{k=1}^R \mu_{i,k} \mu_{j,k} T_k(V\vec{x} + \mu_k | V\vec{x}) \quad (2.16)$$

Using the stoichiometric coefficients and the equations above (Eq. (2.16)), we can write down the stochastic dynamics explicitly in term of an Ito stochastic differential equation as Eq. (2.18) in the main text:

$$\dot{\vec{x}} = \mathcal{D}(\vec{x}) + \frac{1}{\sqrt{V}} \xi(\vec{x}, \tau), \quad (2.17)$$

where $\mathcal{D}(\vec{x})$ will be the deterministic vector field and $\xi(\vec{x}, \tau)$ is a Gaussian random variable with $\langle \xi(\vec{x}, \tau) \rangle = 0$ and $\langle \xi_i(\vec{x}, \tau) \xi_j(\vec{x}, \tau') \rangle = \mathcal{T}_{ij}(\vec{x}) \delta(\tau - \tau')$.

Finally, we can write down the stochastic Lotka-Volterra dynamics. The reaction terms T_α can be written explicitly following the same reasoning of Eq. (2.13). Hence, following Eq. (2.16) the stochastic Lotka-Volterra dynamics reads (Constable & McKane, 2015):

$$\dot{\vec{x}} = \vec{x}(\vec{r} - \mathcal{A}\vec{x}) + \frac{1}{\sqrt{N}} \nu(t), \quad (2.18)$$

where the first term on the right hand side is the deterministic vector field, N is the system size, and $\nu(t)$ is Gaussian white noise (McKane et al., 2014) with zero mean and correlations given by $\langle \nu_i(t) \nu_j(t') \rangle = \mathcal{B}_{ij} \delta(t - t')$. Where $\mathcal{B}_{ij} = \vec{x}(\vec{r} + \mathcal{A}\vec{x}) \forall i = j$ and zero otherwise. The δ function in the last expression characterizes the white noise.

2.4.2 The structural approach in finite populations

Here we ask whether we can apply the structural approach to finite populations, i.e., to Eq. (2.18). For globally stable systems, because the dynamics away from the fixed point are dominated by the deterministic component, the deterministic vector field pushes back the perturbations induced by stochastic fluctuations of the state variables to the feasible solution. Nevertheless, note that the extinction probability is always different from zero because the system size is finite. That is, globally stable fixed points are only metastable since demographic stochasticity will eventually induce a transition into an absorbing state (Rulands et al., 2013). Generally, the smaller the system size, the larger the extinction probability. However, the time for extinction T_{ext} due to stochastic fluctuations is typically very large, and in globally stable SLV systems this time scales as $T_{\text{ext}} \propto e^N$ (Dobrinevski & Frey, 2012). Note that N is finite but assumed to be large, hence T_{ext} can be regarded as infinity for practical purposes. Thus, in this limit, the structural approach can be applied to globally stable SLV systems.

However, recently it has been shown that in neutral and quasi-neutral conditions, the argument provided above does not hold anymore (Dobrinevski & Frey, 2012; Constable & McKane, 2015; Constable et al., 2016). Specifically, the extinction time for neutrally stable SLV models scales proportionally to the system size N , i.e. $T_{\text{ext,neut}} \sim N^\alpha$ rather than exponentially (Dobrinevski & Frey, 2012; Cremer et al., 2009; Rulands et al., 2013). The reason for this change in behavior of the system is that, near neutrality, there exists a new dynamics on the center (CM) and slow (SM) manifold of SLV, that cannot be studied in the deterministic setting. That is, extinctions of species in time proportional to $T_{\text{ext,neut}}$ are driven by a stochastic drift on the CM or the SM (Constable & McKane, 2015; Constable et al., 2016).

Importantly, the existence of a center manifold is a hallmark of *structural instability* given that the dominant eigenvalue of the Jacobian matrix is zero (Duan, 2015). That is, the feasibility domain shrinks to a line. Then, because stochastic fluctuations of the state variables in the vicinity of the fixed point are equivalent to fluctuations of the parameters (Arnoldi et al., 2016), infinitesimally small demographic noise can

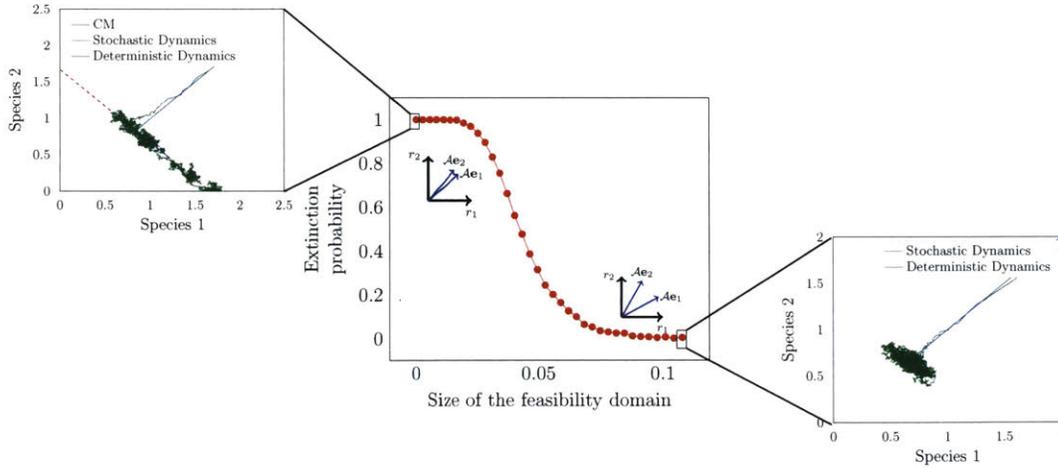


Figure 2.4: Extinction probability as a function of the structural stability of the system. The figure shows the fraction of stochastic extinctions in a two species LV system, that is predicted to be feasible in the deterministic setting, as a function of the size of the feasibility domain. We observe a transition from a state in which the probability of extinction is effectively one (i.e., the structural approach fails) to a region where this probability is approximately zero (i.e., validity of the structural approach). The structural stability of the feasible solution increases with the size of the feasibility domain, reducing the effect of stochasticity on feasibility. We run simulations for a time $\tilde{T} \ll T_{\text{ext}} \sim e^N$, where eventually all finite systems will go extinct.

turn feasible into unfeasible solutions. As discussed above, within this regime of parameters (i.e., near neutrality) $T_{\text{ext}} \sim N^\alpha$, extinction takes place with probability one in finite time. Note that globally stable fixed points correspond to the opposite side of the spectrum, for which $T_{\text{ext}} \sim e^N$. Thus, focusing on a finite time scale, by moving from one side of the spectrum to the opposite we can observe an abrupt transition for the extinction probability in times that are small compared to e^N (see Figure 2.4). Notably, this probability is a function of the dominant eigenvalue of the Jacobian, and therefore, of the size of the feasibility domain.

The statement above has an important implication for the structure of the feasibility domain itself. In fact, the investigation of the explicit stochastic dynamics reveal that, within the feasibility domain, there exists a non-uniform distribution of the extinction probability. This was already noticed in previous work (Rohr et al., 2016) in the context of environmental stochasticity. Importantly, for a given (fixed) population size N and a given (fixed) interaction matrix \mathcal{A} , the distribution of the

extinction probability is non-homogeneous in space (in the parameter space of intrinsic growth rates). As shown in Figure 2.5 panel (a), regions near the border of the feasibility domain have a larger extinction probability than regions that are far away from the border. Therefore, using the size of the feasibility domain to study the structural stability of stochastic systems can be misleading. Instead, one needs to use the size of the region within the feasibility domain for which, in finite times, $P_{\text{ext}} = 0$. Note that other regions are deterministically feasible, but will undergo extinction with probability one in the presence of demographic noise.

While measuring the extinction probability as a function of the parameter space of high-dimensional feasibility domains is a challenging task, we propose that this can be estimated by the difference between the largest and the smallest eigenvalue of the Jacobian matrix of the deterministic system (see Figure 2.5 panel (b)) (Parsons & Rogers, 2017). Indeed, the lack of structural stability is characterized by a strong separation of time scales between the most stable (largest eigenvalue λ_{max}) and the least stable (smallest eigenvalue λ_{min}) dynamics. Importantly, this separation reaches its minimum within the interior of the feasibility domain, where the slow and fast dynamics are not fully separated. On the contrary, along the borders of the feasibility domain, this separation will be maximum (see Figure 2.6). Thus, we can quantify the structural stability of a community by measuring the size of the region, within a feasibility domain, in which the fast and slow dynamics are not well separated. For example, a possible way to perform this analysis is to measure the relative difference of time scales $\Delta\tau = \frac{\lambda_{\text{max}} - \lambda_{\text{min}}}{\lambda_{\text{max}}}$ across the feasible region of the parameter space (see Figure 2.6). Then, we can fix a threshold ϵ on $\Delta\tau$ (typically $\epsilon = \Delta\tau_c$ at which P_{ext} goes from zero to one) and measure the size of the feasibility domain that falls below this threshold. Generally, the larger this region is, the larger the structural stability of a SLV dynamics. Note that, for fixed N and \mathcal{A} , $\Delta\tau$ is a function of the location of a system in the feasibility domain (i.e., a function of \vec{r}), and resembles the distribution of the extinction probability (see Figure 2.5).

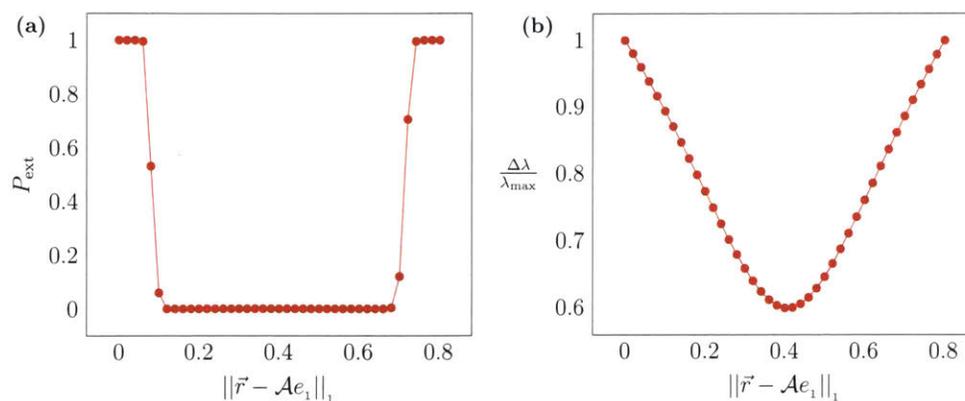


Figure 2.5: **Partition of a two species feasibility domain in regions that undergo extinctions within different time scales.** Panel (a) shows the extinction probability as a function of the position of a vector of intrinsic growth rates inside the feasibility domain. To generate Panel (a), we selected a vector of intrinsic growth rates \vec{r} on one of the columns of the feasibility domain and compute the extinction probability of Eq. (2.18) over an ensemble of 10^3 realizations. Then, we selected a new vector \vec{r} until we reached the opposite border. The x-axis shows the distance (in the L_1 norm) of the new sampled vector \vec{r} from the border of the feasibility domain, which we have fixed as origin. Panel (b) shows how the separation of time scales $\Delta\tau$ changes as a function of space on the surface of a two dimensional feasibility domain (i.e., a line).

2.5 Conclusion

The structural stability of a population dynamics model can be used as a measure to quantify the tolerance of biological communities to environmental perturbations (Thom & Fowler, 1975). For example, the structural approach based on the classic LV model uses the size of the feasibility domain (the set of parameter values that guarantees the existence of a fixed point at which all species coexist) as a measure for structural stability (Rohr et al., 2014; Saavedra et al., 2014). Although this approach was derived to investigate the structural stability of the LV dynamics, its applicability is not limited exclusively to population dynamics models with linear functional responses (Saavedra et al., 2017b). Indeed, in this chapter, we have shown that the structural approach can be applied to investigate the structural stability of biological communities governed by a large class of deterministic nonlinear models. Furthermore, we have shown how to extend this approach to stochastic models.

In particular, we have shown that in order to guarantee the validity of the struc-

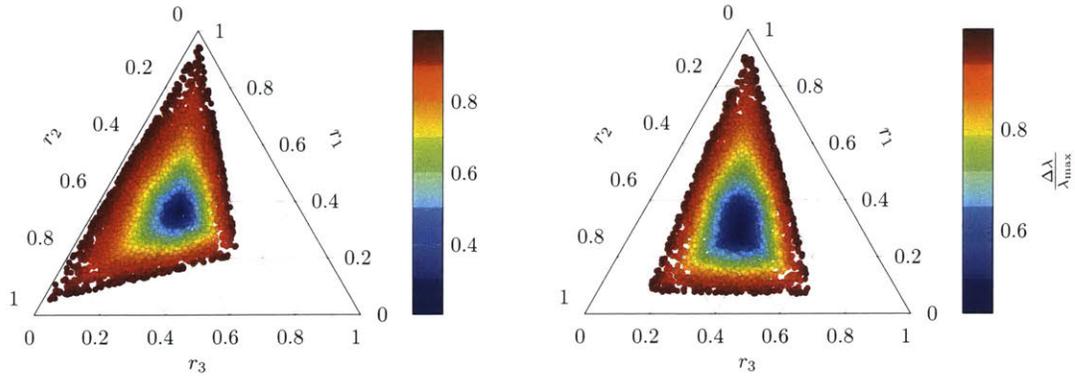


Figure 2.6: **Structural approach applied to stochastic Lotka-Volterra dynamics.** The figure shows the surface of two different feasibility domains in the parameter space of intrinsic growth rates r_i of 3-dimensional systems. The colors within the feasibility domains show the relative difference between the largest and smallest eigenvalue of the corresponding Jacobian matrix. Note that the separation of time scales between the most and least stable direction is proportional to the extinction probability. Then, the structural approach can be applied by measuring the size of the region with minimum time-scale separation (center blue region).

tural approach under a general nonlinear model, the functional response of Eq. (2.4) needs to satisfy Eq. (2.7). Importantly, these conditions are satisfied by many functional responses that have been found to reasonably explain experimental data (Ayala et al., 1973). The results of section 2.3 imply that the structural approach can be applied even when the underlying population dynamics model changes across time, provided that every model satisfies Eq. (2.7). This is relevant for the analysis of empirical data. In fact, because the functional response of biological communities are typically context dependent (Dick et al., 2017), it is advantageous to have methods that do not rely too strictly on the assumption of a particular functional form.

Finally, we have discussed the applicability of the structural approach to investigate the structural stability of the feasible fixed point for populations governed by the SLV dynamics. We have shown that in finite populations, structural stability is strongly correlated with the time-scale separation between fast and slow dynamics. Thus, it is possible to divide regions of the feasibility domain into those where the population exhibits either a zero or non-zero extinction probability in finite times. We

have proposed that the size of the region for which the extinction probability is zero can be used as an appropriate measure of structural stability in finite populations. Note that because the SLV dynamics are derived from a linear noise approximation of a master equation, the extension of the structural approach to Eq. (2.18) applies to finite but small noise. This should not be seen as a strong limitation of this approach. Empirical studies have shown that stochastic effects are not always the main drivers of community dynamics (Stegen et al., 2012). Indeed, recent experiments have shown that population dynamics, in closed communities, exhibit a strong deterministic component with stochasticity mainly driven by environmental variations (Frentz et al., 2015). Overall, we hope that this research will serve as a baseline for future studies investigating the structural stability of biological communities, where the dynamics are typically driven by time-varying, nonlinear, functional responses and by small demographic stochasticity.

CHAPTER 3

SPECIES INTERACTIONS IN CHANGING ENVIRONMENTS

Interactions among species within a community are typically represented in terms of networks of which nodes represent species and links interactions. The structural approach presented in the previous chapter was originally introduced to explain what type of structure of species interaction network maximize the capacity of a community to withstand environmental perturbations. In this chapter however, we show that network structure alone cannot be used as predictor of the robustness of biological populations in changing environments. Specifically, here we demonstrate that such a perspective leads to inconsistent conclusions and that tolerance to environmental changes can only be understood as an interplay between species interactions and environmental perturbations. Thus, we discuss a research agenda to investigate the relative importance of the structure of ecological networks under an environment-dependent framework. Overall, we hypothesize that only by studying systematically the link between network structure and population dynamics under an environment-dependent framework, we can uncover the limits at which biological populations can tolerate environmental changes.

3.1 Introduction

Since the beginnings of modern network theory (Newman, 2003, 2010), studies have assessed the importance of particular network structures (e.g., exponential or scale-free networks) by their capacity to tolerate an external perturbation acting on their structure or dynamics (e.g., a random or targeted sequential removal of nodes, Albert et al. 2000). This has paved the way for a similar research agenda in network and community ecology (Pascual & Dunne, 2005; Bascompte & Jordano, 2013). In particular, theoretical studies have been investigating this importance by quantifying the effects of network structure on community persistence (Montoya et al., 2006). To capture these effects, a typical approach has been centered on randomly removing species (removing interactions or sampling randomly model parameters) and comparing the extent to which different network structures avoid additional species extinctions. This tolerance has then been taken as evidence for a structure's advantage, disadvantage, or lack of any importance over other structural patterns (James et al., 2013; Sales-Pardo, 2017). However, the large number of degrees of freedom involved in these analyses (e.g., parameter values, choice of perturbation) has been a central limitation. In fact, it is unclear the extent to which such conclusions can be generalized (Saavedra et al., 2013; Grilli et al., 2016). Therefore, the question has become whether it is possible at all to infer the importance of a network structure through its capacity to tolerate external perturbations (Rohr et al., 2014).

In this chapter, we use a simple example to demonstrate that the tolerance to external perturbations of different network structures under the same dynamics can quickly change as a function of the type, direction, and magnitude of the perturbations. That is, the importance of a network structure depends on the external perturbations faced by a community at any given point in time (Song et al., 2017; Coulson et al., 2017; Cadotte & Tucker, 2017). Thus, if studies focus on a specific set of external perturbations in order to infer the general importance of a network structure, it would lead to inconsistent conclusions. This implies that the importance of a given network structure in a community should always be understood in relation to local environmental settings. In this line, we propose and discuss a research agenda

to investigate the relative importance of the structure of ecological networks under an environment-dependent framework. We strongly believe that this new synthesis can move the field of ecology towards a more systematic and predictive science (Petchey et al., 2015).

3.2 Linking network structure and persistence

To investigate the tolerance of an ecological network to external perturbations, studies have been linking the structure of interaction networks with community persistence. Traditionally, this structure has been derived from the topology of species interaction networks, i.e., the binary representation of who interacts with whom in a given location and time (Pascual & Dunne, 2005; Bascompte & Jordano, 2013). This topology can be represented by a binary matrix, whose elements denote the presence or absence of a direct interaction between two species. In order to talk about the structure of a network, it has also been necessary to talk about the lack of structure in a network (Newman, 2010). In this context, random networks (or null models) have been the gold-standard benchmark of no structure. Generally, these random networks are simple ensembles of binary matrices with a given number of 1s and 0s randomly shuffled. Statistically significant deviations from these random networks have been taken as a sign of a structure in an observed interaction network (Ulrich et al., 2009). Note that the characterization of a structure is not restricted to its topology, and many different definitions (e.g., weighted instead of binary patterns) and null models can be used (Schupp et al., 2017). Yet, the conceptual framework is exactly the same.

This framework has revealed that many real-world networks exhibit a distribution of interactions that depart from null expectations (Newman, 2010). For example, in network ecology, two of the structures that have captured most of the attention are modular and nested structures (Pascual & Dunne, 2005; Bascompte & Jordano, 2013). Modular structures are those in which groups of species have many interactions among them, but few interactions with the rest of the species in the network. Nested structures are those where highly connected species interact with both highly

connected and poorly connected species, while poorly connected species interact almost exclusively with highly connected species. Thus, many studies have been interested in understanding the existence and importance of such network structures through their potential links with community persistence.

Formally, community persistence corresponds to the capacity of a given community to sustain positive abundances for all its constituent species (Hofbauer & Sigmund, 1998). In many cases, this definition has been relaxed and taken as the fraction of species that can sustain positive abundances in a community when subject to some initial conditions or external perturbations (Pascual & Dunne, 2005; Ives & Carpenter, 2007; Bunin, 2017). In the absence of empirical data, the challenge has been how to model the temporal evolution of species abundances and their response to external perturbations (Rohr et al., 2014). Therefore, some studies have analyzed community persistence by assuming a random or targeted sequential removal of species (or interactions) in an interaction network, and considered extinctions (i.e., when abundances go to zero) whenever a species is left without interactions (Pascual & Dunne, 2005). Other studies have used population dynamics models to investigate the fraction of species that ends up with positive abundances at equilibrium under some random initial conditions (Rohr et al., 2014). Similarly, using population dynamics models, other studies have systematically investigated the range of parameter values (initial conditions) compatible with positive abundances of all species in a community (Saavedra et al., 2017b). Overall, regardless of the method employed, community persistence has been broadly defined as the capacity of a community to avoid species extinctions.

One can now link the two concepts above and ask, for example, to what extent modular or nested structures can increase the capacity of communities to avoid species extinctions under random perturbations. In fact, this has been a recurrent question in network and community ecology (Pascual & Dunne, 2005; Bascompte & Jordano, 2013; Sales-Pardo, 2017). To address this question, studies have fixed a community size, changed the structure systematically by re-arranging the interactions of the networks, adopted a measure of community persistence, introduced some external

perturbation or condition, and investigated how this measure of community persistence changes as a function of a network structure (Stouffer & Bascompte, 2003; Bastolla et al., 2009; Thébault & Fontaine, 2010; James et al., 2013; Rohr et al., 2014). Importantly, these theoretical studies have shown significant associations between network structures and community persistence under given external perturbations.

However, the reading of these findings has many times led to a belief that a structure is either advantageous (important) or not for a community (Sales-Pardo, 2017). For example, modular and nested structures have been positively associated with community persistence in antagonistic and mutualistic communities, respectively (Thébault & Fontaine, 2010). Thus, changes of a presumed important network structure in a community across temporal or environmental gradients have been directly translated to changes in robustness (Dalsgaard et al., 2013; Takemoto & Kajihara, 2016; Welti & Joern, 2015). Moreover, this view has led other studies to suggest that the absence of a given structural pattern in a community can be related to the lack of importance of such network structure overall (James et al., 2013; Strona & Veech, 2015). Yet, all these generalizations are derived from particular scenarios of external perturbations, and currently it is unknown whether these results are consistent under a more systematic analysis (Saavedra et al., 2013; Grilli et al., 2016).

3.3 Inconsistent conclusions about the importance of network structure

To illustrate how naive simulations of external perturbations can lead to inconsistent conclusions about links between network structure and community persistence we follow the structural approach (Thom & Fowler, 1975; Rohr et al., 2014; Cenci & Saavedra, 2018a) discuss in chapter 2. This approach is particularly useful for our purposes as it allows us to focus on how the qualitative behavior of a dynamical system changes as a function of the parameters of the system itself. For example, the dynamics of the system can be approximated by a population dynamics model (Case, 2000). Then, the qualitative behavior of this dynamical system can be trans-

lated into a given measure of community persistence. Thus, one can investigate the extent to which different interaction networks can tolerate external perturbations (changes in parameter values) without pushing species towards extinction.

Let us consider a 2-species community as a conceptual example. The axes of Figure 5.1 represent the 2-dimensional parameter space of species intrinsic growth rates. The colored regions correspond to the set of those intrinsic growth rates compatible with positive species abundances (the necessary condition for community persistence). The size and shape of this region depend upon network structure (Cenci et al., 2018a). This region is what in chapter 2 we called the feasibility domain. (Logofet, 1993; Rohr et al., 2014; Cenci & Saavedra, 2018a). Notice that in chapter 2 we have shown that the structural approach has a precise geometrical interpretation (see Figure 2.1 and Eq. (3.2)). Here we will be somehow less rigorous and we will illustrate our point with a cartoon example. This simplified example make it is easier to generalize the idea in higher dimensions (i.e., larger communities). The open and colored symbols represent some initial and final parameter values, respectively, i.e. the parameters of the system before and after a hypothetical external perturbation. Rows correspond to the same external perturbation under two different network structures (for the moment the reader can think of any type of structures). Columns correspond to the same network structure under two different external perturbations. Note that positive species abundances will be satisfied as long as the parameter values fall inside the feasibility domain. The Figure clearly illustrate that, if we were to focus on the first row only, we would conclude that structure 1 is more robust than structure 2. However, if we were to focus on the second row, then we would conclude the opposite. Similarly, if we were to focus on each column separately, we would arrive to contrasting conclusions. Moreover, these inconsistent conclusions can be repeated by moving the perturbation (parameter values) to almost any other direction. That is, there is no conceptual support to think of either a positive or negative association between structure and persistence without taking into consideration external environmental conditions, i.e., the direction and the magnitude of the perturbation.

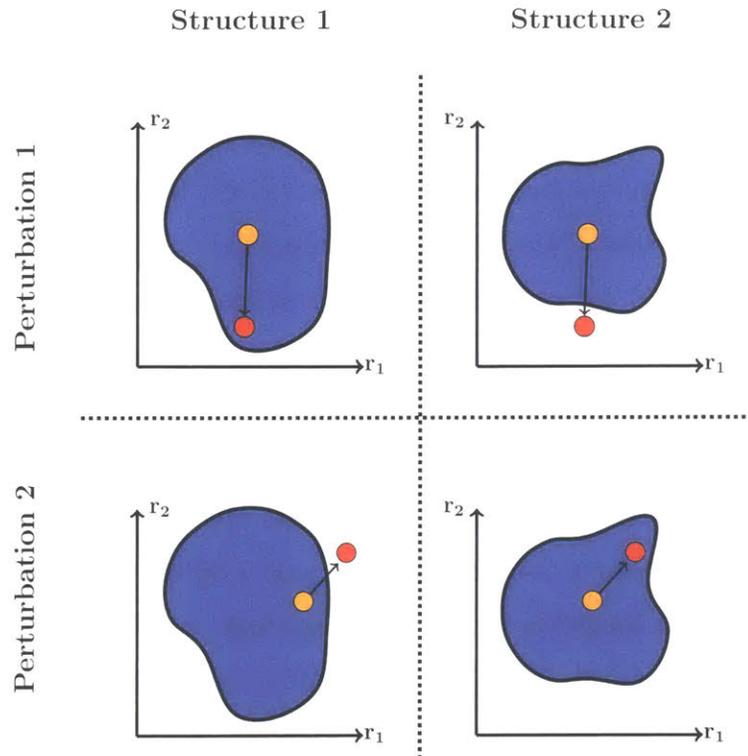


Figure 3.1: **Linking external perturbations, network structures, and community persistence.** The blue region represents the feasibility domain (parameter space compatible with community persistence) of a population dynamics model. The orange and red circles represent a vector of species intrinsic growth rates $\vec{r} = [r_1, r_2]$ before and after a hypothetical perturbation, respectively. The necessary condition for community persistence is to have a vector of intrinsic growth rates within the feasibility domain (as we show on the top-left and bottom-right panels). The cartoon shows that not only the structure of an interaction network is important for community persistence, but also the direction of the perturbation. In fact, just by changing the direction of the perturbation, one may not observe community persistence under the same network structures (as we show on the top-right and bottom-left panels). That is, structure *per se* says little about community persistence if not seen in the light of its local environment.

Figure 5.1 provided a conceptual example. To quantitatively illustrate the inconsistent conclusions about the importance of the structure of interaction networks through their tolerance to external perturbations, we test the association of community persistence with modular and nested structures under LV dynamics. Yet, we need to stress that, in principle, our approach can be applied to any combination of structures, perturbations, and models. Indeed as shown in chapter 2 the following analysis could be carried out using any GLV dynamics belonging to the class of the

structural approach. We measure community persistence as the capacity of a particular structure to avoid extinctions. We build interaction networks on communities of 21 species (this number allows us to easily divide the network into modules, but different dimensions generate the same qualitative results). Interactions are distributed among the species so that there is a clear distinction between the two types of structures analyzed (see 5.2 for a graphical representation). For comparison purposes, the elements of the interaction matrix \mathcal{A} are taken from a normal distribution with parameters chosen such that the resulting interaction matrices for each structure have same mean and standard deviation (so to compare fairly among structures). In the absence of an interaction between two species, the corresponding entry in the interaction matrix is zero. Communities (each with a different type of structure) are initialized inside the feasibility region by fixing a lognormal distribution of species abundances \vec{x}^* (Begon et al., 2009), and then finding the corresponding vectors of intrinsic growth rates i.e., $\vec{r} = \mathcal{A}\vec{x}^*$ (Rohr et al., 2016). Once the communities (with the different structures) are initialized with all species present ($\vec{x}^* > 0$), we introduce random and directional perturbations on either the interaction matrix or the vector of intrinsic growth rates. While random perturbations act on all the elements of the interaction matrix or the vector of intrinsic growth rates, directional perturbations act on one single column or element. These changes are equivalent to random and targeted perturbations either on the interactions or nodes of a network (Saavedra et al., 2013, 2014). After the perturbations, we compute the new equilibrium solution with the changed parameters. Then, we record which community (network structure) avoids extinctions. We repeat this process 5000 times.

Figure 5.2 shows the estimated community persistence (number of times a given structure avoids extinctions) derived for each combination of structure and perturbation. The first row corresponds to the community persistence under random perturbations acting on either the interaction matrix (Panel A) or the vector of intrinsic growth rates (Panel B). The second row corresponds to the community persistence under directional perturbations acting on either the interaction matrix or on the vector of intrinsic growth rates of the most and least connected species. In the same line as in Figure 5.1, if we were to focus on the first row only, we would conclude that

modular and nested structures are more robust under perturbations acting on the interaction matrix and intrinsic growth rates, respectively. However, if we were to focus on the second row, then we would conclude the opposite (Panels C and F), or simply that there is no difference between the structures (Panels D and E). Similarly, if we were to focus on each network structure separately, we would arrive to contrasting conclusions as a function of the perturbations. Note that these inconsistencies are not exclusive to the perturbations here analyzed. Overall, these simple conceptual and quantitative analyses demonstrate that the association of a given network structure with community persistence completely depends on the type, direction, and magnitude of perturbations.

3.4 A plea for an environment-dependent framework

As it is known, species interaction networks are the result of different evolutionary and ecological processes acting at the individual and the collective level (Thompson, 2005). Because these processes typically yield to adaptation to local environments (Grant & Grant, 2014; Valverde et al., 2018), it becomes useful to think about the relative importance of a network structure under a particular environmental setting. That is, the importance of a network structure should be studied under an environment-dependent framework (Chamberlain et al., 2014; Song et al., 2017). This research agenda can be achieved by using environment-dependent parameters as the link between community dynamics and environmental conditions (Rohr et al., 2016; Cadotte & Tucker, 2017). The goal will be to investigate the range of environmental conditions compatible with community persistence under a given network structure, the expected environmental conditions in a given location, and their overlap. Below, we explain this environment-dependent framework in more detail.

The first step is to systematically study the set of environmental conditions tolerated by a community with a given network structure. This requires to link community dynamics and environmental conditions, which can be achieved through specifying environment-dependent parameters in a model (e.g., species carrying capacities)

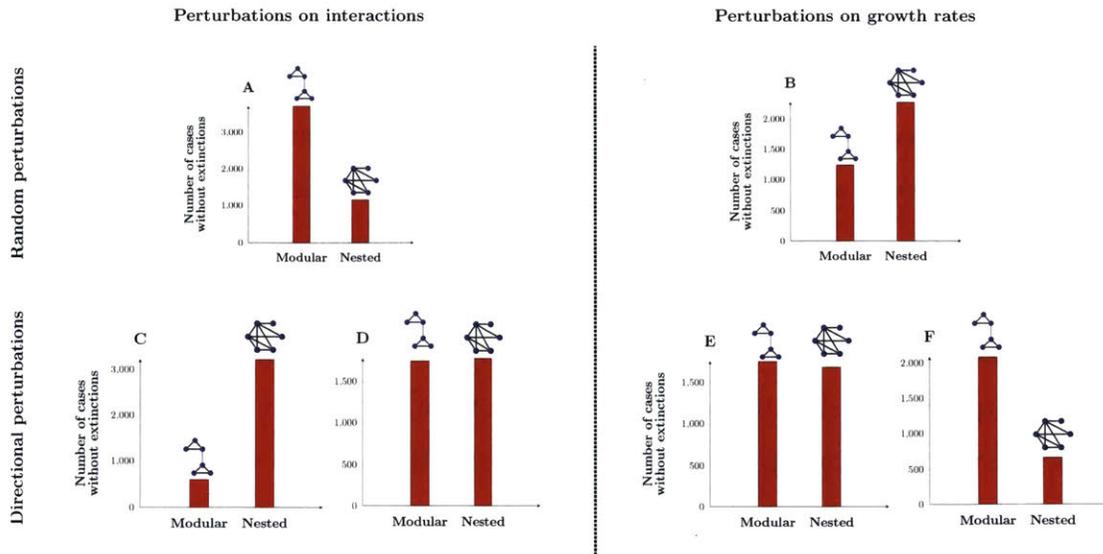


Figure 3.2: Inconsistent conclusions about the importance of network structures. As an example, we show the response of two different network structures (a modular and a nested structure) to different external perturbations (see text for details on the simulations performed using a Lotka-Volterra competition model). All communities are initialized inside the feasibility domain having the same species abundance distribution (see text for details). The y-axis corresponds to the number of times (out of 5000) that the community tolerated a perturbation (i.e., no species goes extinct). That is, the larger the bar, the more tolerant the network structure. Panel **A** corresponds to random perturbations on species interactions. Panel **B** corresponds to random perturbations on species intrinsic growth rates. Panels **C** and **D** correspond to directional perturbations on species interactions. That is, only the values of one column of the interaction matrix are changed in each case. Finally, Panels **E** and **F** correspond to directional perturbations on species intrinsic growth rates. That is, only one growth rate of one species is changed in each case.

(Coulson et al., 2017; Cadotte & Tucker, 2017). That is, instead of studying the tolerance of a network structure as a function of random environmental conditions, we propose to study the set of conditions (range of environment-dependent parameter values) compatible with the persistence of the community. This set is what we have previously called the feasibility domain. Recall that this domain is defined by the particular dynamics and network structure of a community (Rohr et al., 2014). As an illustration, Figure 8.2A shows the feasibility domain of two different network structures under the same population dynamics. Importantly, in many cases, the size and shape of the feasibility domain can be analytically investigated (Saavedra et al., 2017b).

The second step involves estimating the range of environment-dependent parameter values compatible with the environment in a given location (i.e., the local environment over a period of time) (Petchey et al., 2015; Coulson et al., 2017). For example, it may be possible to map the influence of environmental conditions; such as temperature, humidity, or soil composition, into species carrying capacities (Kuang & Chesson, 2009; Levine & HilleRisLambers, 2009; Narwani et al., 2013; Godoy et al., 2014; Germain et al., 2016; Song et al., 2017). Similarly, species carrying capacities can be linked to the total amount of available resources per consumption rate, where monocultures have been used as an experimental ground to estimating such values (Case, 2000). Note that this estimation could also be represented by a probability distribution for the carrying capacities (Cadotte & Tucker, 2017). That is, a measure of which type of carrying capacity are more likely to be observed in a given environment. As an illustration, Figure 8.2B shows, a hypothetical environment in a given location as a function of environment-dependent parameters. Note that this characterization would also require knowledge about climatic conditions or the use of weather models (Schleuning et al., 2016), as well as knowledge about how individual species respond to those changes. While this task could be challenging, both new theoretical (Benadi et al., 2013; Amarasekare & Coutinho, 2013; Poisot et al., 2015; Csergő et al., 2017; Hunter-Cevera et al., 2017; Holt et al., 2017) and empirical studies (Vázquez et al., 2009; Schleuning et al., 2016; CaraDonna et al., 2017) are providing a good guideline towards this goal. Overall, the task is to characterize a representative set of potential environmental conditions for a given location rather than an arbitrary set of random external perturbations.

The third step corresponds to merging steps one and two. Because it is virtually impossible to know the type, direction, and magnitude of environmental conditions acting on a community at every given point in time; it becomes useful to study the extent to which the environmental conditions compatible with a given network structure overlap with the environment faced by a community in a given location. This overlap corresponds to the proportion of the feasibility domain of a community that is inside the set of environment-dependent parameter values expected in a given location (for a graphical example see Figure 8.2C). Formally, this proportion can be

defined by the ratio of the following volumes

$$\Gamma(D_F(\mathcal{A}_i) \cap \Xi_j) = \text{vol}(D_F(\mathcal{A}_i) \cap \Xi_j) / \text{vol}(\Xi_j) \quad (3.1)$$

where we recall from chapter 2 section 2.2 that:

$$D_F(\mathcal{A}_i) = \{\vec{r} \in \mathbb{R}^d \mid \exists \lambda_1, \dots, \lambda_d > 0, r_j = \sum_k \lambda_k a_{kj}\} \quad (3.2)$$

corresponds to the feasibility domain of a community i (network structure) and Ξ_j corresponds to the distribution j of environment-dependent parameter values expected in an environment. In other words, community persistence should always be measured under an environment-dependent framework (Levins, 1968). This definition implies two important concepts: (1) under the same environment Ξ_j , differences among structures (\mathcal{A}_1 and \mathcal{A}_2) can reflect differences among feasibility domains, and, in turn, differences among community persistence (Γ_{ij}). (2) under different environments (Ξ_1 and Ξ_2), differences among feasibility domains do not directly reflect differences between community persistence. That is, a community with a small feasibility domain can have a much greater community persistence than a community with a large feasibility domain if the environment happens to overlap more with the domain of the small one (see 8.2C for an illustration). While this approach certainly does not remove the existence of inconsistent conclusions under different environments (e.g., moving from Ξ_1 to Ξ_2), it allows us to better characterize the link between network structures and environmental settings.

3.5 Final thoughts

We have shown that it is necessary to rethink the importance of the structure of ecological networks under an environment-dependent framework; otherwise, we may miss the forest for the trees. Because local adaptation is the *leitmotiv* of natural selection, a plausible hypothesis is that network structures are emergent responses of

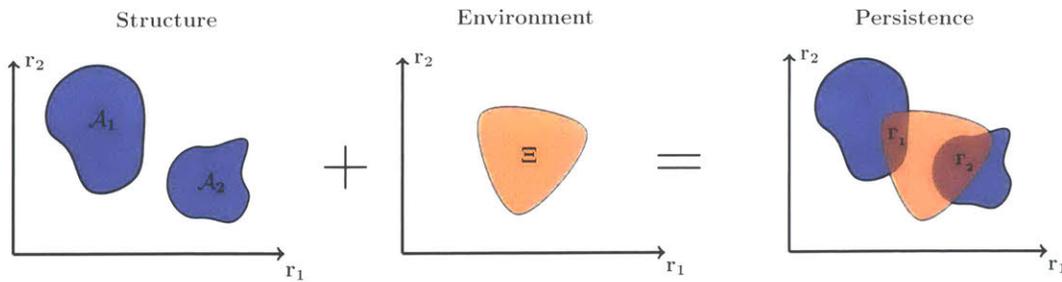


Figure 3.3: **The relative importance of network structures under an environment-dependent framework.** This cartoon illustrates the extent to which the importance of network structures depends on local environmental settings. The left panel shows two different feasibility domains (range of environment-dependent parameter values compatible with community persistence) of a hypothetical 2-dimensional system. These domains are generated by two different network structures characterized by their corresponding interaction matrices. Note that the left domain is larger than the one on the right. The middle panel corresponds to a hypothetical characterization of an environment in a given location (e.g., set of environmental conditions over a period of time) as a function of environment-dependent parameters. This region of parameters can also be represented by a distribution. The right panel corresponds to the overlap between the environment and the two network structures (i.e., community persistence). Note that the smaller feasibility domain A_2 can have a larger overlap with the environment Ξ (i.e., $\Gamma_2 > \Gamma_1$ in the Figure), providing a higher probability of persistence for the community.

communities to their environment (Odum, 1969; Margalef, 1968; Levins, 1968; Levin, 2005; Coulson et al., 2017; Cenci et al., 2018a; Valverde et al., 2018). That is, there is no one better network structure than other in general. However, there can be one network structure more tolerant than other structure to a particular environment. This also implies that a change in network structure is neither advantageous nor detrimental *per se*. Changes in network structure can be the result of different factors, such as: external perturbations, ecological or evolutionary processes, invasions, and extinctions, among others. Similarly, these changes can be adaptive or non-adaptive, revealing that changes of a presumed important network structure in a community across temporal or environmental gradients cannot be directly translated to changes in robustness. In fact, studies have already shown contrasting effects of structural changes on adaptability (Kondoh, 2003; Gilljam et al., 2015). Thus, again, these analyses can only make sense within an environmental context. Therefore, the lack of a structural pattern in one or several communities does not imply that such structure

has no importance at all. It may only imply that such structure is not particularly advantageous under the current environmental settings.

While we have developed a non-exhaustive quantitative exercise of the many possible structures, dynamics, and perturbations, we have shown clear counter-examples of how conclusions about the importance of a network structure derived from particular combinations are inconsistent. Thus, because it is virtually impossible to know *a priori* all the characteristics of future perturbations (Stenseth et al., 2002; Walther, 2010; Scheffers et al., 2016); we need to abandon a dichotomous view, and systematically link network structure and community persistence under an environment-dependent context. While the approach we have outlined in this work is by no means the only possible one, we hope it can be used as a guideline towards a better understanding of both the existence of structural patterns in ecological networks and how communities may respond under future scenarios of climate change.

Building on the ideas developed in this chapter, in the next chapter we will perform a more quantitative analysis. Specifically, we will extend the structural approach to systematically understand the importance of network structure and the reorganization of species interactions under an environment-dependent framework.

CHAPTER 4

ADAPTABILITY IN CHANGING ENVIRONMENTS

Building on the discussions of the previous chapter, here we will derive an extension of the structural approach to estimate the effects of the reorganization of interactions on the adaptability of species to changing environments. Specifically, looking at species interactions under an environment-dependent framework we derive a measure of network structure that can be used as predictor for likelihood of adaptability in fast changing environments. This measure estimate the size of what we call adaptation space, i.e., the set of environmental conditions compatible with the persistence of species in changing environments. To test our theoretical framework we analyze data from several biological populations divided in controlled and artificially perturbed. Overall, we find that the groups of species present in both control and perturbed populations are among the ones with the largest adaptation space. We believe that the results derived from our framework point out towards new directions to understand and estimate the adaptability of species to changing environments.

4.1 Introduction

In chapter 3 we have shown that environmental variations can change the conditions compatible with the persistence of species in biological populations by affecting relevant biological parameters, e.g. the intrinsic growth rates. This is not only a mathematical statement but rather a well known issue in population dynamics studies (Levins, 1968; Odum, 1969). For example, it has been shown that variations in resource availability can have significant effects on species fitness (Dmitriew, 2011), species interactions (Dirzo et al., 2014), and, eventually, community composition (Snyder, 2008; Scheffers et al., 2016). Although biodiversity changes have been well studied and documented (Walther et al., 2002; Atkins & Travis, 2010), predicting them is still difficult (Sala et al., 2000; Dirzo et al., 2014; Scheffers et al., 2016). Yet, understanding and anticipating the impact of environmental disturbances on biological populations is of central importance for developing strategies for sustaining ecosystem services that depend on biodiversity, such as: nutrient cycling, water purification, and soil formation, among others (Dirzo et al., 2014; Scheffers et al., 2016).

Forecasts of biodiversity loss, due to global changes (Kumar Duraiappah & Naeem, 2005), have been criticized for not considering species capacity to adapt to novel environmental conditions (Skelly et al., 2007; Martin et al., 2012). Indeed, this is also a pitfall of the structural approach which depict a static view of community dynamics in changing environments. For example, in chapter 3 we have shown that different network structure can be more or less resilient under different type of perturbations but we did not include in our analysis the possibility of adaptive changes of species interactions. The extension of the structural approach to take into consideration ecological adaptive dynamics is going to be the main focus of this chapter and the first attempt to systematically look at the role of network structure under and environment-dependent framework as proposed in chapter 3.

4.2 Background

It is well known, from both theoretical and observational studies, that not all species respond equally to environmental perturbations (Dobzhansky, 1973; Davis et al., 2005; Saavedra et al., 2013). The adaptability of species to variable environments depends on both evolutionary and ecological processes operating at the species and community level (Pelletier et al., 2009). At the species level, these processes range from genetic to morphological, physiological, and behavioral changes (Hairston et al., 2005; Carlson et al., 2014; Chevin et al., 2010). At the community level, species interaction networks and their dynamics play a crucial role in shaping the tolerance of species to environmental perturbations (Pascual & Dunne, 2005; Saavedra et al., 2013; Alexander et al., 2015). In fact, empirical and theoretical studies have shown that the reorganization of interactions among co-occurring species can modulate the chances of species to persist under variable environmental conditions (Staniczenko et al., 2010; Ramos-Jiliberto et al., 2012; Saavedra et al., 2016a, 2017a; CaraDonna et al., 2017). However, empirical evidence has shown that these ecological dynamics typically facilitate the persistence of groups of species rather than entire communities (Lopezaraiza-Mikel et al., 2007; Walther, 2010; Montero-Castaño & Vilà, 2015; Rafferty et al., 2015; Fukami, 2015; Alexander et al., 2015; Saavedra et al., 2017a). Unfortunately, so far, we have no systematic methodology to investigate when a reorganization of interactions among a group of species can lead to adaptation, and how to identify those groups of species within a community with the highest or lowest chances to adapt. Yet, this could prove extremely relevant to help us improving current forecast models of biodiversity loss (Dirzo et al., 2014; Scheffers et al., 2016).

