



# Forward and inverse modelling of eco-evolutionary dynamics

in ecological and economic systems



Victor BOUSSANGE

**Victor Boussange**

*Forward and inverse modelling of eco-evolutionary dynamics in ecological and economic systems*

October 24, 2022

Cover picture: Top: forest in Sorapiss, Dolomites, Italy. Bottom: New York City, USA. @ Luca  
Bravo / PhotoSpirit

The document format is based on the *Clean Thesis* style developed by Ricardo Langner.

DISS. ETH NO. XX

**Forward and inverse modelling of  
eco-evolutionary dynamics**  
in ecological and economic systems

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zurich)

presented by

VICTOR BOUSSANGE

M.Sc. Energy and environmental sciences, Institut National des Sciences  
Appliquées de Lyon  
born January 10<sup>th</sup>, 1995  
citizen of Bordeaux, France

accepted on the recommendation of

Prof. Dr. Loïc Pellissier, (doctoral thesis supervisor), examiner  
Prof. Dr. Didier Sornette, co-examiner  
Prof. Dr. Arnulf Jentzen, co-examiner  
Prof. Dr. Samir Suweis, co-examiner

2022



# Summary

Ecological and economic systems are complex adaptive systems, composed of multiple entities structured at different organizational levels – from genes to ecological communities and from individual behavioral rules to national economies –, that interact between themselves and their environment, and experience evolutionary processes. The processes operating at the different organizational scales affect each other, generating complex couplings across scales. Yet, despite this astonishing organizational complexity, ecological and economic systems both display invariant patterns. A fundamental challenge in the study of ecological and economic systems is to identify the general organizational principles from which these patterns originate.

Recently, studies have shown that evolutionary processes can occur on similar time scales as ecological processes, generating eco-evolutionary feedbacks which may play an important role on the dynamics of ecological systems. In parallel, studies on economic systems suggest that economic change is determined by analogous eco-evolutionary processes. Yet, our understanding of eco-evolutionary processes and feedback mechanisms in empirical systems is limited, because of the oversimplicity of current eco-evolutionary models and a lack of confrontation with empirical data. By developing novel modelling approaches to improve the modelling of empirical systems, this thesis advances our general understanding of the eco-evolutionary processes and feedbacks shaping the dynamics of ecological and economic systems.

Specifically, Chapter 2 presents and analyses an eco-evolutionary model on spatial graphs to understand how eco-evolutionary processes, together with complex habitat structures, influence the phenotypic distribution of biological populations. Chapter 3 presents an inverse modelling framework to estimate the most likely parameter values of dynamical models from empirical data, granting the opportunity to discriminate between competing eco-evolutionary hypotheses. Chapter 5 tests whether processes operating on economic activities, consisting of positive and negative interactions between them, their spatial dispersal and their transformations, can explain their dynamics at the country level. To reach this goal, the inverse modelling framework, together with data covering 59 years of economic time series over 74 countries, are used. Finally, Chapter 5 presents two numerical methods to efficiently simulate eco-evolutionary models of biological populations that are structured in high dimensional spatial and phenotypic spaces.

This thesis altogether advances our general understanding of the eco-evolutionary processes and feedbacks shaping the dynamics of biological populations and economic activities. A comprehensive mapping of elemental eco-evolutionary feedbacks influencing spatially structured biological populations is established. As regards economic activities, processes involving positive interactions between themselves and spatial dispersal are evidenced to systematically affect their dynamics at the country scale. In parallel to those fundamental

results, novel forward and inverse modelling methods are developed, allowing to better capture the dynamics of empirical systems. In the face of the ongoing climate and biodiversity crisis, there is an urgent need to accelerate our general understanding of the mechanisms shaping our world. Bridging biology, mathematical modelling, machine learning and economics, the approaches developed in this thesis can massively help us to reach this goal.

## Résumé

Les systèmes écologiques et économiques sont des systèmes complexes adaptatifs, composés de multiples entités structurées à différents niveaux d'organisation – des gènes aux communautés écologiques et des règles de comportement individuel aux économies nationales – qui sont en interactions et sont affectés par des processus évolutifs. Les processus d'interactions et d'évolution, opérant à différents niveaux d'organisation, s'affectent entre eux, générant des couplages d'échelle. Pourtant, malgré cette incroyable complexité, les systèmes écologiques et économiques démontrent tous deux des comportements invariants. Des mécanismes d'organisation généraux sont à l'origine de ces derniers, et l'enjeu principal de l'étude des systèmes écologiques et économiques est de les identifier.

Récemment, des études ont montré que certains processus évolutifs peuvent agir à des échelles de temps similaires à celles des processus écologiques, donnant alors lieu à des boucles de rétroaction éco-évolutives qui pourraient jouer un rôle fondamental sur la dynamique des systèmes écologiques. Par ailleurs, dans les systèmes économiques, des études suggèrent que le changement économique est déterminé par des processus éco-évolutifs analogues. Néanmoins, notre compréhension des processus éco-évolutifs et des boucles de rétroaction dans les systèmes empiriques est limitée, du fait de la simplicité des modèles actuels, et à cause du manque de confrontation avec des données empiriques. Développant de nouvelles approches visant à mieux modéliser les systèmes empiriques, cette thèse avance notre compréhension générale des processus et boucles de rétroaction façonnant la dynamique des systèmes écologiques et économiques.

Le chapitre 2 présente et analyse un modèle éco-évolutif sur graphes spatiaux pour comprendre comment des processus éco-évolutifs, en concert avec des structures d'habitat complexes, influencent la distribution phénotypique de populations biologiques. Le chapitre 3 présente une méthode de modélisation inverse pour estimer la valeur la plus probable de paramètres de modèle dynamiques à partir de données empiriques, permettant de distinguer différentes hypothèses éco-évolutives. Le chapitre 4 teste si des processus s'appliquant aux activités économiques, comprenant des interactions positives et négatives entre elles, leurs dispersions spatiales et leurs transformations, peuvent expliquer leur dynamique à l'échelle d'un pays. Dans ce but, la méthode de modélisation inverse, ainsi que des données économiques couvrant 59 années dans 74 pays, sont utilisées. Finalement, le chapitre

5 présente 2 méthodes numériques permettant de simuler des modèles éco-évolutifs de populations biologiques structurées dans des espaces géographiques et phénotypiques de grande dimension.

Cette thèse fait progresser notre compréhension générale des processus et boucles de rétroaction éco-évolutifs impliqués dans la dynamique de populations biologiques et dans celle des systèmes économiques. Une carte des boucles de rétroaction éco-évolutives affectant les populations biologiques structurées est établie. De même, il est démontré que des interactions positives entre les activités économiques, et leur dispersion spatiale, affectent considérablement et de façon généralisée leur dynamique à l'échelle d'un pays. En parallèle de ces résultats fondamentaux, de nouvelles méthodes de simulation et de modélisation inverse sont présentées, permettant de mieux rendre compte des dynamiques de systèmes empiriques. Face à la crise climatique et de la biodiversité en cours, il est urgent d'accélérer notre compréhension générale des mécanismes qui affectent notre monde. Associant biologie évolutive, modélisation mathématique, apprentissage machine et sciences économiques, les approches développées dans cette thèse peuvent nous y aider de façon substantielle.