In this chapter, we aim to shed new light on the conditions under which the reorganization of interactions among a group of species within a community can have a significant effect on their adaptability to changing environments. We define reorganization of interactions as changes of trophic interactions between co-occurring species in a given location and time. We introduce the concept of adaptation space of a group of species by means of a new theoretical framework which is based on the feasibility analysis of a population dynamics model (Rohr et al., 2014, 2016; Saave-

dra et al., 2014, 2017b,a). In a feasibility analysis, one assumes the existence of environmental perturbations on the dynamics of the population, and study the set of conditions under which species persistence is possible. Thus, we define the adaptation space of a group of species as the region of environmental conditions that can be made compatible with its persistence thorough the reorganization of interactions among species within the group.

In particular, we focus on the adaptability of consumers in consumer-resource (CR) interaction networks. CR interactions are arguably the most common type of ecological interaction networks in Nature (Murdoch et al., 2003; Olff et al., 2009). In fact, many direct and indirect trophic interactions, such as: resource competition, predation, and mutualism are recognized as different topological configurations of CR interactions (Holt & Polis, 1997; Pauw, 2013). Mutualism, for example, can be seen as a bi-directional CR dependency (Olff et al., 2009; Pauw, 2013), in which consumers are the resources of their resources. In the reminder of this manuscript, we provide a formal and detailed definition of all our methods and concepts. Then, we provide and discuss a numerical analysis to test the validity of our theoretical framework. Then, we analyze two independent manipulative experiments on pollinator-plant interaction networks. In both experiments, a rapid disturbance was applied on the community and the immediate behavioral response of the pollinator was recorded. We use both the numerical examples and data sets to study and identify the groups of consumers (pollinators) with the largest and smallest adaptation space, i.e., with the highest and lowest chances to adapt to an environmental perturbation following a reorganization of their CR interactions.

4.3 Theoretical Framework

4.3.1 Consumer-resource interaction networks

The standard way of representing CR interactions is through a bipartite network (Pascual & Dunne, 2005). A bipartite network is a graph $\mathcal{G}(Q, K, E)$, where Q and K are disjoint sets of nodes (i.e., species). The separation of the nodes into two

disjoint sets mirrors the separation of species between two different trophic levels: consumers and resources. The set E contains edges that connects species in the set Q with species in the set K (i.e., inter-trophic interactions). Note that the two sets of nodes (i.e., the two trophic levels) do not necessarily have the same number of elements (i.e., species), and the adjacency matrix of a bipartite network is a rectangular binary matrix that we will denote by β . This bipartite representation of the network only provides information about interactions between consumers and their resources (i.e., inter-trophic interactions). Bipartite networks do not describe intra-trophic interactions (e.g., the strength of plant-mediated competition between two pollinators). In the following sections, we will show a method to infer these intra-trophic interactions from the bipartite structure of the network.

We define groups of species in a bipartite interaction network as a subset of (nodes) species with at least one (node) species per (set) trophic level. Formally, a group is a subgraph $g(q, k, e)$ of $\mathcal{G}(Q, K, E)$ with $q \subseteq Q$, $k \subseteq K$ and $e \subseteq E$. To preserve the bipartite structure of the group, we need to assume that in any group there is at least one consumer and one resource (i.e., $\min(|q|) = \min(|k|) = 1$). Also we do not impose constraints on the number of interactions within the group. For example, a group can be a subset of non-interacting species (i.e., $e = \emptyset$). The dimension of a group is the product of the number of species of the two sets of the subgraph (i.e., $d = |q||k|$). Note that this definition of a group simply allows us to identify all the possible combinations of species within a community. This identification is relevant as it is not known a priori which species will be able to adapt to new environmental conditions (Lopezaraiza-Mikel et al., 2007; Staniczenko et al., 2010; Kaiser-Bunbury et al., 2010; Montero-Castaño & Vilà, 2015).

4.3.2 Feasibility analysis

To estimate the effect of the reorganization of interactions on the adaptability of a group of consumers within a community, we focus on the conditions leading to species permanence (Hofbauer & Sigmund, 1988). Specifically, we derive a theoretical framework for consumers based on the feasibility analysis of a generalized

Lotka-Volterra (GLV) model as discussed in chapter 1 (Saavedra et al., 2014; Rohr et al., 2016; Saavedra et al., 2017b; Cenci & Saavedra, 2018a):

$$\dot{\vec{x}} = g(\vec{x})(\vec{r} - \mathcal{A}f(\vec{x})). \quad (4.1)$$

Recall from chapter 2 that a vector of intrinsic growth rates $\vec{r} \in D_F(\mathcal{A})$ is called a feasible vector for the community defined by the interaction matrix \mathcal{A} . Note that feasibility is a necessary condition for persistence but is not a sufficient one (Hofbauer & Sigmund, 1988). From an ecological perspective, one is mainly interested in solutions of Eq. (4.1) that are *permanent* (Sigmund, 2007) (i.e., positive abundances of species that neither vanish nor increase indefinitely). Hence in the following, we focus on solution that are feasible and globally stable, which for Eq. (4.1) corresponds to a sufficient condition for species permanence (Baigent & Hou, 2012) (see Appendix 4.7.1 for a more detailed discussion on global stability and permanence).

4.3.3 Parameterization of the interaction matrix

The interaction matrix \mathcal{A} entirely defines the feasibility domain of a resource-mediated community of consumers. However, empirical data on the strength of interspecific interactions are seldom available because of difficulties in performing experimental measurements in large communities (Trøjelsgaard & Olesen, 2016; Laska & Wootton, 1998; Pekkonen & Laakso, 2012). Thus, it has become necessary to develop methods to investigate the behavior of ecological systems under limited information (Yeakel et al., 2012; Rohr et al., 2014; Saavedra et al., 2014, 2016a, 2017a). A typical approach is to consider random interaction matrices, where the strength of the interactions between consumers are sampled following a niche overlap framework (Logofet, 1993; MacArthur, 1968; May & Mac Arthur, 1972; May, 1974). Under this framework, the strength of interactions (i.e., the elements of the interaction matrix) are generated by defining a resource spectrum for all species and a resource utilization function for each consumer (typically, a set of probability distributions on a resource space). The interaction strength between two consumers then corresponds to the overlap between their niches (or probability distributions). The larger their niche overlap, the

larger their interaction strength. Here, we follow the same rationale but, instead of defining an arbitrary resource spectrum and utilization functions, we approximate the niche of a consumer with the set of resource-species this consumer is interacting with. Then, we derive the level of niche overlap between consumers from the binary matrix β of the CR interaction network. That is, instead of measuring overlap of probability distributions, we compute correlations between binary vectors (i.e., the rows of β). The advantage of this approach is that the binary matrix of the CR interaction network can be inferred from data (Bartlett, 2017), or directly measured from empirical studies (Saavedra et al., 2014; Rohr et al., 2016; Saavedra et al., 2017a).

Specifically, to parameterize the elements of the interaction matrix for consumers, we used the monopartite projections of the binary matrix β of the CR interaction network (Saavedra et al., 2014). The monopartite projections of the binary matrix is defined as $\Lambda_{\text{col}} = \beta^T \beta$ (also called the *Gram matrix*, Saracco et al. (2015)). The off-diagonal entries of the monopartite projection Λ_{ij} count the paths of length two between species i and j , which correspond to the number of resources shared between two consumers. Thus, we can take the pairwise interaction strength to be proportional to Λ_{ij} . Rescaling the entries of the matrix by the sum of their column (i.e., $\Lambda'_{ij} = \frac{\Lambda_{ij}}{\sum_i \Lambda_{ij}}$), we have a left-stochastic matrix Λ' whose elements Λ'_{ij} can be interpreted as the probability of species j of affecting species i . That is, if species i and j share many resources, their probability of affecting each other is high. On the other hand, if the two species do not share any resources, the probability of affecting directly each other is zero.

Following niche theory and without any loss of generality (MacArthur & Levins, 1967; Pauw, 2013; Bastolla et al., 2002), the diagonal entries (the intra-specific competition) can be set to one, i.e., $\Lambda_{ii} = 1$. Fixing the diagonal to one generates a uniform bias in the direction of the spanning vectors of the feasibility domain to all matrices studied. That is, setting the intra-specific competition to one is equivalent to choosing a reference system that we will hold fixed throughout the analysis. The interaction matrix \mathcal{A} is then of the form $\mathcal{A} = \Lambda \circ \mathcal{R}$, where \circ is the Hadamard (entry-wise) product and

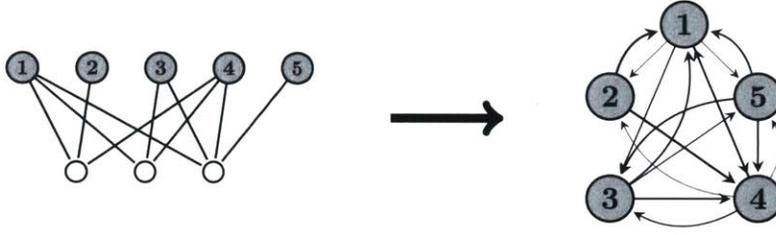


Figure 4.1: **Example of a resource-mediated community of consumers.** The network on the left hand side of the figure is the consumer-resource interaction bipartite network. The width of the arrows in the right figure represents the relative effect of one species on the other.

$$\mathcal{R}_{ij} = \begin{cases} \frac{1}{\Lambda_{ij}} & \text{if } i = j \\ \frac{1}{\sum_i \Lambda_{ij}} & \text{Otherwise} \end{cases} \quad (4.2)$$

We can then normalize the columns of \mathcal{A} to one in the L_2 -norm given that the feasibility conditions, as we mentioned before, are determined by the directions and not by the magnitudes of vectors in the parameter space (Logofet, 1993; Rohr et al., 2014; Saavedra et al., 2017b). Note that this parameterization generates asymmetric interaction matrices. That is, the effect of species i on species j can be different from the effect of species j on species i .

The framework we have defined in this section relies on a number of assumptions. Firstly, to each binary matrix β corresponds one and only one parameterization of the interaction matrix \mathcal{A} . The only free parameter is, therefore, the configuration of the CR interaction network, as a function of which we can study the feasibility conditions of Eq. (4.1). Secondly, a number of reasonable ecological constraints are satisfied (Margalef, 1968; MacArthur & Levins, 1967; Pauw, 2013; Büchi & Vuilleumier, 2014), such as: if two species do not share a common resource they do not affect each other directly, the strength of the interspecific pairwise effect is proportional to the number of shared resources, specialist species are better competitors than generalist, and finally intra-specific effects are stronger than the sum of interspecific effects. See Figure 4.1 for a graphical example.

4.3.4 Environmental changes

As mentioned above, to satisfy the conditions of species permanence, the vectors of intrinsic growth rates of a group of consumers need to be aligned to a feasible vector \vec{r}_f of the corresponding interaction matrix \mathcal{A} (Rohr et al., 2014; Saavedra et al., 2014). Under changing environmental conditions, however, the intrinsic growth rate of species can significantly change (Levins, 1968; Roughgarden, 1975). That is, environmental perturbations can turn feasible vectors \vec{r} into unfeasible vectors by moving them towards one of the borders of the feasibility domain (Saavedra et al., 2014; Rohr et al., 2016). Once a border is reached, the trajectory of Eq. (4.1) in the abundance space will tend to the border of the positive orthant which, in ecological terms, corresponds to one or more extinctions. The number of extinctions depends upon where on the feasibility domain the perturbed \vec{r} is pushed (Saavedra et al., 2014, 2017b; Rohr et al., 2016). For instance, if \vec{r} is on a vertex of the feasibility domain, then all the species but one will go extinct. If instead \vec{r} is on the border but not on a vertex, then there will be n extinctions with $n \in \{1, \dots, N - 1\}$. In our framework, we define changes in environmental conditions as perturbations to the mean $\langle \vec{r} \rangle$ of the vectors of intrinsic growth rates \vec{r} of consumers. Furthermore, we assume that environmental changes can happen randomly in any direction (Saavedra et al., 2017b).

4.3.5 Measuring the size of the adaptation space of a group of species

Following our definition of environmental change, a group of consumers within a resource-mediated community may respond to an external perturbation by reorganizing their interactions (by changing their resources) and moving their feasibility domain of intrinsic growth rates towards the direction of the new perturbed vector \vec{r}_p with mean $\langle \vec{r}_p \rangle$. We assume that the reorganizations take place within a time scale τ_{pert} that is smaller than the time scale of the population dynamics t_{pop} . That is, we assume $\tau_{\text{reorg}} \sim \tau_{\text{pert}} < t_{\text{pop}}$. Note that this reorganization only happens among species inside the group by creating a new feasibility domain for the group. This is a projection of the feasibility domain of the community. Figure 4.2 shows an illustration of

Before the perturbation After the perturbation

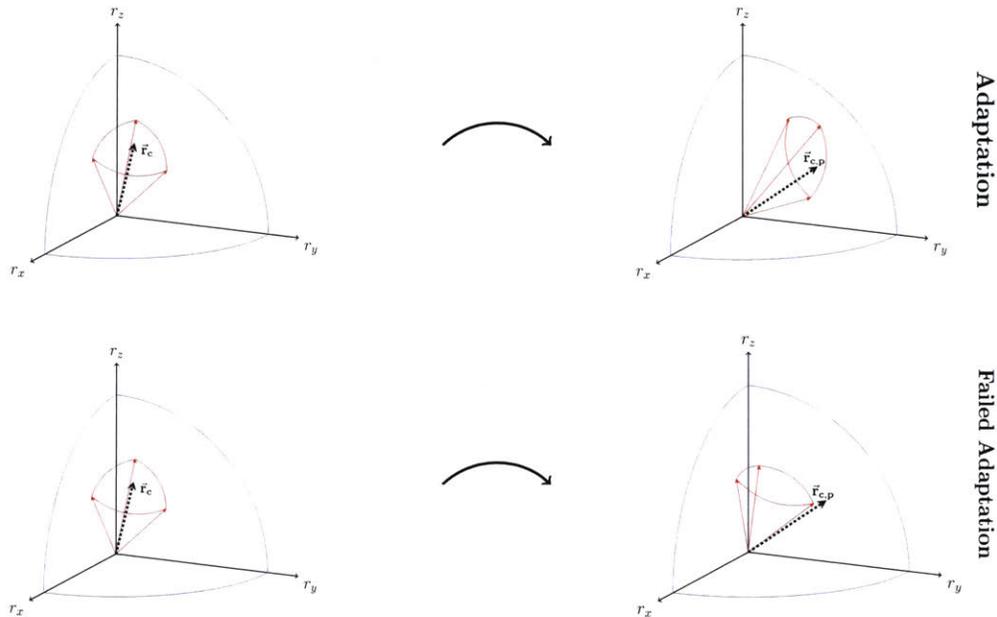


Figure 4.2: **Illustration of the impact of reorganization of interactions on the adaptability of species to environmental perturbations.** The red cones correspond to the feasibility domain $D_F(\mathcal{A})$ of a group of three species within a community defined by the interaction matrix \mathcal{A} . The spanning vector of the feasibility domain are the column vectors of the interaction matrix. The dashed arrows correspond to a hypothetical vector of intrinsic growth rates. Note that when this vector is inside the feasibility domain, the three species can persist, i.e., the environmental conditions are compatible with those necessary to sustain a positive abundance of the three species. The top and bottom panels present a scenario where the vector starts at a given location and then it is pushed away by environmental changes. The top panel corresponds to the case when the feasibility domain of the species is moved following a reorganization of interactions as a response to the perturbation. In contrast, the bottom panel presents a scenario where the feasibility domain does not move following a reorganization of interactions and the species failed to adapt to the new conditions—at least one of the species goes extinct.

this response in a group of three species. Different reorganizations can generate different feasibility domains. The set of the different feasibility domains generated by the combination of all the possible reorganizations of interactions within the group defines a region in the parameter space that is compatible with persistence. We call this region the adaptation space of a group (see Figure 4.3 for a graphical example). The larger the adaptation space, the larger the effect of the reorganization of interac-

tions on the adaptability of a group of consumers. Thus, here we propose to estimate the likelihood of a group to adapt to novel environmental conditions by estimating the size of its adaptation space.

To estimate the size of the adaptation space of a group of species within a community, we compute the fraction of distinct feasibility domains that can be generated through the reorganization of interactions and their average distance. To do so, we approximate the position of a given feasibility domain in the parameter space of intrinsic growth rates with the position of its geometric centroid \vec{r}_c , such that $r_{c,i} = \frac{1}{n} \sum_j^n \mathcal{A}_{ij}$ where n is the number of consumers (Rohr et al., 2014; Saavedra et al., 2014; Rohr et al., 2016) (see Figure 4.3 panel **B** for a graphical illustration). The geometric centroid is equivalent to the center of mass of a convex object with n -vertices all having the same mass. For this reason, the solution of Eq. (4.1) with $r = r_c$ provides an equilibrium with all abundances of the species being the same. This approximation considerably simplifies the computation of the size of the adaptation space, and it does not alter the conclusions of our study, as feasibility domains tend to be rather small for large communities (Saavedra et al., 2017b).

By definition, two feasibility domains are distinct if they are spanned by a different set of column vectors of the interaction matrix \mathcal{A} . Thus, because in our framework there exists a one-to-one correspondence between the binary matrix β of the network and the columns of the interaction matrix \mathcal{A} , two topologically equivalent configurations of the CR interaction network generate two identical feasibility domains. Specifically, two CR interaction networks are topologically equivalent if the monopartite projections of their bipartite binary matrix ($\Lambda = \beta^T \beta$) are the same up to a permutation of the columns. Two CR interaction networks have the same monopartite projection if their binary matrices are the same up to a permutation of the rows. Therefore, a reorganization of interactions among a group of consumers (and their resources) leaves the monopartite projection invariant and the feasibility domain unchanged, if it is equivalent to a permutation of the rows that, together with a relabeling, preserves the adjacency of the nodes (i.e., an automorphism) (Golubitsky & Stewart, 2006; MacArthur et al., 2008; Garlaschelli et al., 2010). Conse-

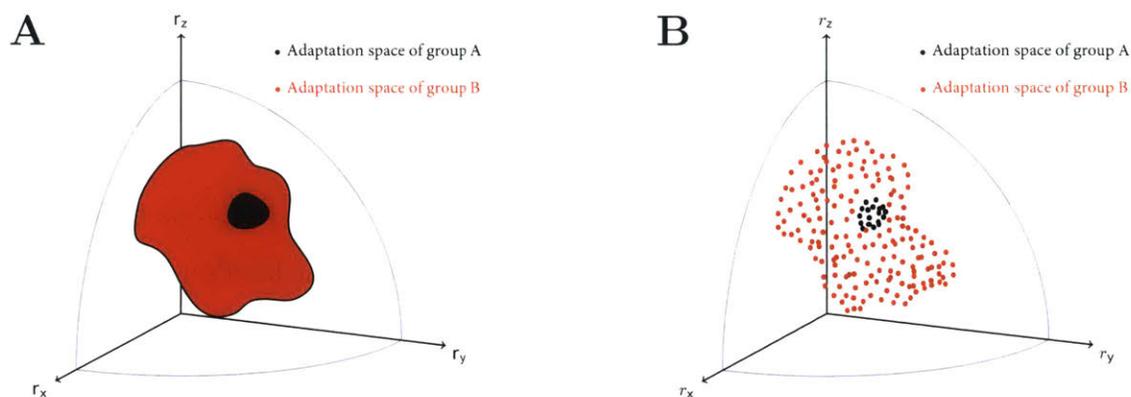


Figure 4.3: **Illustration of the adaptation space of a group of species.** In panel **A** we show a hypothetical adaptation space for two groups (differentiated by colors) of three species within a community: the region representing all the different feasibility domains (set of intrinsic growth rates compatible with positive abundances for the species within a group) that can be generated via the reorganization of interactions within the group. The larger the adaptation space, the larger the likelihood of a group to adapt to a new environment through a reorganization of interactions. Panel **B** is an illustrative example of the approximation that we use in our methods to estimate the size of the adaptation space. Each point in the figure represents the geometric centroid of a feasibility domain. The size of the space is then computed by counting the number of non-overlapping centroids and their average distance. In this hypothetical case, group A (black) would have a lower probability of adapting to novel environmental variations than group B (red). Our framework assumes that environmental changes can happen randomly in any direction.

quently, all of the topologically equivalent configurations of the CR interaction network form a class. The number of classes generated by the $2^{q \times n}$ possible configurations of a $q \times n$ dimensional group is equivalent to the number of distinct feasibility domains. That is, the smaller the number of permutations that preserve adjacency, the larger the number of distinct feasibility domains inside the adaptation space. See Figure 4.4 for an illustration of topologically equivalent and distinct configurations of a CR interaction network.

Let us now consider the set of all topologically distinct configurations of the CR interaction network. We use the average distance among the different feasibility domains inside the adaptation space as a second estimator for its size. Note again that there is a one-to-one correspondence between the monopartite projection of the CR interaction network and the feasibility domain. Therefore, the distance be-

tween two feasibility domains depends upon the amount of change in the structure of the monopartite projection before and after the reorganization. Recall that the monopartite projection is a count of the number of paths of length two between nodes (Saracco et al., 2015). If species outside the group share a large number of interactions, the number of these paths will already be large. Under these conditions, a reorganization of interactions among a group of species can only add a negligible fraction of paths. Therefore, it can only cause small changes in the monopartite projection. If instead, there are few interactions between the group and the other species in the community, the reorganization of interactions can act as bridges creating new interspecific effects, and potentially, significantly different feasibility domains inside the adaptation space.

In sum, the considerations above show that both the number of and the average distance between feasibility domains inside the adaptation space of a group of species must be a function of the interactions of species outside the group. These interactions act as a *structural constraint* inhibiting the capacity of generating new feasibility domains following a reorganization of interactions among the group of species. Formally, we can define a measure of structural constraint of a group as $S = \frac{\mathcal{L}_o}{n - n_g}$, where \mathcal{L}_o is the number of interactions including at least one species of the community outside the group, n is the possible number of interactions in the CR interaction network, and n_g is the possible number of interactions among the group of species. Note that other standard measures of connectivity are significantly different from the one defined here (see Appendix 4.7.2 for a more detailed discussion). Figure 4.5 illustrates how to compute the structural constraint of a group of species within a CR interaction network.

4.4 Data

To test our theoretical framework on empirical data, we compile seven pairs of CR pollinator-plant interaction networks from two independent experiments: one carried out in the UK (replicated 4 times) and one in Spain (replicated 3 times) (Lopezaraiza-

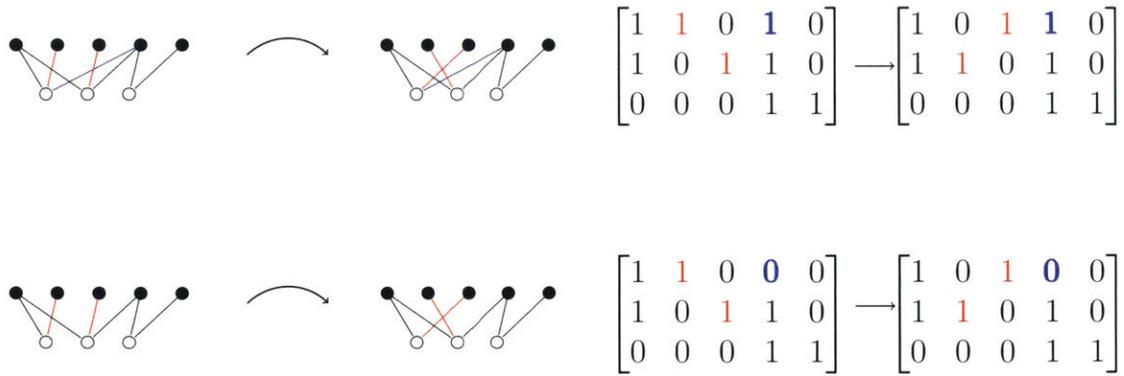


Figure 4.4: **Illustration of equivalent and distinct topological configurations of a (bipartite) CR interaction network.** Each pair of networks present possible changes (in red) of interactions. Black and white nodes correspond to species from different trophic levels (e.g., consumers and resources). The matrices on the right are the binary matrices β of the CR interaction networks. The top panel corresponds to equivalent network configurations. They are equivalent up to a permutation of the first two rows. Because the column vectors of their monopartite projections ($\Lambda = \beta^T \beta$) are equivalent, their feasibility domains are equivalent too. This means that these changes do not generate different feasibility domains. In contrast, the bottom panel shows two distinct network configurations and consequently two distinct feasibility domains.

Mikel et al., 2007; Montero-Castaño & Vilà, 2015). Each of the seven pairs was composed of adjacent plots in the field observed during one flowering season: in 2003 (UK) and 2010 (Spain). The exact location and distance between these adjacent plots were selected in order to maximize the match between plant composition, and to assure foraging capacities of pollinators across plots. From each plot, the interactions between pollinators and plants were recorded and used here as the observed CR interaction network β . Then, these networks are used to derive the interaction matrices \mathcal{A} for pollinators. Note that over the time scale of the experiment, pollinators can be regarded as consumers affecting each other through their shared plant-resources (Lopezaraiza-Mikel et al., 2007; Montero-Castaño & Vilà, 2015; Olf et al., 2009; Pauw, 2013; Holt & Polis, 1997). In one of these two adjacent networks, the flowers of one plant species were continuously cut throughout the experiment—we called this the perturbed network. The plants *Impatiens glandulifera* and *Hedysarum coronarium* were selected for flower removal in the British and Spanish experiments, respectively. Besides of small fluctuations in the intrinsic growth rates due to en-

environmental stochasticity (which is assumed to affect both fields given their adjacency), the flower-removal process is the largest resource perturbation that distinguishes the two fields at the start of the experiment. However, the observed community composition and interaction networks are the result of observations from the beginning until the end of the sampling period.

The dimension of the pollinator-plant interaction networks ranged from 20 to 86 species. A preliminary analysis of the data shows that, in each pair, there is a group of pollinators that is observed in both controlled and perturbed networks. On average, these persistent groups were formed by $\sim 37\%$ of the pollinator species present in the control networks. Furthermore, while the identity of the species in these groups were the same between pairs of networks, only few of the interactions within the groups were the same. Because of the proximity and set up of the adjacent networks, only pollinators could respond to the experiment by altering their foraging choices, supporting further their treatment as consumers affecting each other through their shared plant-resources. Thus, these experiments generate a scenario of rapid external perturbations on plant-resources under which some pollinators reorganized interactions and established in different environments. These manipulative experiments then allow us to test fast collective responses within CR interaction networks, that otherwise would be difficult to monitor at larger temporal scales.

4.5 Results

4.5.1 Simulations

Numerically, we found that our derived measure of structural constraint is a good inverse proxy for the size of the adaptation space of a group of consumers in CR interaction networks. To illustrate our findings, we estimate the average effect of the reorganization of interactions for groups of species with the same dimension. Then, these effects are compared against the structural constraint of each group. To measure the effects of reorganization, we computed the number of and the average distance among distinct feasibility domains inside the adaptation space of a group.



Figure 4.5: **Representative cartoon of the derived measure of structural constraint for six groups of species with same dimension in a (bipartite) consumer-resource interaction network.** The species in the groups and their interactions are represented in shaded symbols. The structural constraint of each group is a function of the number of interactions \mathcal{L}_o that involve at least one species outside the group (represented in darker symbols). Formally, for a $q \times k$ group within a $Q \times K$ network, the structural constraint is given by \mathcal{L}_o divided by the difference between the possible number of interactions among species in the entire CR interaction network $n = Q \times K$ and the possible number of interactions among species in the group $n_g = q \times k$, i.e., $\mathcal{S} = \frac{\mathcal{L}_o}{n - n_g}$. For example, let us focus on the first group on the left hand side: there are six possible interactions (i.e., $n = 3 \times 2 = 6$). The species outside the group are involved in two of them (i.e., $\mathcal{L}_o = 2$) and there are two possible interactions between the three species within the group (i.e., $n_g = 2 \times 1 = 2$). Therefore, the structural constraint of this group is $\mathcal{S} = \frac{2}{6-2} = \frac{1}{2}$. Note that the average structural constraint of the six groups is the connectance of the entire network (see Appendix 4.7.2).

The average effect then is calculated over an ensemble of several randomly-generated CR interaction networks, mimicking the different conditions under which a group of species (with a given structural constraint) can be found.

Specifically, we took a 25×30 CR interaction network and we choose a 3×6 CR dimensional group. The dimensions of both the network and group were chosen in order to explore all the possible configurations in a network, such that the results of our numerical analysis did not need to be tested against null models. We initialized the simulations setting the structural constraint of the group to $\mathcal{S} = 0.15$. This value of structural constraint guaranteed that all species had at least one interaction. Then, to compute the number of topologically distinct configurations of this group, we generated its 2^{n_g} binary matrices through all the possible reorganizations of interactions among species within the group. Note that each binary matrix β generates one unique interaction matrix \mathcal{A} , and therefore, one feasibility domain $D_F(\mathcal{A})$. Then, for each feasibility domain, we computed its geometric centroid \vec{r}_c . Then, we computed the number of topologically distinct feasibility domains as the fraction of distinct centroids (up to the fourth digit) divided by the number of globally stable configurations (typically $\sim 2^{n_g}$). The average distance between distinct

feasibility domains inside the adaptation space is calculated by taking the median of the pairwise distances between the centroids. We repeated the process described above 20 times. Note that at each run, we randomly changed the configuration of the CR interaction network while keeping the structural constraint of the group fixed. Finally, we repeated the numerical analysis for the same specified dimensions for both network and group, but the structural constraint of the group was changed in the interval $[0.15, 1]$ with steps $\Delta S = 0.05$.

Figure 4.6 shows that both the number of and average distance between feasibility domains inside the adaptation space decrease as a function of the structural constraint of a group. Figure 4.6A shows that the structural constraint of a group plays the role of a control parameter. There exists a critical value ($S_c \sim 0.5$ in this particular example) above and below which different reorganization of interactions have a significantly different effect on the adaptability of the group. On the one hand, above the critical level of structural constraint, the group exhibits a strong structural redundancy. That is, many reorganizations lead to CR interaction networks with same monopartite projection (and therefore interaction matrix) and thus to the same feasibility domain. In this example, only one over six reorganizations can actually generate a different feasibility domain. On the other hand, below the critical value of structural constraint, the structural redundancy is lost giving rise to an enhanced structural plasticity (Staniczenko et al., 2010). For groups that have structural constraints below this critical threshold, each reorganization can generate new feasibility domains. Furthermore, Figure 4.6B shows that even in the region where structural redundancy is lost (i.e., low structural constraint), the average distance between feasibility domains increases as the structural constraint of the group is decreased. Overall, these results reveal that the size of the adaptation space of a group is inversely related to its structural constraint.

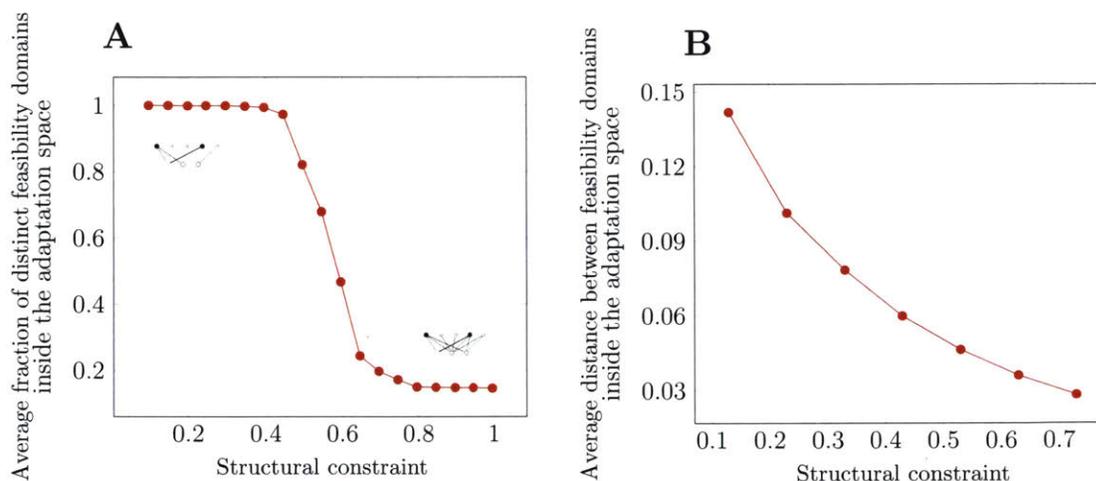


Figure 4.6: Structural constraint as an inverse proxy for the size of the adaptation space. For details about our numerical analysis see section 4.5.1. Panels **A** and **B** show that both the average fraction of distinct feasibility domains and the average pairwise distance between feasibility domains inside the adaptation space decrease as a function of the structural constraint of groups of species in CR interaction networks. See Figure 4.3 for a cartoon representation of a small and a large adaptation space. The small networks inside Panel **A** represent a group (shaded nodes) with low structural constraint (dark edges), top left, and large structural constraint, bottom right. See Figure 4.5 for an explanation of our derived measure of structural constraint.

4.5.2 Data analysis

In this last section, we tested our framework on empirical CR interaction networks. We used seven pairs of control and perturbed pollinator-plant interaction networks to validate the finding of the previous section: the structural constraint can be used as an inverse proxy for the size of the adaptation space of a group of species (pollinators in this specific case). Overall, we found that our derived measure of structural constraint was broadly distributed across groups in pollinator-plant interaction networks, and groups that were observed in both control and perturbed networks had relatively low structural constraint.

Specifically, we used the observed pollinator-plant interaction networks and their binary matrix β to parameterize the interaction matrix \mathcal{A} following our theoretical framework. Then, we partitioned the observed pollinator-plant networks into groups with same dimension (i.e., same number of pollinators and plants) and struc-

tural constraints \mathcal{S} . In the range $[\mathcal{S}_{\min}, \mathcal{S}_{\min} + 0.02]$, we randomly sampled 20 groups and took the average of the average pairwise distance between feasibility domains over the ensemble. Note that \mathcal{S}_{\min} is the minimum value of structural constraint of these groups. Then, we randomly sampled other 20 groups with $\mathcal{S} = \mathcal{S}_{\min} + 0.02$, and we ran the same analysis. We repeated the process until we reached the \mathcal{S}_{\max} in the network. Figure 4.7 confirms that the measure of structural constraint was inversely related to the size of the adaptation space of a group, as expected from our theoretical results.

Additionally, for each pair of observed pollinator-plant interaction networks, we found that the group of pollinators (including their plant-resources) that was present in both control and perturbed networks had a relatively low structural constraint. Specifically, for each pair, we computed the structural constraint of the persistent group, and compared it against other groups that could potentially be formed with the same dimension. To perform the comparison, we needed to partition the binary matrix β into all possible sub-matrices with fixed number of rows and columns (i.e., fixed dimension of the group). Note that the number of these partitions grows rapidly with the size of the pollinator-plant interaction network. That is, a $q \times k$ bipartite network can be partitioned into $N_g = \sum_{i=1}^q \binom{q}{i} \sum_{j=1}^k \binom{k}{j}$ groups. Thus, for five out of the seven observed networks, we could only sample a fraction of all the possible groups.

Figure 4.8 shows the distribution of the structural constraints for the groups of pollinators (including their plant-resources) that can be formed in each of the seven control pollinator-plant interaction networks separately. Note that the mean of the distributions correspond to the connectance of the networks (see Appendix 4.7.2). Importantly, the left tails of the distributions give us information about for which groups adaptability is increased by reorganizations. According to our observations, we found that the groups present in both perturbed and controlled networks were among the ones minimizing the structural constraint (see vertical line in Figure 4.8). Specifically, the three persistent groups in the Spanish networks were among the

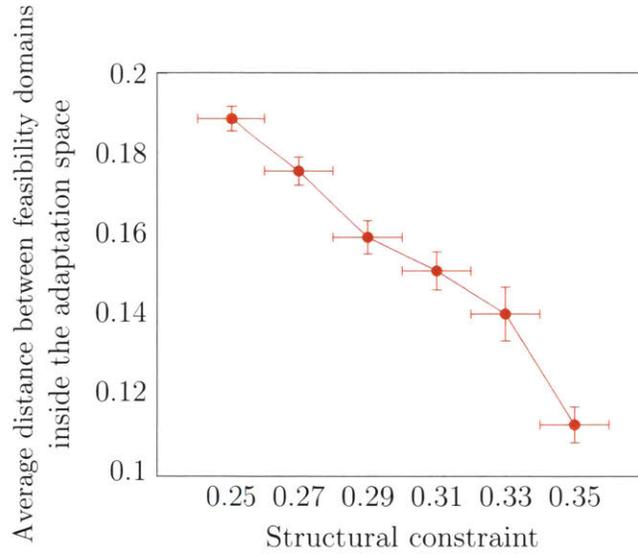


Figure 4.7: **Structural constraint as an inverse proxy of the adaptation space of groups of pollinators on an observed pollinator-plant interaction network.** For one observed pollinator-plant interaction network (namely, field 2 of the Spanish networks), the figure shows the estimated average distance between feasibility domains inside the adaptation space as a function of their structural constraint (x-axis) for all groups with the same dimension of pollinators (including their plant-resources). The error bars on the vertical and horizontal axes correspond, respectively, to the standard errors (over and ensemble of 20 groups) and the size of the bins used to cluster configurations with similar structural constraint S (i.e., $\Delta S = 0.02$). Note that because of the low connectance of all the observed pollinator-plant interaction networks, each group had the same number of distinct feasibility domains inside their adaptation space.

extreme 0.05%, 0.009%, and 5%, respectively. Similarly, the four persistent groups in the British networks were among the extreme 35%, 15%, 2%, and 5%, respectively.

Because in the analysis above we have fixed the dimension of the groups, those groups of pollinators with small structural constraint have a relatively high connectance. Note as well that the most connected species (i.e., generalists) are typically the most abundant (Vázquez et al., 2007). Therefore, one can hypothesize that groups of species present under different environments are simply the groups with the most abundant species. In that case, our framework would provide an additional dynamical explanation for the advantages of being highly connected. To test this, however, it will require to sample interaction networks independently from the abundances of species (Schupp et al., 2017). Interestingly, our Supplementary

Figure A.1 shows that even if we compare groups of species with different dimensions (which allows groups with same structural constraint to have different connectance), the observed persistent groups continue to be the ones with the smallest structural constraint (largest adaptation space). This is also true if we compare the structural constraint of groups with same connectance (see Supplementary Figure A.2). These results then provide support to the idea that the reorganization of interactions can indeed modulate the chances of species persistence under variable environmental conditions (Staniczenko et al., 2010; Ramos-Jiliberto et al., 2012; Saavedra et al., 2017a).

4.6 Discussion

Previous studies have shown that large environmental perturbations can give rise to significant variations in the intrinsic growth rates of species (Roughgarden, 1975; Dmitriew, 2011; Handelsman et al., 2013). It is well known that these variations, if detrimental to the species, can be counterbalanced by fast evolutionary processes (Bell & Gonzalez, 2011). However, empirical evidence has also shown that this evolutionary rescue not always occurs (Gonzalez et al., 2012). In this context, this work provides a theoretical framework to investigate how adaptation of species to novel environmental conditions can emerge from community dynamics.

Empirical evidence as well as ecological theory suggest that changes of species interactions, as a response to variations in environmental conditions, can act as a possible mechanism of ecological or behavioral adaptation (Odum, 1969; Staniczenko et al., 2010; Ramos-Jiliberto et al., 2012; Saavedra et al., 2017a). It has also been shown that the effects of these responses are typically not homogeneous across all the species within a community (Kaiser-Bunbury et al., 2010; Saavedra et al., 2017a; Lopezaraiza-Mikel et al., 2007; Montero-Castaño & Vilà, 2015). That is, within a community, some groups of species may have higher chances to adapt to a novel environment than other groups (Fukami, 2015; Rafferty et al., 2015; Saavedra et al., 2017a). Here, we have developed a feasibility study to quantify the extent to which

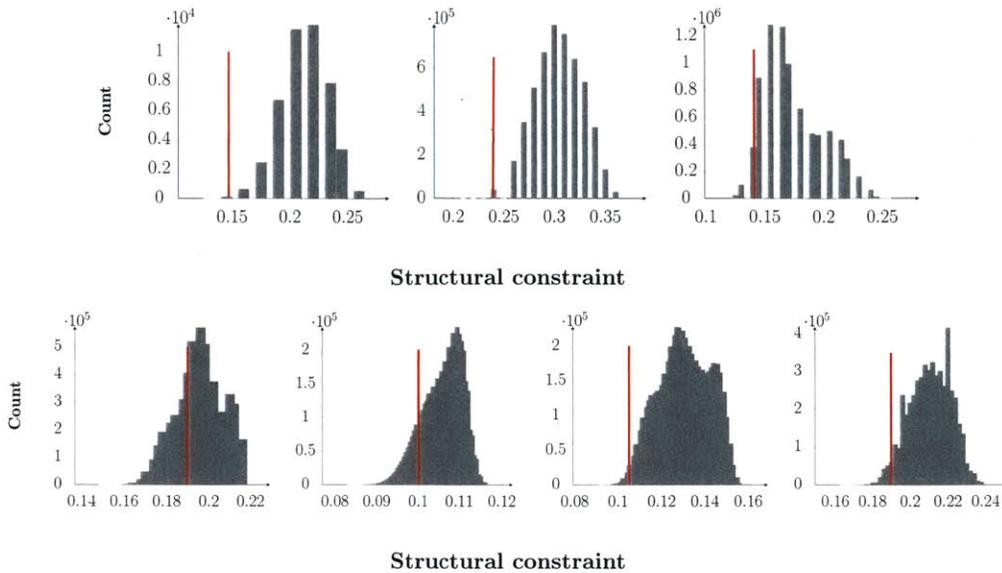


Figure 4.8: Structural constraints as measure of likelihood of adaptation: validation on empirical data. For each of the seven, control, pollinator-plant interaction networks, the figure shows the distribution of structural constraints across all possible groups of pollinators (including their pant-resources) that can be formed with the same dimension as the persistent group (i.e., the group that is observed in both control and perturbed networks, including their plant resources). The red vertical lines show the structural constraint of the groups present in both the control and perturbed networks. Top panel: the three control networks from Spain. Bottom panel: the four control networks from the UK. Recall that the structural constraint is inversely related to the size of the adaptation space of a group.

rapid reorganizations of interactions among a group of consumers can counterbalance the effects of variations in intrinsic growth rates in resource-mediated communities. Note that we are not making any assumptions about the rule of the dynamics or the biological mechanisms driving the reorganization process. Therefore, we are just measuring the potential adaptation space of a group to persist under novel environmental conditions following a reorganization of their interactions.

Our results have revealed that the adaptation space of a group depends upon the structure of the interaction network before the action of the perturbation. In other words, the adaptability of a group of species to novel environmental conditions does not only depend upon their capacity to change interactions, but also on the structural constraint imposed by the species outside the group. To derive the results we have considered ecological adaptation as a search process for a configuration of the

interaction network that matches the conditions for species permanence imposed by the newly perturbed environment. We have provided a quantitative method to estimate the size of the adaptation space, i.e., the search space of the adaptation process, for each group that can be formed in an ecological community. As a result of our investigation, we have derived a new local network measure, what we have called structural constraint, that can be used as an inverse proxy for the effect of reorganization of interactions on the adaptability of a group of species. We have shown that the smaller the structural constraint of a group, the larger the size of its adaptation space. Note that in a fixed observed network, this group corresponds to the set of species with the highest fraction of interactions among them, although this is not the only modulator. Note as well that the mean of the distribution of structural constraints across groups is the standard connectance of the interaction network. This signals important connectivity trade-offs modulating the adaptability of species (Gonzalez et al., 2012).

It is also important to understand the timescales under which our framework operates. The timescale of the reorganization process (which follows a perturbation) needs to be much faster than the characteristic time of the temporal evolution of species abundances defined by Eq. (4.1). Otherwise, extinctions could occur before any reorganization takes place. This separation of timescales has been observed in empirical studies with different types of perturbations or environmental changes (Fukami, 2015; CaraDonna et al., 2017). Because the duration of observations (i.e., of the experiment) is on the same timescale of the reorganization process, in our framework, we are not assuming that the community will relax to the equilibrium point of Eq. (4.1). We have investigated how many distinct feasible equilibrium points are plausible through all the potential new configurations of the group (i.e., the size of the adaptation space). Indeed, if we can guarantee the existence and feasibility of a fixed point, then we know that the transient state is feasible as well (i.e., the boundary of the positive orthant of the abundance space is attractive in Eq. (4.1)). In other words, we are using Eq. (4.1) to investigate the size of the adaptation space as a function of all possible changes in the parameters of the interaction matrix. Note that these changes are not implicit in Eq. (4.1). This equation just provides the set of

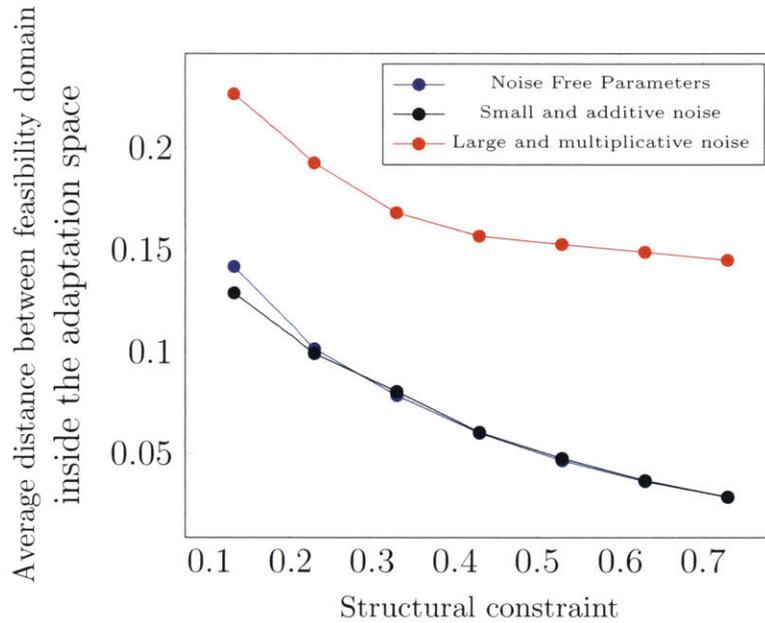


Figure 4.9: **Size of the adaptation space with noise in the parameters of the interaction matrix.** We perturbed the non-zero entry of the Gram Matrix after the parameterization. We perturbed in two different ways: (1) Gaussian noise ($\mathcal{N}(0.1, 0.1)$) added to the non-zero interactions and (2) Log-normal noise ($\mu = 0, \sigma = 1$) multiplied by the non-zero interactions.

conditions compatible with feasible solutions for any possible new configuration.

Importantly, we have empirically tested our theoretical framework. Our analysis of empirical data has shown that groups of species present under variable environmental conditions tend to be the ones for which the reorganization of interactions can have the largest impact on adaptability (i.e., are among the groups with the lowest structural constraint and largest adaptation space). However, we acknowledge that further experiments are necessary to claim any generality about our empirical validation. For instance, an experiment intended to validate our framework has to be designed in a way that the assumptions under which the GLV dynamics are valid (e.g., ecological time scales, large populations and large mobility) are explicitly satisfied (Logofet, 1993). Also, it would be important to develop experiments where the community can be tracked before and after the perturbation. Then, changes in the community can be contrasted against a control community without perturbation. While we have focused our analysis on consumer-resource interaction networks, the

general concept of mapping the reorganization of interactions to a potential adaptation space of a group of species can be extended to other multi-trophic systems. Moreover, the applicability of our approach is not restricted to GLV dynamics as long as the feasibility conditions can be expressed analytically. For example, this extension is straightforward for models with a linear stationary state (Saavedra et al., 2017b).

Finally, we would like to stress that the parameterization that we have chosen is just one among the infinitely many possible ways of parameterizing Eq. (4.1). Yet, there are three important reasons why we adopted this parameterization: we were interested in assuming a set of parameters that could be inferable from (binary) network data, that could be grounded on strong ecological theory (in this case the niche framework), and that minimize the number of free parameters (i.e., with our parameterization the only free parameter is the binary structure of the interaction network). We acknowledge that the results of our work are a direct consequence of our parameterization. Nevertheless, the main result of section 4.3.5—the size of the adaptation space is inversely proportional to the structural constraints—does not change qualitatively if we consider small noise around the interaction coefficients of the interaction matrix (see Figure 4.9). Overall, we believe that given its foundation on a well-established ecological theory and empirical data, our parameterization could be a good candidate under the lack of more information on interaction strengths. This additional information is typically very challenging to collect (Wootton & Emmerson, 2005), but we hope that future experimental studies on community dynamics can test the validity of our parameterization. Overall, we believe that the results of this work provide a guideline to systematically study the potential mechanisms driving the collective adaptation of species to changing environmental conditions.

4.7 Appendix

4.7.1 Appendix A. Stability of the fixed points

Despite Eq. (4.1) may have a fixed point at which all species can coexist, due to intrinsic noise and external perturbations the densities will not be located exactly on it (Lewontin, 1969). However, over short time scales and in the limit of large populations (Constable et al., 2016; Huang et al., 2015), if the fixed point is stable then, on average, the density will be the one predicted by Eq. (4.1) (as long as the perturbation does not lead to extinctions) (Lewontin, 1969). The basin of attraction, B , of a fixed point, \bar{x}^* , is the maximal neighborhood of \bar{x}^* such that for each point $\vec{x} \in B$ we have that $\lim_{t \rightarrow \infty} \vec{x}(t) \rightarrow \bar{x}^*$ (Hofbauer & Sigmund, 1988; Fasano & Marmi, 2006). If the basin of attraction of a fixed point is the whole state space then \bar{x}^* is said to be *globally stable*. The global stability of a fixed point of Eq. (4.1) is guaranteed as long as there exists a positive, diagonal, matrix D such that $DA + A^T D$ is positive definite (Logofet, 1993; Goh & Jennings, 1977; Goh, 1979; Saavedra et al., 2014). A matrix that satisfies this condition is said to be Volterra-dissipative (Logofet, 1993). An important property of the Lotka-Volterra dynamics, is that a feasible, globally stable fixed point is *permanent* (Baigent & Hou, 2012). We recall that for a feasible solution of Eq. (4.1) to be permanent the following two conditions must be satisfied: $\lim_{t \rightarrow \infty} \text{Inf } x_i(t) > \epsilon$ and $\lim_{t \rightarrow \infty} \text{Sup } x_i(t) \leq K \forall i$ and $\epsilon, K > 0$ (Hofbauer & Sigmund, 1988). That is, a feasible and stable solution of Eq. (4.1) is bounded: the density of the species does not go to infinity or to zero. Note that permanence is a stronger condition than species persistence inasmuch the former impose an upper as well as a lower bound in the species density (Hofbauer & Sigmund, 1988). Note as well that while global stability is a sufficient condition for permanence, it is not a necessary one (e.g., one can think about the example of an attracting limit cycle (Baigent & Hou, 2012)). Throughout the paper, to test the global stability of the fixed point, after each reorganization and parameterization, we check for $\frac{A+A^T}{2}$ to be positive defined, i.e., we check that the real parts of its eigenvalues are positive.

4.7.2 Appendix B. Discussions on the structural constraint

Let us now briefly discuss differences and similarities between the structural constraint and other two network properties: *connectance* and *conductance*. Firstly we want to show that for a bipartite network $\mathcal{G}(Q, K, E)$ with connectance \mathcal{C} we have $\langle \mathcal{S} \rangle_{\mathcal{g}} = \mathcal{C}$, where the expectation value is taken over all possible non trivial groups. The number of nontrivial groups \mathcal{C} in a network $\mathcal{G}(Q, K, E)$ is:

$$\tilde{N}_{\mathcal{g}} = \sum_{i=1}^{|Q|} \binom{|Q|}{i} \sum_{j=1}^{|K|} \binom{|K|}{j} - 1 \quad (\text{A1})$$

where $|\cdot|$ is the cardinality of a set, i.e. the number of elements in the set. Each term in the summation of Eq. (A1) is the number of groups of dimension $q \times k$. The structural constraint of each of these groups is, as defined in the main text, $\mathcal{S} = \frac{\mathcal{L}_o}{n - n_g}$ where $n_g = q \times k$, \mathcal{L}_o is the number of interactions with at least one species outside the group and n is the number of possible links in the whole network. Note that $\mathcal{S} \in [0, 1]$. The expectation value of \mathcal{S} is: $\langle \mathcal{S} \rangle_{\mathcal{g}} = \frac{1}{\tilde{N}_{\mathcal{g}}} [\mathcal{S}_1 + \dots + \mathcal{S}_{\tilde{N}_{\mathcal{g}}}]$. Let us call the number of groups of dimension d_i : N_{d_i} , e.g. $N_6 = \binom{|Q|}{2} \binom{|K|}{3}$ is the number of groups of dimension $d_i = 2 \times 3 = 6$. For each iso-dimensional group we have:

$$\sum_i^{N_{d_i}} \mathcal{S}_i = \frac{\mathcal{L} - e_1}{n - n_g} + \dots + \frac{\mathcal{L} - e_{N_{d_i}}}{n - n_g} = \frac{\mathcal{L} N_{d_i}}{n - n_g} - \frac{\sum_i e_i}{n - n_g} \quad (\text{A2})$$

where \mathcal{L} is the number of links in the network and e_i is the number of links in group i . The last term in Eq. (A2) sum up to:

$$\phi = \sum_i^{N_{d_i}} e_i = \frac{n_g}{n} \mathcal{L} N_{d_i} \quad (\text{A3})$$

Note that the last quantity is an integer. In fact, let us call q, k the number of rows and columns of the group:

$$\begin{aligned}
\frac{n_g}{n} \mathcal{L} N_{d_i} &= \frac{qk}{|Q||K|} \mathcal{L} \binom{|Q|}{q} \binom{|K|}{k} = \\
&= \frac{qk}{|Q||K|} \mathcal{L} \frac{|Q|!}{(|Q|-q)!q!} \frac{|K|!}{(|K|-k)!k!} = \\
&= \mathcal{L} \frac{(|Q|-1)!}{(|Q|-q)!(q-1)!} \frac{(|K|-1)!}{(|K|-k)!(k-1)!} = \\
&= \mathcal{L} \binom{|Q|-1}{q-1} \binom{|K|-1}{k-1}
\end{aligned} \tag{A4}$$

which is an integer. Therefore we have:

$$\sum_i^{d_i} \mathcal{S}_i = \frac{\mathcal{L}}{n - n_g} N_{d_i} \left(1 - \frac{n_g}{n}\right) = \frac{\mathcal{L}}{n} N_{d_i} = \mathcal{C} N_{d_i} \tag{A5}$$

If we divide by N_{d_i} both sides of the last equality we have: $\langle \mathcal{S} \rangle_{d_i} = \mathcal{C}$. There are $\tilde{n} = |Q||K| - 1$ sets of non trivial iso-dimensional groups, therefore we have $\sum_i^{\tilde{n}} \mathcal{C} N_{d_i} = \frac{\mathcal{L}}{n} \tilde{N}_g$. Therefore we have:

$$\langle \mathcal{S} \rangle_g = \frac{1}{\tilde{N}_g} [\mathcal{S}_1 + \dots + \mathcal{S}_{\tilde{N}_g}] = \frac{1}{\tilde{N}_g} \frac{\mathcal{L}}{n} \tilde{N}_g = \mathcal{C} \quad \square \tag{A6}$$

Thus, the macroscopic average of the structural constraint is the connectance. Equivalently, the average of the connectance of each group is the connectance itself. However, the distributions of the two over the groups are substantially different and therefore they capture different microscopic properties of the network. One can think about the structural constraint as the connectance of the network as seen from the group perspective. For this reason the methods used in the data analysis (i.e., decompose a network in groups) and the one used in the numerical analysis (i.e., placing one group in different networks) are equivalent. Let us now briefly discuss the conductance. The conductance ϕ of a group $g(q, e)$ within a monopartite network $\mathcal{G}(Q, E)$ with $q \subset Q$ and $e \subset E$ is defined as $\phi = \frac{|E|_{q, \bar{q}}}{\min(d(q), d(\bar{q}))}$ where $|E|_{q, \bar{q}}$ is the number of edges between the nodes in the set q (i.e., nodes in the group) and

its complement $Q \setminus q$ (i.e., nodes outside the group). $d(\bullet)$ instead is the sum of the degrees of the nodes in set \bullet . The conductance of the network is then defined as $\phi = \min_{g_i}(\phi_i)$ (Bollobás, 1998; Naumann & Schenk, 2012). The definition can be extended to bipartite networks by considering q as a union of nodes of the two layers. The main differences between the structural constraint and the conductance is that the latter is well defined only if the sum of the degree of the group is greater than zero. Here instead we are also considering groups with $d(q) = 0$ as we allow for all possible reorganization of the interactions (see the first section of the Theoretical framework). Therefore the two measures are significantly different as they take on different values for each group and their support is different. Overall, the structural constraint is a new simple measure of network structure which can be used as a proxy to capture the impact of the reorganization of interactions on the adaptability of a group of species.

Part II

Nonparametric approaches

PREFACE

In chapter 1 we discussed advantages and disadvantages of using parametric models to explain and predict population dynamics in changing environments. The main advantage is that the parameters of parametric models have a clear biological interpretation. For example, the parameters of the Lotka-Volterra dynamics are the intrinsic growth rates of and the pairwise interaction strengths between species. Therefore, using parametric models it is possible to investigate explicitly what consequences changes parameters will have on the overall dynamics. For example, if we want to understand the effects of micro-plastic debris on marine communities we can use a model to predict how fitness decay, a quantity that can be estimated from observations and experiments, affect the stationary distribution of the population on the long run. Then we can use these results to inform policy-maker about what are the best strategies of intervention to maintain biodiversity.

Surely interpretability is an important feature of a model that is used to analyze empirical data. However, interpretability *per se* does not resolve the important issue discussed in chapter 1. That is, conclusions based on parametric models are conclusions about the models mechanisms and not about nature (Breiman, 2001). Therefore if the results of a study strongly depends on the details of the model then one needs to properly ensure that the system under investigation does actually follow the assumed dynamical rules. Otherwise the conclusion we draw have mathematical but not scientific validity. However, because the mechanisms that drive population dynamics are frequently too complex to be captured by one single model,

parametric approaches rarely offer an accurate description of natural populations and even more rarely can be used to predict dynamics in changing environments.

The structural approach that we used in the previous three chapters offers an interesting alternative to purely parametric models because, as shown in chapter 2, it only requires general assumptions on model structure and not specific assumptions on the details of the model. The results drawn from the structural approach are still conclusions about model's mechanism but this time these mechanisms are shared by a large class of models, i.e. all of those models topologically equivalent to Lotka-Volterra. As long as observational data are generated by processes defined within this class we can use the structural approach to draw conclusions about nature. Here we suggest that the structural approach is able to explain many ecological theories because many biological populations follow dynamics within this class.

Conceptually the structural approach is therefore a very powerful framework to understand and predict community dynamics in changing environments because it is an interpretable parametric approach based on loose assumptions. However, it still suffer of three very important limitations: (1) it still require assumptions, even if loose, on the governing equations, (2) it requires data on species interactions which are typically very hard to collect and (3) it assumes equilibrium dynamics which are rarely observed in natural ecosystems. The goal of the second part of this thesis will be to develop a novel framework to understand and predict the dynamics of biological populations in changing environments that is (1) model-free but interpretable, (2) it require time-series data that are the most abundant type of data in population studies and (3) it does not require equilibrium assumptions. The following chapters are intended to develop a nonparametric structural approach.

CHAPTER 5

STATISTICAL INFERENCE OF JACOBIAN MATRICES FROM TIME-SERIES DATA

In this second part of the thesis we want to derive a nonparametric structural stability analysis. Before deriving the general theory however, there are a number of issues that we need to address. These issues are related to the challenges associated with accurately inferring Jacobian matrices (used as proxy for biotic interactions) from observational data. This and the next chapter are devoted to address these important issues. As we shall see in chapter 7 an accurate inference of Jacobian coefficients is necessary to estimate structural stability from data. An important challenge to overcome to infer Jacobian matrices from observational data is that empirical dynamics are inherently stochastic. Even when, as we assume in this thesis, there exists a deterministic skeleton of the dynamics, the true biological mechanisms (such as death, reproduction and competitions) are stochastic in nature. Hence, to infer Jacobians from empirical data we need algorithms that are robust to both observational noise (uncertainties in the measurements) and process noise (stochasticity of the true dynamics). In this chapter we derive a new algorithm that is an improvement of existing methodologies to infer Jacobian matrices from noisy nonlinear multivariate time series data.

5.1 Introduction

It is well known that purely deterministic ecological dynamics are mathematical idealizations (Bjørnstad & Grenfell, 2001; Kantz & Schreiber, 2004). Stochastic processes affect populations at all scales, from random realizations of birth and death processes to random environmental variations (Ranta et al., 2000; Frenzt et al., 2015; Spanio et al., 2017; Cenci & Saavedra, 2018a). It has been broadly recognized that fluctuations of species abundances observed in ecological time series emerge from the interplay between nonlinear (density-dependent) deterministic dynamics and stochastic forces (Higgins et al., 1997; Bjørnstad et al., 1999; Dixon et al., 1999; Coulson et al., 2004; Mutshinda et al., 2009; Dobrinevski & Frey, 2012), plus unavoidable noise due to measurement errors (Sugihara & May, 1990; Mason et al., 2018; Coulson et al., 2004).

Intrinsic stochasticity such as demographic stochasticity or, more generally, any type of process noise can induce new dynamics that cannot be predicted from a deterministic analysis: from the reversal of deterministic predictions (Constable et al., 2016) to the emergence of alternative stable states (Schooler et al., 2011). Furthermore, noise has important implications on the statistical analysis of time-series data. For example, for many inference techniques, process noise can set qualitative lower bounds to the number of data points needed to fully recover the parameters. That is, the number of neighboring points required in order to fit the real dynamics (rather than just the noise) grows with the level of noise (Bradley & Kantz, 2015). Similarly, process noise can reduce the out-of-sample forecasting skill of learning algorithms by, for example, breaking down the self-similarity of strange attractors in chaotic systems (Kantz & Schreiber, 2004), or, more generally, by increasing the risk of overfitting the training data (Clark et al., 2001; Abu-Mostafa et al., 2012). Thus, to improve inductive analyses of ecological dynamics, it is necessary to develop robust methodologies that can allow us to analyze ecological time-series data under the presence of process noise.

Two important analyses performed on ecological time series are the inference of the effect of species interactions on community dynamics and the forecast of species

abundances. Inferring accurate effects of species interactions from time-series data is important as a first step to find answers to fundamental questions in ecology such as: what is the role of species interactions during community assembly (Enke et al., 2018)? How does the effect of species interactions on community dynamics evolve in changing environments (Lawrence et al., 2012; Cenci et al., 2018a)? Does interaction variability in changing environments have any impact on species co-evolution (Barraclough, 2015)? Equivalently, a good forecast of changes in species abundances is key to improve the planning and decision-making process of ecosystem management (Clark et al., 2001), and to design successful socio-economic developmental strategies (Burkov et al., 2015). However, inference and forecasting from noisy, non-linear, time series is notoriously challenging (Casdagli et al., 1991; Bjørnstad & Grenfell, 2001; Wood, 2010; Perretti et al., 2013b,a).

In standard ecological-inference studies, the local effect of species interactions on community dynamics are approximated by the Jacobian coefficients of the data generating process (Ives et al., 2003; Deyle et al., 2013; Ushio et al., 2018). Formally, the Jacobian coefficients are the partial derivatives of the vector field of a dynamic model with respect to the state variables. In a population dynamics setting, this corresponds to the change of growth of a species as a result of a change in abundance of another species. Jacobian coefficients can be estimated from nonlinear time-series data using either parametric or nonparametric approaches. Using parametric approaches, the Jacobian coefficients are computed analytically from the assumed model after inference of its parameters (performed using for example state-space Bayesian approaches or Kalman filters (So et al., 1994; Meyer & Christensen, 2000; Stein et al., 2013; Bucci et al., 2016)). Using nonparametric approaches instead, the Jacobian coefficients are inferred directly from the data with minimal assumptions (such as stationarity and distribution of the noise) on the data generating process (Ting et al., 2008; Ghosh et al., 2014; Deyle et al., 2016b). Here, we focus on nonparametric approaches for inference and forecasting.

Earlier work has shown that learning algorithms based on nonlinear-state-space reconstruction can be used to infer nonparametrically Jacobian coefficients and to fore-

cast species abundances from noisy ecological time series (Sugihara, 1994). In this context, a commonly used predictive nonlinear function approximation algorithm, known as S-map (a locally-weighted multivariate linear regression model with state-space dependent kernel function (Deyle et al., 2016b)), has been derived to deal with signals generated by a mixture of linear and nonlinear components contaminated by noise. The S-map involves computing Jacobian matrices sequentially as the system travels along its attractor. The coefficients of these Jacobian matrices vary with each location on the attractor, giving rise to state-dependent interactions (Deyle et al., 2016b). The robustness of the S-map to observational noise has been demonstrated, however, its vulnerability to process noise remains a significant open problem that needs to be addressed in order to increase the applicability of this method.

In this chapter we show that by integrating appropriate regularization techniques to the S-map, it is possible to increase significantly the performance of both the in-sample inference of the Jacobian coefficients and the out-of-sample forecast of species abundances in the presence of process noise. To validate our results, we use three standard stochastic population dynamics models, which allow us to represent ecological time series data while knowing the ground truth for inference and forecasting. We further validate our methodology on empirical data.

5.2 Materials and Methods

5.2.1 Noisy nonlinear time series

The goal of this study is to derive an algorithm for an efficient inference of Jacobian matrices and out-of-sample forecast of species abundance from nonlinear time series in the presence of process noise. Hence, the time series used in this study were generated by numerical integration of Ito stochastic differential equations (SDE). To test the proposed algorithm against different levels of noise and nonlinearity, we considered three parameterizations of a stochastic Lotka-Volterra (SLV) dynamics: a 2-species predator-prey model, a 3-species rock-paper-scissor (RPS) model, and a 4-species model that exhibits chaotic trajectories. These models were chosen to 1)

generate fluctuations that resemble the behavior of observed ecological time series (Bjørnstad & Grenfell, 2001; Nahum et al., 2011), 2) produce dynamics that are notoriously challenging to forecast in the presence of process noise (Cheng et al., 2015), and 3) to explicitly write down the analytical form of the noise and systematically study its behavior.

Time series were generated by numerical integration of:

$$\dot{\vec{x}} = \vec{x}(\vec{r}_\gamma - \mathcal{A}_\gamma \vec{x}) + \frac{1}{\sqrt{V}} \xi_\gamma(\vec{x}, t), \quad (5.1)$$

where \vec{x} is the state vector (i.e., species abundance) and \vec{r}_γ and \mathcal{A}_γ are the vector of intrinsic growth rates and the interaction matrix of model γ , respectively. The noise component $\xi_\gamma(\vec{x}, t)$ in Eq. (5.1) is a Gaussian random variable with mean $\langle \xi_\gamma(\vec{x}, t) \rangle = 0$ and variance $\langle \xi_{\gamma,i}(\vec{x}, t), \xi_{\gamma,j}(\vec{x}, t') \rangle = \mathcal{T}_{ij}(\vec{x}) \delta(t - t')$, where \mathcal{T} is the covariance matrix of the noise. Note that the noise is multiplicative and depends on the parameters of the model. The covariance matrix of the noise in Eq. (5.1) was derived from the master equation of the microscopic stochastic process using the Van Kampen System size expansion (Van Kampen, 1992; Gardiner, 2004) as explained in chapter 1. The noise level was set by adjusting the system size V (note that for $V \rightarrow \infty$ we recover a deterministic equation). Specifically, SDE and parameterization are the following:

- Two species predator prey:

$$\dot{\vec{x}} = \vec{x}(\vec{r}_\gamma - \mathcal{A}_\gamma \vec{x}) + \frac{1}{\sqrt{V}} \xi_\gamma(\vec{x}, \tau) \quad (5.2)$$

Where $\vec{r}_1 = [1, -0.5]$. The interaction matrix is:

$$\mathcal{A}_1 = \begin{bmatrix} 0 & -0.3 \\ 0.5 & 0 \end{bmatrix} \quad (5.3)$$

and the covariance matrix of the noise is a diagonal matrix with diagonal elements $\mathcal{T}_{ii}(\vec{x}) = x_i(r_i + \sum_j \mathcal{A}_{ij} x_j)$

- Three species Rock-Paper-Scissor game:

$$\dot{\vec{x}} = -\vec{x}\mathcal{A}_2\vec{x} + \frac{1}{\sqrt{V}}\xi_2(\vec{x}, \tau), \quad (5.4)$$

where the interaction matrix \mathcal{A} reads:

$$\mathcal{A}_2 = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \quad (5.5)$$

The covariance matrix of the Gaussian noise is:

$$\mathcal{T}(\vec{x}) = \begin{bmatrix} x_1(|\mathcal{A}_{12}|x_2 + |\mathcal{A}_{13}|x_3) & -|\mathcal{A}_{21}|x_1x_2 & -|\mathcal{A}_{31}|x_1x_3 \\ -|\mathcal{A}_{12}|x_1x_2 & x_2(|\mathcal{A}_{21}|x_1 + |\mathcal{A}_{23}|x_3) & -|\mathcal{A}_{32}|x_2x_3 \\ -|\mathcal{A}_{13}|x_1x_3 & -|\mathcal{A}_{23}|x_2x_3 & x_3(|\mathcal{A}_{31}|x_1 + |\mathcal{A}_{32}|x_2) \end{bmatrix}$$

- Four species chaotic Lotka Volterra:

$$\dot{\vec{x}} = \vec{x}(\vec{r}_\gamma - \mathcal{A}_\gamma\vec{x}) + \frac{1}{\sqrt{V}}\xi_\gamma(\vec{x}, \tau) \quad (5.6)$$

Where $\vec{r}_3 = [1, 0.72, 1.53, 1.27]$ and the interaction matrix is:

$$\mathcal{A}_3 = \begin{bmatrix} 1 & 1.09 & 1.52 & 0 \\ 0 & 0.72 & 0.316 & 0.98 \\ 3.56 & 0 & 1.53 & 0.72 \\ 1.53 & 0.64 & 0.44 & 1.27 \end{bmatrix} \quad (5.7)$$

The covariance matrix of the noise is, as in the predator prey system, a diagonal matrix with element $\mathcal{T}_{ii}(\vec{x}) = x_i(r_i + \sum_j \mathcal{A}_{ij}x_j)$

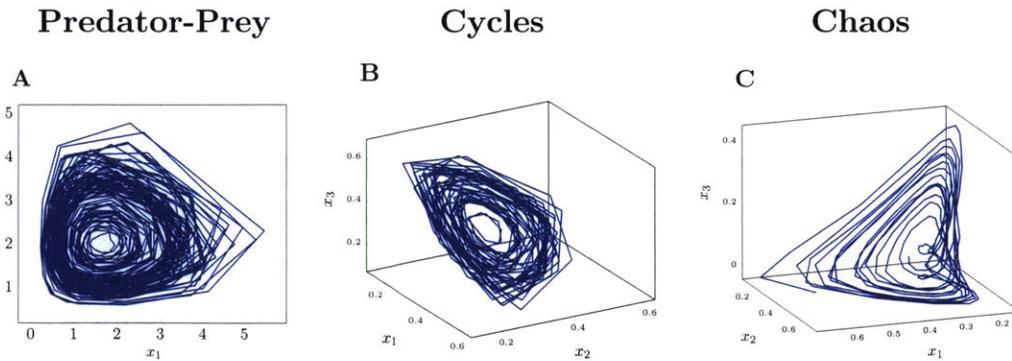


Figure 5.1: Illustrative example of the studied nonlinear, noisy, ecological time series. The figure shows the attractor of a 2-species predator-prey model (Panel A), a 3-species rock-paper-scissor game which exhibits cyclic dynamics (Panel B), and a 4-species chaotic dynamics (Panel C). Note that the chaotic attractor in Panel C is a projection of the 4-dimensional attractor in a 3-dimensional state space. Each axis corresponds to a species abundance.

To integrate the SDE's numerically, we used the python library *sdeint*. To compare inferred and analytical Jacobian coefficients, we simulated the noise with a vector of independent Wiener processes ($d\vec{W}(t)$) with zero mean and standard deviation $\sigma = \sqrt{dt}$, where $dt = 0.01$ is the integration step (i.e., the noise component is $\frac{1}{\sqrt{V}}\mathcal{T}(t)d\vec{W}(t)$). Note that additional noise was introduced by the sampling procedure. That is, after numerical integration of Eq. (5.1), we sampled equidistant points ($\Delta T = 200$) and used this sampled time series for the numerical analysis, rather than the whole solution of Eq. (5.1). This procedure accounts for the fact that observational data are sampled far more sparsely than the numerical solutions of an equation. See Figure 5.1 for a graphical example of the time series generated by integration of these models.

Our proposed methodology for inference and forecasting is an extension of a locally linear regression scheme with state-space-dependent kernel function known as S-map (Sugihara, 1994). For completeness, in the next section, we briefly review the details of the S-map. The readers familiar with the S-map methodology may skip the next section.

The S-map

The S-map is a nonlinear extension of standard linear vector autoregressive models (VAR). It differs from VAR and close linear relatives, such as dynamical linear models, in that the coefficients (or weights) of the regression in the S-map depend upon the position of the predictee on an attractor in the state space (Hsieh et al., 2005), and not on its proximity in time. Therefore, the S-map is particularly useful when the observed time series has been generated by nonlinear processes in which state-space dependence is an important factor.

In particular, the S-map is the SVD solution for C to the linear model:

$$B = A \cdot C, \quad (5.8)$$

with $B_k = w_k x_i(t_k + 1) = w_k Y_i$, $A_{kj} = w_k X_j(t_k)$, and

$$w_k = e^{-\theta \frac{\|x(t_k) - x(t^*)\|}{\bar{d}}} \quad (5.9)$$

are state-dependent weights. Note that in Eq. (5.9), $x(t^*)$ is the predictee variable and $\bar{d} = \frac{1}{n} \sum_i \|x(t_i) - x(t^*)\|$ is the average distance from the target point (Sugihara, 1994; Dixon et al., 2001; Deyle et al., 2016b). The weights then depend upon the distance of the target point in the state space and on a tuning parameter θ that controls the local weighting (Sugihara, 1994; Deyle et al., 2016b). That is, θ controls the degree of nonlinearity of the S-map. For this reason, the forecasting skill of the S-map has been extensively used as a test for nonlinearity of the time series (Perretti et al., 2013b; Sugihara, 1994), i.e., time-series generated by nonlinear processes can be better forecast with $\theta \neq 0$ (Dakos et al., 2017). Note that solving the linear problem above is equivalent to solving a weighted-least-square minimization problem, which reads in matrix form as:

$$\hat{c} = \underset{c \in \mathbb{R}^d}{\operatorname{argmin}} \frac{1}{n} \sum_j w_j (y_j - x_j c)^2 = \min_{c \in \mathbb{R}^d} \frac{1}{n} (Y - Xc)^T W (Y - Xc), \quad (5.10)$$

where X is the data matrix (for the rest of this chapter we will use all the variables of

the time series to reconstruct the attractor), W is a diagonal matrix of the weights in Eq. (5.9), and c is one row of the Jacobian matrix. The solution of this minimization problem (i.e., ordinary least square, OLS) is $\hat{c} = (X^T W X)^{-1} X^T W Y$. Equivalently, the solution of the SVD for C in the linear model in Eq. (5.8) is

$$\hat{c} = (\tilde{V} \Sigma^{-1} \tilde{U}^T) W Y, \quad (5.11)$$

where the solution is written in terms of the weighted data matrix $\tilde{X} = W X$. It is then straightforward to prove that if $X^T X$ is invertible then the two solutions are equivalent.

$$(X^T W X)^{-1} X^T W Y = X^{-1} W^{-1} X^{-T} X^T W Y = X^{-1} Y \quad (5.12)$$

Equivalently, if we call $\tilde{X} = W X_t$, the SVD solution for the linear system is

$$(\tilde{V} \tilde{\Sigma}^{-1} \tilde{U}^T) W Y = \tilde{X}^{-1} W Y = (W X)^{-1} W Y = X^{-1} Y. \quad (5.13)$$

Then, $(X^T W X)^{-1} X^T W Y = (\tilde{V} \tilde{\Sigma}^{-1} \tilde{U}^T) W Y$. However, the inversion of the matrix $X^T X$ is a well known problem in standard statistics as, even if strictly invertible, small singular values can create significant numerical instabilities. To avoid these numerical instabilities in the inversion of the matrix Σ , the S-map methodology removes small singular values before inversion by setting to zero each singular value below a pre-defined threshold, i.e., $\sigma_{ii} < \epsilon$ for some fixed ϵ (typically $\epsilon = 10^{-5}$ (Ye et al., 2017)).

Importantly, as mentioned before, process noise can introduce upper bounds to the forecasting skill of learning algorithms, such as the S-map. Specifically, small perturbations in the input data (induced by noise) can cause large changes in the coefficients of the fit and the forecast values. Indeed, small singular values can drastically increase the prediction variance of the ordinary-least-square optimization, an effect that is amplified by noise. However, the cut-off value of the singular value is not universal, as in the standard S-map, but depends upon the structure of the data to fit.

Generally, depending on the data, one needs to select the cut-off by appropriately balancing the prediction bias and its variance. This is typically done by introducing some form of regularization of the fit. Below, we discuss how to introduce regularization techniques to tackle small singular values, and to improve the performance of the fit along with the out-of-sample generalization in the presence of process noise.

Regularized S-map

Regularized loss minimization is a well-established learning rule, in which both a regularization function (\mathcal{R}) and the empirical risk (\mathcal{L}) are minimized together (Shalev-Shwartz & Ben-David, 2014):

$$\hat{c} = \operatorname{argmin}_{c \in \mathbb{R}^d} \{ \mathcal{L}(\mathcal{D}, \vec{c}) + \mathcal{R}(\vec{r}, \vec{c}) \}, \quad (5.14)$$

where \mathcal{D} are the data, \vec{c} is the vector of parameters to estimate, and \vec{r} is a vector of regularization parameters. Regularized loss minimization is a typical approach used to constrain the coefficients of a regression model in order to make the algorithm more stable in the presence of noise in the data (recall that a learning algorithm is stable if small changes in the input data do not generate large changes in the output (Abu-Mostafa et al., 2012; Shalev-Shwartz & Ben-David, 2014)).

There are a variety of possible regularization functions to choose from, here we discuss the most common ones. Tikhonov regularization requires to substitute Eq. (5.10) with the following minimization problem called regularized least square or ridge regression (Shalev-Shwartz & Ben-David, 2014):

$$\hat{c} = \operatorname{argmin}_{c \in \mathbb{R}^d} \frac{1}{n} (Y - Xc)^T W (Y - Xc) + \lambda \|c\|_2^2. \quad (5.15)$$

The regularization function $R(\lambda, c) = \lambda \|c\|_2^2$ constrains the hypothesis space \mathcal{H} (i.e., controls its complexity) or, equivalently, adjusts the least-square solution to avoid numerical instability during matrix inversion. In fact, the solution of Eq. (5.15) is:

$$\hat{c} = (X^T W X + \lambda n I)^{-1} X^T W Y, \quad (5.16)$$

where the regularization parameter λ , selected by running cross-validation on a training set data, avoids possible instabilities during matrix inversion in Eq. (5.16). We chose Tikhonov regularization because the solution of Eq. (5.15) can also be written in terms of singular values, and it can be shown that the regularization parameter λ damps any term in the solution with $\sigma_i < \lambda_i$. Therefore, Tikhonov regularization is strictly related to a truncated singular value decomposition (Gorodnitsky & Rao, 1994). Indeed, Tikhonov regularization introduces the balance between prediction bias and variance that we discussed in the previous section (Abu-Mostafa et al., 2012). This means that Tikhonov regularization is, in fact, the natural type of regularization of the S-map. Note that regularizations of local linear fits have already been considered in settings different than the S-map, and for purposes different than controlling the quality of the inference of the effect of species interactions.

Generally, the choice of the regularization function is heuristic (Abu-Mostafa et al., 2012). The choice depends upon the structure of the available data. For example, ridge regression filters the small singular values that can create instabilities when they are contaminated by noise. However, it may produce suboptimal results if most of the information is contained in these values. Other types of regularizations, such as LASSO penalization (Tibshirani, 1996), can be more suitable when dealing with sparse data sets. For comparison purposes, here we also use elastic-net regularization (Zou & Hastie, 2005), which is a more flexible function than Tikhonov. Elastic net is a combination of Tikhonov (L_2 norm) and LASSO (L_1 norm) regularization with an additional parameter α that controls the relative strength of the two norms:

$$\hat{c} = \operatorname{argmin}_{c \in \mathbb{R}^d} \frac{1}{n} (Y - Xc)^T W (Y - Xc) + \lambda ((1 - \alpha) \|c\|_1 + \alpha \|c\|_2^2). \quad (5.17)$$

Note that when $\alpha \rightarrow 1$, one recovers Tikhonov regularization. Whereas when $\alpha \rightarrow 0$, one recovers LASSO. Importantly, elastic-net regularization has the advantage that it can be adapted to the particular structure of the time series. For example, if the time series is sparse then α should be small. However, if small constraints are enough to avoid both over-fitting and instability, then α should be close to one. In order to properly compare two different algorithms, one can always keep $\alpha < 1$.

Finally, we discuss the choice of weights in the locally-weighted regression (Eq. (5.10)). The standard S-map uses exponentially decay kernels, such as Eq. (5.9). The choice, however, is not unique and generally depends upon the characteristics of the analyzed dataset. For example, exponentially decay kernels are motivated by Lyapunov instability, where information decays exponentially. Other kernels, such as the Epanechnikov or the tri-cubic kernel (Hastie et al., 2001), may be more suitable than Eq. (5.9) for different types of nonlinearity in the time series. In general, a preliminary analysis of the data is needed in order to choose the most appropriate regression model. Here we test the following kernels:

$$K(x, x') = \left(1 - \left(\frac{\|x - x'\|}{\theta}\right)^3\right)^3 \Theta\left(1 - \frac{\|x - x'\|}{\theta}\right), \quad (5.18)$$

$$K(x, x') = \frac{3}{4} \left(1 - \left(\frac{\|x - x'\|}{\theta}\right)^2\right)^2 \Theta\left(1 - \frac{\|x - x'\|}{\theta}\right), \quad (5.19)$$

$$K(x, x') = e^{-\theta \frac{\|x - x'\|}{d}}. \quad (5.20)$$

These kernels are the tri-cubic, Epanechnikov, and exponential respectively. Note that $\Theta(\cdot)$ is the Heaviside theta (i.e., $\Theta(n) = 1$ if $n \geq 0$ and zero otherwise). Below we specify how to measure and compare the inference and forecasting skills of the standard (i.e., non-regularized) and regularized S-maps.

Inference of Jacobian coefficients

To infer Jacobian coefficients from the stochastic time series using the standard and regularized S-maps, we ran the following analysis. Firstly, we generated time series of species abundances by integrating numerically Eq. (5.1) and by sampling data points at constant rate as discussed above (see Figure 5.1 for a graphical example of these time series). Then, we randomly sampled a block of length $L = 100$ of the time series. The block is standardized to have zero mean and standard deviation one. Finally, we fitted the data using the standard and the regularized S-map. Regularization and kernel parameters were chosen by running leave-one-out cross

validation (Hastie et al., 2001), i.e., by choosing those parameters that minimize the validation error on the training set. Note as well that we fit each species independently so regularization and kernel parameters are d dimensional vectors (where d is the number of species in the system). Because we want to show the effect of different regularization functions on the performance of the S-map, we fixed the parameter α in the elastic-net regularization to be strictly less than one. Specifically, we fixed $\alpha = 0.9$ for the rock-paper-scissor model, and $\alpha = 0.95$ for the chaotic and predator-prey dynamics. These values were chosen after a preliminary analysis of the time series. We did not run cross-validation on α because, contrary to the parameters λ and θ , its optimum value did not vary significantly across the realizations.

The Jacobian coefficients are the coefficients of the fit (Deyle et al., 2016b; Ushio et al., 2018). To measure the quality of the inference, we computed the Pearson correlation coefficient (Sugihara, 1994) and the root mean square error (RMSE) between analytical and inferred Jacobian matrices. To obtain a reliable statistics, we repeated this process for 80 random blocks (i.e., for 80 times we sample a random block, standardized it, run leave-one-out cross validation to select regularization and kernel parameters and estimate the Jacobian coefficients). Finally, notice that, given that we are dealing with synthetic data for which we know the true dimensionality, we used the dimension of the model as the embedding dimension for the inference of the Jacobian coefficients. However, recall that for the analysis of empirical data, for which the true embedding dimension is generally unknown, this parameter must be chosen together with the regularization and kernel parameters.

Forecasting of species abundances

Using the kernel and regularization parameters estimated with leave-one-out cross validation in the training set, we built a predictive model and forecast 30 data points out-of-sample. To make out-of-sample forecast we ran the following analysis: after training over L -data points (i.e., over the whole training set), the S-map provided $L - 1$ coefficients given that the last data point of the training set was not trained (i.e., we did not have information about the $L + 1$ abundance which was the first

data in the test set). Therefore, we used the $(L - 1)$ th coefficient and the L th data point to make predictions about the first data point in the test set (i.e., the $(L + 1)$ th point in the time series). Predictions were made with a locally-linear model (Ye et al., 2017):

$$x_i(t + 1) = c_0 + \sum_{j=1}^d J_{ij}(t - 1)x_j(t), \quad (5.21)$$

where c_0 is the intercept fitted adding a column of ones to the data matrix (Hastie et al., 2001) and d is the number of species in the system. The matrix J_{ij} is the Jacobian matrix, whose rows are the vectors c_i solutions of the minimization problem in Eq. (5.15) and Eq. (5.17) (and the standard S-map). After each forecast, we used the predicted point to fit the L th interaction coefficient, and we repeated the operation for as many times as numbers of data points we had in the test set (i.e., 30). The forecast is purely out-of-sample because this procedure never uses the original test set to select the model parameters or to make predictions. In addition, the test set is standardized using the mean and the standard deviation of the training set. Hence, we truly tested for the generalization skills of the algorithms on unseen data.

To measure the quality of the forecasting skill, we used the Pearson correlation coefficient (Sugihara & May, 1990) and the RMSE as we did for the coefficients of the Jacobian matrix. Moreover, as benchmark for the value of the RMSE in the test set, we compared the quality of the forecast with the RMSE of the naive predictor (i.e., the predictor that uses the last point in the training set as prediction for any data in the test set).

To test the robustness of the algorithm, we repeated the inference and forecasting analyses with different training-set lengths ($L = \{20, 40, 80, 160\}$) and different levels of (Gaussian) observational error with standard deviation $\sigma \in [0.05, 0.2]$ and $\mu = 0$ added to the training data. In order to make the noise comparable to the perturbed value we added multiplicative noise to the data, i.e. $\vec{x}_{\text{perturbed}} = \vec{x}_{\text{unperturbed}} + \vec{x}_{\text{unperturbed}} \times \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma)$ (notice that \vec{x} come for the integration of an SDE so it already has a intrinsic noise component).

Analysis on empirical data

Finally, we tested the performance of the regularized S-map on two empirical time series of marine microbial communities: the Bermuda Atlantic Time Series (BATS) and the Hawaii Ocean Time Series (HOTS). Because in empirical data we lack the ground truth of Jacobian coefficients, we only tested the forecasting skills of the algorithm. The time series can be freely downloaded from http://batsftp.bios.edu/BATS/bottle/bval_bottle.txt (BATS) and <http://hahana.soest.hawaii.edu/hot/hot-dogs/index.html> (HOTS).

The BATS data include the abundance of four species: *Prochlorococcus*, *Synechococcus*, *Picoeukariotes* and *Nanoeukariotes*. Here we focused on the time period that goes from September to October 2011 for which data were collected approximately twice per day. This time series has 30 data points. We split our data in 28 points for training (i.e., for cross validation) and 2 points for testing. The HOTS data include species abundances of: *Prochlorococcus*, *Synechococcus*, *Picoeukariotes* and *heterotrophic bacteria*. These data were collected monthly from 2006 to 2016 (93 data points). We used 91 points for training and 2 points for testing.

To analyze the empirical time series, we ran a standard variable selection (Hastie et al., 2001) to subset only those variables and their time lags that provide the best validation error. Then, we ran the same analysis described above to choose the optimal regularization parameters and kernel function of the regularized S-map. We only report the RMSE of the analysis (note the Pearson correlation) given that the test set only includes 2 data points.

5.3 Results

5.3.1 Results on Synthetic data

Overall, we found that the regularized S-map consistently outperforms the standard S-map in both tasks: the inference of Jacobian coefficients and the forecast of species

Inference Regression scheme	Predator Prey		Cyclic		Chaotic	
	ρ	RMSE	ρ	RMSE	ρ	RMSE
Standard S-map	0.61 \pm 0.01	0.62 \pm 0.01	0.51 \pm 0.02	0.37 \pm 0.01	0.43 \pm 0.05	0.61 \pm 0.02
Regularized S-map	0.64 \pm 0.01	0.58 \pm 0.01	0.61 \pm 0.01	0.31 \pm 0.007	0.5 \pm 0.03	0.36 \pm 0.01

Table 5.1: **Inference of the Jacobian coefficients using synthetic data.** For each of the three models analyzed (columns), the table shows the performance of the standard and regularized S-map (rows) in inferring the analytic Jacobian coefficients. The performance is computed as the Pearson correlation (ρ) and root mean square error (RMSE), i.e., smaller (larger) RMSEs (correlations) correspond to better performances. Note that the regularized S-map consistently outperforms the inference skills of the standard S-map.

Forecasting Regression scheme	Predator Prey		Cyclic		Chaotic	
	ρ	RMSE	ρ	RMSE	ρ	RMSE
Standard S-map	0.75 \pm 0.18	0.88 \pm 0.26	0.41 \pm 0.23	1.08 \pm 0.22	0.03 \pm 0.04	NA
Regularized S-map	0.88 \pm 0.08	0.49 \pm 0.06	0.60 \pm 0.06	0.72 \pm 0.08	0.8 \pm 0.07	0.65 \pm 0.11
Naive predictor	NA	1.32 \pm 0.06	NA	1.4 \pm 0.08	NA	1.35 \pm 0.1

Table 5.2: **Out-of-sample forecasting skills using synthetic data.** For each of the three models analyzed (columns), the table shows the performance of the standard S-map, regularized S-map, and naive predictor (rows) in predicting species abundances. The performance is computed as the Pearson correlation (ρ) and root mean square error (RMSE), i.e., smaller (larger) RMSEs (correlations) correspond to better performances. Note that the regularized S-map consistently outperforms the standard S-map (and the naive predictor).

abundances. Tables 5.1 and 5.2 provide the summary results of the best inference and forecasting skills of the two algorithms. Performance is expressed in terms of median RMSE and correlation coefficient (ρ) over an ensemble of 80 random chunks of the time series. Errors around the median are the 95% confidence intervals computed with nonparametric bootstrapping (full results are shown in Figs. 5.2 and 5.3). The tables show the results of the fit that provides the best trade-off between maximum correlation coefficients and minimum RMSE using the same kernel and regularization function. The RMSE of the standard S-map on the test set of the chaotic model is out of scale and is reported as NA.

Importantly, Figure 5.2 and Table 5.3 illustrate that the quality of the inference and the forecasting strongly depends upon the choice of the kernel function used in the weighted regression scheme. For example, Figure 5.2 B shows that for the periodic predator-prey time series, the best forecast is achieved using the Epanechnikov ker-

nel with an elastic net regularization function. Instead, Figure 5.2 **F** shows that for the chaotic time series, the exponential kernel exhibits the best inference and forecasting skills.