# Acknowledgement

What a journey! Four years ago, I have decided to embark on a ship to live a big adventure. Lacking sailing experience, I still envisioned a voyage in unknown seas, at the crossroads between biology, economics, and applied mathematics. "*À vaincre sans périls, on triomphe sans joie*" (Maufrais, 1997), and I purposely wanted to experience storms and dangers. I have obtained what I expected, and more. I could only make it safely back home thanks to the support of the people that have supported me all along this journey. Here are some words that cannot express all the gratitude I have towards them.

I thank my supervisor, Loïc Pellissier, who made this expedition possible by providing the necessary ingredients: a boat and a map. Loïc, thank you for allowing me the freedom to draw the exact route, while kindly indicating me the North when I got lost. I thank Arnulf Jentzen for guiding me through the strange sea of applied mathematics, besides my lack of mathematical rigor. Arnulf, thank you for your forgiveness and patience, I have come a long way! I thank Didier Sornette for welcoming ideas with enthusiasm, and give it substance with his experience and knowledge. I thank Thomas Poulet for his mentorship throughout the voyage. Thomas, you were the one who taught the most important rudiments required to sail! I thank Heike Lischke for her help in crucial times. I thank Sebastian Becker and Pau Vilimelis Aceituno for their goodness, skills and precious hands.

I naturally thank all my friends and colleagues from the Landscape Ecology group. Special thank-you to Giulia Donati and Camille Pitteloud for introducing me to Zürich, and to the world of research. The times when you were here in the group were the most delightful moments of the voyage. Special thank-you to Charles de Novaes and Benjamin Flück, who respectively introduced me to Julia and helped me with amazing efficiency when I was about to throw my computer in the bin (which I eventually did). And special thank-you to Conor Waldock, Marc Grünig and Alexander Skeels for their help and kindness. You guys always had the right advice at the right time! Special thank-you to Yi Liu who's been, for reasons I still do not understand, constantly harassing me with cheers!

I am also thankful to all of my wonderful friends, whose company gave me precious half-times to further continue the journey. Special thank-you to Agathe Trouette, Théo Lafitte, Matthieu Demolin, Yoann Roncajolo, Diego Parlange and Luc Thiebaut for sharing wine, waves and delicious idiocies. "*La connerie, c'est la décontraction de l'intelligence*", this saying by Serge Gainsbourg, has never been so meaningful than with you. Special thank-you to Simon Ducroz, Valentin Mathieu, Mathieu Duthoit, Baptiste Le Maire, Guillaume Chabert and Alexandre Martineau for sharing moments of pure beauty, friendship and happiness in the high alpine. Special thank-you to my riding buddies Morgan Gueuning and Daniel Grogg for introducing me to enduro mountainbiking, and for sharing sheer moments of adrenaline and euphoria. Life in Zürich would have been very different without you guys! And special

thank-you to my climbing buddies Anita Herger, Eike Köhn and Jonas von Wartburg, for fabulous moments of verticality, bliss and philosophy, and German-climbing sessions.

A massive thank-you to my family for their support. Sabine, Clémentine and Marc, thank you so much for always having been there with me.

I thank again Camille Pitteloud, Marc Grünig, François Rousset, Merin Chacko, Thomas Keggin, Marc Boussange and Sabine Charru for useful comments and feedbacks on earlier versions of this manuscript.

And last but not least, I thank Flora for making every second of this life joyful, with or without perils. I'll never be as good of a sailor as you are, but we'll become two captains very soon.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Context . . . . .	1
1.1.1	Ecological and economic systems as complex adaptive systems . . . . .	1
1.1.2	Ecological and evolutionary processes drive the dynamics of ecological systems . . . . .	2
1.1.3	Drivers of economic change . . . . .	4
1.2	Modeling eco-evolutionary dynamics . . . . .	4
1.2.1	Forward modelling of eco-evolutionary processes . . . . .	4
1.2.2	Inverse modelling . . . . .	6
1.2.3	Artificial intelligence to leverage forward and inverse modelling . . . . .	7
1.2.4	Scientific programming and computational environments . . . . .	8
1.3	Thesis outline . . . . .	9
<b>2</b>	<b>Eco-evolutionary model on spatial graphs reveals how habitat structure affects phenotypic differentiation</b>	<b>11</b>
2.1	Introduction . . . . .	13
2.2	Results . . . . .	15
2.2.1	Eco-evolutionary model on spatial graphs . . . . .	15
2.2.2	Deterministic approximation of the population dynamics under no selection . . . . .	16
2.2.3	Effect of graph topology on neutral differentiation under no selection	18
2.2.4	Deterministic approximation of the population dynamics and adaptation under heterogeneous selection . . . . .	21
2.2.5	Effect of graph topology on adaptive differentiation under heterogeneous selection . . . . .	23
2.2.6	Effect of habitat assortativity on neutral differentiation under heterogeneous selection . . . . .	24
2.3	Discussion . . . . .	28
2.4	Methods . . . . .	30
2.4.1	Mean field approximation . . . . .	30
2.4.2	Adaptive dynamics on graphs . . . . .	31
2.4.3	Numerical simulations . . . . .	32
2.4.4	Statistics and reproducibility . . . . .	33
2.5	Data availability . . . . .	33
2.6	Code availability . . . . .	33
2.7	Acknowledgements . . . . .	33

2.A	Supplementary Note . . . . .	34
2.A.1	Mathematical construction of the model . . . . .	34
2.A.2	Deterministic approximation . . . . .	35
2.A.3	Trait-dependent competition . . . . .	38
2.A.4	Derivation of the habitat assortativity metric $r_\Theta$ in binary environments	38
2.B	Supplementary Figures . . . . .	40
2.C	Supplementary Tables . . . . .	50
<b>3</b>	<b>Mini-batching ecological data to improve ecosystem models with machine learning</b>	<b>53</b>
3.1	Introduction . . . . .	55
3.2	Machine learning framework for ecosystem models . . . . .	57
3.2.1	Ecosystem model parametrization as a learning problem . . . . .	57
3.2.2	ML framework for ecosystem models . . . . .	59
3.3	Simulated food-web model as a case study . . . . .	62
3.3.1	Three-compartment food-web ecosystem . . . . .	63
3.3.2	Parameter learning in a perfect-model setting . . . . .	64
3.3.3	Elucidating mechanistic pathways . . . . .	68
3.4	Discussion . . . . .	69
3.5	Conclusion . . . . .	72
3.6	Code availability . . . . .	72
3.7	Acknowledgements . . . . .	72
3.A	Supplementary Information . . . . .	72
3.A.1	Dynamics under perturbations . . . . .	73
3.A.2	Consequences for the shape of the loss surface . . . . .	75
3.A.3	Regularizing the loss surface with mini-batches . . . . .	76
3.B	Three-compartment food-web models . . . . .	78
3.B.1	Reference food-web model . . . . .	78
3.B.2	Omnivory variant food-web model . . . . .	78
3.C	Supplementary Figures . . . . .	78
3.D	Supplementary Tables . . . . .	81
<b>4</b>	<b>Analogous processes to ecological interactions and dispersal shape the dynamics of economic activities</b>	<b>83</b>
4.1	Introduction . . . . .	85
4.2	Methods . . . . .	87
4.2.1	Eco-evolutionary model to characterize the dynamics of economic activities . . . . .	87
4.2.2	Empirical data and model likelihood . . . . .	91
4.2.3	Maximum likelihood estimation . . . . .	91
4.3	Results . . . . .	92
4.3.1	Validation with synthetic data . . . . .	92
4.3.2	Discrepancies in dynamical regime across countries . . . . .	95
4.3.3	Evidence for eco-evolutionary processes . . . . .	97
4.4	Discussion . . . . .	100

4.5	Code availability . . . . .	102
4.6	Acknowledgements . . . . .	102
4.A	Supplementary Figures . . . . .	103
4.B	Supplementary Tables . . . . .	109
<b>5</b>	<b>Deep learning approximations for non-local nonlinear PDEs with Neumann boundary conditions</b>	<b>123</b>
5.1	Introduction . . . . .	125
5.2	Machine learning-based approximation method in a special case . . . . .	128
5.2.1	Partial differential equations (PDEs) under consideration . . . . .	128
5.2.2	Reflection principle for the simulation of time discrete reflected processes . . . . .	129
5.2.3	Description of the proposed approximation method in a special case .	129
5.3	Machine learning-based approximation method in the general case . . . . .	132
5.3.1	PDEs under consideration . . . . .	132
5.3.2	Description of the proposed approximation method in the general case	132
5.4	Multilevel Picard approximation method for non-local PDEs . . . . .	135
5.4.1	Description of the proposed approximation method . . . . .	135
5.4.2	Examples for the approximation method . . . . .	136
5.5	Numerical simulations . . . . .	138
5.5.1	Fisher–KPP PDEs with Neumann boundary conditions . . . . .	141
5.5.2	Non-local competition PDEs . . . . .	141
5.5.3	Non-local sine-Gordon type PDEs . . . . .	143
5.5.4	Replicator-mutator PDEs . . . . .	144
5.5.5	Allen–Cahn PDEs with conservation of mass . . . . .	151
5.6	Code availability . . . . .	152
5.7	Acknowledgements . . . . .	152
<b>6</b>	<b>Discussion</b>	<b>155</b>
6.1	Contributions . . . . .	155
6.1.1	Linking eco-evolutionary processes to patterns of differentiation .	155
6.1.2	Linking economic patterns to eco-evolutionary processes . . . . .	156
6.1.3	Advances in the modelling of realistic spatial and phenotypic structures	158
6.1.4	Advances in inverse modelling for identifying eco-evolutionary processes in empirical systems . . . . .	160
6.2	Limitations . . . . .	161
6.2.1	Computational complexity for modelling the evolution of full phenotypic densities . . . . .	161
6.2.2	Uncertainty estimation and model selection bias . . . . .	163
6.3	Perspectives . . . . .	164
6.3.1	Development opportunities in inverse modelling . . . . .	164
6.3.2	Confronting eco-evolutionary model on spatial graphs and empirical data . . . . .	165
6.3.3	Econobiology, a new venue to understanding economic systems, and designing more appropriate governance . . . . .	166

6.4 Concluding remarks . . . . .	166
<b>7 CV</b>	<b>169</b>
<b>Bibliography</b>	<b>175</b>
<b>List of Figures</b>	<b>205</b>

# Introduction

”

*Nature loves to hide.*

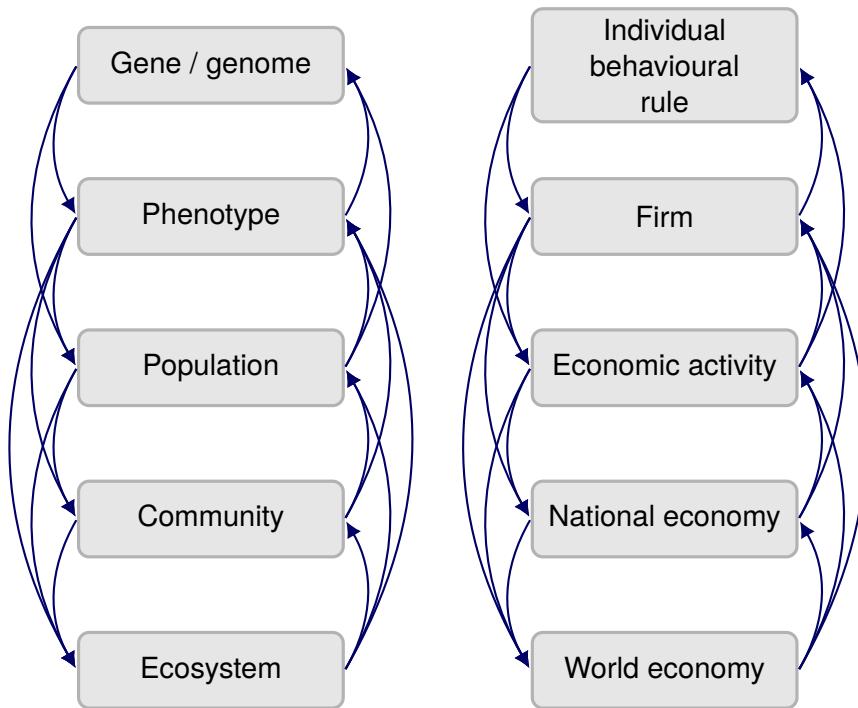
— Heraclitus (c.6th-5th century BCE)

## 1.1 Context

### 1.1.1 Ecological and economic systems as complex adaptive systems

What are the similarities between the dynamics of ecological and economic systems? Think of an ecological system as a community of interacting biological organisms (Chapin et al., 2002), and think of an economic system as a community of interacting economic agents (Dopfer and Potts, 2007). The dynamics of an ecological system depend on fluxes of matter and energy between organisms, and the dynamics of an economic system depend on fluxes of capital between economic agents. *A priori*, the underlying processes strongly differ, because the behavior of economic agents is motivated by rationality, where economic agents maximize utility (Lawson, 2013). Nonetheless, economic agents are faced with uncertainty (Foster and Metcalfe, 2012) and their rationality is bounded (Veblen, 1898; Nelson, 1985). As a result, economic agents adopt a variety of behavioral rules (Foster and Metcalfe, 2012) through trial-and-errors, which are themselves subject to natural selection through competition processes (Schumpeter, 1934). Along this perspective, both ecological and economic systems are complex adaptive systems (Levin, 2002), composed of heterogeneous entities that interact in nonlinear ways and experience evolutionary processes. The processes of interaction and evolution operate at different organizational level (Levin, 1998), from genes to ecological communities, and from individual behavior rules to economies, affecting each other across the organizational levels through different couplings (see Fig. 1.1). Interestingly, the stochasticity of the processes at stake, and their couplings, do not necessarily lead to unpredictable structures and dynamics, but rather induce organized structural properties and invariant patterns (Olff et al., 2009; Mitchell, 2009). In ecological systems, invariant patterns include, for instance, patterns of species richness, where e.g. montane regions are often associated with a disproportionately high number of species (Rahbek et al., 2019b). In economic systems, invariant patterns include, for instance, the bimodal shape of the distribution of international income, where some countries have systematically developed much more rapidly than others (Acemoglu et al., 2001). A common direction on the research agenda in biology and economics is to understand general organizational principles, i.e. to underpin the fundamental processes

and couplings that generate invariant patterns (Levin, 2002; Olff et al., 2009; Veldhuis et al., 2018). In ecological systems, the fundamental processes affecting biodiversity are identified (Rahbek et al., 2019a; Rangel et al., 2018; Hagen, 2022), and the current challenge is to underpin the mechanisms resulting from their couplings (Hagen, 2022). In economic systems, we still do not exactly understand the fundamental processes at stake.