Moreover, Figure 5.2 shows that one can obtain different reconstructions of the Jacobian coefficients (left panels) simply by using different regularization schemes or kernel functions for the weights of the linear regression (see for example panel **E**). The figure also shows that the quality of the reconstruction is not necessary informative of the generalization skills of the algorithm (right panels). This effect is clearly illustrate in Figure 5.2. For example one can compare the performance of the Exponential kernel and an elastic net regularization function for the inference of Jacobian coefficients in panel **A** with its generalization skills in panel **B**. Overall, these results pose serious challenges to the problem of model selection. That is, using empirical data one can only select parameters, kernels, and regularization functions based on validation or prediction errors, yet this may not be the optimal solution for inference purposes.

Finally, Figure 8.2 shows the quality of the inference and forecast as a function of the level of observational noise added to the training data. Performance is again expressed in terms of median error over 60 random chunks of the time series and errors are the 95% confidence interval around the median error. The figure shows that the median RMSE, while increasing, remains significantly lower than the naive prediction and the standard S-map (compare the Figure with Table 5.1 and 5.2). Figure 5.4 in the Supplementary Information shows the robustness of the algorithm to different lengths of the training set. Interestingly, in our example, smaller training sets do not necessarily correspond to larger errors. This is particularly true for the rock-paper-scissor dynamics which is the most noisy model we used. This counter-intuitive result could be explained by the fact that, because the time series we are analyzing are stochastic, training on more data increases the probability of training on noisy regions of the attractor. Overall, Figures 8.2 and 5.4 show that the regularized S-map is robust to different levels of noise and sizes of training data.

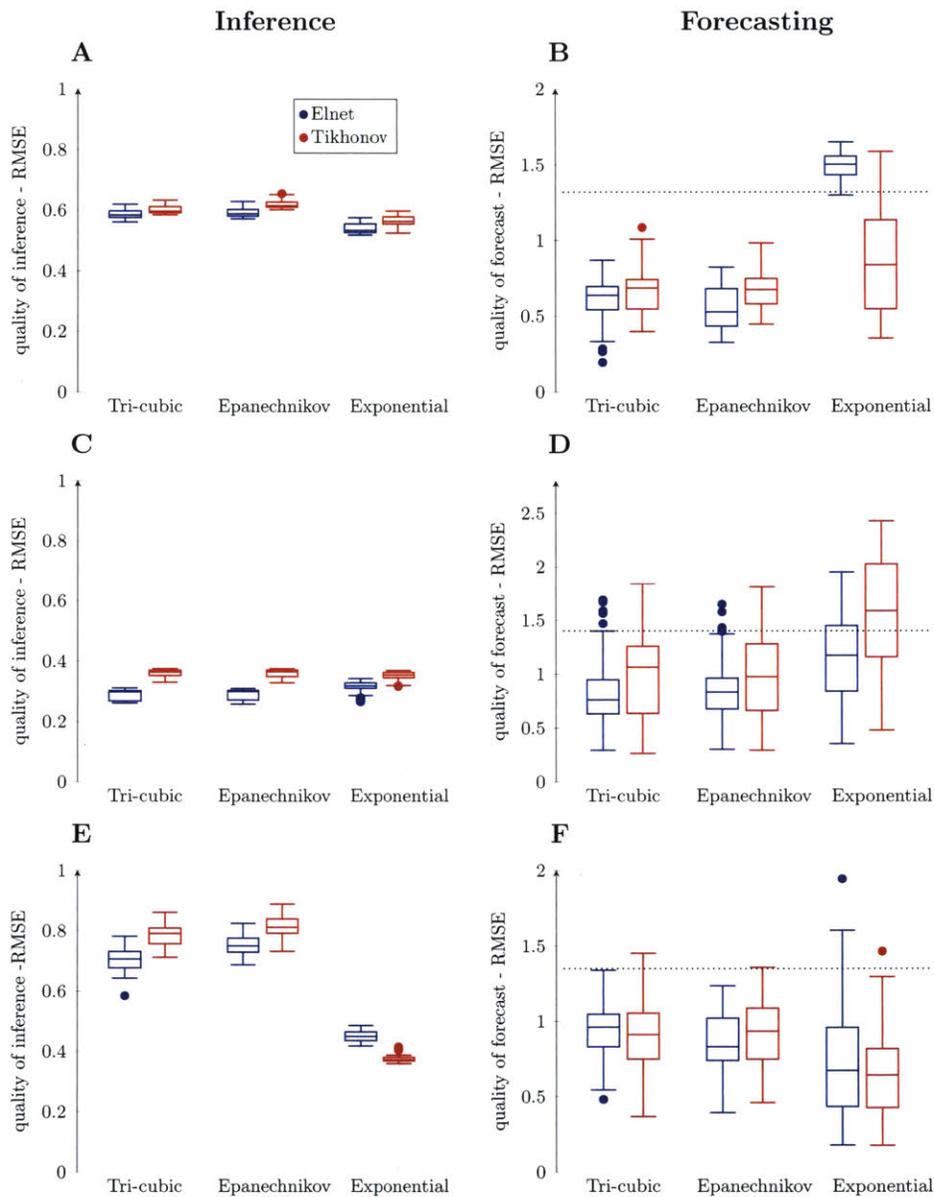


Figure 5.2: **Importance of the kernel function.** The figure illustrates the full distribution of the inference and forecasting skills of the regularized S-map using the tri-cubic kernel, the Epanechnikov kernel, and the exponential kernel (used in the standard S-map). Models go from top to bottom: Predator prey, Rock-Paper-Scissor and chaotic Lotka-Volterra. The left panels show the results of the inference, whereas the right panels show the results of the forecast. The dotted lines in the right panels depict the value of the naive error. All performances are reported as the root mean square error (RMSE), i.e., smaller RMSEs correspond to better performances. Overall, the figure illustrates that the choice of the kernel function is critical for the performance of the algorithm.

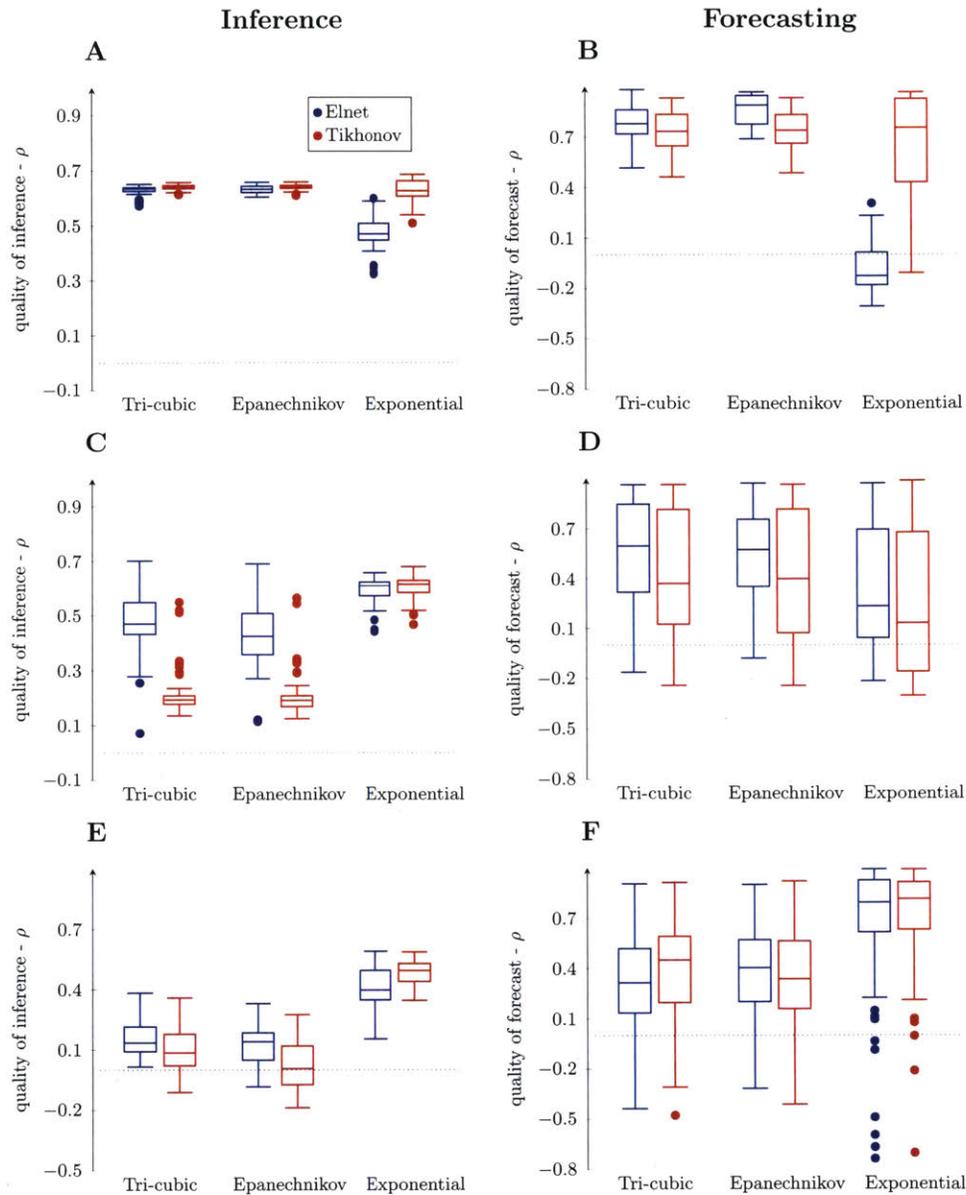


Figure 5.3: **Inference and forecasting skills in terms of correlation coefficients.** The figure is equivalent to Figure 5.2 in the main text, but the results are expressed in terms of ρ .

Results on Empirical data

Focusing on the BATS data, we found that the exponential kernel and elastic-net regularization function with $\alpha = 0.85$ provided the best out-of-sample skills (in terms of RMSE) for predicting the abundances of the three species of bacteria. In the analysis, we included the abundance of *Prochlorococcus* (embedding dimension, $E = 1$)

	Tikhonov	Elastic net
Exponential	Chaos, Chaos, Predator-Prey	Cycles
Epanechnikov		Predator-Prey
Tri-cubic		Cycles

Table 5.3: Combinations of kernels and regularization functions that provide the best inference and forecasting skills. For each of the three tested models, the table shows which is the best combination of kernels and regularization functions to infer the Jacobian coefficients (red) and to forecast species abundance (blue). Note that the best kernel is always chosen to be the one that provides the best trade-off between maximum correlation coefficient and minimum RMSE.

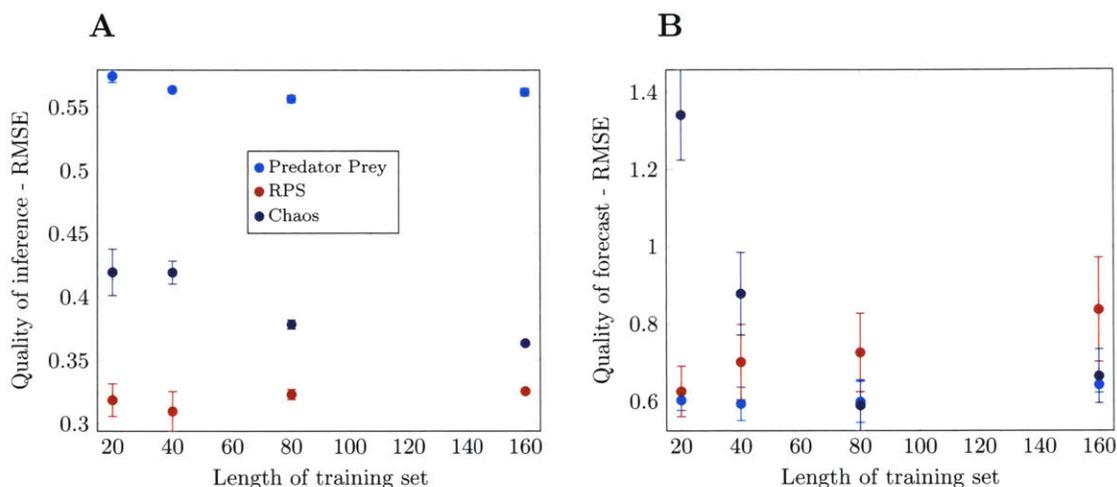


Figure 5.4: Robustness of the regularized S-map to the length of the training set. For each of the three models analyzed, the figure shows that both the inference and the forecasting skills of the regularized S-map are robust to the number of data points used to train the algorithm. The performance is reported as RMSE.

Synechococcus ($E = 2$) and *Picoeukariotes* ($E = 1$). Moving to the HOTS data, we found that the Epanechnikov kernel with $\alpha = 0.9$ provided the best performance. In the analysis, we included the abundance of *Prochlorococcus* ($E = 2$) and *Picoeukariotes* ($E = 1$).

Table 5.4 shows the performance comparisons between the standard and regularized S-maps. We used the naive forecast as a benchmark for the quality of performances (see Methods section). As expected from the results derived from synthetic data, we found that the regularized S-map consistently outperforms the standard S-map on the empirical time series.

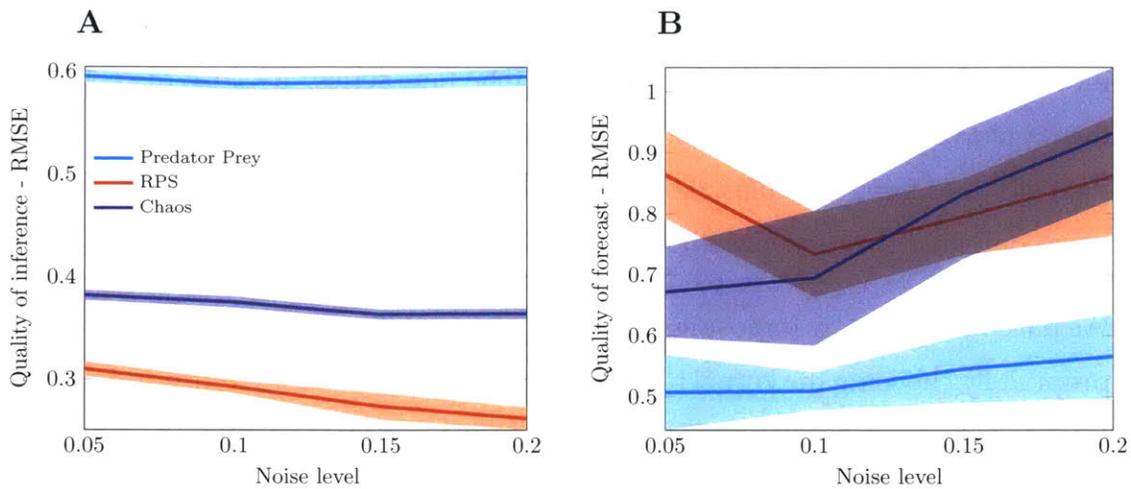


Figure 5.5: **Robustness of the regularized S-map to observational noise.** The figure shows that both the inference and the forecasting skills of the regularized S-map are robust to observational noise. Note that while the RMSE increases as we increase the level of noise, it still remains significantly lower than the naive error, see Table 5.2 and Figure 5.2.

Method	BATS	HOTS
Standard S-map	0.21 (0.19,0.36,0.07)	0.20 (0.24,0.16)
Regularized S-map	0.18 (0.17, 0.3, 0.07)	0.15 (0.23,0.07)
Naive Forecast	0.32	0.93

Table 5.4: **Performance of the algorithms on the test set of the empirical time series.** Performance is reported as the root mean square error (RMSE) on the test set, i.e., smaller RMSEs correspond to better performances. As a benchmark, we also report the error of the naive predictor (see Methods section). The numbers between brackets are the errors on each of the single species analyzed. The table illustrates that the regularized S-map outperforms the standard S-map also on the empirical data.

5.4 Discussion

In this chapter, we have studied the application of regularized loss minimization methods to a locally-weighted linear fit known as S-map developed to perform inference and forecasting from nonlinear ecological time series. The goal of this study has been to investigate if regularization applied to S-maps can improve the inference of the Jacobian coefficients (that are typically used as proxy for the local effect of biotic interactions on community dynamics) and the out-of-sample forecasting skill of species abundances when the time series are contaminated with process noise. In particular, to study the effect of regularization on different types of dynamics and

with different levels of process noise, we have investigated three standard nonlinear models: a predator-prey, a cyclic, and a chaotic time series. Because the choice of the regularization function is not unique (Abu-Mostafa et al., 2012), we have investigated two different regularization algorithms (Tikhonov and elastic net) and we have compared their performance on simulated and empirical data.

We have found that by imposing different types of regularization functions and appropriately choosing the kernel functions, we can improve the inference and the out-of-sample forecasting skills of the S-map. Generally, we recommend testing different regularization schemes and kernels in a preliminary analysis of a new data set to determine which one performs best. We speculate that the type of kernel that yields the best forecast could be informative of the type of dynamics generating the time series (this comes directly from the fact that kernels are a way to express prior knowledge about the data). For example, as expected, we have found that exponentially decay kernels perform better when forecasting with chaotic time series—where information decays exponentially (although more research is needed to validate or reject this hypothesis). Moreover, we suggest to choose the kernel based on a trade-off between different performance metrics such as correlation coefficients and the RMSE.

Most of the results of this study have been based on the analysis of synthetic data. We have used synthetic data to show the quality of the inference by comparison with analytic results (this is not possible with empirical data where the true Jacobian coefficients are unknown). Furthermore, the use of synthetic data have allowed us to systematically introduce process noise in the time series. However, to further test the validity of our approach, we have also provided an application on two empirical data sets. We have found that, as expected, the regularized S-map outperforms the standard S-map also on these empirical time series.

It is also worth mentioning that while the S-map has been and can be used on empirical data, there are a number of limitations for its applicability. These limitations apply equivalently to both the standard and the regularized S-maps. Firstly, it is important to stress that for an accurate reconstruction of Jacobian coefficients the S-map needs densely sampled neighborhoods on the attractor. Therefore, the data

requirement for an accurate inference increases exponentially with the system size (Hastie et al., 2001). Generally, if the number of variables (i.e., dimensionality of the system) is large and the time series is short we discourage the use of the S-map for inference studies. The exact definition of large, however, strongly depends upon the length of the time series. Here, for example, with four species and 50 data points, we have already obtained accurate reconstruction of Jacobian coefficients even in the presence of process and observational noise. Moreover, the dimensionality of the system may be reduced by excluding variables that are not causally related with the rest of the system (Sugihara et al., 2012). Secondly, as for many learning algorithms used in nonlinear time-series analysis, the stationarity of the data is an important requirement. Thus, to guarantee the statistical significance of the inferred Jacobian coefficients, a preliminary check for stationarity is needed (Kantz & Schreiber, 2004). Finally, because of the sparsity and nonlinearity of empirical time series, coefficients inferred with statistical algorithms, as the regularized S-map, may not be practically identifiable (meaning that more than one set of parameters can explain the data approximately equally well) (Saccomani, 2013; Saccomani & Thomaseth, 2016). Hence, inference studies should also include uncertainty quantification analysis (Cenci & Saavedra, 2018b). The applicability of the S-map for predictive studies is, on the other hand, less of a concern because the validity of the results can always be checked by means of cross validation.

Overall, the increasing amount of available time-series data is paving the way for moving theoretical population biology from being an assumption-driven towards a data-driven science. Because ecological time series are the product of an interplay between nonlinear deterministic and stochastic dynamics, the integration of methods of nonlinear time series analysis (based on dynamical system theory) with machine learning algorithms (developed to deal with observational data contaminated by noise) is an important step forward towards this goal.

CHAPTER 6

UNCERTAINTY QUANTIFICATION OF THE JACOBIAN COEFFICIENTS

In the previous chapter we have proposed an algorithm to efficiently infer Jacobian matrices from noisy time series data. Following standard approaches Jacobian matrices are then used as proxy for the effects of biotic interactions on community dynamics. We have shown the inconsistent link between explanatory and predictive studies and therefore that using prediction error to select the best parameters may lead to inconsistent conclusions. In this chapter we tackle this issue. Specifically, we show that one of the main limitations of existing methods to infer Jacobian matrices is that parameters inferred from noisy, sparsely sampled, nonlinear data are seldom uniquely identifiable. That is, many different parameters can be compatible with the same dataset and can generalize to independent data equally well. Hence, it is difficult to justify conclusive assertions about the effect of biotic interactions without information about their associated uncertainty. Here, we develop an ensemble method based on model averaging to quantify the uncertainty associated with Jacobian matrices inferred from non-equilibrium ecological time series data. Our method is able to detect the most informative time intervals for each jacobian element within a multivariate time series and can be easily adapted to different regression schemes.

6.1 Introduction

Biotic interactions are important regulators of the dynamics of natural systems (Tyc et al., 2014; Crowther et al., 2015; Abrudan et al., 2015; Ho et al., 2016; Enke et al., 2018). For example, interactions between primary producers in the ocean can alter the chemistry of their environment (Amin et al., 2015). Similarly, microbial interactions influence evolutionary responses to new environments (Lawrence et al., 2012), and facilitate emergent phenomena such as interaction-induced production of metabolites (Watrous et al., 2012). Yet, the relative importance of the effects of biotic and abiotic factors on community dynamics still remains a matter of debate (Fargione et al., 2003; Houlahan et al., 2007; Mutshinda et al., 2009).

As discussed in chapter 5, one of the main limiting factors in our understanding of the effects of biotic interactions on community dynamics is that, contrary to abiotic parameters, reliable experimental measurements of biotic parameters are seldom available (Wootton & Emmerson, 2005; Fuhrman et al., 2015). Hence, a large number of data-driven statistical methods have been proposed to infer the effect of biotic interactions on community dynamics from species abundance data (Lamon et al., 1998; Ives et al., 2003; Jiang et al., 2015; Deyle et al., 2016b; Marbach et al., 2012). Broadly speaking, these methods can be divided into three categories: (1) statistical or mechanistic parametric approaches, (2) nonparametric approaches based on correlations and co-occurrence of species (or taxa) in several equilibrium samples (Faust et al., 2012; Friedman & Alm, 2012; Kurtz et al., 2015), and (3) data-driven nonparametric approaches based on the theory of nonlinear-state-space reconstruction such as the regularized S-map developed in the previous chapter (Deyle et al., 2016b).

The first two approaches have presented a number of limitations: models are either too simplistic to capture the true dynamics (small deviations from the assumed to the true model can have significant impacts on the inference outcome (Xiao et al., 2017)) or too complex to be structurally identifiable (i.e., it is not possible to associate a unique value with the model parameters from empirical measurements) (Bellman & Aström, 1970; Villaverde et al., 2016; Saccomani & Thomaseth, 2016). Similarly, correlation or co-occurrence based approaches have long been criticized for not

being able to capture causation (Xiao et al., 2017), and for being prone to identify mirage correlations ubiquitous in nonlinear systems (Deyle et al., 2016b).

Instead, nonparametric state-space approaches based on attractor reconstruction have offered a promising alternative (Deyle et al., 2016b): they do not require linear equilibrium or mechanistic assumptions, and their parameters (used as proxy for the effects of biotic interactions on community dynamics (Ives et al., 2003; Deyle et al., 2016b; Xiao et al., 2017)) are structural identifiable. The goal of chapter 5 was to improve an existing nonparametric state-space approach to deal with the unavoidable stochasticity of natural systems. Here we deal with a different issue that arise when attempting to infer interactions from data. That is, while the coefficient of the S-map are structurally identifiable, structural identifiability does not imply that parameters can be uniquely identified given that noise, missing data, or sampling rates in empirical data can introduce confounding effects (Angulo et al., 2017). Specifically, distinct sets of parameters can be equally compatible with the observed data and can be used to construct models that generalize equally well on independent data, even if the model parameters themselves are structurally identifiable. Hence, there is unavoidable uncertainty associated with the parameters inferred from empirical data that, if caused by the data-collection process or inherent to the data-generating process, cannot be avoided regardless of the inference method used (Hines et al., 2014; Angulo et al., 2017).

Importantly, uncertainty on inferred parameters (the existence of more than one single explanation of the empirical data) translates into an uncertainty of the scientific conclusions that are drawn based on them. This may have significant consequences when, for example, conclusions are used to inform policy (Milner-Gulland, 2012), to develop novel drugs (Bollenbach, 2015), or to validate theories (Mutshinda et al., 2009). Yet, current nonparametric state-space approaches do not provide a methodology to quantify the level of uncertainty associated with the inferred interactions, or a measure of how this uncertainty changes in time when dynamics are nonlinear and interactions are time-dependent (Deyle et al., 2016b). To fill this gap, we develop a methodology to associate an uncertainty level with the temporal effect of

biotic interactions on community dynamics when inferred from multivariate non-linear time-series data using nonparameteric approaches.

Following our proposed methodology, we assess temporal variations of uncertainty about the effect of biotic interactions on community dynamics by measuring how many different explanations (set of parameters) are equally compatible with the same observational data at any given point in time. To validate our methodology, we study a chaotic synthetic time series for which we know the ground truth and use it to test the validity of the methodology. Then, as a case study, we investigate the uncertainty associated with the effect of biotic interactions in the two time series analyzed in the previous chapter: the Bermuda Atlantic time series (BATS) and the Hawaii ocean time series (HOTS). We focus on the same data for clarity of exposition and also show that different analysis (e.g. model selection, data-preprocessing) needs to be performed if the goal of the study is inference or forecasting.

6.2 Methods

6.2.1 Background

As we did in chapter 5, here we assumed that data are non-equilibrium time series of species abundances generated by a nonlinear population dynamics models of the form:

$$\frac{d\vec{x}}{dt} = \mathcal{F}(\vec{x}, \vec{\beta}), \quad (6.1)$$

where \mathcal{F} is an unknown and unspecified vector field (defining the dynamics of the system), $\vec{x} \in \mathbb{R}^d$ is the state vector (e.g., the abundances of d species), and $\vec{\beta} \in \mathbb{R}^q$ are the q parameters of the model (e.g., the birth and death rates of species). Importantly, the vector field (or model) \mathcal{F} does not need to be purely deterministic, but we assumed the existence of a manifold attractor (Casdagli et al., 1991; Sugihara, 1994). That is, Eq. (6.1) has a steady state (e.g., chaotic, fixed point, limit cycle) to which any initial trajectory converges (this is the only assumption on the data-generating process).

The goal of this chapter is to infer the Jacobian (\mathcal{J}) of \mathcal{F} from a realization of Eq. (6.1) using the regularized S-map and to associate an uncertainty level with its coefficients. To make the core of this chapter (section 6.2.4) more understandable here we provide a succinct summary of the regularized S-map developed in chapter 5. The reader familiar with the regularized S-map can skip the next section.

6.2.2 Statistical inference of the effect of biotic interactions from multivariate time series

The S-map is a weighted local regression model (Deyle et al., 2016b). The weights are state-space-dependent kernel functions. To limit potential problems of overfitting and singularities in the regression procedure of the S-map, we imposed an elastic-net regularization function—a convex mixture of L_1 and L_2 penalty terms (Zou & Hastie, 2005; Li & Liny, 2010). Specifically, for each point X^* on the manifold attractor (i.e., for each $t^* \in \{1, \dots, n\}$ with n number of observations) and for each $i \in \{1, \dots, d\}$ (with d number of dimension of the system i.e., number of species), we solved the following minimization problem:

$$\min_{\vec{\mathcal{J}}_i \in \mathbb{R}^d} (Y_i - X \vec{\mathcal{J}}_i)^T K(X, X^*, \theta) (Y_i - X \vec{\mathcal{J}}_i) + \lambda (\alpha \|\vec{\mathcal{J}}_i\|_2^2 + (1 - \alpha) \|\vec{\mathcal{J}}_i\|_1), \quad (6.2)$$

In Eq. (6.2) \mathcal{J}_i is row i of the Jacobian \mathcal{J} , and X is an $(n - 1) \times d$ data matrix where the point removed from the data matrix is the target point $x(t^*)$. Notice that for the analysis of empirical data, where the true dimension of the attractor is unknown, d is not necessarily equal to the number of observed species in the system but to the number of species plus an embedding. In addition, Y_i is the variable to be predicted (i.e. $Y_i = X_{t+1,i} \forall X_i \neq X_i^*$), $K(X, X^*, \theta)$ is a kernel function, and λ and α are two regularization parameters. The coefficients $\vec{\mathcal{J}}_i$ that minimize Eq. (6.2) are the rows of the Jacobian matrix—vectors on the tangent space of the manifold attractor of the data-generating process (i.e., Eq. (6.1)). That is, Eq. (6.2) provides the parameters of the linearization of the dynamics along each point on the manifold attractor. This linearization depends upon the curvature of the attractor at each point through the

kernel function. Motivated by the analysis of chaotic time series (Sugihara, 1994), a typical choice for the entries $K(\vec{x}, \vec{x}^*, \theta)$ of the kernel matrix (K) is an exponentially decaying function with a tuning parameter θ (chosen with cross-validation) that set the level of nonlinearity of the fit:

$$K(\vec{x}, \vec{x}^*, \theta) = e^{-\theta \frac{\|x-x^*\|}{\bar{d}}}, \quad (6.3)$$

where x^* is the target point and \bar{d} is the average distance of each point x to x^* . Note that the parameter θ measures the nonlinearity of \mathcal{F} (Sugihara, 1994). The choice of the kernel function (e.g., tri-cubic or Epanechnikov kernel) is not unique but depends on the structure of the time series. Recent work has shown that the S-map provides a good approximation of the Jacobian coefficients in nonlinear dynamical systems (Deyle et al., 2016b; Ushio et al., 2018).

6.2.3 Identifiability issues of the regularized S-map

It is important to notice that the solution of Eq. (6.2) is unique for any given set of λ, α , and kernel parameters. Indeed, Eq. (6.2) is a strictly convex problem for any $\alpha > 0$ (Zou & Hastie, 2005). Hence, the parameters inferred using the regularized S-map are structurally identifiable and their optimum value can be found, for example, by means of cross-validation (Hastie et al., 2001). However, the existence of a unique minimum of the loss function (Eq. (6.2)) is a weak condition for identifiability. In fact, while cross-validation applied to a strictly convex loss function selects one single model, the training-error and validation-error landscapes of Eq. (6.2) can be degenerate around the minimum error. That is, generally, different sets of λ, α , and kernel parameters can provide solutions (inferred parameters) with validation and training-errors close to the minimum. However, as we will see in the next section, it cannot be guaranteed that degeneracy in training and validation errors necessarily translates into a degeneracy of the inferred parameters. Hence, different sets of parameters may explain the observational data equally well, leaving the uncertainty of which set of parameters to choose as a true explanation of the data. In the next

section, we develop an algorithm to associate an uncertainty level with the Jacobian coefficients inferred from the S-map for each point along the manifold attractor.

6.2.4 A model average algorithm to associate an uncertainty level with the effect of biotic interactions

To associate an uncertainty level with the Jacobian matrix of Eq. (6.1) inferred from the S-map at any given time t , we proposed an algorithm based on model averaging. Specifically, for each point in time, we fixed the ratio α of L_1 to L_2 norms and solved Eq. (6.2) using leave-one-out cross-validation. Then, we changed α and solved again Eq. (6.2). We repeated this for $\alpha \in [0, 1]$ with steps of $\delta\alpha = 0.01$. From this ensemble of solutions (i.e., ensemble of Jacobian coefficients), we selected the subset \mathcal{M} of parameters that can be used to construct local linear models that exhibit minimum training and test errors (within a threshold that we fixed at 95% of the minimum training and test error, respectively). If the intersection is empty, we only considered models with optimum test error. This selection procedure allowed us to discard parameters (and therefore models) that either do not fit or overfit the data.

The fitting procedure above has a Bayesian interpretation. In fact, a particular choice of α in Eq. (6.2) corresponds to the assumption of a specific prior distribution on the parameters (Zou & Hastie, 2005; Li & Liny, 2010). Then, the subset \mathcal{M} is an ensemble of models equally compatible with the observed data, but with both prior and posterior distributions of the parameters that are not necessarily the same. In other words, we assumed uncertainty on the prior distribution of the parameters, which is translated into an uncertainty in the posterior. Finally, from the optimal ensemble of parameters \mathcal{M} , we calculated an expected value of the Jacobian coefficients with a weighted average of the parameters in the ensemble, i.e., \mathcal{L} :

$$\mathbb{E}[\mathcal{J}(t)|X] = \sum_{m \in \mathcal{M}} \mathbb{E}[\mathcal{J}(t)|\mathcal{L}_m, X] \mathbb{P}[\mathcal{L}_m|X] \quad \forall t \in [t_o, t_f]. \quad (6.4)$$

Note that the choice of the weights is arbitrary. For example, in a Bayesian framework, one typically includes in the weighted average all the models and uses the

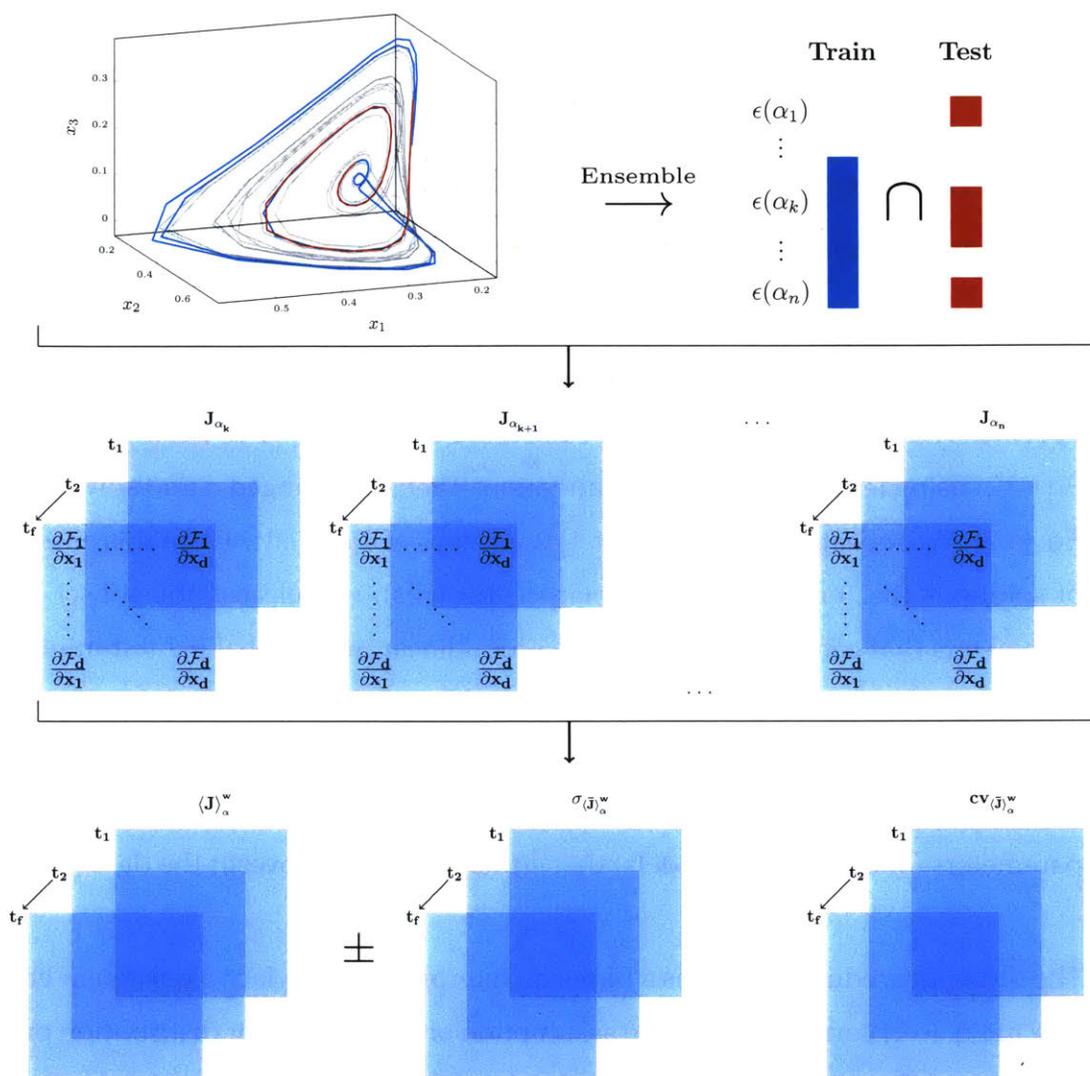


Figure 6.1: **Schematic picture of the proposed ensemble methodology used to associate a level of uncertainty with the effects of biotic interactions.** Based on a multivariate time series (top left panel), we inferred the Jacobian matrices at each point $\vec{x}(t_i)$ of a training set (light blue) on the manifold attractor \mathcal{M} . The quality of the model is then tested on an independent test set (dark red). We used Eq. (6.2) to perform the inference of the Jacobian coefficients. Then, we repeated the inference by changing the ratio α of L_2 to L_1 norms and running cross-validation to select the optimal λ and bandwidth of the kernel. This generated an ensemble of optimal models, from which we selected those with the best performance on the training and test sets—i.e., minimum $\epsilon(\alpha_i)$ within a threshold (colored squares in the top right panels). Finally, using these time series of Jacobians (one for each model in the intersection, middle panel), we computed their weighted average, standard error, and coefficient of variation (bottom panel).

posterior probability of each model as weight (Hastie et al., 2001). Instead, in an information-theoretic framework, one can use the likelihood of a model given the data applying any information-theoretic measure, such as the Akaike's information criterion (Burnham & Anderson, 2002). Here, we restricted the weighted average to the subset \mathcal{M} . We assumed that the weights are proportional to the probability that the model explains the data using a normalized fraction of explained variance in the training set (recall that we have already selected those models with optimal generalization skills). This procedure allowed us to take into account all the best solutions of Eq. (6.2) in the computation of the the effect of biotic interactions, and to discard solutions that either do not explain the data or generalize poorly on independent data. Using the weighted average, we constructed a new Jacobian matrix at each time t as $\mathcal{J}_{\text{ens}}(t) = \mathbb{E}[\mathcal{J}(t)|X]$. Notice that, while the training and test errors in the ensemble are all approximately the same, the coefficients over which we averaged can be significantly different.

An important advantage of computing the Jacobian from an ensemble of models is that we can associate an uncertainty level with the elements of \mathcal{J} . That is, we can compute the standard errors of the Jacobian coefficients and use them to construct confidence intervals. Then, using these errors, one can associate an uncertainty level with any quantity expressed as a function of the Jacobian elements (e.g., eigenvalues, Lyapunov exponents). In the following, we used the coefficient of variation (i.e., standard deviation over the mean) in order to compare the uncertainty across Jacobian coefficients with different means. It is important to stress that the coefficient of variation cannot be used to construct confidence intervals, but it provides a more intuitive measure about how many different explanations (Jacobian coefficients) are compatible with the same dataset. Figure 6.1 shows a schematic cartoon of the proposed methodology. In the next sections, we describe the application of our methodology on synthetic and empirical time series data.

6.2.5 Analysis of synthetic time-series data

Firstly, we tested our approach on synthetic data for which we know the ground truth of the Jacobian coefficients. As an illustrative example of nonlinear dynamics, we generated synthetic data using a chaotic 4-species Lotka-Volterra population dynamics model:

$$\dot{\vec{x}} = \vec{r}\vec{x}(1 - \mathcal{A}\vec{x}), \quad (6.5)$$

where \vec{r} is a vector of intrinsic growth rates and \mathcal{A} is the interaction matrix. Following previous work (Vano et al., 2006), we set the parameters used to generate chaotic trajectories from Eq. (6.5). To mimic sparsity in the time series, we sampled data every 200 data points after a numerical integration of Eq. (6.5). Then, we compared the parameters inferred using Eq. (6.2) and the model-averaging method with the analytical Jacobian of Eq. (6.5) computed at each point along the manifold attractor. To perform this analysis, we fixed the length of the training set to 100 data points and the length of the test set to 10 data points. The number of data points is chosen accordingly to the standard length of ecological time series. To obtain a reliable statistics, we repeated this process for 20 randomly sampled time intervals of the generated time series. Finally, we explored how the uncertainty level associated with different parameters changes in time from this purely deterministic nonlinear dynamical system.

6.2.6 Analysis of empirical time-series data

To illustrate the applicability of our methodology, we performed our uncertainty analysis on the effect of biotic interactions from the two empirical datasets analyzed in the previous chapter. For completeness here we describe the data again. Also, different from the previous chapter, here the goal is not to forecast species abundance but rather to accurately infer biotic interactions. Therefore model selection and data pre-processing will be different from before. The first dataset is a time series from a marine microbial community located at 19.225°N-39.455°N, 59.649°W-74.6°W. The data are publicly available at <http://batsftp.bios.edu/BATS/bottle/>

bval_bottle.txt. The dataset includes the abundance of four species (*Prochlorococcus*, *Synechococcus*, *Picoeukariotes* and *Nanoeukariotes*) as well as environmental parameters, such as temperature, depth of sampling, and salinity. Other ecological variables such as nutrients have been omitted from the analysis as they included many missing values. We analyzed data collected approximately twice per day from September 28th 2011 to October 18th 2011 across the ocean surface down to $\sim 200m$. The second dataset is a time series of a similar microbial community with *Prochlorococcus*, *Synechococcus*, *Picoeukariotes* and heterotrophic bacteria located at $22^{\circ} 45'N$, $158^{\circ} 00'W$ —approximately 100 km north of Oahu, Hawaii. The data are publicly available at <http://hahana.soest.hawaii.edu/hot/hot-dogs/index.html>. We analyzed data collected approximately once per month from 2006 to 2016.

In both datasets, we focused on the upper layer of the ocean, i.e., from the surface to a depth of 50m, which is the region of the ocean dominated by high light-adapted clades of *Prochlorococcus* (i.e., eMIT9312 and eMED4) (Johnson et al., 2006). After sub-setting the data within this region, we aggregated the collected variables (i.e., abundance of species and environmental parameters) and excluded the remaining missing points. This subdivision of data yielded a time series of 30 data points for the Bermuda dataset, out of which 28 were used for training and 2 for testing. Similarly, this subdivision yielded a time series of 93 data points for the Hawaii dataset, out of which 91 were used for training and 2 for testing.

Using the pre-processed data, we performed a variable selection: We standardized variables to have zero mean and unitary variance in the training set. Following standard approaches, we standardized the test set using the mean and variance of the original training set (Shmueli et al., 2017). Then, differently from the previous chapter where the optimal model was selected based on correlations as the goal was to forecast, here we performed a causality test to reduce the risk of including spurious correlations and to select a model system within which variables were causally related. Specifically, we used a convergent cross-mapping (CCM) test (Sugihara et al., 2012). Using this new subset of variables, we selected the kernel function (from a list of exponential, Epanechnikov, tri-cubic, and Matern), the combination of predictors,

and the time lags (i.e., embedding dimension) that maximized the out-of-sample forecasting skills of the regularized S-map (Eq. (6.2)). We measured the forecasting skills by the root mean squared error (RMSE), which we compared against the RMSE of the naive forecast (i.e., the forecast that assumes that the variables on the test set are equal to the last point in the training set (Shmueli et al., 2017)). Finally, we ran the same analysis described in the previous sections to analyze the uncertainty level associated with the effect of biotic interactions (Jacobian coefficients) for these two microbial communities.

6.3 Results

6.3.1 Analysis of synthetic data

We tested our proposed methodology on a (synthetic) chaotic time series. We found that, for this particular time series, our model-averaging method provides a better inference of the Jacobian coefficients than simple cross-validation. We measured the quality of the inference using the Pearson correlation coefficient (Sugihara, 1994) between inferred and analytical Jacobians (see Fig. 6.2). Then, we looked at the distribution of the coefficient of variation of the Jacobian coefficients within the ensemble of optimum models (recall that, as discussed in the Method section, we look at the coefficient of variation to fairly compare uncertainty across coefficients with different means). Surprisingly, Figure 6.3 (top panel) shows that while the value of some elements of the Jacobian matrix was consistently estimated across each model in the ensemble (i.e., we observed a small coefficient of variation), the value of other coefficients changed significantly from model to model (i.e., large coefficient of variation). These findings reveal that not all the inferred effects of biotic interactions (Jacobian coefficients) can be equally trusted (St John & Doyle, 2013). Recall that we are inferring coefficients from a purely deterministic, noise-free, time series. Hence, this effect is not due to the quality of the data, but to intrinsic properties of the dynamics.