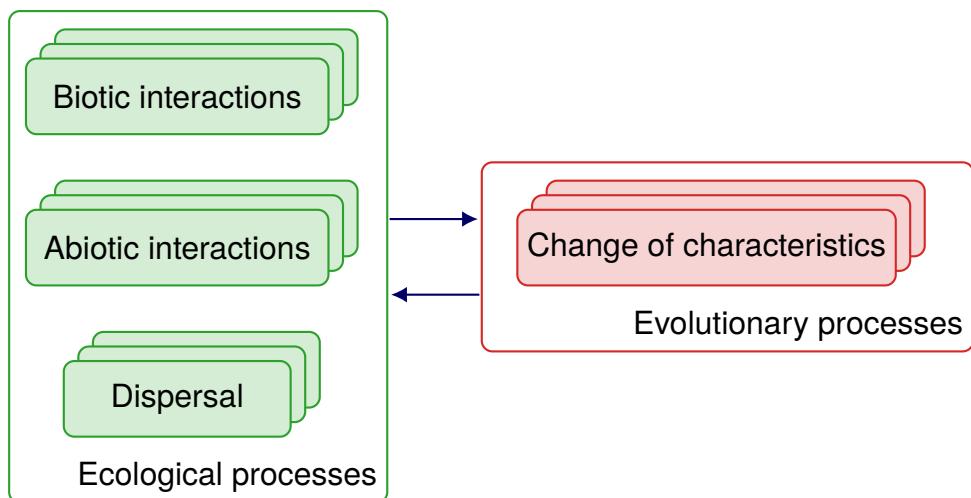
**Fig. 1.1:** Graphical representation of organizational levels and their couplings in ecological and economic systems. An arrow indicates that the organizational level at its tail can influence the organizational level at its head. Left diagram is inspired from Hendry, 2016.

### 1.1.2 Ecological and evolutionary processes drive the dynamics of ecological systems

In ecological systems, interaction processes are more commonly designated as ecological processes, and encompass the processes of interaction between organisms (biotic interactions) and between organisms and their environment (abiotic interactions), as well as dispersal processes (movement of individual across space) (Vellend, 2010a, see Fig. 1.2 for a graphical representation). Evolutionary processes designate those processes responsible for the change of heritable characteristics (DNA, genes, phenotypes) over successive generations (Hall, 2013, Fig. 1.2). The coupling between ecological and evolutionary processes is acknowledged since the very birth of the theory of evolution. During his voyage on the Beagle, Darwin documented a link between the different ecological opportunities across the Galápagos Islands and the different beak shapes in the finches he found on each island (Darwin,

1859). He reasoned that the variations in ecological opportunities lead to a differential in survival for certain phenotypes, which over time resulted in the evolution of different beak shapes. Since then, we know that ecological processes interact with evolutionary processes, and they together shape the long term dynamics of ecological systems (Urban et al., 2016; Rahbek et al., 2019a; Rangel et al., 2018; Hagen, 2022).

Empirical studies have now demonstrated that evolution can be rapid and occur on similar time scales as ecology (Hairston et al., 2005; Pelletier et al., 2009). As a result, evolutionary processes have quantifiable effects on ecological dynamics (Ezard et al., 2009), leading to feedbacks between ecological and evolutionary processes, so-called eco-evolutionary feedbacks (Pelletier et al., 2009; Schoener, 2011; Govaert et al., 2019). Eco-evolutionary feedbacks involve situations where an ecological process (e.g., replication, competition, dispersal) influences an evolutionary process (e.g. phenotypic change), which then feeds back to an ecological process, or vice versa (Govaert et al., 2019, Fig. 1.2). Examples are feedbacks between population dynamics (replication and competition) and phenotypic change, which can lead to evolutionary branching through the effect of competition (Dieckmann and Doebeli, 1999). In spatially structured populations, another classical example of eco-evolutionary feedbacks is the mechanism of local adaptation (Savolainen et al., 2007), where feedbacks between population dynamics, dispersal and trait evolution can facilitate or prevent populations to adapt to local environmental conditions (Meszéna et al., 1997; Doebeli and Dieckmann, 2003). Importantly, the eco-evolutionary feedbacks involved in adaptation mechanisms are expected to affect the dynamics of ecosystems in the coming decades (Norberg et al., 2012; Urban et al., 2016), because of the expected rapid changes in environmental conditions due to anthropogenic pressure and climate change (Ellis, 2011; Midgley and Hannah, 2019). Nevertheless, our understanding of eco-evolutionary feedbacks in realistic ecological scenarios is limited (Lion et al., 2022).



**Fig. 1.2:** Graphical representation of the eco-evolutionary processes determining eco-evolutionary dynamics in ecological systems. By extension, I use this terminology to designate interaction and evolutionary processes in economic systems.

### 1.1.3 Drivers of economic change

In economic systems, the fundamental processes determining economic change are controversial (Dopfer and Potts, 2007; Nelson et al., 2014; Hodgson, 2019). To explain economic change, Neoclassical theory (Lawson, 2013) assumes that economic systems are in equilibrium, in the sense that the demand and supply of goods and services are balanced on all relevant markets. Firms are rational in maximizing profits by adapting to demand and supply, and the observed economic change is driven by exogenous forces, such as technological innovations (Romer, 1986). Evolutionary economics, promoted by the seminal work of Nelson et al., 2014, criticizes this view and seeks to explain economic change by focusing on endogenous forces. Evolutionary economics suggests that eco-evolutionary processes, acting upon firms and their organizational routines, are major processes contributing to economic change (Hodgson, 2019). Organizational routines designate the ensemble of behavioral or technological rules that defines the cohesive identity of the firms, and play the same roles as genes, experiencing evolutionary processes and determining how firms engage in interactions with others (Nelson, 1985). Firms are the elemental units of economic systems Nelson, 1985, and through couplings across the different organizational levels, these eco-evolutionary processes should affect economic change at the local, regional, national, and international scale (see Fig. 1.1).

The deep analogy between genes and organizational routines, together with the subsequent analogies between firms and phenotypes, between economic activities and biological populations, and between the processes acting upon them (Fig. 1.1), invite understanding economic change with biological concepts. Accordingly, a number of modelling approaches have borrowed concepts and methods from biology, aiming at underpinning the fundamental processes underlying invariant patterns in economic systems (Tacchella et al., 2018; Saavedra et al., 2009; Scholl et al., 2021; Zhang et al., 2018; Modis, 1997; Saavedra et al., 2014; Farmer and Lo, 1999; Michalakelis et al., 2011; Marasco et al., 2016; Gatabazi et al., 2019; Cauwels and Sornette, 2012; Applegate and Lampert, 2021; Suweis et al., 2015). For instance, Saavedra et al., 2009 has successfully used a model of mutualistic interaction to explain structural patterns in industrial cooperation. Also, Scholl et al., 2021 uses the concepts of food webs and population dynamics to explain market malfunctions and excess volatility in financial markets. However, those studies did not seek to understand how eco-evolutionary processes may affect economic change at the national scale. Biologically inspired eco-evolutionary models may help to disentangle the effect of eco-evolutionary processes on the dynamics of national economic systems, and could explain differences in economic change across countries.

## 1.2 Modeling eco-evolutionary dynamics

### 1.2.1 Forward modelling of eco-evolutionary processes

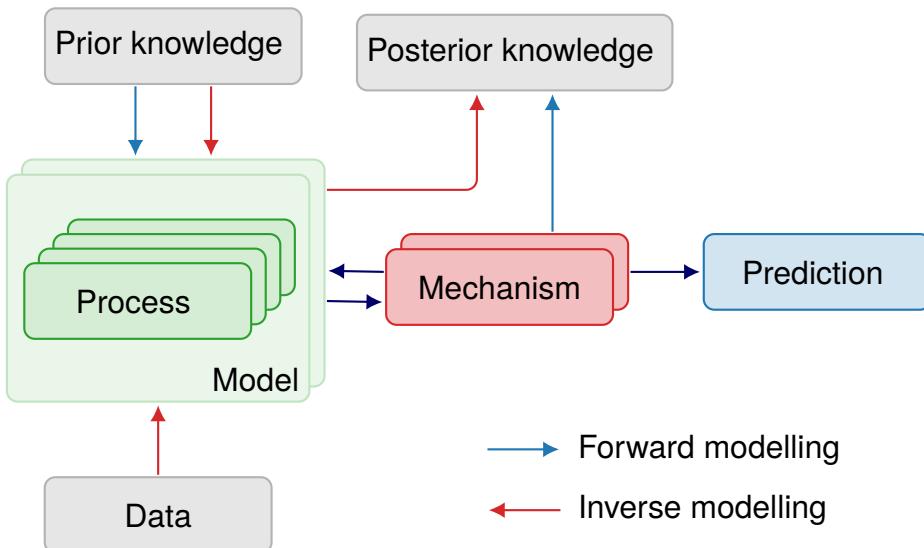
The complex interplay between ecological and evolutionary processes can hardly be studied with experimental approaches (Pontarp et al., 2019; Hagen, 2022, but see Becks et al., 2012 for an experimental study with rotifer-alga microcosms). As such, deductive reasoning, rely-

ing on forward modelling, has traditionally been put forward to underpin the mechanisms underlying invariant patterns in ecological systems (Brummitt et al., 2020). Along this approach, hypotheses about causal processes are embedded in a model, whose simulation (forward integration) generates predictions (see Fig. 1.3). A fundamental aspect of forward modelling is to underpin the underlying mechanisms, i.e. to disentangle how the interplay between the processes generate the predictions. In the early 1930s to 1940s, by formulating tractable mathematical models implementing the processes of reproduction, dispersal and mutations, the work of Fisher, Wright and Haldane has greatly contributed to the modern synthesis of evolutionary biology (Huxley, 1942), generally accepted as the basis of our current understanding of evolutionary dynamics. Yet in order to obtain tractable mathematical model, Fisher, Wright and Haldane have neglected eco-evolutionary feedbacks (Govaert et al., 2019). In particular, their works have relied on strong simplifications concerning the ecological processes, and have neglected the effect of evolutionary processes on population dynamics (Lion et al., 2022).

With the increase in computational capacity, novel modelling approaches relying on individual-based models have appeared (DeAngelis and Mooij, 2005). These models require less simplifying assumptions than traditional mathematical models (DeAngelis and Mooij, 2005), and can unveil more realistic mechanisms by allowing to capture processes acting at the individual level. However, the lack of analytical tractability of individual-based models is a shortcoming, because it challenges the ability of the modeler to underpin general principles from the simulations (Lion, 2016; May, 2004). The recent development of mathematical techniques, such as moment closure approximations (Law and Dieckmann, 1999; Gandhi et al., 2000; Nordbotten et al., 2020; Lion, 2016), adaptive dynamics theory (Metz et al., 1995), and probability theory (Champagnat et al., 2006), are generating novel pathways by filling the gap between individual-based models and mathematical models. Analogous to renormalization group analysis developed in quantum and statistical physics (Sayama, 2015), they form a toolbox to rigorously identify the couplings between the individual and the population level. As such, these mathematical techniques allow an analytical underpinning to individual-based model simulations, and can generate a general understanding of the key mechanisms at stake (Lion, 2016).

The combination of numerical simulations and, e.g., adaptive dynamics theory, has successfully shed light on the emergence of evolutionary branching under feedbacks between population dynamics and phenotypic change (Dieckmann and Doebeli, 1999; Doebeli and Dieckmann, 2003). Another example is the work of Meszéna et al., 1997; Débarre et al., 2013; Mirrahimi and Gandon, 2020, that has provided new insights on the effect of habitat heterogeneity on local adaptation. However, our current understanding of eco-evolutionary feedbacks neglects specificities of real biological populations that may significantly alter the resulting mechanisms, such as the organization of populations over complex spatial structures (Nowak and Komarova, 2001) and highly dimensional phenotypic space (Doebeli and Ispolatov, 2010a).

Investigating whether the features of realistic spatial and phenotypic structures affect eco-evolutionary feedbacks is important to advance our understanding, but raises challenging methodological issues. In particular, the consideration of complex spatial structures may hinder the underpinning of general eco-evolutionary mechanisms in structured populations.



**Fig. 1.3:** Forward and inverse modelling approaches for the understanding of complex adaptive systems. A forward modelling approach consists in deriving a model, embedding a set of processes inspired from prior knowledge. The objective is to understand how the interplay between the processes considered transforms in (feedback) mechanisms that are associated with a prediction. An inverse modelling approach integrates empirical observation within the modelling process. Posterior knowledge on the empirical system is obtained whether from the interpretation of the estimated parameter values, or through model selection.

Also, the consideration of multiple traits leads to an increase in the dimensionality of the models, which in turn leads to an exponential increase in the computational cost associated with the numerical simulations (Bellman, 2010). In order to advance our understanding of eco-evolutionary dynamics, we need to understand how the features of realistic spatial and phenotypic structures affect eco-evolutionary feedbacks, which in turn requires methodological developments, in order to cope with the extra complexity and computational cost.

## 1.2.2 Inverse modelling

Another approach to underpin processes and mechanisms in ecological systems consists in inverse modelling, where empirical data is used to constrain the model (Clermont and Zenker, 2015, see Fig. 1.3 for a graphical illustration). Inverse modelling can take the form of parameter estimation (Schartau et al., 2017) or model selection (Johnson and Omland, 2004), both involving the use of inference methods to estimate the probability of the empirical data given the model parameters (i.e., the likelihood). Provided that they are inferred together with uncertainties, parameters can be interpreted to better understand the strengths and effects of the processes considered (Pontarp et al., 2019). For instance, Higgins et al., 2010; Curtsdotter et al., 2019 infer the parameters of population dynamic models to understand the processes involved in ecosystem functions. In model selection,

candidate models embedding competing hypotheses about causal processes are derived, and the relative support of each model given the data is computed to discriminate between the hypotheses (Johnson and Omland, 2004). For instance, using inverse modelling and eco-evolutionary models embedding competing evolutionary speed hypotheses, Skeels et al., 2022 shows that temperature-dependent evolutionary speed most likely explains variations in biodiversity patterns.