Moreover, we found that the uncertainty level associated with the elements of the Jacobian matrix changes in time. For example, Figure 6.3 (bottom panel) shows how

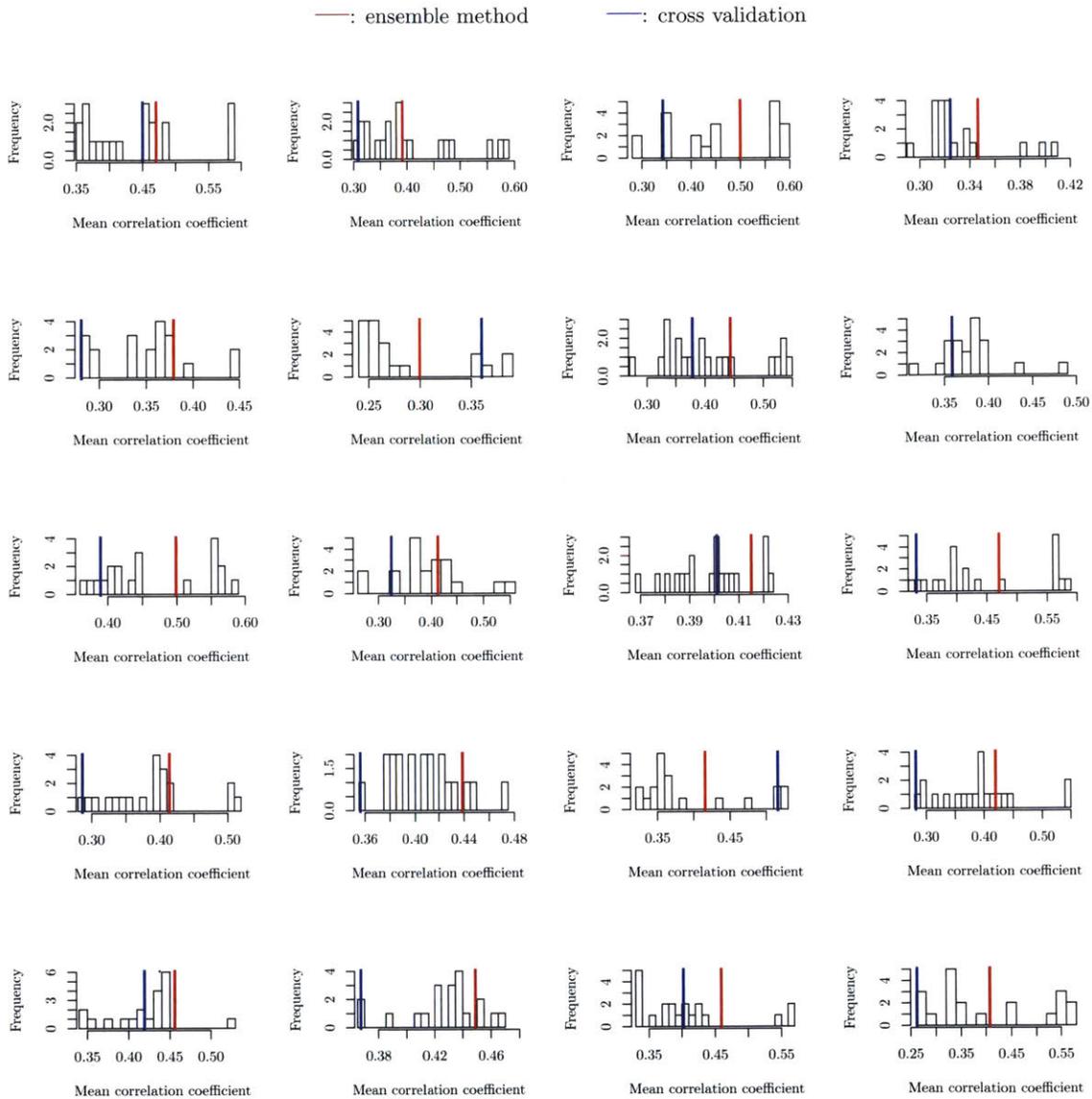


Figure 6.2: Quality of inference across models The figure shows the distribution of the mean Pearson correlation coefficient between analytical and inferred interactions (elements of the Jacobian matrix). Each panel is one of the 20 samples of the chaotic time series discussed in the main text. The blue and red horizontal lines correspond to the correlations generated by the single best model and the averaged model, respectively. Importantly, this figure shows that the model-averaging procedure tends to move the inference towards models that could not be selected by cross-validation, but exhibit a better correlation with the analytical coefficients.

the uncertainty level associated with the least consistently inferred Jacobian coefficients changes along each of the points on the manifold attractor. In this figure, one can observe regions of small uncertainty (i.e., small coefficients of variation)

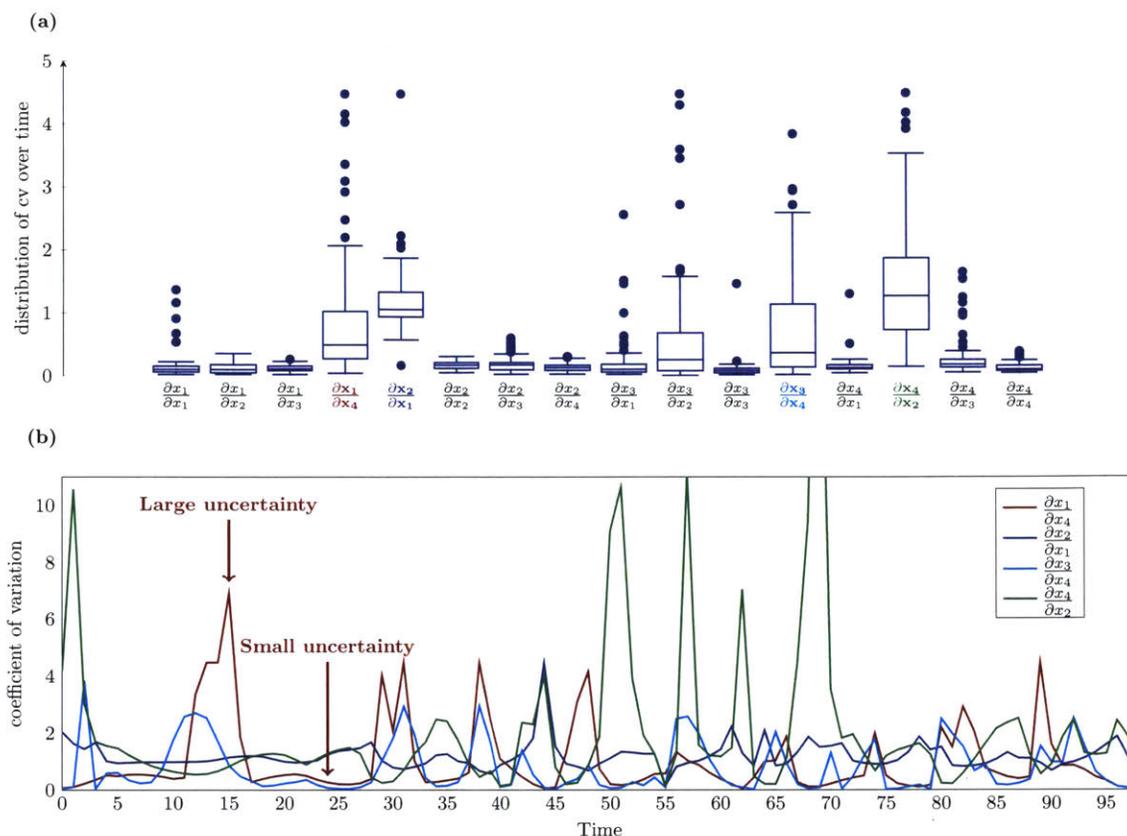


Figure 6.3: **Uncertainty analysis on synthetic data.** The figure illustrates that the uncertainty level associated with the Jacobian coefficients of nonlinear time series exhibits a strong time-dependency even in a perfect deterministic setting. The top panel shows the distribution of the coefficient of variation of the Jacobian coefficients. We used the coefficient of variation to compare the level of uncertainty across coefficients with a different mean. The distribution is computed for each element of the Jacobian over the time interval of the training set. The colored labels correspond to the four elements with the largest coefficients of variation. For these four elements, the bottom panel shows how their uncertainty level changes across time, i.e., it is weak or strong depending on the position on the manifold attractor.

followed by peaks of strong uncertainty (i.e., large coefficients of variation). These findings suggest that the inferred effects of biotic interactions can be trusted more in certain periods of time than in other periods. This also suggests that our method can be used to detect when the data are more informative about the inferred effects of biotic interactions.

6.3.2 Analysis of empirical data

6.3.2.1 Bermuda Atlantic time series

Focusing on the BATS dataset, the CCM test showed a significantly low causal relation between *Prochlorococcus* and *Nanoekariotes*, as well as between *Picoekariotes* and *Nanoekariotes*. The strongest causal relation was found within the sub-system made of *Prochlorococcus*, *Synechococcus*, and *Picoekariotes*. Figure 6.4 shows the CCM analysis. Using the time series of *Prochlorococcus*, *Synechococcus*, and *Picoekariotes*, the model system with the smallest out-of-sample forecast error ($\text{RMSE}_{\text{test,smap}} = 0.17$, $\text{RMSE}_{\text{test,naive}} = 0.31$) was found for exponentially decaying kernels and a combination of *Prochlorococcus* (one time lag or embedding dimension $E = 1$) and *Picoekariotes* (two time lags). We excluded the abundance of *Synechococcus* from our analysis since adding it as a predictor variable (with a suitable change in the embedding dimensions) increased both the in- and out-of-sample errors of *Prochlorococcus* and *Picoekariotes*. Using this 2-species model system, we inferred the interactions using Eq. (6.2) and the procedure described in section 6.2. This resulted in an ensemble of 21 models with the lowest test and training errors ($R^2_{\text{training}} \sim 0.82$), which were used for the uncertainty analysis.

Figure 6.5 (top left panel) shows the distribution of the coefficient of variation of the Jacobian coefficients of this 2-species model system. Similar to the results shown for the synthetic data, we found results consistent with theoretical expectations. Specifically, some Jacobian coefficients (effect of biotic interactions on community dynamics) are inferred with more confidence than others. Moreover, in line with the results of the analysis on synthetic data, the uncertainty level associated with each coefficient also changed across time. That is, there were regions of the manifold attractor of this model system in which all Jacobian coefficients had equally low uncertainty (resulting in a strong confidence about the selected average model), and other regions in which some coefficients had large uncertainty (providing a low confidence) (see bottom left panel of Figure 6.5). Figure B.3 shows the Jacobian coefficients with confidence intervals constructed using the standard error.

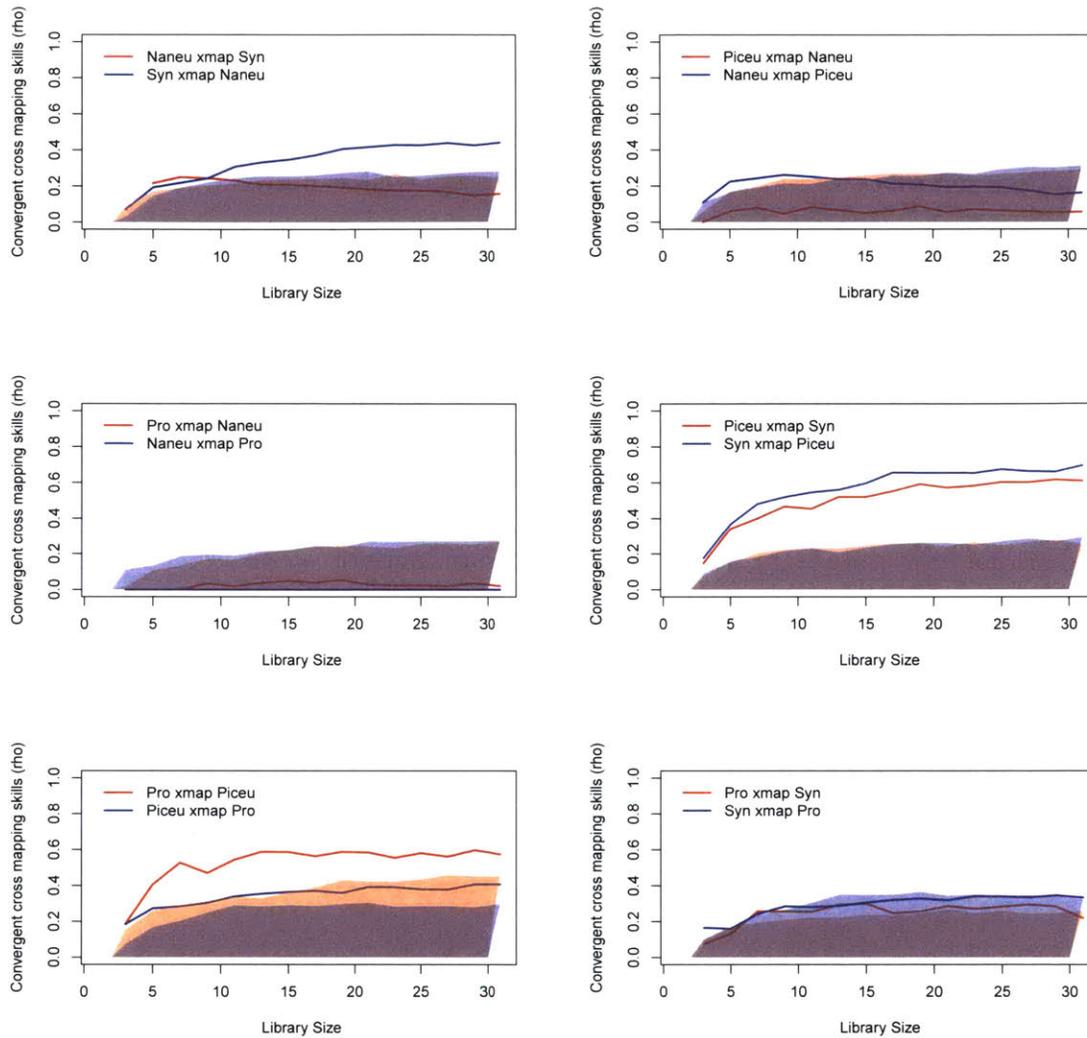


Figure 6.4: **Convergent cross mapping.** Results of the convergent cross mapping analysis of the Bermuda Atlantic time series. The shaded area corresponds to the analysis of a surrogate time series. Specifically, to measure the significance of the causal link between variables, we used 50 surrogate time series and the shaded area represents the 5th and 95th percentile of the ensemble. Long-term correlation coefficients larger than the respective shaded area indicate significant causation.

6.3.2.2 Hawaii ocean time series

Shifting our focus to the HOTS dataset, we centered our uncertainty analysis on the interactions between *Prochlorococcus* and *Picoeukariotes* in order to compare their effects against the BATS dataset. In this time series, the causal relation of the two phytoplankton species (computed with the CCM test) was significantly lower than in

the BATS dataset. Nevertheless, this model system provided the best out-of-sample forecast ($\text{RMSE}_{\text{test,smap}} = 0.17$, $\text{RMSE}_{\text{test,naive}} = 0.92$), which was found for an embedding dimension (time lag) of $E = 2$ for each variable using a tri-cube kernel function (Hastie et al., 2001). The maximum explained variance in the training set was $R^2_{\text{training}} \sim 0.7$. This resulted in an ensemble of 8 optimal models, which were used for the uncertainty analysis.

Figure 6.5 (top right panel) shows the distribution of the coefficient of variation of the Jacobian coefficients of this 2-species model system. Different from the BATS dataset, most of the Jacobian coefficients have a significantly low level of uncertainty. Specifically, only the effect of *Prochlorococcus* on *Picoeukariotes* exhibited a large coefficient of variation, but because this coefficient is very close to zero, the actual standard error is very low. Additionally, Figure 6.5 (bottom right panel) shows that, contrary to the BATS dataset, the temporal pattern of uncertainty on the coefficients was relatively homogeneous across time. Figure B.4 shows the Jacobian coefficients with confidence intervals constructed using the standard error.

Finally, we observed that the temporal pattern of uncertainty did not change dramatically when we included additional variables in the regression analysis. For example, in Figure B.2, we show that the distribution of the coefficient of variation of only one of the Jacobian elements changed significantly by adding the abundance of *Synechococcus*. Recall, however, that *Synechococcus* was excluded from the original analysis because its inclusion reduced the performance on both the training and test sets.

6.4 Conclusions

Understanding the effects of biotic interactions on community dynamics has been challenging because of the difficulty in accurately estimating interaction parameters and their associated uncertainty from empirical observations. To tackle this problem, standard approaches have used conditional least squares estimates of MAR(1) parameters as a proxy for biotic interactions, and have used either parametric bootstrapping or profile likelihood methods to estimate their uncertainty (Ives et al.,

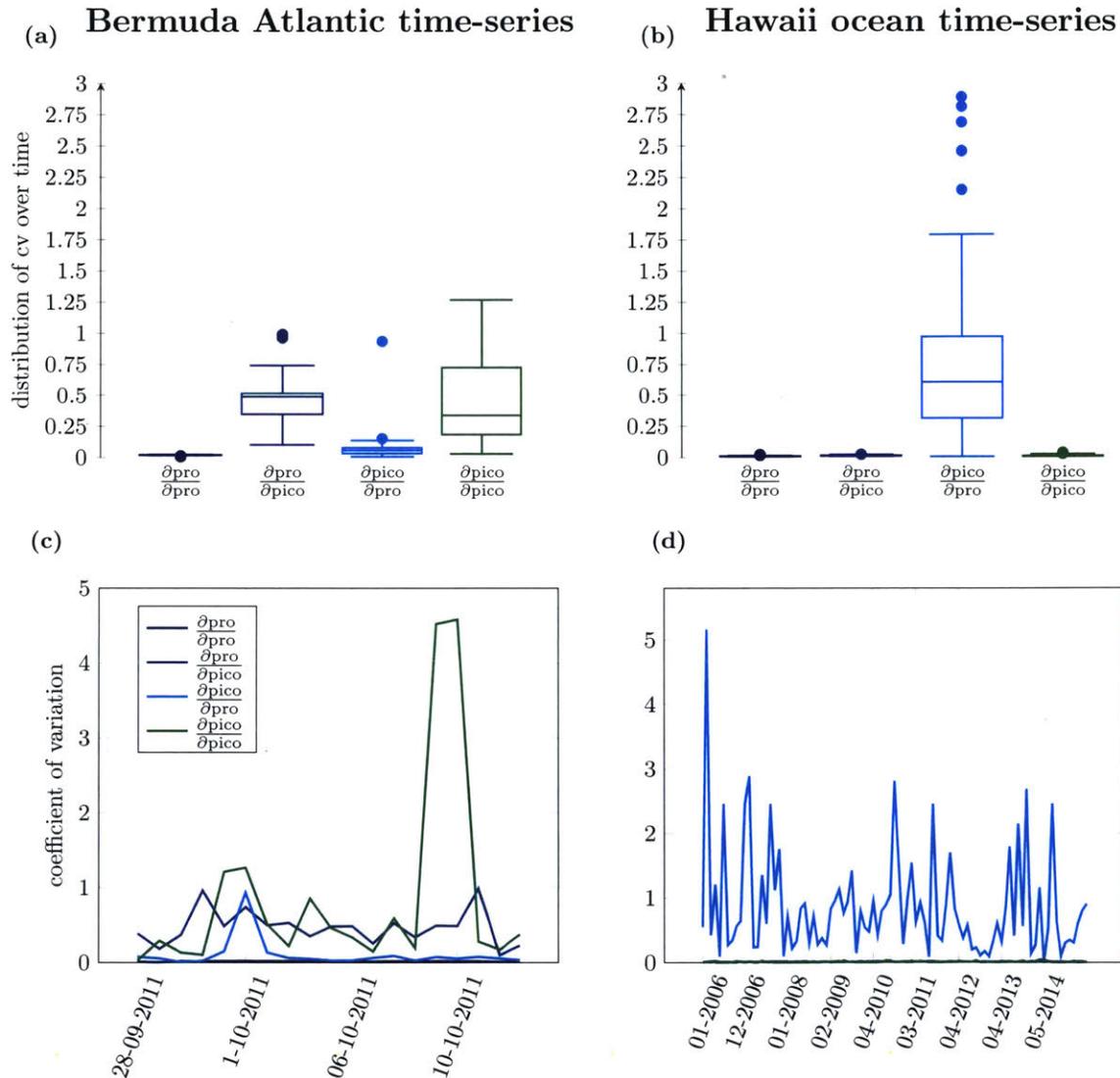


Figure 6.5: **Uncertainty analysis on empirical data.** The top panels show the coefficient of variation of the empirical Jacobian coefficients from each of the top models from the ensemble (see Fig. 1). The distributions were computed for each element of the Jacobian over the time interval of the training set. The bottom panels show how the coefficient of variation changes across the manifold attractor.

2003; Knight, 2000). However, because these approaches rely on equilibrium assumptions and equilibrium dynamics are rarely observed in natural ecosystems (Sugihara, 1994; Benincà et al., 2008; Sugihara et al., 2012; Perretti et al., 2013b; Ushio et al., 2018), their applicability on empirical data has been limited. Similarly, other approaches have used nonparameteric estimates of Jacobian coefficients as a proxy for the effects of biotic interactions on community dynamics (Holmes et al., 2012; Deyle et al., 2016b). However, while these approaches can deal with non-equilibrium dynamics, they lack of a consistent framework for quantifying the uncertainty associated with their results.

To address the limitations above, here we have developed a novel data-driven approach based on model averaging to quantify the uncertainty level associated with the local effect of biotic interactions (Jacobian coefficients) on community dynamics across time from a multivariate nonlinear time series. We have quantified the uncertainty of these interactions based on the number of equally valid explanations compatible with a particular dataset. Importantly, the confidence intervals constructed using our approach, which is local in time, are time-dependent. This is an important advantage because even in the presence of noise, strong nonlinearities, or poor quality of the data, our approach provides a clear intuitive methodology to identify regions of the data that are strongly identifiable and from which conclusions can be asserted with stronger confidence.

We have found three main results derived from our proposed methodology. Firstly, by averaging out different posterior distributions, our methodology can provide better inference of the effects of biotic interactions than previous methodologies (Deyle et al., 2016b) (see Figure 6.2). Note that the inference quality of our methodology, or any other statistical inference algorithm, cannot be tested from empirical data as the ground truth is unknown. Cross validation provides a mean to choose a particular set of parameters, but low validation errors do not necessarily guarantee a good accuracy on the inferred parameters as inferring and forecasting are two separate tasks (Shmueli, 2010; Shmueli et al., 2017).

Secondly, using two marine microbial communities as case studies, we have found

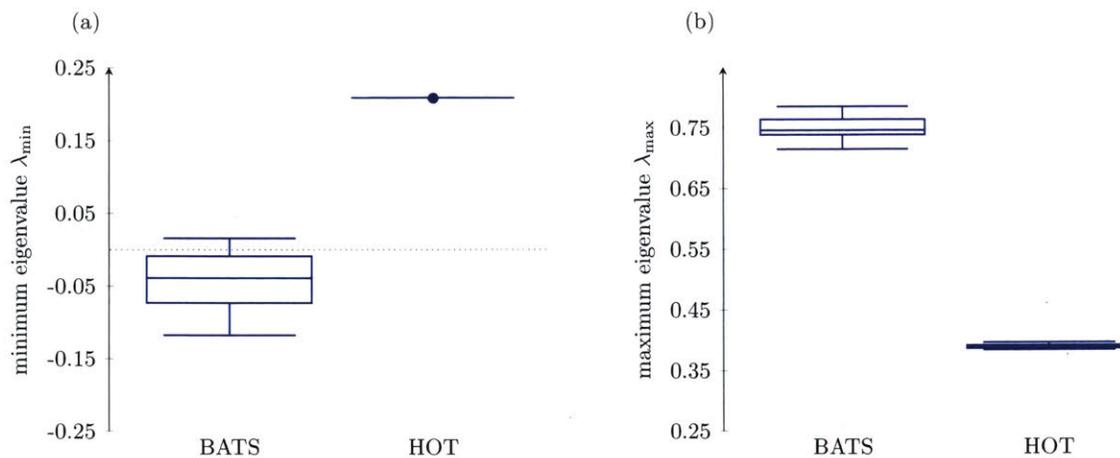


Figure 6.6: **Uncertainty associated with the eigenvalues of the empirical Jacobian coefficients** The Figure illustrate the distribution of the minimum (panel (a)) and maximum (panel (b)) eigenvalues of the empirical Jacobian matrices within the ensemble used to constructed the average Jacobian coefficients. As we were expecting based on the result of Figure 6.5 a large number of eigenvalues (even with different sign) are compatible with the data of the Bermuda Atlantic time series.

that the uncertainty associated with the effect of biotic interactions changes significantly across time (Figure 6.5 top panels). Moreover, we have shown that different interactions can have significantly different levels of uncertainty. This result implies that while some of the single interactions can be trusted, the whole Jacobian matrix can have a large associated uncertainty. In fact, we have shown that this can also be true for noise-free synthetic datasets (Figure 6.3). This is an important point to bear in mind because if the aim of a study is, for example, to investigate the stability of a community from the inferred Jacobian matrix (Ives et al., 2003; Ushio et al., 2018), then even a small uncertainty associated with an element of the Jacobian can translate into a large uncertainty on its eigenvalues (see Fig. 6.6).

Thirdly, we have found that in both synthetic and empirical time series the pattern of uncertainty can be considerably different across time. These differences happen within the same model system sampled at different locations and at different time scales. Hence, our method can also be used to choose from a number of datasets the one from which parameter inference is more reliable.

Overall, we have proposed a methodology to associate a level of uncertainty with

Jacobian coefficients of nonlinear systems inferred from empirical data. As a case study, we have analyzed data from population biology for which Jacobian coefficients can be used as a proxy for the local effect of biotic interactions on the dynamics of the community. However, because our approach only relies on the assumption that data are generated by processes that are not purely stochastic, it can be of practical use for other disciplines where time-series data are expected to align with this assumption, such as in finance or economics.

CHAPTER 7

NONPARAMETRIC STRUCTURAL STABILITY ANALYSIS

Using the concepts and the methods developed in the previous chapters, here we introduce a novel nonparametric framework to estimate the effect of environmental changes on the dynamics of biological populations. The chapter is divided in two main sections. Firstly we introduce the theoretical framework and we validate it on several synthetic datasets. Then using empirical data from a long term observational experiment on a rocky intertidal community we show that our approach can unveil an interesting relation between the structural stability of the community and relevant environmental variables. Overall, this chapter is an important step forward towards the foundation of data-driven, nonparametric, ecological theories.

7.1 Introduction

It is well known that changes in environmental conditions affect biological populations through perturbations of ecological parameters (Levins, 1968; Thomas et al., 2004). For example, fluctuations of temperature, salinity, and pH significantly affect the growth rate of bacteria in marine communities (Chase & Harwood, 2011). Similarly, it has been shown that climate change has strong impacts on species interactions (Montoya & Raffaelli, 2010; Thackeray et al., 2016). However, predicting

the effects that the variability of environment-dependent parameters (e.g., species interactions and growth rates) have on the dynamics of ecological communities has proved to be far more challenging due to the ubiquitous nonlinearities and state-dependent behavior of natural ecosystems (Sugihara et al., 2012). Yet, this knowledge is key for understanding and anticipating the response of ecological communities to environmental changes (Dai et al., 2012).

In the previous chapters and in numerous other research papers it has been shown that a theoretical framework to estimate the effect of parameter perturbations on nonlinear dynamics can be built based on the concept of structural stability (Thom & Fowler, 1975; Allen, 1976; Rohr et al., 2014; Arnoldi & Haegeman, 2016; Cenci & Saavedra, 2018a; Song et al., 2018), i.e., stability of the qualitative behavior of a dynamical system against perturbations of its parameters. Yet, the generalization of these approaches to empirical data is limited by the fact that current approaches assume either a specific parametric model for the dynamics (see (Rohr et al., 2014) and chapter 2-4 of this thesis), or the convergence to locally or globally stable equilibrium (fixed) points, see for example reference (Arnoldi & Haegeman, 2016). As discussed in the preface of this second part of the thesis, assuming parametric models is limiting because the validity of the results relies heavily on the correctness of the assumed equations and their parameters (Wood & Thomas, 1999; Perretti et al., 2013b; Ye et al., 2015). Similarly, assuming convergence to equilibrium points can be problematic because equilibrium dynamics are rarely observed in natural ecosystems (Benincà et al., 2008; Sugihara et al., 2012; Ushio et al., 2018).

To overcome these limitations, here we develop a data-driven nonparametric approach to study the structural stability of non-equilibrium dynamical systems. In this chapter, we focus on chaotic systems that we use as prototype for nonlinear dynamics. This choice is motivated by the growing evidence that ecological communities can exhibit chaotic dynamics not only in closed controlled experiments but also in the field (Costantino et al., 1997; Turchin & P. Ellner, 2000; Becks et al., 2005; Benincà et al., 2008, 2015). However, our approach is valid for any type of nonlinear dynamics. Using several synthetic datasets, we show that structural stability in

chaotic systems can vary significantly across time. That is, the tolerance of ecological communities to equivalent perturbations can change depending on the timing of their occurrence. Finally, using empirical data from a rocky intertidal community that exhibits dynamics at the edge of chaos (Benincà et al., 2015), we show that the structural stability of this community follows a clear seasonal pattern mediated by changing environmental conditions.

7.2 Theoretical framework

To illustrate our approach, we considered chaotic time series generated by some unknown dynamical system:

$$\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{\eta}), \quad (7.1)$$

where $\vec{f} \in C^1$ is a vector field (or model), $\vec{x} \in \mathbb{R}^d$ is the state vector (i.e., species abundance), and $\vec{\eta}$ a vector of environment-dependent parameters (e.g., rate of interactions, growth rates). In an empirical setting, the solution of the unknown system in Eq. (7.1) provides the observational time series that can be collected from an experiment or field observations. Here, we assumed that the time series is stationary and has a deterministic component stronger than the noise level (i.e., the trajectories lie on an attractor, \mathcal{M}). We also assumed that the observational noise on empirical data is normally distributed. Note that these are standard assumptions in nonlinear time-series analysis (Kantz & Schreiber, 2004; Mees, 2011).

7.2.1 Time-dependent response to changing environments

To understand the derivation of our proposed measure of structural stability, recall that, loosely speaking*, a system is structurally stable if small changes to the parame-

*More formally, a dynamical system is structurally stable if the topology of the phase portrait is preserved under smooth changes of the vector field (Strogatz, 2014). Here we consider the definition of structural stability mostly adopted in the ecological literature (Rohr et al., 2014; Saavedra et al., 2017b).

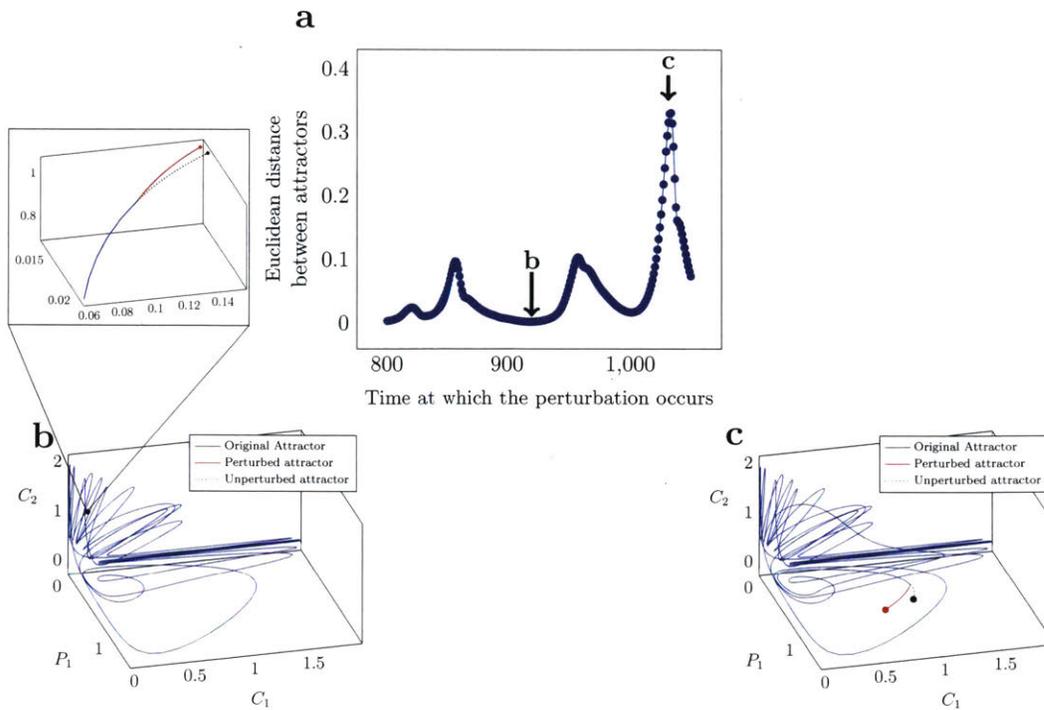


Figure 7.1: **Parameter perturbations occurring at different points in time on a chaotic dynamical system can have different effects on its dynamics.** Panel a shows the divergence of perturbed and unperturbed trajectories (measured by the Euclidean distance between attractors) when one parameter (ν_1 in Eq. (7.2)) of a consumer-resource chaotic model is changed at different points in time (i.e., different points on the attractor). The x-axis is the time at which the perturbation occurs, whereas the y-axis illustrates the effect of the perturbation. Panels b and c show the effect of the perturbation at the least and most affected points, respectively. The panels show a 3-dimensional projection of the 5-dimensional original attractor (i.e., the projection on the space of two consumers and one predator). The full circles in the two lower panels illustrate the state of the system after a time $\tau = 3$ after the perturbation (i.e., 300 points separated by a spacing $dt = 0.01$). The distance between the two states (perturbed, red circle, and unperturbed, black circle) is much larger in Panel c than in Panel b even though the perturbation to the parameters was exactly the same. Overall, the figure illustrates how in nonlinear systems, perturbations acting at different points in time can have different impacts on the dynamics.

ters (or the model itself) do not induce large changes in the dynamics (Mayo-Wilson, 2015). While structural stability is a global property of a model (or vector field), a striking feature of chaotic systems (and nonlinear systems in general) is that the same parameter perturbation can have a different impact on the dynamics depending on the time of its occurrence. This phenomenon follows directly from local properties of chaotic attractors (Nese, 1989) and has been observed in natural ecosystems (Song

& Saavedra, 2018). For example, it is well known that the effect of temperature on sardine stocks depends on the state of the system (Sugihara et al., 2012). Thus, there is a need to study and understand how the structural stability of nonlinear system changes locally along the model's trajectories.

To illustrate the state-dependent response of nonlinear dynamics to parameter perturbations, we ran the following analysis. (1) We integrated the dynamics of a 5-species chaotic dynamical system:

$$\begin{aligned}\frac{dP_i}{dt} &= \nu_i \lambda_i \frac{P_i C_i}{C_i + C_i^*} - \nu_i P_i \\ \frac{dC_i}{dt} &= \mu_i \kappa_i \frac{C_i R}{R + R^*} - \nu_i \lambda_i \frac{P_i C_i}{C_i + C_i^*} - \mu_i C_i \\ \frac{dR}{dt} &= R \left(1 - \frac{R}{k}\right) - \sum_{i=\{1,2\}} \mu_i \kappa_i \frac{C_i R}{R + R^*},\end{aligned}\tag{7.2}$$

with $i = \{1, 2\}$ and $\nu_1 = 0.1, \nu_2 = 0.07, \lambda_1 = 3.2, \lambda_2 = 2.9, C_1^* = C_2^* = 0.5, \mu_1 = \mu_2 = 0.15, \kappa_1 = 2.5, \kappa_2 = 2, R^* = 0.3, k = 1.2$. Following previous work (Deyle et al., 2016b), we chose the initial conditions such that we ensured that $\vec{x}(t_1)$ is on the attractor. (2) We run the dynamics up to a time t_1 and at this point we perturbed the parameter ν_1 from $\nu_1 = 0.1$ to $\tilde{\nu}_1 = 0.15$ (3) Then, we integrated the model with the perturbed parameters ($\tilde{\nu}_1$) and initial conditions $\vec{x}(t_1)$. After the perturbation, we ran the dynamics for 300 integration steps with spacing $dt = 0.01$. (4) Then, we measured the root mean squared distance of the perturbed and unperturbed trajectories. (5) Finally, we repeated steps (2)-(4) n -times to produce each point in Figure 7.1 Panel a. The values on the x -axis of the figure are the times t_1, \dots, t_n ; while the values on the y -axis are the Euclidean distances between perturbed and unperturbed trajectories. Overall, the figure shows the state-dependent effect of a parameter perturbation on nonlinear systems.

Figure 7.1 (Panel a) shows that the extent to which the perturbed and unperturbed trajectories diverge from each other depends on the time at which the perturbation occurs. Panel b and c show a visual example of a small and a large effect of the

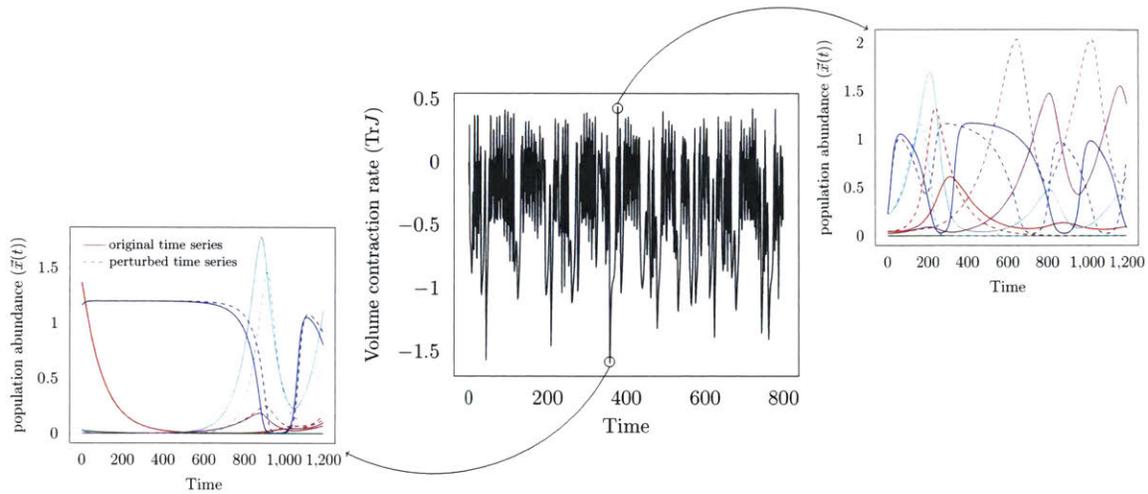


Figure 7.2: Perturbations acting on the parameters of a nonlinear system with large and small volume contraction rate induce different effects on the trajectories. The central figure shows the temporal changes of the volume contraction rate (i.e., trace of the Jacobian) for the same model discussed in Figure 7.1. The left and right panels show the perturbed (dotted line) and unperturbed (solid line) trajectories (of all the 5 species) in the most (bottom left) and least (top right) structurally stable states. The figure illustrates that the volume contraction rate can be used as an inverse measure of local structural stability. That is, the larger the volume contraction rate, the smaller the structural stability of the system.

perturbation on the trajectories, respectively. This example illustrates that in systems away from equilibrium points, the level of structural stability (measured here as the effect of a perturbation on the model's future trajectory) changes depending on the state of the system. Clearly, in a real setting, testing the effects of all possible perturbations is not a possibility. Therefore, to study the structural stability of a chaotic (non-equilibrium) system more systematically, we need to investigate the local properties of the vector field along the attractor (i.e., model's trajectories).

7.2.2 Structural stability of non-equilibrium dynamics

We propose to use the volume contraction rate (VCR) of a non-equilibrium dynamical system as a local inverse measure of structural stability. The VCR is the divergence of the vector field or, equivalently, the trace of the Jacobian matrix of \vec{f} , i.e., $\nabla \cdot \vec{f}(\vec{x}, \vec{\eta}) = \text{Tr}J$. Because the Jacobian matrix is a function of the state vector (i.e.,

species abundances, $J : J(\vec{x})$, the VCR changes with the state of the system. At each point on the attractor \mathcal{M} , the VCR can be negative or positive, or more generally, can be larger or smaller than in other points on \mathcal{M} . For small (but finite) perturbations in the parameters, models generated by perturbations acting on states with a small VCR will continue to have a small VCR. Hence, these perturbed models will stay closer to the original model than models generated by perturbations acting on states with a large VCR. This effect is shown in Figure 7.2. A small (large) deviation between perturbed and unperturbed trajectories after a perturbation of the parameters captures the idea that vector fields with a small (large) VCR are weakly (strongly) affected by a perturbation to the parameters. Thus, we can associate the VCR at each point on the attractor to the inverse of the structural stability of the vector field at that precise point. That is, the smaller the VCR, the larger the structural stability. Importantly, this association must hold true on average because there is always an uncertainty associated with the direction toward which the parameter perturbation pushes the model and its trajectories. Note that this is an heuristic argument in favor of the VCR as a probabilistic measure of structural stability and not a proof. Thus, in the next section we provide an extensive numerical validation of our hypothesis.

7.2.3 Validation on synthetic data

In this section, using synthetic data, we want to show that the VCR is a reliable probabilistic measure of structural stability. To test systematically the validity of our proposed measure, we run simulations using 8 different chaotic dynamical systems, namely:

- Model 1 is a competitive chaotic 4-species Lotka-Volterra model(Vano et al., 2006):

$$\dot{\vec{x}} = \vec{r}\vec{x}(1 - \mathcal{A}\vec{x}), \quad (7.3)$$

with $\vec{r} = [1, 0.72, 1.53, 1.27]$ and :

$$A = \begin{bmatrix} 1 & 1.09 & 1.52 & 0 \\ 0 & 1 & 0.44 & 1.36 \\ 2.33 & 0 & 1 & 0.47 \\ 1.21 & 0.51 & 0.35 & 1 \end{bmatrix}. \quad (7.4)$$

- Model 2 is a 3-species chaotic food chain(Upadhyay, 2000):

$$\begin{aligned} \frac{dx}{dt} &= rx\left(1 - \frac{x}{k}\right) - \frac{a_1xy}{1 + b_1x} \\ \frac{dy}{dt} &= -sy + hxy - \frac{a_2yz}{1 + b_2y} \\ \frac{dz}{dt} &= -lz + nyz, \end{aligned} \quad (7.5)$$

with $r = 4.3, k = 50, a_1 = b_1 = a_2 = b_2 = 0.1, s = 1, h = 0.05, l = 1, n = 0.03$

- Model 3 is Eq. (7.2)
- Model 4(Lai et al., 2018):

$$\begin{aligned} \frac{dx}{dt} &= ax - yz \\ \frac{dy}{dt} &= -by + xz \\ \frac{dz}{dt} &= -cz + xyz + k \end{aligned} \quad (7.6)$$

with $a = 4, b = 9, c = 3.6, k = 4.$

- Model 5(Pham et al., 2017):

$$\begin{aligned}\frac{dx}{dt} &= z \\ \frac{dy}{dt} &= -z(ay + by^2 - xy) \\ \frac{dz}{dt} &= x^2 - |x|y + y^2 - 1\end{aligned}\tag{7.7}$$

with $a = 5, b = 3.5$

- Model 6(Sprott, 1994):

$$\begin{aligned}\frac{dx}{dt} &= xy - z \\ \frac{dy}{dt} &= x - y \\ \frac{dz}{dt} &= x + az\end{aligned}\tag{7.8}$$

with $a = 0.3$

- Model 7(Sprott, 1994):

$$\begin{aligned}\frac{dx}{dt} &= y \\ \frac{dy}{dt} &= -x + ayz \\ \frac{dz}{dt} &= 1 - y^2\end{aligned}\tag{7.9}$$

with $a = 1$

- Model 8(Sprott, 1994):

$$\begin{aligned}
\frac{dx}{dt} &= -y \\
\frac{dy}{dt} &= x + z \\
\frac{dz}{dt} &= xz + ay^2
\end{aligned} \tag{7.10}$$

with $a = 3$.

Using these models we run the following numerical experiment. We integrated a chaotic dynamical system until the trajectories converge on the attractor \mathcal{M} . Convergence is ensured by integrating the differential equation for a long period of time before starting the simulation. Moreover, for many of the tested models initial conditions that converge quickly to the attractor were taken from the literature. Then, we perturbed a random number of parameters with normally distributed noise with zero mean and standard deviation proportional to a random fraction of the perturbed parameter. For example, for a given parameter μ we have $\mu_{\text{perturbed}} = \mu_{\text{unperturbed}} + \eta$ with $\eta \sim \mathcal{N}(0, \mu_{\text{unperturbed}}/\kappa)$ with $\kappa \sim \mathcal{U}(2, 10)$. The number of perturbed parameters was a random number drawn from a uniform distribution. Then, we randomly selected two points on the attractor by randomly sampling below (above) the 15th lower (upper) percentile of the trace of the Jacobian. This value is chosen, by studying how the RMSD between perturbed and unperturbed trajectories varies as function of the percentile as shown in Figure C.1. Then, we integrated again the model with the perturbed parameters and we measured the RMSD between perturbed and unperturbed trajectories. Finally, we took the logarithm of the ratio between the RMSD of the trajectories perturbed at the point in time with maximum structural stability to the one with minimum structural stability. For our measure to be valid, this ratio needs to be less than zero, i.e., perturbations to the parameters occurring in structurally stable states have weaker effects on trajectories than perturbations to the parameters occurring in states that are less structurally stable. Then, we repeated the same procedure for 1000 times over eight distinct chaotic models. The results are shown in Figure 7.3. The figure shows both the distribution

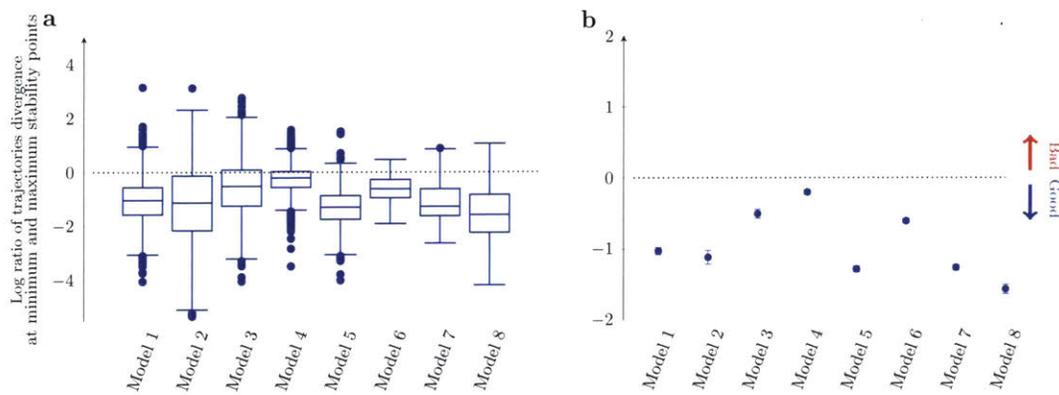


Figure 7.3: The volume contraction rate is a robust inverse measure of structural stability. The figure shows the logarithm of the ratio between the divergence of trajectories when the parameters are perturbed in the least and most structurally stable states (i.e., $\log\left(\frac{d(x, \tilde{x})_{\max \text{ struct stab}}}{d(x, \tilde{x})_{\min \text{ struct stab}}}\right)$) over 8 different chaotic models. Panel **a** shows the distribution over 1000 realizations, while Panel **b** shows the median with its 95% confidence interval computed with a two-sided Wilcoxon test. Note that a median ratio less than zero indicates that the volume contraction rate is a good measure of structural stability. That is, the larger the volume contraction rate, the smaller the structural stability of the system.

over all the realizations (Panel **a**) and the median with its 95% confidence interval computed with a two-sided Wilcoxon signed rank test (Panel **b**).

Figure 7.3 shows that, on average, perturbations on the parameters acting on points with large structural stability (i.e., small VCR) tend to generate trajectories that remain closer to the unperturbed trajectories than perturbations acting on points with small structural stability (i.e., large VCR). The closeness between trajectories was measured as the root mean squared distance between perturbed and unperturbed state vectors on the attractor after the perturbation. The results of this analysis provide support to the hypothesis that the VCR can be used as a reliable probabilistic indicator of how much a trajectory will be perturbed by a parameter perturbation. Therefore, following the definitions of this chapter the VCR is a reliable measure of structural stability.

7.3 Structural stability of empirical populations

Now that we have a measure of time-varying structural stability of non-equilibrium population dynamics we can analyze empirical data and try to obtain some fundamental understanding of what are the factors that influence the structural stability of empirical populations

In the last section we have shown that the VCR is a reliable probabilistic measure of structural stability. Because the VCR is computed by simply taking the trace of the Jacobian matrix we can now use the results from chapter 5 and 6 to compute the structural stability of empirical population without specifying explicitly the parametric form of the dynamics.

As a case study, we investigated the temporal evolution of the structural stability of a rocky intertidal community in New Zealand that exhibits dynamics at the edge of chaos (Benincà et al., 2015). The dataset used in this work can be downloaded from <http://www.pnas.org/content/112/20/6389/tab-figures-data>. The dataset consists on interpolated time series of the rocky intertidal community, the monthly sea temperature, monthly wind speed, and monthly wave height. The community is composed by three species, namely: the honeycomb barnacle *Chamaesipho columna*, the crustose brown alga *Ralfsia cf confusa*, and the little black mussel *Xenostrobus pulex*. The species interact on bare rock. The time series of species abundance has 250 data points sampled monthly from 1987 to 2008. A detailed analysis of the time series, can be found in the original paper (Benincà et al., 2015). Following the reference all species contributed to the dynamics (this is also illustrated in Figure C.2 in the Appendix where we show that all species are included in the causal network).

To estimate non-parametrically the Jacobian coefficients from the time series generated by the rocky intertidal community, we used the regularized S-map developed in chapter 5. To account for uncertainty in model selection, we used the model-averaging technique developed in chapter 6. Then, using the inferred Jacobian coefficients, we computed the value of the VCR (i.e., $\text{Tr}(J)$) at each point in time. Note

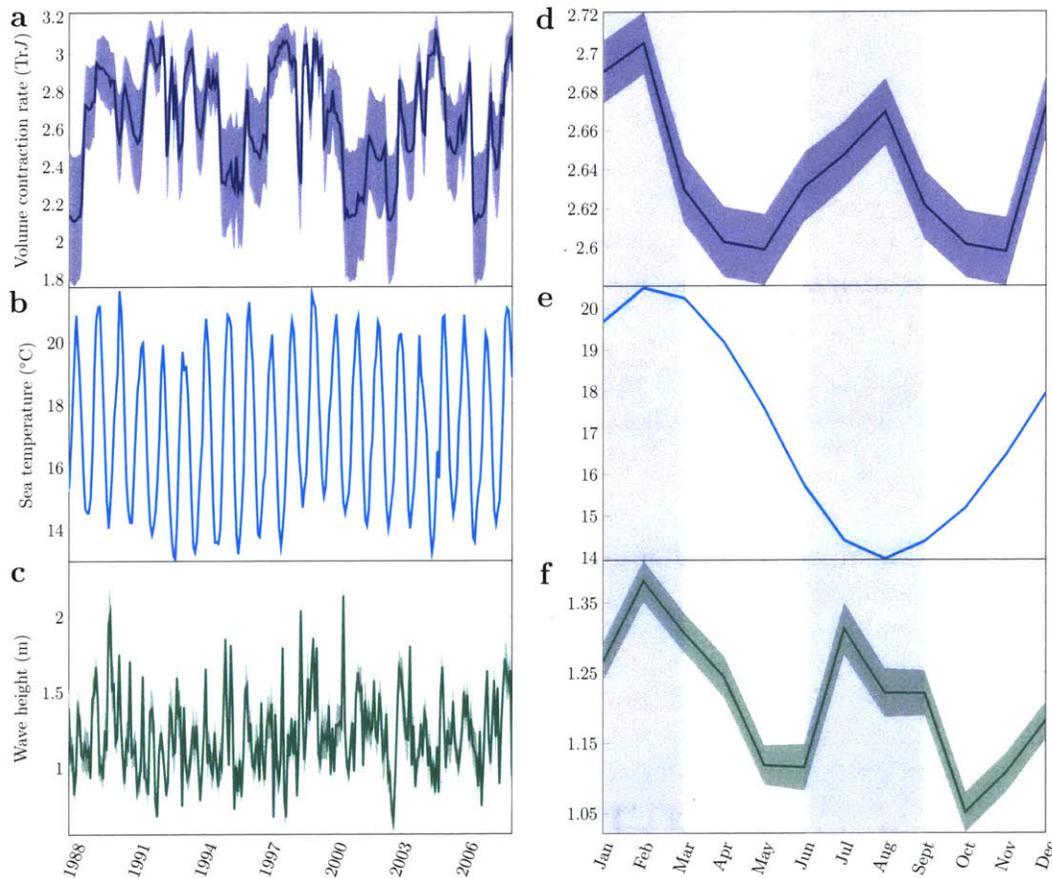


Figure 7.4: **Structural stability pattern of a rocky intertidal community at the edge of chaos.** Panels a-c show the fluctuations of the trace of the Jacobian (inverse measure of structural stability, and inferred from the S-map), sea temperature, and wave height, respectively. Panels d-f show, respectively, the monthly average of these quantities across the observation period. Importantly, Panel d shows that the structural stability of this community follows a clear seasonal pattern. That is, the maximum structural stability (i.e., minimum volume contraction rate) is observed in the spring and autumn seasons when waves are smaller and sea temperatures are milder.

that using the S-map algorithm, we can recover the trace of the Jacobian up to a constant. However, because we are interested in both the local minima and maxima of the trace, only the trend (not the sign of the trace) should be considered as meaningful.

Figure 7.4 (Panel a) shows that the structural stability of the community fluctuates without showing any particular trend across the observation period. Importantly, these observed changes of structural stability have no clear associations with envi-

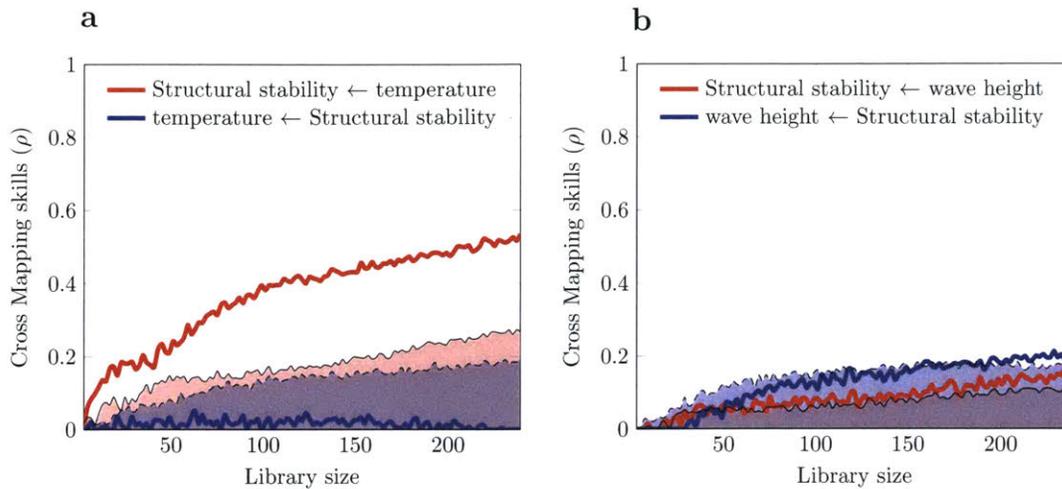


Figure 7.5: **Convergent cross mapping test.** The figure shows the results of the causal analysis discussed in the Methods section. The shaded area corresponds to the 95th percentile of the causality test run on the surrogate time series (providing a null expectation). A causal link is statistically significant if the convergent cross mapping skills (correlation coefficients ρ) increase with the library size and are above the null expectations. Panel **a** shows that sea temperature is a causal driver of the structural stability (i.e., the volume contraction rate) of the rocky intertidal community. Instead, Panel **b** shows that there is no statistically significant causal relation between structural stability and mean wave height.

ronmental parameters (Panels **b** and **c**). This result is not surprising as environmental variables were not included in the reconstruction of the Jacobian matrix (which is a proxy only for biotic interactions) nor in the computation of the structural stability (which is only a function of the Jacobian coefficients). Yet, looking at the monthly average of structural stability throughout the years, Panel **d** shows a clear seasonal pattern. That is, the community exhibits its minimum structural stability (i.e., maximum VCR) in the summer and winter seasons, particularly in the warmest and coldest months (February and August, respectively, see Panel **e**). In addition, we found that fluctuations of structural stability are strongly anti-correlated with the seasonal changes of mean wave height (see Panel **f**), which are known to play a significant role in the dynamics of rocky intertidal communities (Helmuth & Hofmann, 2001; Harley & Helmuth, 2003). Overall, our these finding suggest that environmental forcing can have an important effect on biotic interactions and, consequently, on the structural stability of this community.

To support our argument above however, we need to distinguish between causal associations and spurious correlations formed between environmental variables and structural stability. For this purpose, we used convergent cross mapping (CCM) (Sugihara et al., 2012): a nonlinear causality test that is used to infer causal relations among variables in nonlinear dynamical systems. CCM is an extension of standard Granger causality. However, differently from Granger causality, CCM can be used to test for causation in nonlinear dynamical systems, and requires convergence to large forecasting skills as the number of points included in test increases. Importantly, because CCM is based on Taken's theorem, the results of the CCM test depend strongly on the choice of the embedding dimension used to compute the pairwise causal links. Hence, we have put particular attention in choosing the correct embedding dimension by running numerous tests with different embedding dimensions (from 1 to 20), and picking the one that maximizes the cross mapping skills. To ensure causality, we need to show convergence and significance against a randomized version of the time-series data. To test for the statistical significance of the results, we ran the test with an ensemble of randomized versions of the time series, and we plotted the results from the true time series against the 95th percentile of the random analysis. This provided a 95% significance level on the causal links. Surrogate time series are generated with the *surrogate* package in R. Using standard approaches (Theiler et al., 1992; Kantz & Schreiber, 2004), we generated surrogate data so that the original and randomized time series have the same power spectrum (i.e., the Fourier transform of the original data and the surrogate ensemble only differ in their phase which is randomized but the amplitude is preserved). CCM skills (measured as correlation coefficients ρ) are considered significant if they are above the random expectations. In addition, we have also computed the significance of the observed causal links against different null models. This was performed using the *make_surrogate_data()* function in the R packages *rEDM* with *method = seasonal* and *T_period = 12*.

Figure 7.5 (as well as Figs. C.2 and C.3 in the Appendix) shows the result of this analysis. Firstly, we found that there is a strong causal relation between species abundances, confirming that species interactions are important drivers of the dynamics

of this system (Fig. C.2). Secondly, we found a weakly statistically significant causal effect of sea temperature on the individual species abundances (apart from the abundance of algae, Fig. C.3). Instead, we found that sea temperature has a significant causal effect on the structural stability of the system (Fig. 7.5 Panel a). That is, we found that sea temperature drives the dynamics of a collective property of the community. Note that previous work (Benincà et al., 2015) has shown that temperature is a driver of the dynamics by including it as a covariate in the analysis. Interestingly, here we found the same results without this explicit added constraint (i.e., temperature is not included in the estimation of the Jacobian coefficients). Finally, we found no causal relation between mean wave height and structural stability (Fig. 7.5 Panel b) hence the correlation observed in Figure 7.4 panel d and f are spurious.

7.4 Conclusion

The increasing amount of available data in natural sciences is providing the scientific community with a unique opportunity to gain new insights into the complex dynamics of ecological communities. Nonparametric approaches provide a framework to leverage these new opportunities by overcoming the inherent limitations of generalizability on empirical datasets of equilibrium or model-based approaches (Perretti et al., 2013b). In this context, the main motivation of this chapter and the thesis as a whole, has been to derive a generalizable, data-driven, nonparametric methodology to monitor the tolerance of ecological communities to environmental changes.

We have demonstrated the generalizability of our proposed methodology by showing that it can be successfully applied on several synthetic time-series data. Then, we have illustrated our approach on a long-term ecological study of a community that exhibits dynamics at the edge of chaos (Benincà et al., 2015). An important result derived from this analysis is that we have been able to give evidence to the hypothesis established in previous work (Benincà et al., 2015) that sea temperature is a causal driver of the dynamics of this community. Importantly, we have obtained

these results without including temperature as a covariate in our analysis and without formulating any parametric model to explain the data.

Another important application of our approach could be the possibility of using our derived measure of structural stability in order to estimate the local predictability of nonlinear ecological time series in changing environments. This is an application that may be of practical relevance for practitioners and we will explore briefly in the next chapter.

Overall we have found that, despite relying on minimal assumptions, our approach provides results on empirical data that match with intuition. That is, the structural stability of ecological communities changes as function of environmental conditions. Given the generalizability of our approach to nonlinear systems, we foresee important applications on many different empirical studies. Specifically, we believe that this study can set out the foundations for a data-driven, model-free, research to monitor and anticipate the response of ecological communities to the pressing increasing rate of environmental changes.

CHAPTER 8

STRUCTURAL STABILITY AND PREDICTABILITY OF NONLINEAR TIME SERIES IN CHANGING ENVIRONMENTS

In this brief chapter we discuss an application of the nonparametric structural stability framework developed in the previous chapter. Specifically, we show that a nonparametric structural stability analysis can be used to assess the predictability of empirical time series in changing environments. We foresee that this result can have important practical applications, particularly in ecosystem management.

8.1 Application of nonparametric structural stability analysis for predictive studies

Short-term forecasts of biological time series are important for risk-assessment studies and to inform sustainable decision making in the development of conservation policies. In the past, the main limitation that prevented the achievement of consistently accurate forecast of biological systems was the lack of an exact knowledge of their governing equations and the uncertainty associated with their parameters (Lorenz,

1963; Brunton et al., 2016). Then, in the early 1980's, the pioneering work of Takens (Takens, 1981) showed that important properties of non-equilibrium systems could be reconstructed solely based on empirical observations, allowing later on the development of novel model-free methods for time-series forecasting (Sugihara, 1994; Kantz & Schreiber, 2004). Since Takens original work, model-free prediction skills have advanced significantly thanks to a combination of Takens's embedding theorem and deep-learning algorithms (Jaeger & Haas, 2004; Brunton et al., 2016; Pathak et al., 2018). However, the possibility to predict non-equilibrium biological dynamics in empirical settings has remained fundamentally limited by the fact that, in changing environments, model parameters are in constant change (e.g., species competition within biological communities increases in periods of food scarcity (Deyle et al., 2016b)) and these changes can have significant effects on the system's dynamics (Serquina et al., 2008).

Unfortunately, the effects of environmental variability on a system's trajectory cannot be easily predicted solely based on past observations. This is simply because predictive models built on properties learned from already observed data cannot generalize well on new data generated by unseen processes. Put it differently, minimizing a cost function over training data generated by a given model does not ensure that the cost is also minimized over data generated by a different model (Bickel et al., 2007). This is a fundamental limitation arising in the applicability of modern statistical learning theories for the predictability of non-equilibrium dynamics in changing environments—where model parameters are in constant change.

A possible way to estimate the effects of environmental variability on a system's dynamics is using nonparametric scenario exploration (Deyle et al., 2016a): an approach developed to assess the effect of a small change in a physical driver on empirically reconstructed dynamical systems. However, in this approach, the environmental drivers need to be explicitly included in the reconstructed dynamics and, therefore, its applicability may be limited when the drivers of the dynamics are only partially or not observed. To overcome these limitations, in this short chapter we propose to assess the predictability of a nonlinear time series in changing environ-

ments more fundamentally, by estimating, nonparametrically, the local structural stability of the data-generating process.

8.1.1 Structural stability and predictability in changing environments

In chapter 7 we have shown that the VCR is a reliable probabilistic measure of structural stability. Now we want to show that, in changing environments, structurally stable states are the easiest to predict. To illustrate the relation between structural stability and predictability we run the following numerical experiment: we considered again the 5-species chaotic population dynamics model (Eq. (7.2)). We integrated the dynamics over 10^5 time points and we sampled equally spaced points along the trajectories (one every 150 points). Then, we computed the VCR along the attractor numerically and we sampled a pair of points in the lower and upper 15th percentile of this measure. At this point, we perturbed a random number of parameters with Gaussian random variables at zero mean and with standard deviation proportional to the value of the parameters themselves. Finally, we integrated again the dynamics with the new parameters, taking as initial condition the two perturbed points (i.e., lower and upper percentile). Note that we applied the very same perturbation to each pair of points. This provided two unperturbed trajectories (i.e., before the two perturbation points) over which we trained a Long Short-Term Memory (LSTM) artificial neural network (Hochreiter & Schmidhuber, 1997; Gers et al., 1999) and we then made forecasts over the two perturbed trajectories (see Figure D.1 for a comparison of the performance of the LSTM and the regularized S-map derived in chapter 5). Training was performed over the 500 data points preceding the perturbation time, i.e., $t_{\text{training}} \in [t_p - 500, \dots, t_p]$ where t_p is the time of perturbation. Forecasts were performed up to 20 time steps in the future (i.e., we were interested in short-term forecasts given that long-term predictability of chaotic systems in changing environments is practically unfeasible). Finally, we computed the RMSE at different forecast horizons in the two test sets, i.e., one after the perturbation at the structurally stable points and the second one after the perturbation at the

structurally unstable points. We repeated this numerical experiments 10^3 times at different levels of perturbations.

Figure 8.1 top panels confirms that the VCR is an effective estimator of short-term predictability of a dynamical system. Notably, this is true also in the lack of external perturbations (Figure 8.1) indicating that the VCR capture important information about the intrinsic predictability of nonlinear dynamical systems. Now we want to analyze in further depth the effects of changing environments on the predictability of structurally stable and structurally unstable states. Therefore, to further explore the difference between structurally stable and structurally unstable points we plot the complement of the cumulative distribution function of the random variable $Y = \frac{\epsilon_{\text{large}} \nu}{\epsilon_{\text{small}} \nu}$ and $X = \frac{\epsilon_{\text{small}} \nu}{\epsilon_{\text{large}} \nu}$ (see Figure 8.1 bottom panels. The Figure shows that $\forall z > 0$, $\Pr(Y > z) \gg \Pr(X > z)$ and the difference between the two probability increases as a function of z . This reveals that the impact of perturbations occurring at points with large VCR is much larger than the impact of perturbations taking place at states with low VCR. This results has important practical applications because it illustrate the extent of the importance of monitoring the structural stability of the a system if any decision is made based on forecasts of population abundances.

8.1.2 Predictability of empirical populations

Now we want to test whether structural stability (i.e., VCR) can be used to estimated the predictability of the empirical data presented in the previous chapter. Therefore, we run the following simulation: first, we train the regularized S-map (chapter 5) using leave-one-out cross validation over the first 150 data points. Then, from the optimal reconstruction of the Jacobian coefficients, we computed the VCR. Then we used the last point in the training set ($t_f = 150$) to make predictions in the test set. With the new predicted value we fit the new coefficients and we make a new prediction. Finally, we repeated the prediction up to 5 months ahead (i.e. $t_{f,\text{pred}} = 155$). Then we acquire the points that we have just predicted from the true data (i.e. from $t_f = 150$ to $t_{f,\text{pred}} = 155$) and computed the RMSE in the test set. Then we discard

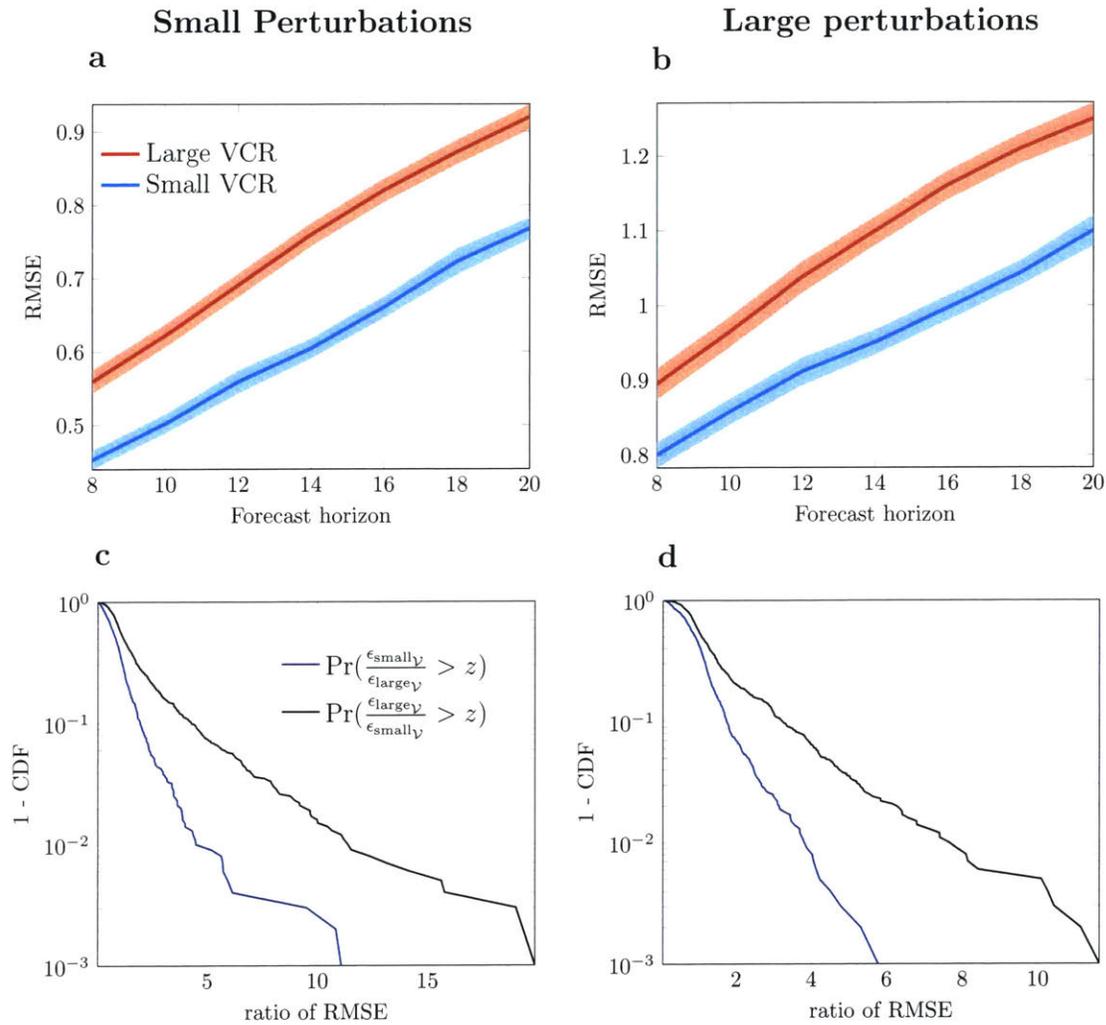


Figure 8.1: Prediction error as a function of forecast horizon at different perturbation points. The upper panels show the root mean square error (RMSE) in the test set at different forecast horizons. The average is taken over 10^3 realizations of the perturbations of the parameters at different points along the model's trajectories and errorbands show the standard error. The level of perturbation (small and large) refers to both the number of perturbed parameters and the magnitude of the noise at each realization. Points at small (large) VCR are sampled from the lower (upper) 15th percentile of the trace of the Jacobian matrix, which is computed numerically at each point on the attractor. The lower panels show that the same parameter perturbation has a much greater impact on the predictability of points with large VCR than points with small VCR. Overall, the figure shows that the VCR is a good probabilistic measure of predictability in changing environments.

the points in the test set, we read a new point ($t_f + 1 = 151$) from the true data, fit the new Jacobian coefficient, compute the new VCR and make predictions again as described above. Finally, we sequentially repeat the process up to the last point of the

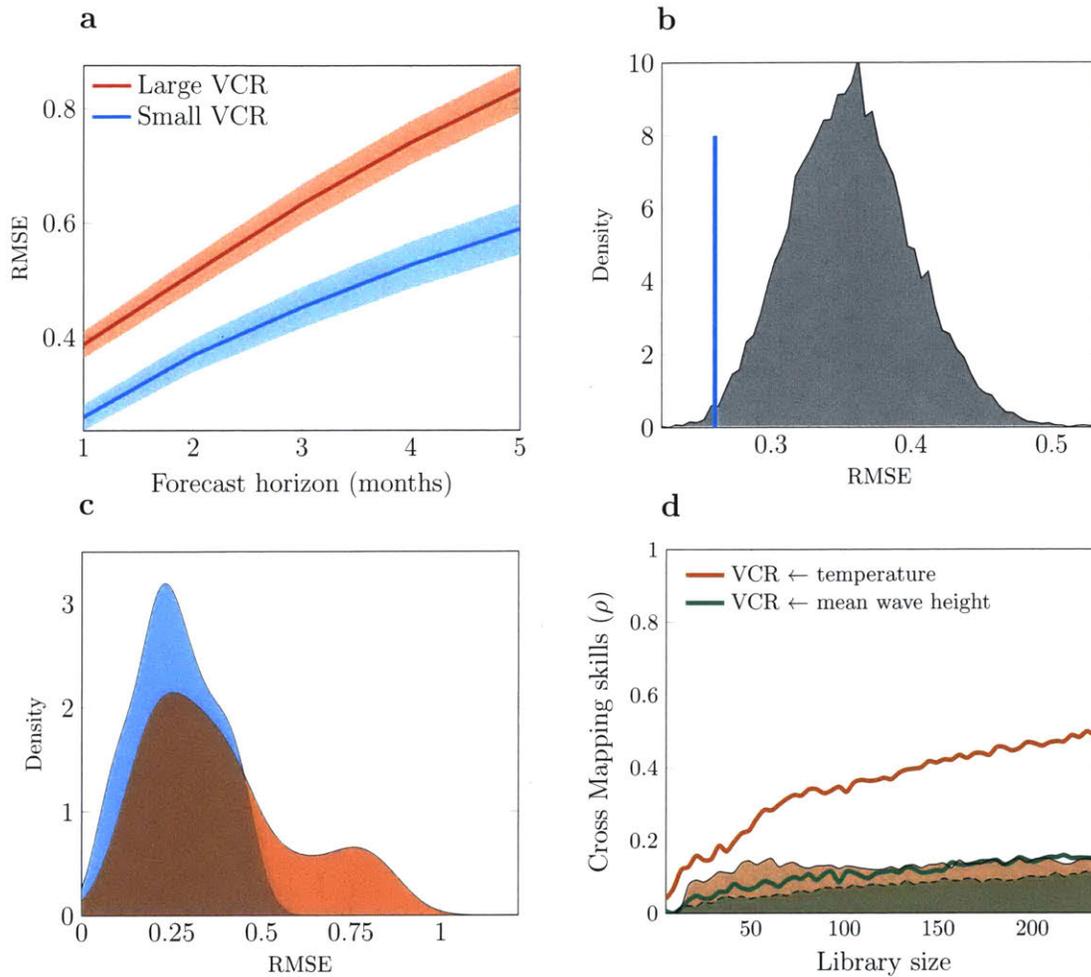


Figure 8.2: | **Validation on an empirical time series** Panel **a** shows the root mean square error (RMSE) of the predictions as function of the forecast horizon (months) for the empirical time series. Accordingly with the theoretical expectations the VCR is a good probabilistic measure for the predictability (error band are standard errors). Panel **b** shows the results of the significance test described in Methods and it illustrate that the VCR is a statistically significant measure of predictability. Panel **c** illustrate the kernel density estimation of the distribution of the RMSE of the predictions at state with small and large VCR. The panel illustrate that the two distribution do not only different in mean but also in shape. Panel **d** shows the results of the causality test between VCR and temperature and mean wave height. The panel clearly illustrate that sea temperature has a much greater effect on the VCR than mean wave height.

time series at $t_{\text{final}} = 247$ (with new re-training of the S-map every 50 data points). All the forecast are completely out-of-sample and the VCR is always computed using training data only.

Figure 8.2 panel **a** is a confirmation of the theoretical results shown in Figure 8.1. In particular, the panel illustrates that predictions after points with small VCR have, on average, a smaller RMSE than prediction after points with a large VCR. Hence, similarly to the results found using synthetic data, the VCR computed empirically from the time series can be used to estimate the local predictability of the data. Predictions were made up to five months ahead and we set a threshold in the VCR at the 15th percentile (results at different threshold are shown in Figure D.3). Because we have one single realization of this dynamics we test for the statistical significance of our results by comparing the observed RMSE in the lower percentile of the VCR with the distribution of possible RMSE's one would have obtained by randomly sampling points from the test set (see Methods). Results are shown in Figure 8.2 panel **b** and Figure D.4 in the Supplementary Information where we show that results are robust to the selection of the threshold of the VCR up to the 30th percentile. Above this threshold the VCR is not a statistically significant measure of predictability. Finally, Figure 8.2 panel **c** and D.5 in the Supplementary Information show the kernel density estimation of the distribution of the RMSE in the test set in the lower and upper percentile of the VCR. The panel clearly illustrate that the two distribution do not only differ in mean but also in shape. Specifically, the distribution of errors at large VCR (red) is significantly more skewed towards larger errors than the distribution of errors in the lower percentile of the VCR (cyan).

In conclusion, panel **d** shows the same causality test performed in the previous chapter (Figure 7.5). For completeness we recall that our findings (Figure 8.2 panel **d**) reveal that sea temperature has a causal influence on the VCR (structural stability and predictability) however we find no significant causal relationship between mean wave height and VCR. This result is interesting because it illustrates that environmental variables (e.g. mean wave height) can play a crucial role for the dynamics (e.g. regulate the mortality rate (Helmuth & Hofmann, 2001)) but have no effects on the overall community predictability (uncertainty) that is associated with the dynamics.

8.2 Discussion

Forecasting with nonlinear time series is a central challenge in science and engineering (Lorenz, 1963; Kantz & Schreiber, 2004; Bradley & Kantz, 2015). The issue is particularly hard to address when true dynamics are unknown and parameters are likely to change in response to environmental changes. In this context, it would be of particular practical relevance to know when predictions can be trusted the most. To address this issue, here we have provided theoretical arguments as well as synthetic and empirical evidences in support to the hypothesis that the volume contraction rate, a local property of the deterministic skeleton of the dynamics that can be robustly inferred from observational data, provides a reliable probabilistic measure of the local predictability of non-equilibrium dynamical systems in changing environments.

Before concluding we would like to discuss some shortcoming of our approach. One evident pitfall of our approach is that, theoretically, there could be cases in which the VCR stays constant throughout time. A clear example is the Lorenz system: a chaotic dynamical system whose predictability is notoriously state dependent (Nese, 1989) but, at the same time, it has a constant VCR. However, while theoretically possible, nonlinear systems with VCR are more the exception than the rule in empirical settings (e.g. already having process noise which is typically proportional to the square root of the state variable would generate state-dependent VCR's). A problem of more practical relevance is the accuracy of the reconstruction of the VCR from empirical data. In the example provided in this and the previous chapter we had detailed information about the system under investigation. Specifically, all the relevant variables were observed allowing us to use a fully multivariate embedding from which we now we can reconstruct the VCR accurately using the regularized S-map (see chapter 5). However, there are instances in which not all variables are available from experiments or observational studies and therefore the dynamics need to be reconstructed from the available observations. In this cases, results are sensitive to the details of the attractor reconstruction, i.e., the selection of the optimal time lag and embedding dimension. This is a delicate issue that could be addressed prag-

matically, by choosing, if it exists, the state space reconstruction that provides the optimal balance between good predictions and separation of points based on their predictability. On the other hand, however, sensitivity to time lags and embedding dimensions could be a reason to doubt about the practical relevance of the approach all together, i.e. the trace of the Jacobian cannot be associated to the VCR if the reconstructed attractor is embedded in a lower or large dimensional space. This is a critical issue that needs to be address in details when one attempts to use our method on empirical data. Overall, we believe that the question discussed in this chapter, i.e., how to asses the reliability of time-series forecasts in continuously changing environments, is of primary importance and more research is needed to better address this issue.

CHAPTER 9

CONCLUSION

In this chapter I summarize the results and main findings of the previous chapters. Then, I discuss important future directions of research that can be pursued building upon the results and methodology developed in this thesis. Furthermore, because this chapter focus on my personal views and beliefs I will switch the language of the text from a collective 'we' to an individual 'I'.

9.1 Conclusion and future research directions

Many aspects of human societies, from the sustenance of national economies to the control of population health, strongly depends on the dynamics of biological populations within a given environment. Therefore, understanding and predicting the effects of changing environments on the dynamics of biological populations evolving in a continuously changing world is, nowadays, one of the most important challenges in the life sciences.

In this context, this thesis brought two important contributions:

- Firstly, I have demonstrated that it is possible to build a parametric approach to study the effect of changing environments on the dynamics of biological

populations that relies on minimal assumptions on the data generating processes (chapter 2). Importantly, I have then used this result to study the effect that species interactions (chapter 3) and their dynamics (chapter 4) have on community coexistence under fast environmental changes.

- Secondly, I have laid the foundation of a nonparametric framework to estimate the tolerance of biological populations to environmental variability. Specifically, I have improved/developed computational methods (chapter 5 and 6) to infer biotic interactions and their uncertainty from noisy nonlinear time series data. Then I have used the inferred biotic interactions to construct a framework to study the possible effects of environmental perturbations on the nonlinear dynamics of empirical populations.

The explosion in data availability, computational power and statistical learning theories is opening the doors to a new era for population dynamics. I believe that, in this new era, a full transition from parametric to nonparametric theory construction is not only possible but desirable. I believe that a fully data-driven nonparametric scientific theory, if realizable, is desirable because it would be free from the necessary simplifications adopted to circumnavigate the complexity of the mechanisms that drive the dynamics of biological populations. In the light of this belief in the next section I discuss some possible steps toward this direction.

9.1.1 What's next?

In this last section I briefly discuss what I believe are the possible steps to take towards the foundation of nonparametric population dynamics theories. Most of the following discussions are speculative and based on methodological advances in other nonlinear sciences, such as physics and economics.

9.1.1.1 Artificial neural networks for model discovery

At the time of writing this thesis we live a revolution in what is known as deep learning. Broadly speaking, deep learning is a field at the intersection of statistics and computer science which uses nonparametric statistical models, known as neural networks, to learn complex structure in large datasets (Dechter, 1986; Aizenberg et al., 2000). Typically neural networks are considered as black boxes that are of little use for the development of scientific theories. Yet, in recent years there has been growing interest in using neural networks for discovering the structure, or at least the most important features, of the data-generating processes. Examples of applications of deep learning for model discovery can be found in (Schmidt & Lipson, 2009; Iten et al., 2018; Bakker et al., 2000). In particular, in (Bakker et al., 2000) the authors show how to use neural networks to obtain reliable reconstructions of dynamic attractors from empirical data. Interestingly, in their work the authors show that attractors reconstructed with deep neural networks provide ground not only for the possibility of short term predictions but also for the opportunity to generate time series with similar chaotic characteristics as the measured data (Bakker et al., 2000) so to open up the possibility of nonparametric long-term predictions. This is an important advantage since an important critics against nonparametric approaches is that their applicability to long-term studies is limited.

At the moment, the main challenge for the applicability of deep learning for the analysis of ecological systems, is data availability. Nevertheless, this problem can be relatively easily overcome in certain areas of ecology, such as for example microbial ecology, where the development of new technologies for data extraction and the fast time scales over which ecological and evolutionary process take place, create the perfect conditions to develop and test nonparametric theories. I therefore foresee important applications of artificial neural network in nonparametric microbial population dynamics studies.

9.1.1.2 Causal inference

An important conceptual limitation for the applicability of deep learning for the analysis of population dynamics data, is that neural networks are statistical models constructed based on inferred correlations between features within a dataset. However, in biology and nonlinear sciences in general, many of the observed correlations are spurious, meaning that they do not reflect true causation. This is not an issue for the development of predictive methods as correlations, even if spurious, can be leveraged to build predictive models. On the other hand the presence of spurious correlations is an important problem for the development of scientific theories.

The development of nonparametric causal inference algorithms for the analysis of nonlinear time series is therefore an essential ingredient for the advancement of model-free population dynamics theories. Example of nonparametric causal inference algorithms can be found in (Sugihara et al., 2012) (an approach that was also discussed in chapter 7 of this thesis), in (Krakovská & Hanzely, 2016) and in (Tank et al., 2018) in which the authors show how deep learning can be used to extend the standard concept of Granger causality for the analysis of nonlinear data. A interesting review of the state of the art of causal inference from nonlinear time series data can be found in (Runge, 2018) in which limitations and opportunities are extensively discussed. Overall, I believe that nonparametric causal inference is an exciting field of research that surely deserved more attention from the scientific community within and beyond the realm of theoretical population biology. In this context I propose that there are two important challenges with causal inference that, to our opinion, deserve further exploration:

- **Subsampling.** An important issue in causal inference is that if the causal mechanisms and the data-sampling process take place on two separate time scales than the true causal links are not identifiable (Runge, 2018). There is growing effort, among computer scientists, in developing new algorithms to identify true causal links in the presence of subsampling (Gong et al., 2015; Hyttinen et al., 2016) but the problem still remains an open challenge particularly when multiple time-scales coexists and data are scarce. In biology, I

believe that a starting point to address this issue should come, when possible, from the design of experiments that carefully address the interplay between data collection and the fundamental time-scales of the observed dynamics.

- **Temporal variability of causal structures.** Generally in causal inference one assumes stationarity in the underlying causal relations among variables, i.e. if there exists a causal link $X_t \rightarrow Y_t$ at time t than this link exists for any $\bar{t} < t^*$. Yet, in highly environment dependent systems, such as biological populations, this assumption may not hold as environmental changes or evolutionary processes can significantly change biological mechanisms. I am not aware of any causal inference algorithm that can deal with time-varying causality. This problem has also been discussed in (Eichler, 2013) and I believe is of pivotal importance.

9.2 Concluding remarks

Overall, the increasing amount of available time-series data collected in biological studies is paving the way for moving theoretical population dynamics from being a science build on assumptions and model simplifications towards a data-driven, model-free, science. As discussed thoroughly this thesis, because populations dynamics are the product of an interplay between nonlinear deterministic and stochastic dynamics, the integration of methods of nonlinear time series analysis (based on dynamical system theory) with machine learning algorithms is an important step forward towards this goal. In this context, I believe and hope that the work of this thesis can serve as basis for further developments in this direction.

*Notice that this definition of stationarity differ from the standard assumption of stationarity in mean and variance of a process.

Appendix

APPENDIX A

SUPPLEMENTARY FIGURES FOR CHAPTER 4

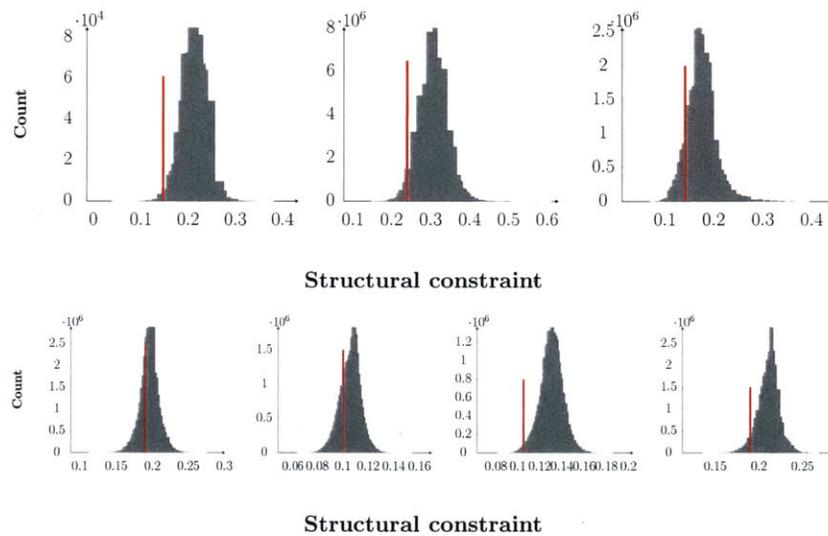


Figure A.1: [Supplementary Figure] Equivalent to Figure 7 in main text, but here groups of pollinators are allowed to be of different dimension. Quantile from top left to bottom right: 0.5%, 1%, 9%, 38%, 18%, 2%, 8%.

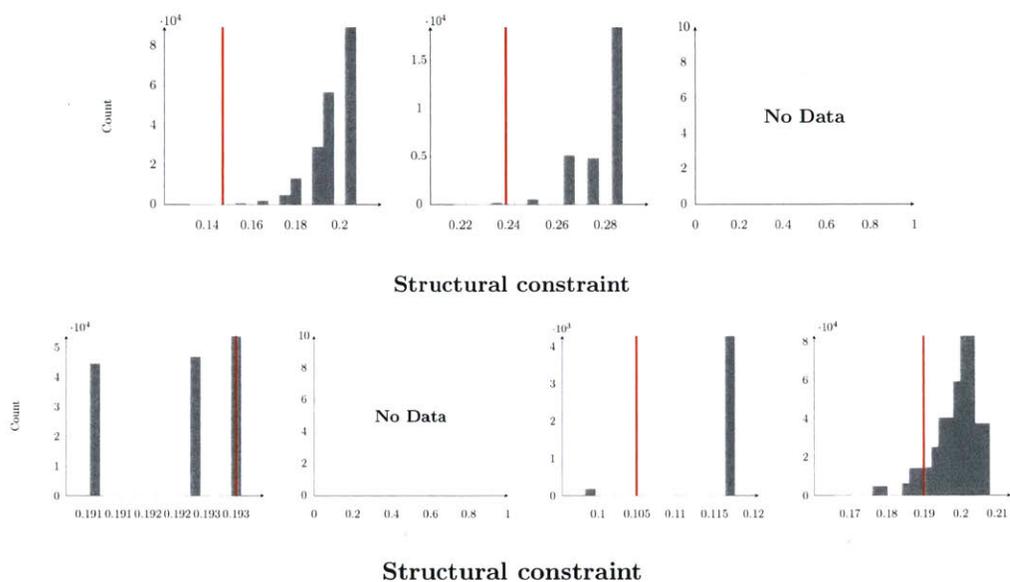


Figure A.2: [Supplementary Figure] Equivalent to Figure 7 in main text, but here groups of pollinators have the same connectance as the observed persistent group. Panels with *No Data* represent communities where we could not find groups with same connectance and different structural constraints within the specified range of species variation. Quantile from top left to bottom right: 0.1%, 0.05%, NaN, 100%, NaN, 4%, 20%.

APPENDIX B

SUPPLEMENTARY FIGURES FOR CHAPTER 6

Model averaging

0.82	0.43	0.59	NA
NA	0.6	0.67	0.06
0.89	NA	0.4	0.33
0.2	0.15	-0.37	0.73

Cross-validation

0.81	0.42	0.56	NA
NA	0.41	0.43	-0.06
0.88	NA	0.41	0.3
0.12	0.04	-0.39	0.7

Figure B.1: **Inference of Jacobian coefficients and the importance of the model-averaging process.** The figure shows the correlation matrix (Pearson's correlation coefficient) between inferred and analytical coefficients of the Jacobian for the chaotic LV model with 4 species discussed in the main text. The left panel shows the correlation matrix between inferred and analytical Jacobians. The inferred Jacobian is constructed with the model-averaging process (ensemble method) proposed in this work. The right panel shows the correlation matrix when the best Jacobian is selected with cross-validation. NA indicates constant coefficients for which the correlation is not defined. The values shown in both panels are average correlations over the 20 random samples of the time series. Importantly, this figure shows that using model averaging, the overall quality of the inferred Jacobian coefficients is higher than the one generated by the best single model.