The computation of the most probable model parameter values, or the computation of the different model supports, critically involves inference methods. Inference methods commonly demand many forward integration of the model, resulting in a computational cost that can be prohibitively expensive (Schneider et al., 2017). The number of forward integration required may dramatically increase with the number of model parameters (Csilléry et al., 2010), and the number of model parameters, together with the model nonlinearities, can eventually lead to false estimates of the most probable model parameter values (Gábor and Banga, 2015). Consequently, inverse modelling methods have mostly been used with simple evolutionary models (Csilléry et al., 2010). Eco-evolutionary models are dependent on numerous parameters (Boyd, 2012), are strongly nonlinear (Hastings et al., 1993; Huisman and Weissing, 1999; Benincà et al., 2008), and their integration is computationally expensive (Fisher et al., 2018), challenging the use of inverse modelling to underpin eco-evolutionary processes. Advances in the field of artificial intelligence could circumvent these issues, allowing to advance our knowledge of eco-evolutionary dynamics in empirical systems.

### 1.2.3 Artificial intelligence to leverage forward and inverse modelling

In the recent years, the field of artificial intelligence has made enormous progresses in computer vision (Voulodimos et al., 2018) and natural language processing (Young et al., 2018). At the backbone of this success are key computational techniques that could leverage the forward and inverse modelling of eco-evolutionary dynamics. Advances in computer vision and natural language processing rely on deep learning methods, that allow neural networks to learn abstract representation of mechanisms from large datasets (LeCun et al., 2015). These abstractions can hardly be interpreted to generate scientific theories (Karpatne et al., 2017), and their prediction ability is limited by the information contained in the training datasets. As such, neural networks cannot be used *per se* to gain scientific insights and extrapolate beyond observed trends (Barnosky et al., 2012; Urban et al., 2016). Nevertheless, their traditional applications and associated methods have been successfully derived in other scientific fields for this purpose (Kashinath et al., 2021; Schneider et al., 2017; Yazdani et al., 2020; Rolnick et al., 2022). For instance, neural networks have been used to reduce the cost of the forward integration of climate models, learning more efficient representations of physical mechanisms (Kashinath et al., 2021). They have also been used to approximate the solution of partial differential equation (PDE) models (Sirignano and Spiliopoulos, 2018; Han et al., 2018), with the major advantage of approximating high dimensional problems at a lower computational cost than traditional methods. Beyond accelerating model simulations, Rackauckas et al., 2020a suggests that advances in artificial intelligence can be used to infer scientific knowledge. The authors argue that generalized variants of the backpropagation algorithm – used for the training

of neural network (LeCun et al., 2015)– can be used to train any scientific model against data (Rackauckas et al., 2020a), with the potential to leverage inverse modelling techniques (Frank, 2022). Together, artificial intelligence techniques may offer unique opportunities for advancing our understanding of eco-evolutionary dynamics, but require a profound shift in the scientific programming paradigm.

## 1.2.4 Scientific programming and computational environments

Combining artificial intelligence techniques with scientific models requires a computational environment that allows to easily develop scientific models, while ensuring simulation performance, and providing composability between artificial intelligence and other scientific libraries (Rackauckas et al., 2020b). Unfortunately, performance and composability are features that are poorly represented in mainstream programming languages used by the scientific community, such as Python, Matlab or R. Those languages are naturally attractive because they are dynamically typed (Bezanson et al., 2017), allowing convenient development iterations. Nonetheless, prototypes written in Python, Matlab or R need to be rewritten in low level, compiled languages such as C, C++ or Fortran for speed and predictable mapping to hardware (Perkel, 2019; Bezanson et al., 2017). This conversion requires significant efforts, leading to a problem commonly designated as the "two-language problem" (Bezanson et al., 2017). In order to circumvent performance issues, most libraries in Python, Matlab or R rely on bindings with low level languages. For instance, the most used deep learning libraries in Python, TensorFlow and PyTorch, are internally written in C++ (see *TensorFlow* 2015; The Linux Foundation, 2016). However, bindings with low level languages come with major negative externalities. First, they restrict the understandability of their source code to computer scientists – prohibiting potential development contributions from the scientific community. Second, they prevent the composability of, e.g., traditional scientific computing libraries and deep learning libraries (Innes et al., 2019). This absence of composability arises because to be trained, numerical models must be differentiated with respect to their parameters. Yet, deep learning libraries such as TensorFlow or PyTorch are only able to differentiate models written in their own internal source code (Innes et al., 2019).

Julia is a recently developed programming language that addresses the issue of the two-language problem (Bezanson et al., 2017; Bezanson et al., 2018). Julia was built over a type-specializing, just-in-time compiler, which makes it easy to generate performant programs in pure Julia, while preserving the essential features of Python, Matlab or R, such as dynamic typing and automatic memory management (Perkel, 2019). The source code of most Julia libraries is consequently written in pure Julia, guaranteeing understandability and composability. In particular, Julia is an automatic differentiation pervasive language (Innes et al., 2019), which allows to differentiate any model written in pure Julia without any modification. As a result, deep learning libraries can be used on any scientific model written in Julia (Rackauckas et al., 2020a). Solving the two-language problem, Julia permits scientists to prototype a program which is readily generic and performant, benefitting not only the development process but also the entire research community (Bezanson et al.,

2017). Overall, the composability and productivity granted by Julia makes it an ideal computational environment to accelerate research.

## 1.3 Thesis outline

In summary, while it is increasingly acknowledged that feedbacks between ecological and evolutionary processes play an important role in ecological systems (Pelletier et al., 2009; Urban et al., 2016), our understanding of eco-evolutionary dynamics in biological populations structured in realistic phenotypic and geographic space is limited (Lion, 2016; Lieberman et al., 2005; Doebeli and Ispolatov, 2010b). Under increasing anthropogenic pressure, advancing this understanding is essential (Urban et al., 2016) but raises challenging methodological issues. Further, while analogous processes to eco-evolutionary processes have been suggested to influence the dynamics of economic systems (Hodgson, 2019), we do not know their effect on economic dynamics at the scale of a country. Here, I develop novel forward and inverse modelling methods, that I use to shed light on general eco-evolutionary processes and feedbacks shaping the dynamics of ecological and economic systems.

In Chapter 2, I investigate how eco-evolutionary processes, in combination with complex habitat structures, influence the phenotypic distribution of biological populations. I proceed using a forward modelling approach: I derive a stochastic eco-evolutionary individual-based model where individuals are structured over a spatial graph, and experience the fundamental processes of reproduction, competition, mutation and migration. Seeking to understand how those processes result in phenotypic differentiation at the population level, I derive analytical approximations of the individual-based model. Together with extensive numerical simulations, they provide insights into how the graph properties affect the population size and phenotypic differentiation. In particular, I show that three main graph properties, relating to landscape connectivity, heterogeneity in connectivity, and habitat spatial autocorrelation, shape phenotypic differentiation. These results establish mechanistic links between landscape features and the eco-evolutionary dynamics of biological populations.