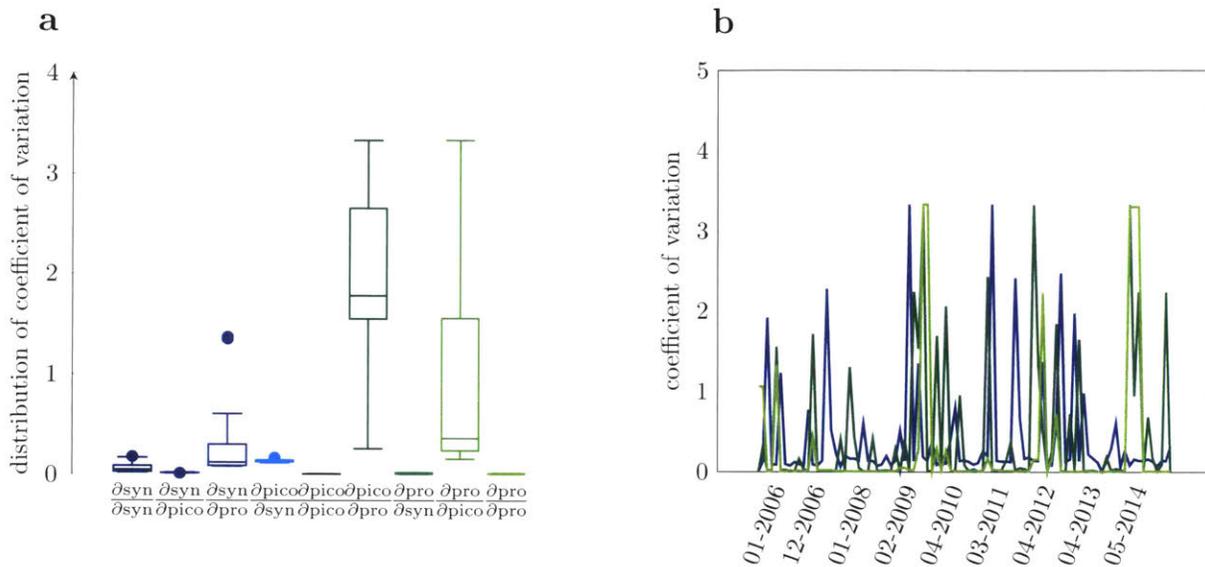


Figure B.2: **Uncertainty analysis on the Hawaii Ocean time series including the abundance of *Synenochoccus*.** The figure shows the pattern of uncertainty on the Jacobian coefficients of the Hawaii Ocean time series when including the abundance of *Synenochoccus* in the regression. Panel a shows that the uncertainty on $(\frac{\partial \text{pico}}{\partial \text{pico}})$, $(\frac{\partial \text{pico}}{\partial \text{pro}})$, $(\frac{\partial \text{pro}}{\partial \text{pico}})$ is left qualitatively unchanged while $(\frac{\partial \text{pro}}{\partial \text{pico}})$ shows a different pattern (compare with Figure 6.5 in the main text). Panel b shows the temporal changes of uncertainty on the three least consistently estimated interactions. Note that we did not include the abundance of *Synenochoccus* in the analysis discussed in the main text because the quality of the fit in the training and test sets is significantly reduced. In fact, by including *Synenochoccus* in the analysis, we have $R^2_{\text{training}} = 0.55$ and $\text{RMSE}_{\text{test}} = 0.3$, $\text{RMSE}_{\text{naive}} = 0.65$ (to be compared against the values in the main text).

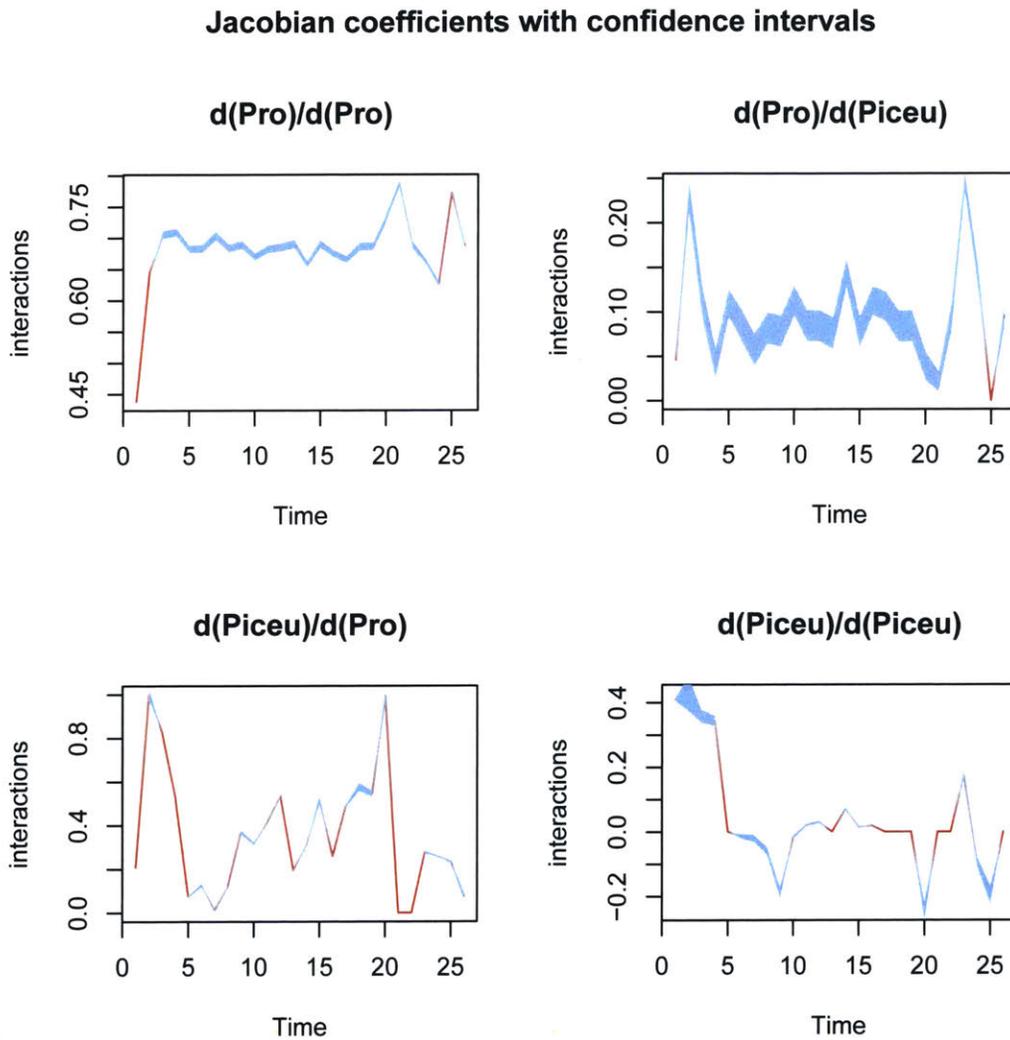


Figure B.3: **Jacobian coefficients of the Bermuda Atlantic time series with confidence intervals.** Shaded area correspond to the 95% confidence intervals computed as discussed in the Method section using the weighted standard error on the coefficient in the optimum ensemble. x-axis are the time points

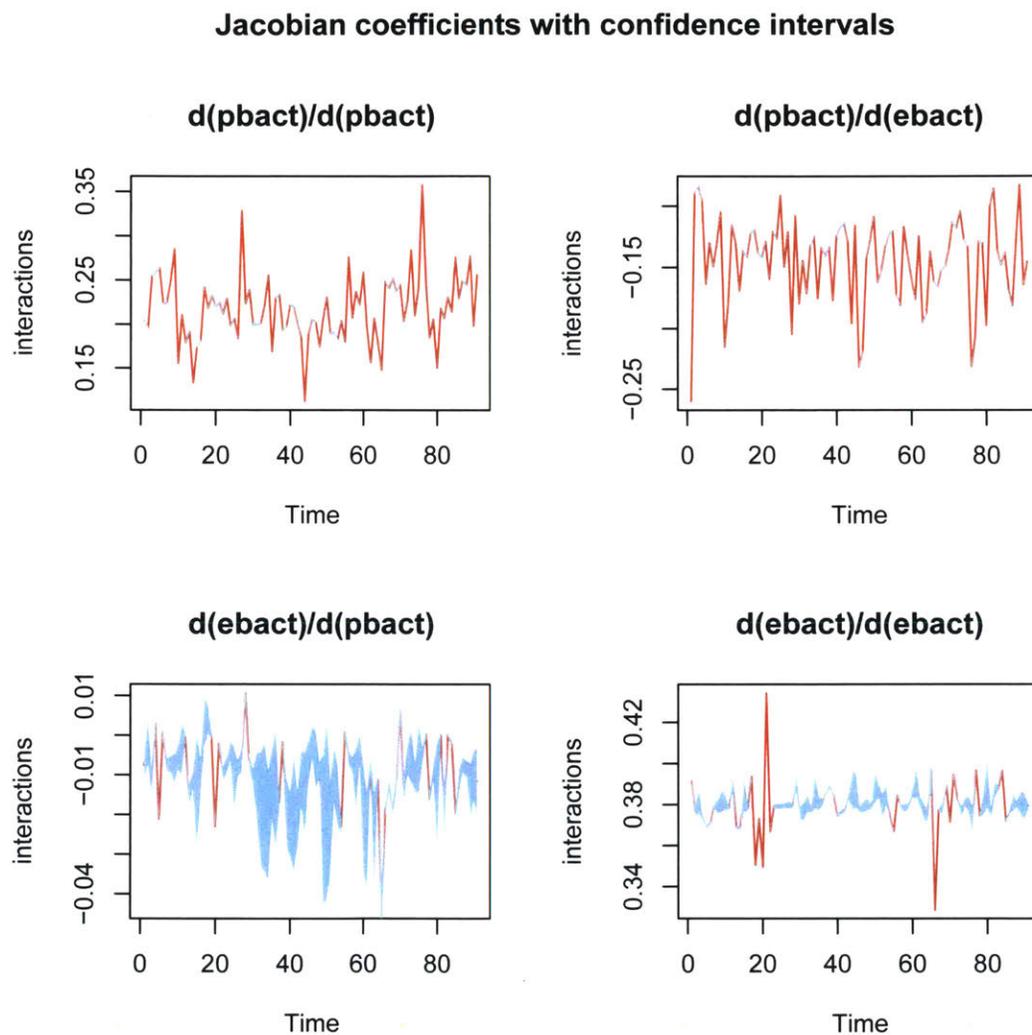


Figure B.4: **Jacobian coefficients of the Hawaii ocean time series with confidence intervals.** Shaded area correspond to the 95% confidence intervals computed as discussed in the Method section using the weighted standard error on the coefficient in the optimum ensemble. x-axis are the time points

APPENDIX C

SUPPLEMENTARY FIGURES FOR CHAPTER 7

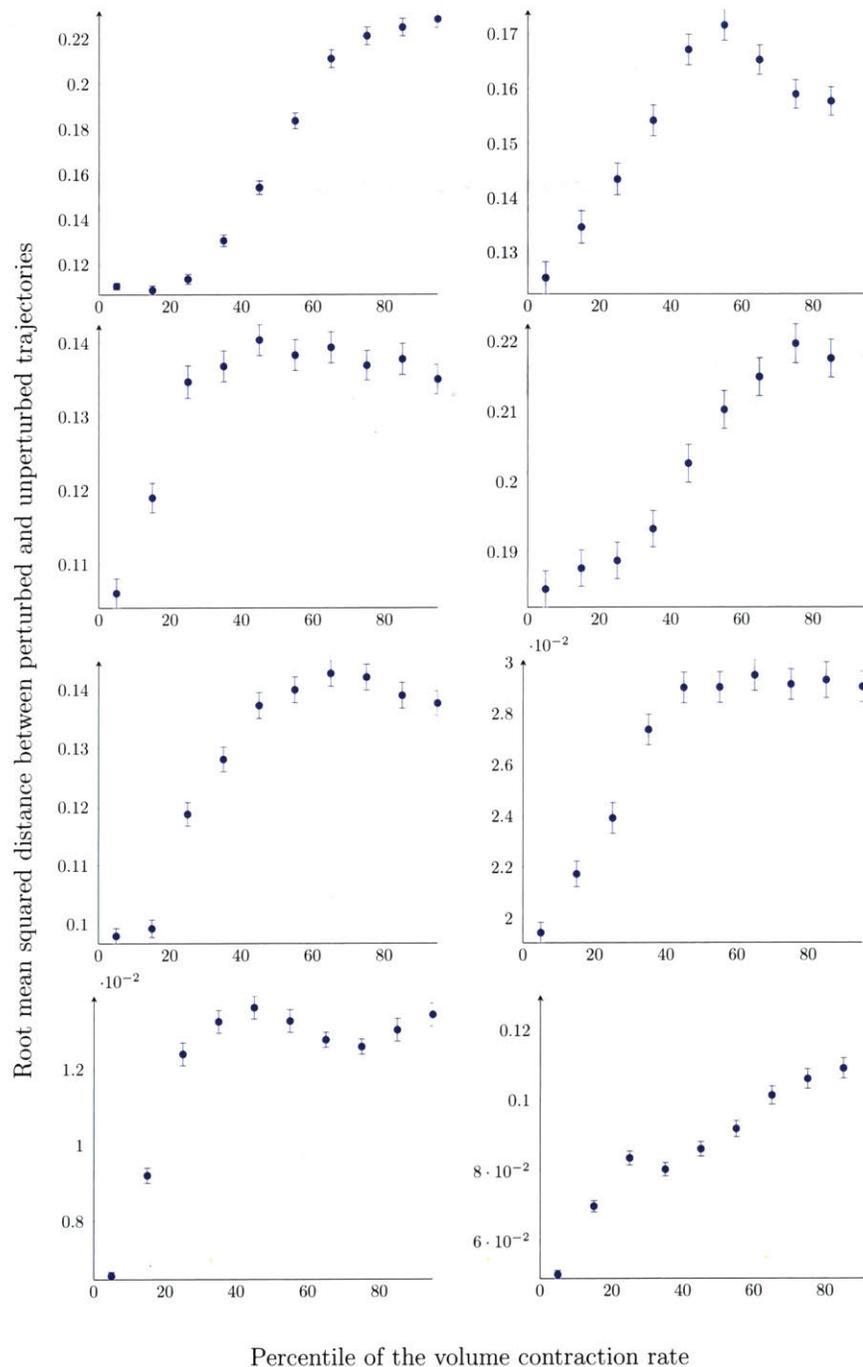


Figure C.1: **Root mean square distance between perturbed and unperturbed trajectories as a function of the percentile of the volume contraction rate for each of the 8 chaotic models.** Each panel shows a similar analysis as the one discussed in section 7.3. Different from the main analysis, rather than taking the ratio of root mean squared distances between trajectories perturbed at structurally stable and unstable points, here we just look at the root mean square distance as a function of the percentile of the volume contraction rate included in the analysis. Each panel (from top left to bottom right) represents one of the 8 models used in the main text. The figure shows that the 15th lower percentile includes structurally stable states, which corresponds to the analysis shown in Figure 7.3. Overall, the figure also shows that there is a positive correlation between effect of a perturbation of the parameters on the dynamics and the volume contraction rate.

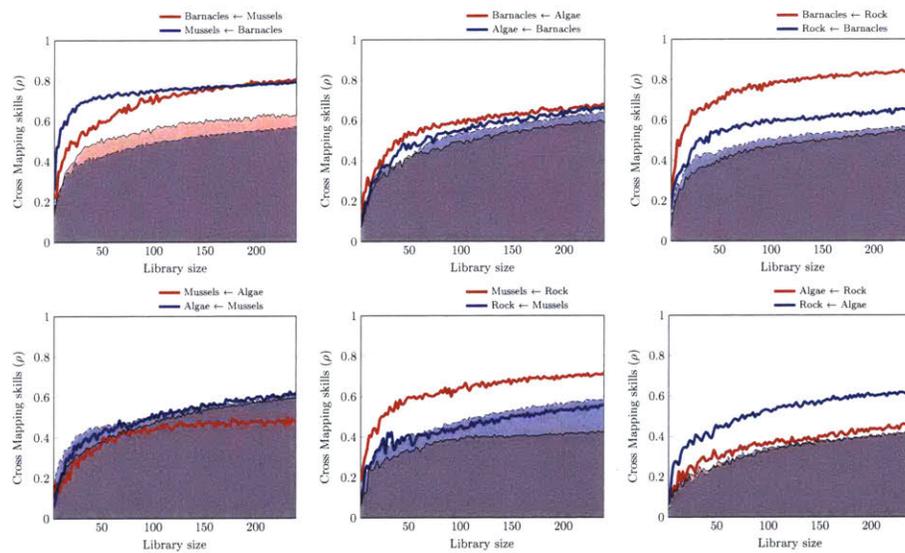


Figure C.2: **Convergent cross mapping test between species abundances.** The figure clearly illustrates that the abundances of species are driven by interactions with the other species in the system.

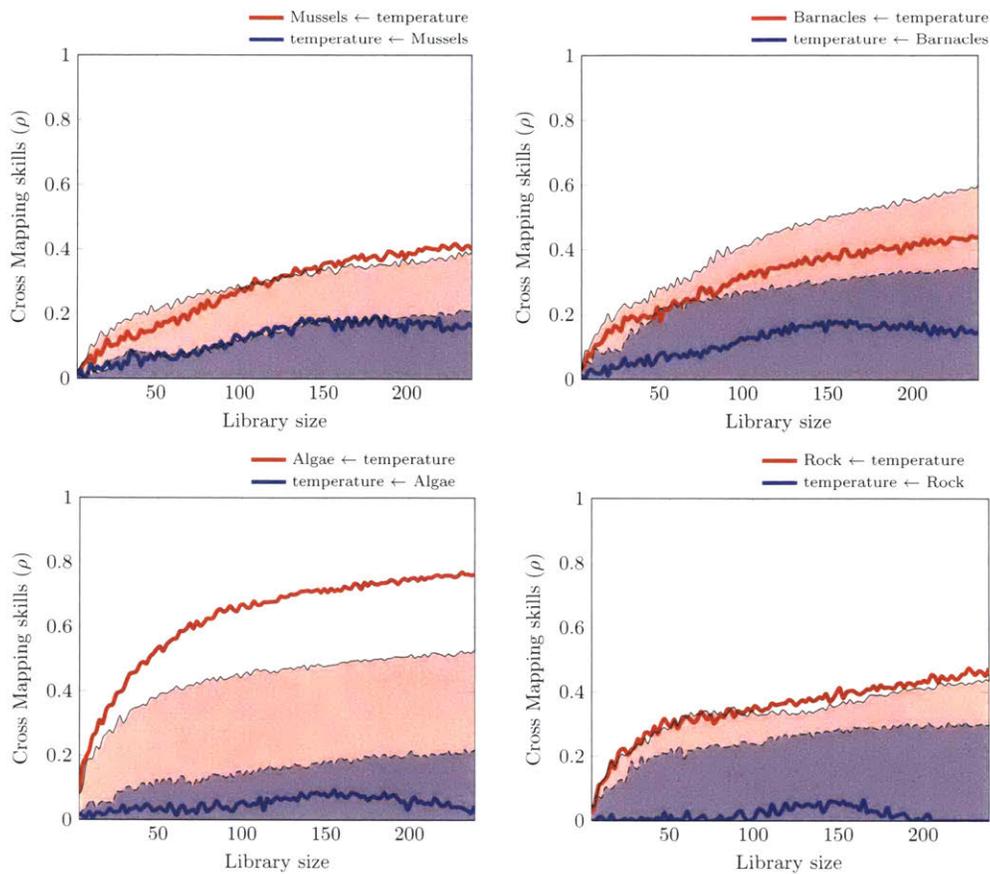


Figure C.3: **Convergent cross mapping test between temperature and species abundances.** The figure shows that the only statistically significant causal link is the one between sea temperature and the abundance of Algae.

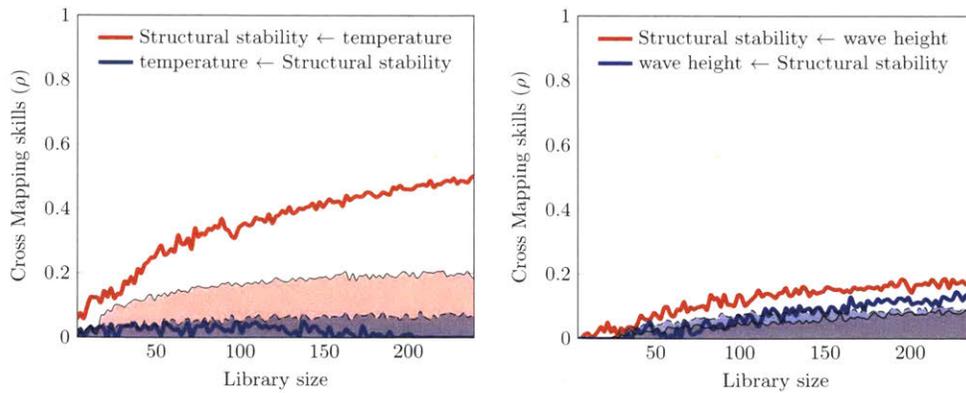


Figure C.4: **Convergent cross mapping test under a seasonal null model.** The figure shows that the statistical significance of the causal link between temperature and structural stability is preserved also under a null model that takes into account possible seasonal bias.

APPENDIX D

SUPPLEMENTARY FIGURES FOR CHAPTER 8

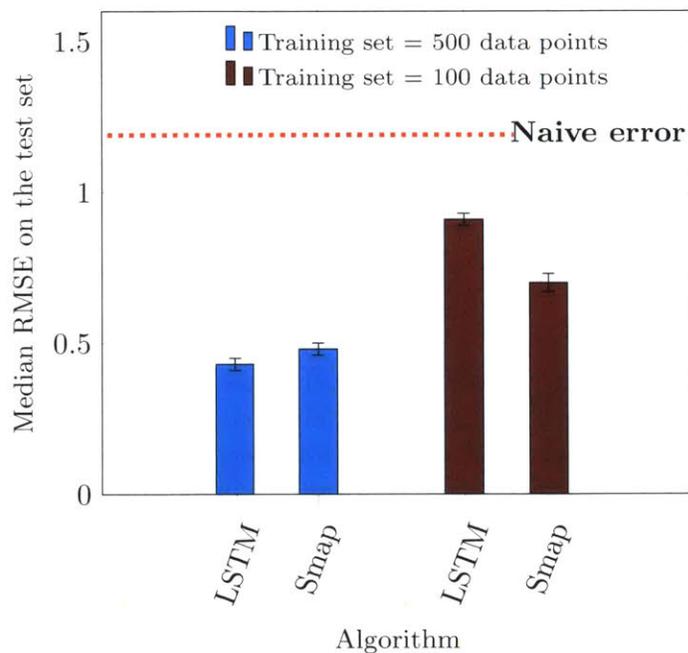


Figure D.1: **Comparison of the performance of the S-map and the LSTM artificial neural network as a function of the length of the training set.** Performance is measured as the median of the RMSE in the test set over 200 predictions of the population dynamics model used in the main text. The error bars are computed with nonparametric bootstrapping. The figure illustrates that for short training sets, the S-map outperforms the LSTM artificial neural network. Thus, we have used the S-map to forecast the empirical time series. Notice that the median naive error for this time series is ~ 1.2 ; hence, both algorithms perform significantly better than naive forecasts even when trained over as little as 100 data points.

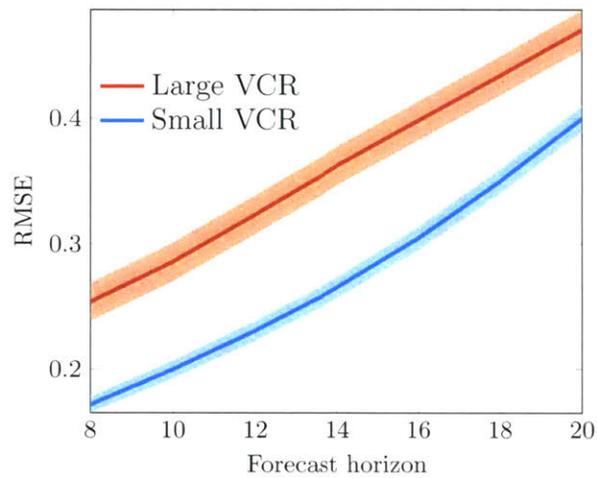


Figure D.2: **Forecast with unperturbed chaotic time series.** This Figure is produced similarly to Figure 8.1 upper panels in the main text. However, here we simply forecast at point with small and large VCR without perturbing the system. The Figure shows that the VCR is an effective estimator of predictability even in the lack of perturbations.

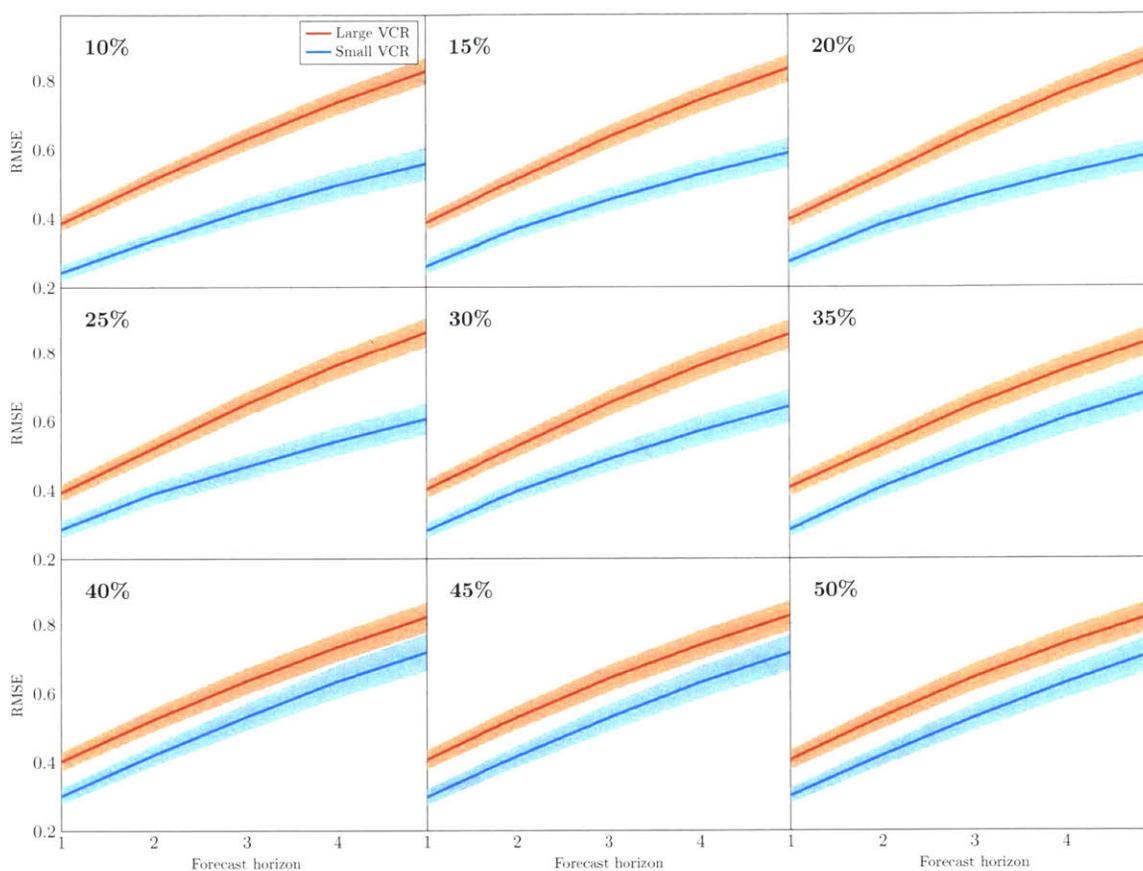


Figure D.3: | **Validation on the biological time series at different thresholds.** Similarly to Figure 8.2 in the main text, the figure shows the root mean square error (RMSE) of the predictions as function of the forecast horizon (months) for the biological time series. Here we also show that the results are consistent as we change the threshold for what we consider small and large VCR. Specifically, we used as criterion of separation the n th and $(1 - n)$ th percentile with $n = [10, 15, \dots, 50]$ (from top to bottom). Overall results are robust to the choice of the separation criterion up to the lower 30th percentile. This result is made more quantitative in Figure D.4 where we show the statistical significance of the VCR as estimator of predictability

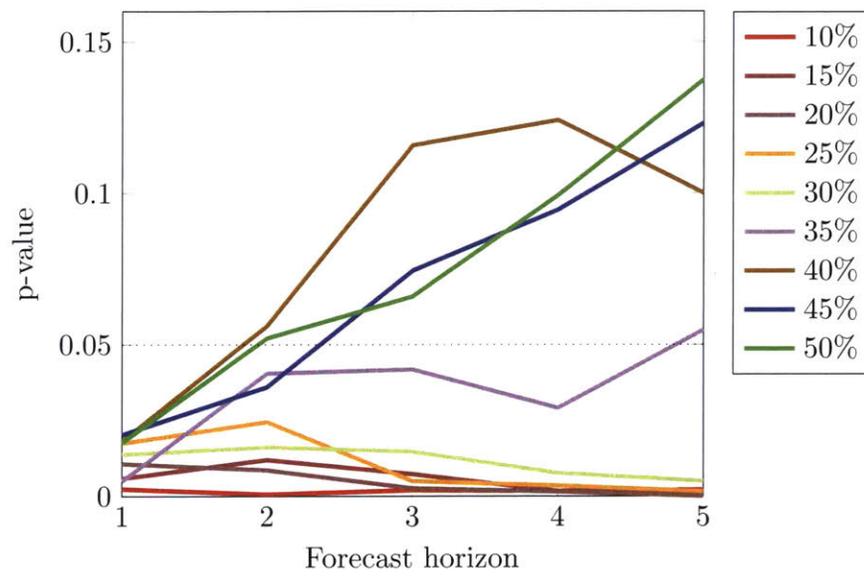


Figure D.4: | Statistical significance of the VCR as estimator of predictability

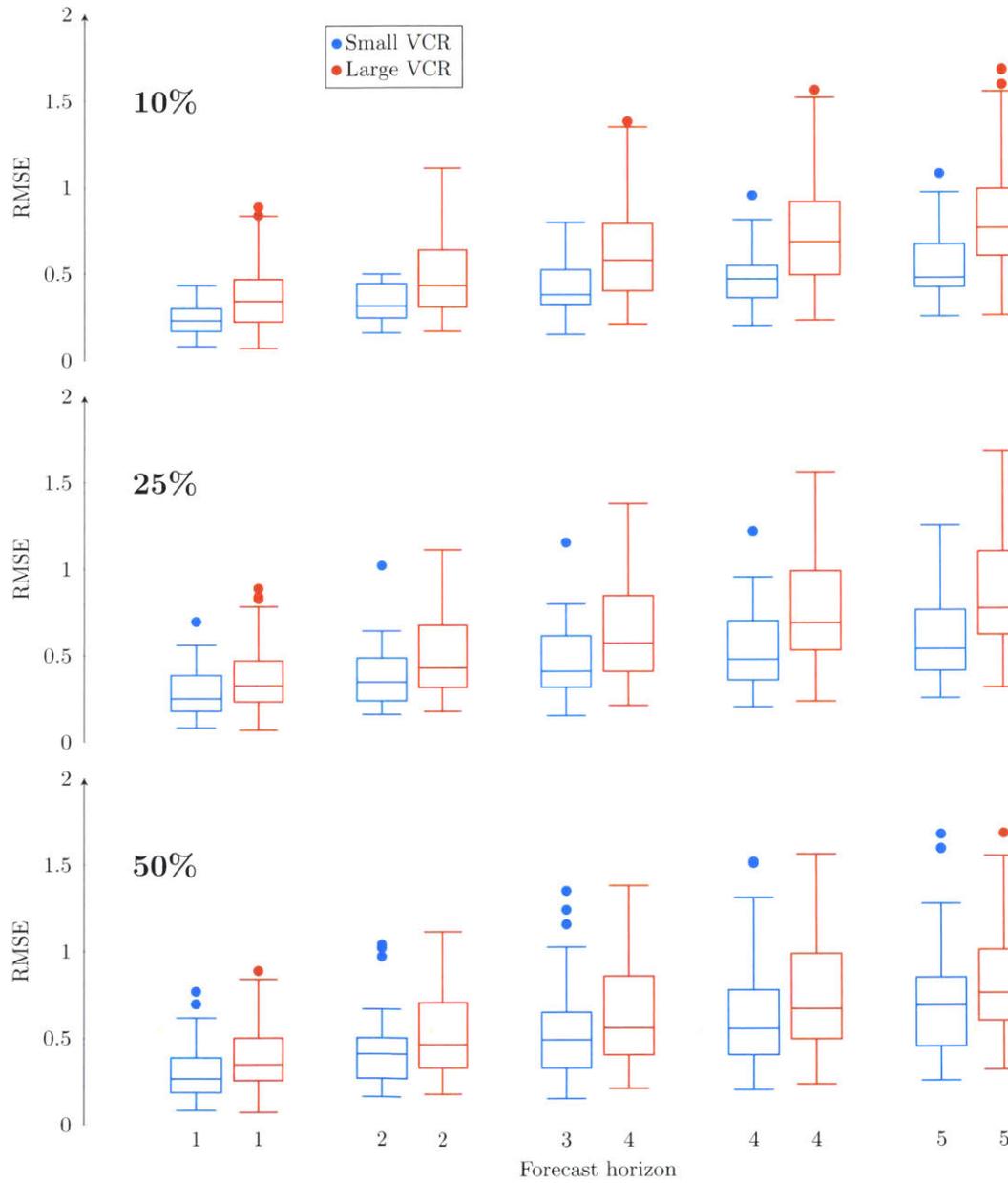


Figure D.5: | Difference in distribution of error for predictions at small and large VCR.

Bibliography

- Abrudan, M. I., Smakman, F., Grimbergen, A. J., Westhoff, S., Miller, E. L., van Wezel, G. P., & Rozen, D. E. (2015). Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proceedings of the National Academy of Sciences*, *112*(35), 11054–11059.
- Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H.-T. (2012). *Learning From Data*. AMLBook.
- Aizenberg, I., Aizenberg, N., & Vandewalle, J. (2000). *Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications*. Springer US.
- Albert, R., Jeong, H., & Barabási, A. L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*, 378–482.
- Alexander, J. M., Diez, J. M., & Levine, J. M. (2015). Novel competitors shape species, responses to climate change. *Nature*, *525*, 515–518.
- Allen, P. M. (1976). Evolution, population dynamics, and stability. *Proceedings of the National Academy of Sciences of the United States of America*, *73*(3), 665–668.
- Amarasekare, P., & Coutinho, R. M. (2013). The intrinsic growth rate as a predictor of population viability under climate warming. *J. of Animal Ecology*, *82*, 1240–1253.
- Amin, S. A., Hmelo, L. R., van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., Morales, R. L., Berthiaume, C. T., Parker, M. S., Djunaedi, B., Ingalls, A. E., Parsek, M. R., Moran, M. A., & Armbrust, E. V. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature*, *522*, 98–101.
- Angulo, M. T., Moreno, J. A., Lippner, G., Barabási, A.-L., & Liu, Y.-Y. (2017). Fundamental limitations of network reconstruction from temporal data. *Journal of The Royal Society Interface*, *14*(127).
- Arnoldi, J.-F., & Haegeman, B. (2016). Unifying dynamical and structural stability of equilibria. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *472*(2193).
- Arnoldi, J.-F., Loreau, M., & Haegeman, B. (2016). Resilience, reactivity and variability: A mathematical comparison of ecological stability measures. *Journal of Theoretical Biology*, *389*, 47 – 59.
- Atkins, K., & Travis, J. (2010). Local adaptation and the evolution of species ranges under climate change. *Journal of Theoretical Biology*, *266*(3), 449 – 457.
- Ayala, F. J., Gilpin, M. E., & Ehrenfeld, J. G. (1973). Competition between species: Theoretical models and experimental tests. *Theoretical Population Biology*, *4*(3), 331 – 356.
- Babtie, A. C., Kirk, P., & Stumpf, M. P. H. (2014). Topological sensitivity analysis for systems biology. *Proceedings of the National Academy of Sciences*, *111*(52), 18507.
- Baigent, S., & Hou, Z. (2012). Global stability of interior and boundary fixed points for lotka–volterra systems. *Differential Equations and Dynamical Systems*, *20*(1), 53–66.
- Bakker, R., Schouten, J. C., Giles, C. L., Takens, F. C., & Van Den Bleek, C. M. (2000). Learning chaotic attractors by neural networks. *Neural Comput.*, *12*(10), 2355–2383.
- Barraclough, T. G. (2015). How do species interactions affect evolutionary dynamics across whole communities? *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 25–48.
- Bartlett, M. (1960). *Stochastic population models in ecology and epidemiology*. Methuen's monographs on applied probability and statistics. Methuen.
- Bartlett, T. E. (2017). Network inference and community detection, based on covariance matrices, correlations, and test statistics from arbitrary distributions. *Communications in Statistics - Theory and Methods*, *46*(18), 9150–9165.
- Bascompte, J., & Jordano, P. (2013). *Mutualistic Networks*. Princeton Univ. Press, NJ.
- Bastolla, U., Fortuna, M. A., Pascual-Garcia, A., Ferrera, A., Luque, B., & Bascompte, J. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, *458*(7241), 1018–1020.

- Bastolla, U., Lässig, M., Manrubia, S. C., & Valleriani, A. (2002). *Dynamics and topology of species networks*, chap. 17, (pp. 299–311). Springer, Berlin-Heidelberg.
- Becks, L., Hilker, F. M., Malchow, H., Jürgens, K., & Arndt, H. (2005). Experimental demonstration of chaos in a microbial food web. *Nature*, *435*, 1226–1229.
- Begon, M., Townsend, C., & Harper, J. (2009). *Ecology: From Individuals to Ecosystems*. Wiley-blackwell, Oxford.
- Bell, G., & Gonzalez, A. (2011). Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. *Science*, *332*(6035), 1327.
- Bellman, R., & Aström, K. (1970). On structural identifiability. *Mathematical Biosciences*, *7*(3), 329 – 339.
- Benadi, G., Blüthgen, N., Hovestadt, T., & Poethke, H.-J. (2013). When can plant-pollinator interactions promote plant diversity? *The American Naturalist*, *182*(2), 131–146.
- Benincà, E., Ballantine, B., Ellner, S. P., & Huisman, J. (2015). Species fluctuations sustained by a cyclic succession at the edge of chaos. *Proceedings of the National Academy of Sciences*, *112*(20), 6389.
- Benincà, E., Huisman, J., Heerkloss, R., Jöhnk, K. D., Branco, P., Van Nes, E. H., Scheffer, M., & Ellner, S. P. (2008). Chaos in a long-term experiment with a plankton community. *Nature*, *451*, 822–825.
- Bickel, S., Brückner, M., & Scheffer, T. (2007). Discriminative learning for differing training and test distributions. In *Proceedings of the 24th International Conference on Machine Learning*, (pp. 81–88). New York, NY, USA: ACM.
- Bjørnstad, O. N., Fromentin, J.-M., Stenseth, N. C., & Gjøsater, J. (1999). Cycles and trends in cod populations. *Proceedings of the National Academy of Sciences*, *96*(9), 5066–5071.
- Bjørnstad, O. N., & Grenfell, B. T. (2001). Noisy clockwork: Time series analysis of population fluctuations in animals. *Science*, *293*(5530), 638–643.
- Bollenbach, T. (2015). Antimicrobial interactions: mechanisms and implications for drug discovery and resistance evolution. *Current Opinion in Microbiology*, *27*, 1 – 9.
- Bollobás, B. (1998). *Modern Graph Theory*. Springer, New York.
- Bongard, J., & Lipson, H. (2007). Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *104*(24), 9943.
- Borrelle, S. B., Rochman, C. M., Liboiron, M., Bond, A. L., Lusher, A., Bradshaw, H., & Provencher, J. F. (2017). Opinion: Why we need an international agreement on marine plastic pollution. *Proceedings of the National Academy of Sciences*, *114*(38), 9994.
- Bradley, E., & Kantz, H. (2015). Nonlinear time-series analysis revisited. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *25*(9), 097610.
- Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199–231.
- Brown, J. H., Whitham, T. G., Morgan Ernest, S. K., & Gehring, C. A. (2001). Complex species interactions and the dynamics of ecological systems: Long-term experiments. *Science*, *293*(5530), 643.
- Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, *113*(15), 3932–3937.
- Bucci, V., Tzen, B., Li, N., Simmons, M., Tanoue, T., Bogart, E., Deng, L., Yeliseyev, V., Delaney, M. L., Liu, Q., Olle, B., Stein, R. R., Honda, K., Bry, L., & Gerber, G. K. (2016). Mdsine: Microbial dynamical systems inference engine for microbiome time-series analyses. *Genome Biology*, *17*(1), 121.
- Büchi, L., & Vuilleumier, S. (2014). Coexistence of specialist and generalist species is shaped

- by dispersal and environmental factors. *The American Naturalist*, 183(5), 612–624.
- Buckland, S., Newman, K., Thomas, L., & Koesters, N. (2004). State-space models for the dynamics of wild animal populations. *Ecological Modelling*, 171(1), 157 – 175.
- Bunin, G. (2017). Ecological communities with lotka-volterra dynamics. *Physical Review E*, 95(4), 042414.
- Burkov, V., Novikov, D., & Shchepkin, A. (2015). *Control Mechanisms for Ecological-Economic Systems*. Studies in Systems, Decision and Control. Springer International Publishing, Switzerland.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, New York, Inc., 2nd ed.
- Cadotte, M. W., & Tucker, C. M. (2017). Should environmental filtering be abandoned? *Trends in Ecology and Evolution*, 32, 429–437.
- 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.
- Carlson, S. M., Cunningham, C. J., & Westley, P. A. (2014). Evolutionary rescue in a changing world. *Trends in Ecology & Evolution*, 29(9), 521 – 530.
- Casdagli, M. (1989). Nonlinear prediction of chaotic time series. *Physica D: Nonlinear Phenomena*, 35(3), 335 – 356.
- Casdagli, M., Eubank, S., Farmer, J., & Gibson, J. (1991). State space reconstruction in the presence of noise. *Physica D: Nonlinear Phenomena*, 51(1), 52 – 98.
- Case, T. J. (2000). *An Illustrated Guide to Theoretical Ecology*. Oxford University Press, Oxford.
- Cauwenberghe, L. V., & Janssen, C. R. (2014). Microplastics in bivalves cultured for human consumption. *Environmental Pollution*, 193, 65 – 70.
- Cenci, S., Montero-Castaño, A., & Saavedra, S. (2018a). Estimating the effect of the reorganization of interactions on the adaptability of species to changing environments. *Journal of Theoretical Biology*, 437, 115 – 125.
- Cenci, S., & Saavedra, S. (2018a). Structural stability of nonlinear population dynamics. *Phys. Rev. E*, 97, 012401.
- Cenci, S., & Saavedra, S. (2018b). Uncertainty quantification of the effects of biotic interactions on community dynamics from nonlinear time-series data. *Journal of The Royal Society Interface*, 15(147), 20180695.
- Cenci, S., & Saavedra, S. (In press). Nonparametric estimation of the structural stability of non-equilibrium community dynamics. *Nature Ecology and Evolution*.
- Cenci, S., Song, C., & Saavedra, S. (2018b). Rethinking the importance of the structure of ecological networks under an environment-dependent framework. *Ecology and Evolution*, 8(14), 6852–6859.
- Cenci, S., Sugihara, G., & Saavedra, S. (2019). Regularized s-map for inference and forecasting with noisy ecological time series. *Methods in Ecology and Evolution*, <https://doi.org/10.1111/2041-210X.13150>.
- Chamberlain, S. A., Bronstein, J. L., & Rudgers, J. A. (2014). How context dependent are species interactions? *Ecology Letters*, 17, 881–890.
- Chase, E., & Harwood, V. J. (2011). Comparison of the effects of environmental parameters on growth rates of vibrio vulnificus biotypes i, ii, and iii by culture and quantitative pcr analysis. *Appl. Environ. Microbiol*, 77(12), 4200–4207.
- Chave, J., Muller, L. H., & Levin, S. A. (2002). Comparing classical community models: Theoretical consequences for patterns of diversity. *The American Naturalist*, 159(1), 1–23.
- Cheng, C., Sa-Ngasoongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z. J., & Bukkapatnam,

- S. T. (2015). Time series forecasting for nonlinear and non-stationary processes: a review and comparative study. *IIE Transactions*, 47(10), 1053–1071.
- Chevin, L.-M., Lande, R., & Mace, G. M. (2010). Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory. *PLOS Biology*, 8(4), 1–8.
- Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., Pascual, M., Pielke, R., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A., Sala, O., Schlesinger, W. H., Wall, D. H., & Wear, D. (2001). Ecological forecasts: An emerging imperative. *Science*, 293(5530), 657–660.
- Constable, G. W. A., & McKane, A. J. (2015). Models of genetic drift as limiting forms of the lotka-volterra competition model. *Phys. Rev. Lett.*, 114, 038101.
- Constable, G. W. A., & McKane, A. J. (2017). Mapping of the stochastic lotka-volterra model to models of population genetics and game theory. *Phys. Rev. E*, 96, 022416.
- Constable, G. W. A., Rogers, T., McKane, A. J., & Tarnita, C. E. (2016). Demographic noise can reverse the direction of deterministic selection. *Proceedings of the National Academy of Sciences*, 113(32), E4745–E4754.
- Costantino, R. F., Desharnais, R. A., Cushing, J. M., & Dennis, B. (1997). Chaotic dynamics in an insect population. *Science*, 275(5298), 389–391.
- Coulson, T., Kendall, B. E., Barthold, J., Plard, F., Schindler, S., Ozgul, A., & Gaillard, J. M. (2017). Modeling adaptive and nonadaptive responses of populations to environmental change. *The American Naturalist*, 190, 313–336.
- Coulson, T., Rohani, P., & Pascual, M. (2004). Skeletons, noise and population growth: the end of an old debate? *Trends in Ecology & Evolution*, 19(7), 359 – 364.
- Cremer, J., Reichenbach, T., & Frey, E. (2009). The edge of neutral evolution in social dilemmas. *New Journal of Physics*, 11(9), 093029.
- Crowther, T. W., Thomas, S. M., Maynard, D. S., Baldrian, P., Covey, K., Frey, S. D., van Diepen, L. T. A., & Bradford, M. A. (2015). Biotic interactions mediate soil microbial feedbacks to climate change. *Proceedings of the National Academy of Sciences*.
- Csergő, A. M., Salguero-Gómez, R., Broennimann, O., Coutts, S. R., Guisan, A., Angert, A. L., Welk, E., Stott, I., Enquist, B. J., McGill, B., Svenning, J. C., Violle, C., & Buckley, Y. M. (2017). Less favourable climates constrain demographic strategies in plants. *Ecology Letters*, 20, 969–980.
- Cunningham, A. A., Dobson, A. P., & Hudson, P. J. (2012). Disease invasion: impacts on biodiversity and human health. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1604), 2804–2806.
- Dai, L., Vorselen, D., Korolev, K. S., & Gore, J. (2012). Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science*, 336(6085), 1175–1177.
- Dakos, V., Glaser, S. M., Hsieh, C.-h., & Sugihara, G. (2017). Elevated nonlinearity as an indicator of shifts in the dynamics of populations under stress. *Journal of The Royal Society Interface*, 14(128).
- Dalsgaard, B., Trøjelsgaard, K., González, A. M. M., Nogués-Bravo, D., Ollerton, J., Petanidou, T., Sandel, B., Schleuning, M., Wang, Z., Rahbek, C., Sutherland, W. J., Svenning, J. C., & Olesen, J. M. (2013). Historical climate-change influences modularity and nestedness of pollination networks. *Ecography (Cop)*, 36, 1331–1340.
- Davis, M. B., Shaw, R. G., & Etterson, J. R. (2005). Evolutionary responses to changing climate. *Ecology*, 86(7), 1704–1714.
- Dechter, R. (1986). Learning while searching in constraint-satisfaction-problems. In *Proceedings of the Fifth AAAI National Conference on Artificial Intelligence, AAAI'86*, (pp. 178–183). AAAI Press.

- Descalzi, O., & Rosso, O. A. (2018). Introduction to focus issue: Nonlinear dynamics of non-equilibrium complex systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7), 075401.
- Desforages, J.-P. W., Galbraith, M., & Ross, P. S. (2015). Ingestion of microplastics by zooplankton in the northeast pacific ocean. *Archives of Environmental Contamination and Toxicology*, 69(3), 320–330.
- Deyle, E. R., Fogarty, M., Hsieh, C.-h., Kaufman, L., MacCall, A. D., Munch, S. B., Perretti, C. T., Ye, H., & Sugihara, G. (2013). Predicting climate effects on pacific sardine. *Proceedings of the National Academy of Sciences*, 110(16), 6430–6435.
- Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S., & Sugihara, G. (2016a). Global environmental drivers of influenza. *Proceedings of the National Academy of Science*, 113(46), 13081–13086.
- Deyle, E. R., May, R. M., Munch, S. B., & Sugihara, G. (2016b). Tracking and forecasting ecosystem interactions in real time. *Proceedings of the Royal Society of London B: Biological Sciences*, 283(1822).
- Deyle, E. R., & Sugihara, G. (2011). Generalized theorems for nonlinear state space reconstruction. *PLOS ONE*, 6(3), 1–8.
- Dick, J. T. A., Alexander, M. E., Ricciardi, A., Lavery, C., Downey, P. O., Xu, M., Jeschke, J. M., Saul, W.-C., Hill, M. P., Wasserman, R., Barrios-O'Neill, D., Weyl, O. L. F., & Shaw, R. H. (2017). Functional responses can unify invasion ecology. *Biological Invasions*, 19(5), 1667–1672.
- Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). *Science*, 345, 401–406.
- Dixon, P., Milicich, M. J., & Sugihara, G. (1999). Episodic fluctuations in larval supply. *Science*, 283(5407), 1528–1530.
- Dixon, P., Milicich, M. J., & Sugihara, G. (2001). *Noise and nonlinearity in an ecological system*, (pp. 339–364). Boston: Birkhauser.
- Dmitriew, C. M. (2011). The evolution of growth trajectories: what limits growth rate? *Biological Reviews*, 86(1), 97–116.
- Dobrinevski, A., & Frey, E. (2012). Extinction in neutrally stable stochastic lotka-volterra models. *Phys. Rev. E*, 85, 051903.
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher*, 35, 125–129.
- Drake, J. M. (2005). Density-dependent demographic variation determines extinction rate of experimental populations. *PLOS Biology*, 3(7).
- Duan, J. (2015). *An Introduction to Stochastic Dynamics*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge.
- Eichler, M. (2013). Causal inference with multiple time series: principles and problems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1997), 20110613.
- Enke, T. N., Leventhal, G. E., Metzger, M., Saavedra, J., & Cordero, O. X. (2018). Microscale ecology regulates particulate organic matter turnover in model marine microbial communities. *Nature Communications*, 9(1), 2743.
- Estrela, S., & Brown, S. P. (2018). Community interactions and spatial structure shape selection on antibiotic resistant lineages. *PLOS Computational Biology*, 14(6), 1–21.
- Fargione, J., Brown, C. S., & Tilman, D. (2003). Community assembly and invasion: An experimental test of neutral versus niche processes. *Proceedings of the National Academy of Sciences*, 100(15), 8916–8920.

- Fasano, A., & Marmi, S. (2006). *Analytical Mechanics: An Introduction*. Oxford University Press, Oxford.
- Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., & Huttenhower, C. (2012). Microbial co-occurrence relationships in the human microbiome. *PLOS Computational Biology*, 8(7), 1–17.
- Freitas, U. S., Letellier, C., & Aguirre, L. A. (2009). Failure in distinguishing colored noise from chaos using the “noise titration” technique. *Phys. Rev. E*, 79, 035201.
- Frentz, Z., Kuehn, S., & Leibler, S. (2015). Strongly deterministic population dynamics in closed microbial communities. *Phys. Rev. X*, 5, 041014.
- Friedman, J., & Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLOS Computational Biology*, 8(9), 1–11.
- Fuhrman, J. A., Cram, J. A., & Needham, D. M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology*, 13, 133–146.
- Fukami, T. (2015). Historical contingency in community assembly: Integrating niches, species pools, and priority effects. *Annu. Rev. Ecol. Evol. Syst.*, 46, 1–23.
- Fussmann, G. F., & Blasius, B. (2005). Community response to enrichment is highly sensitive to model structure. *Biology Letters*, 1(1), 9–12.
- Gallo, F., Fossi, C., Weber, R., Santillo, D., Sousa, J., Ingram, I., Nadal, A., & Romano, D. (2018). Marine litter plastics and microplastics and their toxic chemicals components: the need for urgent preventive measures. *Environmental Sciences Europe*, 30(1), 13.
- Galloway, T. S., Cole, M., & Lewis, C. (2017). Interactions of microplastic debris throughout the marine ecosystem. *Nature Ecology & Evolution*, 1, 0116.
- Gamelon, M., Grøtan, V., Nilsson, A. L. K., Engen, S., Hurrell, J. W., Jerstad, K., Phillips, A. S., Røstad, O. W., Slagsvold, T., Walseng, B., Stenseth, N. C., & Sæther, B.-E. (2017). Interactions between demography and environmental effects are important determinants of population dynamics. *Science Advances*, 3(2).
- Gardiner, C. W. (2004). *Handbook of stochastic methods for physics, chemistry and the natural sciences*, vol. 13 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, third ed.
- Garlaschelli, D., Ruzzenenti, F., & Basosi, R. (2010). Complex networks and symmetry: A review. *Symmetry*, 2(3), 1683–1709.
- Germain, R. M., Weir, J. T., & Gilbert, B. (2016). Species coexistence: macroevolutionary relationships and the contingency of historical interactions. *Proceedings of the Royal Society of London B: Biological Sciences*, 283(1827), 0047.
- Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12, 2451–2471.
- Gharajeh, M. S. (2018). Chapter eight - biological big data analytics. In P. Raj, & G. C. Deka (Eds.) *A Deep Dive into NoSQL Databases: The Use Cases and Applications*, vol. 109 of *Advances in Computers*, (pp. 321 – 355). Elsevier.
- Ghosh, A., Mukhopadhyay, S., Roy, S., & Bhattacharya, S. (2014). Bayesian inference in non-parametric dynamic state-space models. *Statistical Methodology*, 21, 35 – 48.
- Gilljam, D., Curtsdotter, A., & Ebenman, B. (2015). Adaptive rewiring aggravates the effects of species loss in ecosystems. *Nature Communications*, 6, 8412.
- Godoy, O., Kraft, N. J., & Levine, J. M. (2014). Phylogenetic relatedness and the determinants of competitive outcomes. *Ecol. Lett.*, 17(7), 836–844.
- Goh, B., & Jennings, L. (1977). Feasibility and stability in randomly assembled lotka-volterra models. *Ecological Modelling*, 3(1), 63 – 71.
- Goh, B. S. (1979). Stability in models of mutualism. *The American Naturalist*, 113(2), 261–275.
- Golubitsky, M., & Stewart, I. (2006). Nonlinear dynamics of networks: the groupoid formal-

- ism. *Bull. Amer. Math. Soc*, 43, 305–364.
- Gong, M., Zhang, K., Schoelkopf, B., Tao, D., & Geiger, P. (2015). Discovering temporal causal relations from subsampled data. In F. Bach, & D. Blei (Eds.) *Proceedings of the 32nd International Conference on Machine Learning*, vol. 37 of *Proceedings of Machine Learning Research*, (pp. 1898–1906). PMLR.
- Gonzalez, A., & Bell, G. (2013). Evolutionary rescue and adaptation to abrupt environmental change depends upon the history of stress. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1610), 20120079.
- Gonzalez, A., Ronce, O., Ferriere, R., & Hochberg, M. E. (2012). Evolutionary rescue: an emerging focus at the intersection between ecology and evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 368(1610).
- Gorodnitsky, I., & Rao, B. (1994). Analysis of error produced by truncated svd and tikhonov regularization methods. *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, (pp. 25 – 29 vol.1).
- Grant, P. R., & Grant, B. R. (2014). *40 Years of Evolution. Darwin's Finches on Daphne Major Island*. Princeton Univ. Press, NJ.
- Grilli, J., Rogers, T., & Allesina, S. (2016). Modularity and stability in ecological communities. *Nature Communications*, 7, 12031.
- Gunawardena, J. (2014). Models in biology: 'accurate descriptions of our pathetic thinking'. *BMC Biology*, 12(1), 29.
- Hairston, N. G., Ellner, S. P., Geber, M. A., Yoshida, T., & Fox, J. A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. *Ecology Letters*, 8(10), 1114–1127.
- Handelsman, C. A., Broder, E. D., Dalton, C. M., Ruell, E. W., Myrick, C. A., Reznick, D. N., & Ghalambor, C. K. (2013). Predator-induced phenotypic plasticity in metabolism and rate of growth: Rapid adaptation to a novel environment. *Integrative and Comparative Biology*, 53(6), 975–988.
- Hardin, G. (1960). The competitive exclusion principle. *Science*, 131(3409), 1292.
- Harley, C. D. G., & Helmuth, B. S. T. (2003). Local and regional scale effects of wave exposure, thermal stress, and absolute versus effective shore level on patterns of intertidal zonation. *Limnology and Oceanography*, 48(4), 1498–1508.
- Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer-Verlag New York, Inc.
- Helmuth, B. S. T., & Hofmann, G. E. (2001). Microhabitats, thermal heterogeneity, and patterns of physiological stress in the rocky intertidal zone. *The Biological Bulletin*, 201(3), 374–384.
- Higgins, K., Hastings, A., Sarvela, J. N., & Botsford, L. W. (1997). Stochastic dynamics and deterministic skeletons: Population behavior of dungeness crab. *Science*, 276(5317), 1431–1435.
- Hiltunen, T., Virta, M., & Laine, A.-L. (2017). Antibiotic resistance in the wild: an eco-evolutionary perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1712), 20160039.
- Hines, K. E., Middendorff, T. R., & Aldrich, R. W. (2014). Determination of parameter identifiability in nonlinear biophysical models: A bayesian approach. *The Journal of General Physiology*, 143(3), 401–416.
- Ho, A., Angel, R., Veraart, A. J., Daebeler, A., Jia, Z., Kim, S. Y., Kerckhof, F.-M., Boon, N., & Bodelier, P. L. E. (2016). Biotic interactions in microbial communities as modulators of biogeochemical processes: Methanotrophy as a model system. *Frontiers in Microbiology*, 7, 1285.

- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hofbauer, J., & Sigmund, K. (1988). *The theory of evolution and dynamical systems : mathematical aspects of selection*. London Mathematical Society student texts. Cambridge University Press, Cambridge. Translation of: Evolutions theorie und dynamische System.
- Hofbauer, J., & Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Princeton Univ. Press, NJ.
- Hofbauer, J., & So, J.-H. (1994). Multiple limit cycles for three dimensional lotka-volterra equations. *Applied Mathematics Letters*, 7(6), 65 – 70.
- Holmes, E. E., Ward, E. J., & Wills, K. (2012). Marss: Multivariate autoregressive state-space models for analyzing time-series data. *R J.*, 4, 11–19.
- Holt, G., Costa, G. C., Penone, C., Lessard, J. P., Brooks, T. M., Davidson, A. D., Hedges, S. B., Radeloff, V. C., Rahbek, C., Rondinini, C., & Graham, C. H. (2017). Environmental variation is a major predictor of global trait turnover in mammals. *J. of Biogeography*, do:10.1111/jbi.13091.
- Holt, R. D., & Polis, G. A. (1997). A theoretical framework for intraguild predation. *The American Naturalist*, 149(4), 745–764.
- Houlahan, J. E., Currie, D. J., Cottenie, K., Cumming, G. S., Ernest, S. K. M., Findlay, C. S., Fuhlendorf, S. D., Gaedke, U., Legendre, P., Magnuson, J. J., McArdle, B. H., Muldavin, E. H., Noble, D., Russell, R., Stevens, R. D., Willis, T. J., Woiwod, I. P., & Wondzell, S. M. (2007). Compensatory dynamics are rare in natural ecological communities. *Proceedings of the National Academy of Sciences*, 104(9), 3273–3277.
- Hsieh, C.-h., Glaser, S. M., Lucas, A. J., & Sugihara, G. (2005). Distinguishing random environmental fluctuations from ecological catastrophes for the north pacific ocean. *Nature*, 435(7040), 336–340.
- Huang, W., Hauert, C., & Traulsen, A. (2015). Stochastic game dynamics under demographic fluctuations. *Proceedings of the National Academy of Sciences*, 112(29), 9064–9069.
- Huang, Y., Kou, G., & Peng, Y. (2017). Nonlinear manifold learning for early warnings in financial markets. *European Journal of Operational Research*, 258(2), 692 – 702.
- Huisman, J., & Weissing, F. J. (1999). Biodiversity of plankton by species oscillations and chaos. *Nature*, 402, 407–410.
- Hunter-Cevera, K. R., Neubert, M. G., Olson, R. J., Solow, A. R., Shalapyonok, A., & Sosik, H. M. (2017). Physiological and ecological drivers of early spring blooms of a coastal phytoplankter. *Science*, 354, 326–329.
- Hyttinen, A., Plis, S. M., Jarvisalo, M., Eberhardt, F., & Danks, D. (2016). Causal discovery from subsampled time series data by constraint optimization. *JMLR workshop and conference proceedings*, 52, 216–227.
- Iten, R., Metger, T., Wilming, H., Del Rio, L., & Renner, R. (2018). Discovering physical concepts with neural networks. arXiv:1807.10300.
- Ives, A. R., & Carpenter, S. R. (2007). Stability and Diversity of Ecosystems. *Science*, 317(5834), 58–62.
- Ives, A. R., Dennis, B., Cottingham, K. L., & Carpenter, S. R. (2003). Estimating community stability and ecological interactions from time series data. *Ecological Monographs*, 73(2), 301–330.
- Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667), 78.
- James, A., Pitchford, J. W., & Plank, M. J. (2013). James et al. reply. *Nature*, 500, E2–E3.
- Jiang, X., Hu, X., Xu, W., & Park, E. K. (2015). Predicting microbial interactions using vector

- autoregressive model with graph regularization. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 12(2), 254–261.
- Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., & Chisholm, S. W. (2006). Niche partitioning among prochlorococcus ecotypes along ocean-scale environmental gradients. *Science*, 311(5768), 1737–1740.
- Kaiser-Bunbury, C. N., Muff, S., Memmott, J., Müller, C. B., & Caflisch, A. (2010). The robustness of pollination networks to the loss of species and interactions: a quantitative approach incorporating pollinator behaviour. *Ecology Letters*, 13(4), 442–452.
- Kantz, H., & Schreiber, T. (2004). *Nonlinear time series analysis*. Cambridge University Press, Cambridge.
- Kauffman, S. (1993). *The Origins of Order: Self-organization and Selection in Evolution*. Oxford University Press, Oxford.
- Knight, K. (2000). *Mathematical Statistics*. Chapman and Hall/ CRC, Boca Raton, Florida.
- Kondoh, M. (2003). Foraging adaptation and the relationship between food-web complexity and stability. *Science*, 299(5611), 1388–1391.
- Korolev, K. S., Xavier, J. B., & Gore, J. (2014). Turning ecology and evolution against cancer. *Nature Reviews Cancer*, 14, 371–380.
- Krakovská, A., & Hanzely, F. (2016). Testing for causality in reconstructed state spaces by an optimized mixed prediction method. *Phys. Rev. E*, 94, 052203.
- Kuang, J. J., & Chesson, P. (2009). Coexistence of annual plants: generalist seed predation weakens the storage effect. *Ecology*, 90, 170–182.
- Kumar Duraiappah, A., & Naem, S. (2005). *Ecosystems and Human Well-Being: Biodiversity Synthesis*. Island Press, Washington, DC.
- Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. *PLOS Computational Biology*, 11(5), 1–25.
- Lai, Q., Akgul, A., Li, C., Xu, G., & Ünal, C. (2018). A new chaotic system with multiple attractors: Dynamic analysis, circuit realization and s-box design. *Entropy*, 20(1).
- Lamon, E. C., Carpenter, S. R., & Stow, C. A. (1998). Forecasting pcb concentrations in lake michigan salmonids: A dynamic linear model approach. *Ecological Applications*, 8(3), 659–668.
- Laska, M. S., & Wootton, J. T. (1998). Theoretical concepts and empirical approaches to measuring interaction strength. *Ecology*, 79(2), 461–476.
- Lawrence, D., Fiegna, F., Behrends, V., Bundy, J. G., Phillimore, A. B., Bell, T., & Barraclough, T. G. (2012). Species interactions alter evolutionary responses to a novel environment. *PLOS Biology*, 10(5), 1–11.
- Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, 33(3), 183–212.
- Levin, S. A. (2005). Self-organization and the emergence of complexity in ecological systems. *BioScience*, 55, 1075–1079.
- Levine, J. M., & HilleRisLambers, J. (2009). The importance of niches for the maintenance of species diversity. *Nature*, 461(7261), 254–257.
- Levins, R. (1968). *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton University Press, NJ.
- Lewontin, R. C. (1969). The meaning of stability. *Brookhaven Symp. Biol.*, 22, 13–24.
- Li, Q., & Liny, N. (2010). The bayesian elastic net. *Bayesian Analysis*, 5(1), 151–170.
- Logofet, D. (1993). *Matrices and graphs: stability problems in mathematical ecology*. CRC Press.
- Lopezaraiza-Mikel, M. E., Hayes, R. B., Whalley, M. R., & Memmott, J. (2007). The impact of

- an alien plant on a native plant-pollinator network: an experimental approach. *Ecol. Lett.*, 10, 539–550.
- Lorenz, E. N. (1963). Section of planetary sciences: The predictability of hydrodynamic flow. *Transactions of the New York Academy of Sciences*, 25(4 Series II), 409–432.
- Lotka, A. J. (1920). Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences*, 6(7), 410–415.
- Lusher A, M.-H. J., Hollman P (2017). Microplastics in fisheries and aquaculture. status of knowledge on their occurrence and implications for aquatic organisms and food safety. *FAO Fisheries and Aquaculture Technical Paper*, (p. 615).
- MacArthur, B. D., Sánchez-García, R. J., & Anderson, J. W. (2008). Symmetry in complex networks. *Discrete Applied Mathematics*, 156(18), 3525 – 3531.
- MacArthur, R., & Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *Am Nat*, 101, 377–385.
- MacArthur, R. H. (1968). *The theory of niche*, chap. 11, (pp. 159–176). Syracuse University Press, Syracuse, NY.
- Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., Allison, K. R., Consortium, T. D., Kellis, M., Collins, J. J., & Stolovitzky, G. (2012). Wisdom of crowds for robust gene network inference. *Nature Methods*, 9, 796–804.
- Margalef, R. (1968). *Perspectives in Ecological Theory*. University of Chicago Press, Chicago.
- Martin, B. T., Munch, S. B., & Hein, A. M. (2018). Reverse-engineering ecological theory from data. *Proceedings of the Royal Society B: Biological Sciences*, 285(1878).
- Martin, G., Aguilée, R., Ramsayer, J., Kaltz, O., & Ronce, O. (2012). The probability of evolutionary rescue: towards a quantitative comparison between theory and evolution experiments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 368(1610).
- Marx, V. (2013). The big challenges of big data. *Nature*, 498, 255–260.
- Mason, N. W. H., Holdaway, R. J., & Richardson, S. J. (2018). Incorporating measurement error in testing for changes in biodiversity. *Methods in Ecology and Evolution*, 9(5), 1296–1307.
- May, R. M. (1972). Will a large complex system be stable? *Nature*, 238(5364), 413–414.
- May, R. M. (1974). On the theory of niche overlap. *Theoretical Population Biology*, 5(3), 297 – 332.
- May, R. M., & Mac Arthur, R. H. (1972). Niche overlap as a function of environmental variability. *Proc. Natl. Acad. Sci. U.S.A.*, 69(5), 1109–1113.
- Mayo-Wilson, C. (2015). Structural chaos. *Philosophy of Science*, 82(5), 1236–1247.
- McCann, K. S. (2000). The diversity–stability debate. *Nature*, 405, 228–233.
- McKane, A. J., Biancalani, T., & Rogers, T. (2014). Stochastic pattern formation and spontaneous polarisation: The linear noise approximation and beyond. *Bulletin of Mathematical Biology*, 76(4), 895–921.
- Mees, A. (2011). *Nonlinear Dynamics and Statistics*. Birkhäuser Boston.
- Meyer, R., & Christensen, N. (2000). Bayesian reconstruction of chaotic dynamical systems. *Phys. Rev. E*, 62, 3535–3542.
- Michaelian, K. (2005). Thermodynamic stability of ecosystems. *Journal of Theoretical Biology*, 237(3), 323 – 335.
- Milner-Gulland, E. J. (2012). Interactions between human behaviour and ecological systems. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1586), 270–278.
- Mols, C. M. M., van Oers, K., Witjes, L. M. A., Lessells, C. M., Drent, P. J., & Visser, M. E.

- (2004). Central assumptions of predator–prey models fail in a semi–natural experimental system. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(Suppl 3), S85.
- Montero-Castaño, A., & Vilà, M. (2015). Direct and indirect influence of non-native neighbours on pollination and fruit production of a native plant. *PLoS One*, 10, e0128595.
- Montoya, J., & Raffaelli, D. (2010). Climate change, biotic interactions and ecosystem services. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1549), 2013–2018.
- Montoya, J. M., Pimm, S. L., & Solé, R. (2006). Ecological networks and their fragility. *Nature*, 442, 259–264.
- Murdoch, W., Briggs, C., & Nisbet, R. (2003). *Consumer-resource Dynamics*. Princeton University Press, Princeton, NJ.
- Mutshinda, C. M., O'Hara, R. B., & Woiwod, I. P. (2009). What drives community dynamics? *Proceedings of the Royal Society of London B: Biological Sciences*.
- Mysterud, A., Stenseth, N. C., Yoccoz, N. G., Langvatn, R., & Steinheim, G. (2001). Nonlinear effects of large-scale climatic variability on wild and domestic herbivores. *Nature*, 410, 1096 EP –.
- Nahum, J. R., Harding, B. N., & Kerr, B. (2011). Evolution of restraint in a structured rock–paper–scissors community. *Proceedings of the National Academy of Sciences*, 108(Supplement 2), 10831–10838.
- Narwani, A., Alexandrou, M. A., Oakley, T. H., Carroll, I. T., & Cardinale, B. J. (2013). Experimental evidence that evolutionary relatedness does not affect the ecological mechanisms of coexistence in freshwater green algae. *Ecol. Lett.*, 16(11), 1373–1381.
- Naumann, U., & Schenk, O. (2012). *Combinatorial scientific computing*. CRC Press, Boca Raton, FL.
- Nese, J. M. (1989). Quantifying local predictability in phase space. *Physica D: Nonlinear Phenomena*, 35(1), 237 – 250.
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167–256.
- Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford Univ. Press, Oxford.
- Odum, E. P. (1969). The strategy of ecosystem development. *Science*, 164, 262–270.
- Olf, H., Alonso, D., Berg, M. P., Eriksson, B. K., Loreau, M., Piersma, T., & Rooney, N. (2009). Parallel ecological networks in ecosystems. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1524), 1755–1779.
- Parsons, T. L., & Rogers, T. (2017). Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology. *Journal of Physics A: Mathematical and Theoretical*, 50(41), 415601.
- Pascual, M., & Dunne, J. A. (2005). *Ecological Networks: Linking Structure to Dynamics in Food Webs*. Oxford Univ. Press, Oxford.
- Pathak, J., Hunt, B., Girvan, M., Lu, Z., & Ott, E. (2018). Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. *Phys. Rev. Lett.*, 120, 024102.
- Patrão Neves, M., & Druml, C. (2017). Ethical implications of fighting malaria with crispr/cas9. *BMJ Global Health*, 2(3), e000396.
- Pauw, A. (2013). Can pollination niches facilitate plant coexistence? *Trends in Ecology & Evolution*, 28(1), 30 – 37.
- Pekkonen, M., & Laakso, J. T. (2012). Temporal changes in species interactions in simple aquatic bacterial communities. *BMC Ecology*, 12(1), 18.

- Pelletier, F., Garant, D., & Hendry, A. (2009). Eco-evolutionary dynamics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1523), 1483–1489.
- Pérez-Muñuzuri, I. R., V. and Gelpi (2000). Application of nonlinear forecasting techniques for meteorological modeling. *Annales Geophysicae*, 18(10), 1349–1359.
- Perretti, C. T., Munch, S. B., & Sugihara, G. (2013a). Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. *Proceedings of the National Academy of Sciences of the United States of America*, 110(13), 5253–5257.
- Perretti, C. T., Sugihara, G., & Munch, S. B. (2013b). Nonparametric forecasting outperforms parametric methods for a simulated multispecies system. *Ecology*, 94(4), 794–800.
- Petchey, O. L., Pontarp, M., Massie, T. M., Kéfi, S., Ozgul, A., Weilenmann, M., Palamara, G. M., Altermatt, F., Matthews, B., Levine, J. M., Childs, D. Z., McGill, B. J., Schaepman, M. E., Schmid, B., Spaak, P., Beckerman, A. P., Pennekamp, F., & Pearse, I. S. (2015). The ecological forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18, 597–611.
- Pham, V.-T., Jafari, S., & Volos, C. (2017). A novel chaotic system with heart-shaped equilibrium and its circuitual implementation. *Optik - International Journal for Light and Electron Optics*, 131, 343 – 349.
- Poisot, T., Stouffer, D. B., & Gravel, D. (2015). Beyond species: why ecological interaction networks vary through space and time. *Oikos*, 124(3), 243–251.
- Quade, M., Abel, M., Shafi, K., Niven, R. K., & Noack, B. R. (2016). Prediction of dynamical systems by symbolic regression. *Phys. Rev. E*, 94, 012214.
- Rafferty, N. E., CaraDonna, P. J., & Bronstein, J. L. (2015). Phenological shifts and the fate of mutualisms. *Oikos*, 124(1), 14–21.
- Ramos-Jiliberto, R., Valdovinos, F. S., Moisset de Espanes, P., & Flores, J. D. (2012). Topological plasticity increases robustness of mutualistic networks. *Journal of Animal Ecology*, 81(4), 896–904.
- Ranta, E., Lundberg, P., Kaitala, V., & Laakso, J. (2000). Visibility of the environmental noise modulating population dynamics. *Proceedings of the Royal Society of London B: Biological Sciences*, 267(1455), 1851–1856.
- Ravetti, M. G., Carpi, L. C., Goncalves, B. A., Frery, A. C., & Rosso, O. A. (2014). Distinguishing noise from chaos: Objective versus subjective criteria using horizontal visibility graph. *PLOS ONE*, 9(9), 1–15.
- Rochman, C. M., Browne, M. A., Underwood, A. J., Franeker, J. A., Thompson, R. C., & Amaral-Zettler, L. A. (2016). The ecological impacts of marine debris: unraveling the demonstrated evidence from what is perceived. *Ecology*, 97(2), 302–312.
- Rodó, X., Pascual, M., Fuchs, G., & Faruque, A. S. G. (2002). Enso and cholera: A nonstationary link related to climate change? *Proceedings of the National Academy of Sciences*, 99(20), 12901.
- Rohr, R. P., Saavedra, S., & Bascompte, J. (2014). On the structural stability of mutualistic systems. *Science*, 345, 1253497.
- Rohr, R. P., Saavedra, S., Peralta, G., Frost, C. M., Bersier, L.-F., Bascompte, J., & Tylianakis, J. M. (2016). Persist or produce: A community trade-off tuned by species evenness. *The American Naturalist*, 188(4), 411–422.
- Rosso, O. A., Larrondo, H. A., Martin, M. T., Plastino, A., & Fuentes, M. A. (2007). Distinguishing noise from chaos. *Phys. Rev. Lett.*, 99, 154102.
- Roughgarden, J. (1975). A simple model for population dynamics in stochastic environments. *The American Naturalist*, 109, 713–736.
- Rousk, J., & Bengtson, P. (2014). Microbial regulation of global biogeochemical cycles. *Front-*

- tiers in *Microbiology*, 5, 103.
- Rulands, S., Zielinski, A., & Frey, E. (2013). Global attractors and extinction dynamics of cyclically competing species. *Phys. Rev. E*, 87, 052710.
- Runge, J. (2018). Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7), 075310.
- Saavedra, S., Cenci, S., del Val, E., Boege, K., & Rohr, R. P. (2017a). Reorganization of interaction networks modulates the persistence of species in late successional stages. *J. of Animal Ecology*, 86, 1136–1146.
- Saavedra, S., Rohr, R. P., Bascompte, J., Godoy, O., Kraft, N. J. B., & Levine, J. M. (2017b). A structural approach for understanding multispecies coexistence. *Ecological Monographs*, 87(3), 470–486.
- Saavedra, S., Rohr, R. P., Dakos, V., & Bascompte, J. (2013). Estimating the tolerance of species to the effects of global environmental change. *Nat. Commun.*, 4, 2350.
- Saavedra, S., Rohr, R. P., Fortuna, M. A., Selva, N., & Bascompte, J. (2016a). Seasonal species interactions minimize the impact of species turnover on the likelihood of community persistence. *Ecology*, 97(4), 865–873.
- Saavedra, S., Rohr, R. P., Gilarranz, L. J., & Bascompte, J. (2014). How structurally stable are global socioeconomic systems? *Journal of The Royal Society Interface*, 11(100).
- Saavedra, S., Rohr, R. P., Olesen, J. M., & Bascompte, J. (2016b). Nested species interactions promote feasibility over stability during the assembly of a pollinator community. *Ecology and Evolution*, 6(4), 997–1007.
- Saccomani, M. P. (2013). Structural vs practical identifiability in system biology. In *WBBIO 2013 Proceedings*, (p. pp. 305–313).
- Saccomani, M. P., & Thomaseth, K. (2016). *Structural vs Practical Identifiability of Nonlinear Differential Equation Models in Systems Biology*, (pp. 31–41). Springer International Publishing.
- Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L. F., Kinzig, R. B. J. A., Leemans, R., & Lodge, D. M. (2000). Global biodiversity scenarios for the year 2100. *Science*, 287, 1770–1774.
- Sales-Pardo, M. (2017). The importance of being modular. *Science*, 357, 128–129.
- Saracco, F., Clemente, R. D., Gabrielli, A., & Squartini, T. (2015). Randomizing bipartite networks: the case of the world trade web. *Scientific Reports*, 5, 10595.
- Scheffers, B. R., De Meester, L., Bridge, T. C. L., Hoffmann, A. A., Pandolfi, J. M., Corlett, R. T., Butchart, S. H. M., Pearce-Kelly, P., Kovacs, K. M., Dudgeon, D., Pacifici, M., Rondinini, C., Foden, W. B., Martin, T. G., Mora, C., Bickford, D., & Watson, J. E. M. (2016). The broad footprint of climate change from genes to biomes to people. *Science*, 354(6313).
- Schleuning, M., Frund, J., Schweiger, O., Welk, E., Albrecht, J., Albrecht, M., Beil, M., Benadi, G., Bluthgen, N., Bruelheide, H., Bohning-Gaese, K., Dehling, D. M., Dormann, C. F., Exeler, N., Farwig, N., Harpke, A., Hickler, T., Kratochwi, A., Kuhlmann, M., Kuhn, I., Michez, D., Mudri-Stojnic, S., Plein, M., Rasmont, P., Schwabe, A., Settele, J., Vujic, A., Weiner, C. N., Wiemers, M., & Hof, C. (2016). Ecological networks are more sensitive to plant than to animal extinction under climate change. *Nature Comm.*, 7, 13965.
- Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, 324(5923), 81.
- Schooler, S. S., Salau, B., Julien, M. H., & Ives, A. R. (2011). Alternative stable states explain unpredictable biological control of salvinia molesta in kakadu. *Nature*, 470, 86–89.
- Schupp, E. W., Jordano, P., & Gómez, J. M. (2017). A general framework for effectiveness

- concepts in mutualisms. *Ecology Letters*, 20, 577–590.
- Serquina, R., Lai, Y.-C., & Chen, Q. (2008). Characterization of nonstationary chaotic systems. *Phys. Rev. E*, 77, 026208.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. New York, NY, USA: Cambridge University Press.
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310.
- Shmueli, G., Bruce, P. C., & Patel, N. R. (2017). *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*. Wiley Publishing.
- Sigmund, K. (2007). *Kolmogorov and population dynamics*, (pp. 177–186). Springer Berlin, Heidelberg.
- Skelly, D. K., Joseph, L. N., Possingham, H. P., Freidenburg, L. K., Farruggia, T. J., Kinnison, M. T., & Hendry, A. P. (2007). Evolutionary responses to climate change. *Conservation Biology*, 21(5), 1353–1355.
- Snyder, R. E. (2008). When does environmental variation most influence species coexistence? *Theoretical Ecology*, 1(3), 129–139.
- So, P., Ott, E., & Dayawansa, W. P. (1994). Observing chaos: Deducing and tracking the state of a chaotic system from limited observation. *Phys. Rev. E*, 49, 2650–2660.
- Song, C., Rohr, R. P., & Saavedra, S. (2017). Why are some plant–pollinator networks more nested than others? *Journal of Animal Ecology*, 86, 1417–1424.
- Song, C., Rohr, R. P., & Saavedra, S. (2018). A guideline to study the feasibility domain of multi-trophic and changing ecological communities. *Journal of Theoretical Biology*, 450, 30–36.
- Song, C., & Saavedra, S. (2018). Structural stability as a consistent predictor of phenological events. *Proc. Roy. Soc. B*, 285, 20180767.
- Spanio, T., Hidalgo, J., & Muñoz, M. A. (2017). Impact of environmental colored noise in single-species population dynamics. *Phys. Rev. E*, 96, 042301.
- Sprott, J. C. (1994). Some simple chaotic flows. *Phys. Rev. E*, 50, R647–R650.
- St John, P. C., & Doyle, F. J. (2013). Estimating confidence intervals in predicted responses for oscillatory biological models. *BMC Systems Biology*, 7(1), 71.
- Staniczenko, P. P. A., Lewis, O. T., Jones, N. S., & Reed-Tsochas, F. (2010). Structural dynamics and robustness of food webs. *Ecology Letters*, 13(7), 891–899.
- Stegen, J. C., Lin, X., Konopka, A. E., & Fredrickson, J. K. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME Journal*, 6, 1653–1664.
- Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Räscher, G., Pamer, E. G., Sander, C., & Xavier, J. B. (2013). Ecological modeling from time-series inference: Insight into dynamics and stability of intestinal microbiota. *PLOS Computational Biology*, 9(12), 1–11.
- Stenseth, N. C., Mysterud, A., Ottersen, G., Hurrell, J. W., Chan, K. S., & Lima, M. (2002). Ecological effects of climate fluctuations. *Science*, 297, 1292–1296.
- Stouffer, D. B., & Bascompte, J. (2003). Compartmentalization increases food-web persistence. *Proc. Natl. Acad. Sci. U.S.A.*, 108, 3648–3652.
- Strogatz, S. (2014). *Nonlinear Dynamics And Chaos. With applications in Physics Biology, Chemistry and Engineering*. CRC Press, Boca Raton.
- Strona, G., & Veech, J. A. (2015). A new measure of ecological network structure based on node overlap and segregation. *Methods in Ecology and Evolution*, 6, 907–915.
- Stuart, R. K., Mayali, X., Lee, J. Z., Craig Everroad, R., Hwang, M., Bebout, B. M., Weber, P. K., Pett-Ridge, J., & Thelen, M. P. (2015). Cyanobacterial reuse of extracellular organic carbon in microbial mats. *The ISME Journal*, 10, 1240–1251.
- Sugihara, G. (1994). Nonlinear forecasting for the classification of natural time series. *Philo-*

- sophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 348(1688), 477–495.
- Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, 338(6106), 496.
- Sugihara, G., & May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344, 734–741.
- Sun, J., Todd, J. D., Thrash, J. C., Qian, Y., Qian, M. C., Temperton, B., Guo, J., Fowler, E. K., Aldrich, J. T., Nicora, C. D., Lipton, M. S., Smith, R. D., De Leenheer, P., Payne, S. H., Johnston, A. W. B., Davie-Martin, C. L., Halsey, K. H., & Giovannoni, S. J. (2016). The abundant marine bacterium pelagibacter simultaneously catabolizes dimethylsulfoniopropionate to the gases dimethyl sulfide and methanethiol. *Nature Microbiology*, 1, 16065.
- Sussarellu, R., Suquet, M., Thomas, Y., Lambert, C., Fabioux, C., Pernet, M. E. J., Le Goïc, N., Quillien, V., Mingant, C., Epelboin, Y., Corporeau, C., Guyomarch, J., Robbins, J., Paul-Pont, I., Soudant, P., & Huvet, A. (2016). Oyster reproduction is affected by exposure to polystyrene microplastics. *Proceedings of the National Academy of Sciences*, 113(9), 2430–2435.
- Takemoto, K., & Kajihara, K. (2016). Human impacts and climate change influence nestedness and modularity in food-web and mutualistic networks. *PLoS One*, 11, e0157929.
- Takens, F. (1981). *Detecting strange attractors in turbulence*, (pp. 366–381). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Takeuchi, Y. (1996). *Global Dynamical Properties of Lotka-Volterra Systems*. World Scientific, Singapore.
- Tank, A., Cover, I., Foti, N., Shojaie, A., & Fox, E. (2018). An interpretable and sparse neural network model for nonlinear granger causality discovery. arXiv:1711.08160.
- Täuber, U. C. (2011). Stochastic population oscillations in spatial predator-prey models. *Journal of Physics: Conference Series*, 319(1), 012019.
- Teh, L. C. L., & Sumaila, U. R. (2011). Contribution of marine fisheries to worldwide employment. *Fish and Fisheries*, 14(1), 77–88.
- Thackeray, S. J., Henrys, P. A., Hemming, D., Bell, J. R., Botham, M. S., Burthe, S., Helaouet, P., Johns, D. G., Jones, I. D., Leech, D. I., Mackay, E. B., Massimino, D., Atkinson, S., Bacon, P. J., Brereton, T. M., Carvalho, L., Clutton-Brock, T. H., Duck, C., Edwards, M., Elliott, J. M., Hall, S. J. G., Harrington, R., Pearce-Higgins, J. W., Høye, T. T., Kruuk, L. E. B., Pemberton, J. M., Sparks, T. H., Thompson, P. M., White, I., Winfield, I. J., & Wanless, S. (2016). Phenological sensitivity to climate across taxa and trophic levels. *Nature*, 535, 24–245.
- Thébault, E., & Fontaine, C. (2010). Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329, 853–856.
- Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1), 77–94.
- Thom, R., & Fowler, D. (1975). *Structural stability and morphogenesis: an outline of a general theory of models*. The Advanced Book Program. W. A. Benjamin.
- Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Townsend Peterson, A., Phillips, O. L., & Williams, S. E. (2004). Extinction risk from climate change. *Nature*, 427, 145–148.
- Thompson, J. N. (2005). *The Geographic Mosaic of Coevolution*. Chicago Univ. Press, Chicago.

- Thompson, R. C., Olsen, Y., Mitchell, R. P., Davis, A., Rowland, S. J., John, A. W. G., McGonigle, D., & Russell, A. E. (2004). Lost at sea: Where is all the plastic? *Science*, 304(5672), 838–838.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58, 267–288.
- Tilman, D. (1987). The importance of the mechanisms of interspecific competition. *The American Naturalist*, 129(5), 769–774.
- Ting, J., D'Souza, A., Vijayakumar, S., & Schaal, S. (2008). A bayesian approach to empirical local linearization for robotics. *Proceeding of the IEEE International Conference on Robotics and Automation (ICRA'08)*, (Pasadena, CA).
- Trøjelsgaard, K., & Olesen, J. M. (2016). Ecological networks in motion: micro- and macroscopic variability across scales. *Func. Ecol.*, 30, 1926–1935.
- Turchin, P., & P. Ellner, S. (2000). Living on the edge of chaos: Population dynamics of fennoscandian voles. *Ecology*, 81, 3099–3116.
- Tyc, O., van den Berg, M., Gerards, S., van Veen, J. A., Raaijmakers, J. M., de Boer, W., & Garbeva, P. (2014). Impact of interspecific interactions on antimicrobial activity among soil bacteria. *Frontiers in Microbiology*, 5, 567.
- Ulrich, W., Almeida-Neto, M., & Gotelli, N. J. (2009). A consumer's guide to nestedness analysis. *Oikos*, 118, 3–17.
- Upadhyay, R. (2000). Chaotic behaviour of population dynamic systems in ecology. *Mathematical and Computer Modelling*, 32(9), 1005–1015.
- Ushio, M., Hsieh, C.-h., Masuda, R., Deyle, E. R., Ye, H., Chang, C.-W., Sugihara, G., & Kondoh, M. (2018). Fluctuating interaction network and time-varying stability of a natural fish community. *Nature*, 554, 360–363.
- Valverde, S., Montoya, J. M., Joppa, L., & Solé, R. (2018). The architecture of mutualistic networks as an evolutionary spandrel. *Nature Ecology and Evolution*, 2, 94–99.
- Van Kampen, N. (1992). *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library. Elsevier Science, Amsterdam.
- Vandermeer, J. H. (1969). The competitive structure of communities: An experimental approach with protozoa. *Ecology*, 50(3), 362–371.
- Vano, J. A., Wildenberg, J. C., Anderson, M. B., Noel, J. K., & Sprott, J. C. (2006). Chaos in low-dimensional lotka–volterra models of competition. *Nonlinearity*, 19(10), 2391.
- 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.
- Vázquez, D. P., Melián, C. J., Williams, N. M., Blüthgen, N., Krasnov, B. R., & Poulin, R. (2007). Species abundance and asymmetric interaction strength in ecological networks. *Oikos*, 116, 1120–1127.
- Verhulst, P. F. (1838). Notice sur la loi que la population suit dans son accroissement. *Curr. Math. Phys*, 10, 113.
- Villaverde, A. F., Barreiro, A., & Papachristodoulou, A. (2016). Structural identifiability of dynamic systems biology models. *PLOS Computational Biology*, 12(10), 1–22.
- Volterra, V. (1928). Variations and fluctuations of the number of individuals in animal species living together. *ICES Journal of Marine Science*, 3(1), 3–51.
- Volterra, V., & Brelot, M. (1931). *Leons sur la thorie mathmatique de la lutte pour la vie*. Cahiers scientifiques. Gauthier-Villars.
- Walther, G. R. (2010). Community and ecosystem responses to recent climate change. *Phil. Trans. R. Soc. B*, 365, 2019–2024.
- Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin,

- J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, 416(6879), 389–395.
- Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J. M., Moore, B. S., Laskin, J., Bandeira, N., & Dorrestein, P. C. (2012). Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences*, 109(26), E1743–E1752.
- Welti, E., & Joern, A. (2015). Structure of trophic and mutualistic networks across broad environmental gradients. *Ecol Evol.*, 5, 326–334.
- Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466, 1102–1104.
- Wood, S. N., & Thomas, M. B. (1999). Super-sensitivity to structure in biological models. *Proceedings of the Royal Society B: Biological Sciences*, 266(1419), 565–565.
- Wootton, J. T., & Emmerson, M. (2005). Measurement of interaction strength in nature. *Annual Review of Ecology, Evolution, and Systematics*, 36(1), 419–444.
- Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., Jackson, J. B. C., Lotze, H. K., Micheli, F., Palumbi, S. R., Sala, E., Selkoe, K. A., Stachowicz, J. J., & Watson, R. (2006). Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314(5800), 787–790.
- Xiao, Y., Angulo, M. T., Friedman, J., Waldor, M. K., Weiss, S. T., & Liu, Y.-Y. (2017). Mapping the ecological networks of microbial communities. *Nature Communications*, 8(1), 2042.
- Ye, H., Beamish, R. J., Glaser, S. M., Grant, S. C. H., Hsieh, C.-h., Richards, L. J., Schnute, J. T., & Sugihara, G. (2015). Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences*, 112(13), E1569–E1576.
- Ye, H., Clark, A., Deyle, E., Munch, S., Keyes, O., Cai, J., White, E., Cowles, J., Stagge, J., Daon, Y., & Sugihara, G. (2017). *rEDM: Applications of Empirical Dynamic Modeling from Time Series*. R package version 0.6.5.
URL <https://CRAN.R-project.org/package=rEDM>
- Yeakel, J. D., Guimarães, P. R., Novak, M., Fox-Dobbs, K., & Koch, P. L. (2012). Probabilistic patterns of interaction: the effects of link-strength variability on food web structure. *Journal of The Royal Society Interface*, 9(77), 3219–3228.
- Yool, A., & Tyrrell, T. (2003). Role of diatoms in regulating the ocean's silicon cycle. *Global Biogeochemical Cycles*, 17(4), 1103.
- Zettler, E. R., Mincer, T. J., & Amaral-Zettler, L. A. (2013). Life in the plastisphere: Microbial communities on plastic marine debris. *Environmental Science & Technology*, 47(13), 7137–7146.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.
- Zunino, L., Soriano, M. C., & Rosso, O. A. (2012). Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach. *Phys. Rev. E*, 86, 046210.