In Chapter 3, I develop an inverse modelling method to estimate the parameters of eco-evolutionary models and perform model selection. The method is based on a machine learning framework and involves the combination of state-of-the-art artificial intelligence techniques and a novel learning strategy. The learning strategy consists in training the model against mini-batches of data with short time horizon, which I analytically show to solve convergence problems arising from model nonlinearities. I implement the ML framework in a new Julia library released under the name of **MiniBatchInference.jl**, and demonstrate through numerical experiments that it can efficiently and accurately estimate model parameters and provide model support from noisy, incomplete and independent time series. Altogether, the proposed ML framework is a workhorse for inverse modelling and can elucidate mechanistic pathways in ecological and economic systems.

In Chapter 4, I quantify the effect of eco-evolutionary processes on the dynamics of economic systems. I employ the ML framework developed in Chapter 3 to investigate how alternative eco-evolutionary population models can explain the dynamics of economic activities in 77 of the world's richest countries, relying on 59 year of economic data. The

models embed processes acting upon economic activities, including ecological interactions between them, their spatial dispersal, and their transformations. The statistical support of each model is compared to a simple logistic growth model, considered as a null model. I find strong statistical evidence for positive interactions and spatial dispersal. To my knowledge, this is the first study providing quantitative evidences that eco-evolutionary processes shape the dynamics of economic systems at the country level.

In Chapter 5, I extend two recent methods to solve high dimensional non-local nonlinear PDEs. This class of PDEs can be used to construct generic eco-evolutionary models of biological populations structured over complex phenotypic spaces, but up to now, could only be simulated in low dimensions. The first method presented relies on Picard iterations, while the second is based on deep learning and involves neural networks to approximate the PDE model simulations. I implement both methods in a new Julia library released under the name of **HighDimPDE.jl**, and evaluate their performance on high dimensional eco-evolutionary PDE models, and on PDE models arising in physics. The methods yield good results with short run times, opening up new venues to further our understanding of eco-evolutionary dynamics in realistic populations.

# 2

## Eco-evolutionary model on spatial graphs reveals how habitat structure affects phenotypic differentiation

by Victor Boussange<sup>1,2</sup> and Loïc Pellissier<sup>1,2</sup>

<sup>1</sup> Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland

<sup>2</sup> Landscape Ecology, Institute of Terrestrial Ecosystems, Department of Environmental System Science, ETH Zürich, CH-8092 Zürich, Switzerland

Published in Communications Biology (2022)

doi:10.1038/s42003-022-03595-3

“ The Lord is subtle, but He is not malicious.

— Albert Einstein (1879-1955)

Differentiation mechanisms are influenced by the properties of the landscape over which individuals interact, disperse and evolve. Here, we investigate how habitat connectivity and habitat heterogeneity affect phenotypic differentiation by formulating a stochastic eco-evolutionary model where individuals are structured over a spatial graph. We combine analytical insights into the eco-evolutionary dynamics with numerical simulations to understand how the graph topology and the spatial distribution of habitat types affect differentiation. We show that not only low connectivity but also heterogeneity in connectivity promotes neutral differentiation, due to increased competition in highly connected vertices. Habitat assortativity, a measure of habitat spatial auto-correlation in graphs, additionally drives differentiation under habitat-dependent selection. While assortative graphs systematically amplify adaptive differentiation, they can foster or depress neutral differentiation depending on the migration regime. By formalising the eco-evolutionary and spatial dynamics of biological populations on graphs, our study establishes fundamental links between landscape features and phenotypic differentiation.

## 2.1 Introduction

Biodiversity results from differentiation processes influenced by the features of the landscape over which populations are distributed (Hubbell, 2001). The documentation of high levels of species diversity in mountain regions and riverine systems suggests that complex connectivity patterns and habitat heterogeneity foster differentiation (Rahbek et al., 2019a; Ding et al., 2020; Dias et al., 2014; Guégan et al., 1998). However, hypotheses formulated based on empirical evidence should be complemented by mechanistic models to crystallize a causal understanding between processes and patterns (Levin, 2002). While the number of simulation studies is growing steadily (Cabral et al., 2017), such studies often lack a mathematical formalism to facilitate the interpretation of the model outcomes by providing an analytical underpinning to the simulation results (Lion, 2016).

Phenotypic differentiation processes emerge as a result of mutation, selection and migration and can be classified as neutral or adaptive (Holderegger et al., 2006). Neutral differentiation is initiated by the stochastic drift of local phenotypes when spatial isolation and limited dispersal create barriers to gene flow, allowing distinct phenotypes to emerge in spatially structured populations (Slatkin, 1993). In contrast, adaptive differentiation results from heterogeneous selection, which promotes distinct, locally well-adapted phenotypes in populations occupying patches with different habitat conditions (Dieckmann and Doebeli, 1999). The evolution of neutral phenotypes and of adaptive phenotypes are not independent, as selective forces can indirectly select for those neutral phenotypes that happen to be linked to the fittest adaptive phenotypes, a mechanism called the hitchhiking effect (Kaplan et al., 1989). Moreover, selection can generate barriers to gene flow between populations in heterogeneous habitat landscapes (Orsini et al., 2013; Wang and Bradburd, 2014), a phenomenon coined isolation by environment, which can amplify neutral differentiation. How neutral processes, adaptive processes and their interplay are affected by landscape features is difficult to comprehend without a formalized mechanistic model (Garant et al., 2007).

Models link patterns to processes (Levin, 2002), and the explicit representation of the landscape within an eco-evolutionary model can lead to a causal understanding of how landscape features shape differentiation. Spatial graphs provide a convenient mathematical representation of landscapes, where vertices represent suitable habitats hosting populations, and edges capture the connectivity between habitats (Dale and Fortin, 2010). Under ecological dynamics, metapopulation models have been used to study the role of graph topology in the persistence and stability of metapopulation (Holland and Hastings, 2008; Gilarranz and Bascompte, 2012; Mari et al., 2014; Gravel et al., 2016) and community diversity (Carrara et al., 2012; Thompson et al., 2017; Suzuki and Economo, 2021). Evolutionary mechanisms are nevertheless fundamental drivers of diversity, and should therefore be explicitly integrated into models (Pelletier et al., 2009). Evolutionary game theory explores how graph topology impacts the fixation probability and the fixation time of a mutated phenotype (Tkadlec et al., 2019). However, the framework does not consider the continuous accumulation of mutations, and is therefore not suited to addressing the emergence of phenotypic differentiation. By combining a metapopulation model with a model of neutral evolution, (Economo and Keitt, 2007; Economo and Keitt, 2010) investigated how graph

topology affects neutral diversity. Their approach demonstrated the key role of topological properties in shaping diversity, and its predictions could be matched with empirical data from e.g. river basins (Muneepeerakul et al., 2008). Nonetheless, diversity results from the combination of neutral and adaptive processes developing at the population level. A first principles modelling approach considering spatial graphs, but also building upon the elementary processes of ecological interactions, reproduction, mutation and migration may therefore be promising to investigate the emergence of diversity.

Stochastic models for structured populations, rooted in the microscopic description of individuals, offer a generic framework for modelling eco-evolutionary dynamics (Champagnat et al., 2006; Bansaye and Méléard, 2015). In particular, these models can capture the interplay between population dynamics, spatial dynamics and phenotypic evolution, while providing a rigorous set-up for analytical investigation. By anchoring this modelling paradigm in a mathematical framework, the work of Champagnat et al., 2006 generalizes models of population genetics (Bürger, 2000) (investigating the evolution of the frequencies of alleles) and quantitative genetics (Slatkin, 1978; Lande, 1991; Nagylaki, 1994) (investigating the evolution of phenotypic traits), which stimulated research into the link between spatial population structure and neutral differentiation. The framework embraces density-dependent selection, which could explain the emergence of phenotypic differentiation from competition processes (Dieckmann and Doebeli, 1999), and how spatial segregation can emerge as a byproduct of these adaptive processes along environmental gradients (Doebeli and Dieckmann, 2003). Related models have addressed the effects of landscape dynamics and habitat heterogeneity on adaptive differentiation, providing mathematical insights into the dynamics (Meszéna et al., 1997; Aguilée et al., 2012; Débarre et al., 2013; Wickman et al., 2017; Polechová, 2018; Mirrahimi and Gandon, 2020). Because it accounts for finite population size, the baseline model of Champagnat et al., 2006 can also capture neutral differentiation dynamics and therefore the coupling between neutral and adaptive processes (Billiard et al., 2015; Anceschi et al., 2019). Nonetheless, the aforementioned studies were not spatially explicit (Billiard et al., 2015; Anceschi et al., 2019) or they assumed regular spatial structures (regular graphs (Meszéna et al., 1997; Aguilée et al., 2012; Débarre et al., 2013; Mirrahimi and Gandon, 2020) or continuous space (Doebeli and Dieckmann, 2003; Wickman et al., 2017; Polechová, 2018)), therefore not addressing the role of the spatial complexity of landscapes. A stochastic individual-based model using spatial graphs as a representation of the landscape could help formalize fundamental links between landscape features and phenotypic differentiation.

A key challenge is to understand how individual dynamics result in the emergence of differentiation in complex landscapes (Manel et al., 2003). Here, we investigate how complex connectivity patterns and habitat heterogeneity affect both neutral and adaptive phenotypic differentiation by constructing an individual-based model (IBM) that accounts for eco-evolutionary dynamics on spatial graphs. The individuals disperse between habitat patches and possess co-evolving neutral and adaptive traits. The finite size of local populations generates neutral differentiation by inducing a stochastic drift in the neutral trait evolution, while heterogeneous selection gives rise to adaptive differentiation. Macroscopic properties of the model are analytically tractable, and we obtain a deterministic approximation of population size and adaptive trait dynamics which connects the emerging patterns to the

graph properties that generate them. However, neutral differentiation is stochastic by nature, which complicates its analytical underpinning. We therefore rely on numerical simulations of the IBM to measure the effect of graph topology on neutral differentiation. In the case where heterogeneous selection is absent, we investigate how graph topology affects neutral differentiation. In the case of heterogeneous selection, we investigate how the graph topology, in combination with the spatial distribution of habitat types, affects levels of (i) adaptive and (ii) neutral differentiation. By combining analytical methods with numerical simulations, we expect to identify graph properties that determine the level of differentiation. Overall, our study establishes causal links between landscape properties and population differentiation and contributes to a fundamental understanding of how landscape features promote biodiversity.

## 2.2 Results

### 2.2.1 Eco-evolutionary model on spatial graphs

We establish an individual-based model (IBM) where individuals are structured over a trait space and a graph representing a landscape. For the sake of simplicity, we consider the case of asexual reproduction and haploid genetics (Champagnat et al., 2006). Individuals die, reproduce, mutate and migrate in a stochastic fashion, which together results in macroscopic properties. The formulation of the stochastic IBM allows an analytical description of the dynamics at the population level, which links emergent properties to the elementary processes that generate them.

The trait space  $\mathcal{X} \subseteq \mathbb{R}^d$  is continuous and can be split into a neutral trait space  $\mathcal{U}$  and an adaptive trait space  $\mathcal{S}$ . We refer to neutral traits  $u \in \mathcal{U}$  as traits that are not under selection, in contrast to adaptive traits  $s \in \mathcal{S}$ , which experience selection. The graph denoted by  $G$  is composed of a set of vertices  $\{v_1, v_2, \dots, v_M\}$  that correspond to habitat patches (suitable geographical areas), and a set of edges that constrain the movement of individuals between the habitat patches. We use the original measure of genetic differentiation for quantitative traits  $Q_{ST}$  (standing for  $Q$ -statistics) in the case of haploid populations (Lande, 1992; Whitlock, 2008). We denote the neutral trait value of the  $k$ -th individual on  $v_i$  as  $u_k^{(i)}$ , the number of individuals on  $v_i$  as  $N^{(i)}$ , the mean neutral trait on  $v_i$  as  $\bar{u}^{(i)}$ , and the mean neutral trait in the metapopulation as  $\bar{u}$ . It follows that we quantify neutral differentiation  $Q_{ST,u}$  as

$$Q_{ST,u} = \sigma_{B,u}^2 / (\sigma_{B,u}^2 + \sigma_{W,u}^2) \quad (2.1)$$

where  $\sigma_{B,u}^2 = \mathbb{E} \left[ \frac{1}{M} \sum_i (\bar{u}^{(i)} - \bar{u})^2 \right]$  denotes the expected neutral trait variance between the vertices and  $\sigma_{W,u}^2 = \frac{1}{M} \sum_i \mathbb{E} \left[ \frac{1}{N^{(i)}} \sum_k (u_k^{(i)} - \bar{u}^{(i)})^2 \right]$  denotes the average expected neutral trait variance within vertices. We similarly quantify adaptive differentiation  $Q_{ST,s}$ .

Following the Gillespie update rule (Gillespie, 1976), individuals with trait  $x_k \in \mathcal{X}$  on vertex  $v_i$  are randomly selected to give birth at rate  $b^{(i)}(x_k)$  and die at rate  $d(N^{(i)}) = N^{(i)}/K$ , where  $K$  is the local carrying capacity. The definition of  $d$  therefore captures competition, which is proportional to the number of individuals on a vertex and does not depend on

the individuals' traits (we relax this assumption later on). The offspring resulting from a birth event inherits the parental traits, which can independently be affected by mutations with probability  $\mu$ . A mutated trait differs from the parental trait by a random change that follows a normal distribution with variance  $\sigma_\mu^2$  (corresponding to the continuum of alleles model (Kimura and Weiss, 1964)). The offspring can further migrate to neighbouring vertices by executing a simple random walk on  $G$  with probability  $m$ . A schematic overview of the two different settings considered is provided in Fig. 2.1. Under the setting with no selection, individuals are only characterized by neutral traits so that  $\mathcal{X} = \mathcal{U}$ . For individuals on a vertex with trait  $x_k \equiv u_k$  we define  $b^{(i)}(x_k) \equiv b$ , so that the birth rate is constant. This ensures that neutral traits do not provide any selective advantage. Under the setting with heterogeneous selection, each vertex of the graph  $v_i$  is labelled by a habitat type with environmental condition  $\Theta_i$  that specifies the optimal adaptive trait value on  $v_i$ . It follows that, for individuals with traits  $x_k = (u_k, s_k) \in \mathcal{U} \times \mathcal{S}$  on  $v_i$ , we define

$$b^{(i)}(x_k) \equiv b^{(i)}(s_k) = b(1 - p(s_k - \Theta_i)^2) \quad (2.2)$$

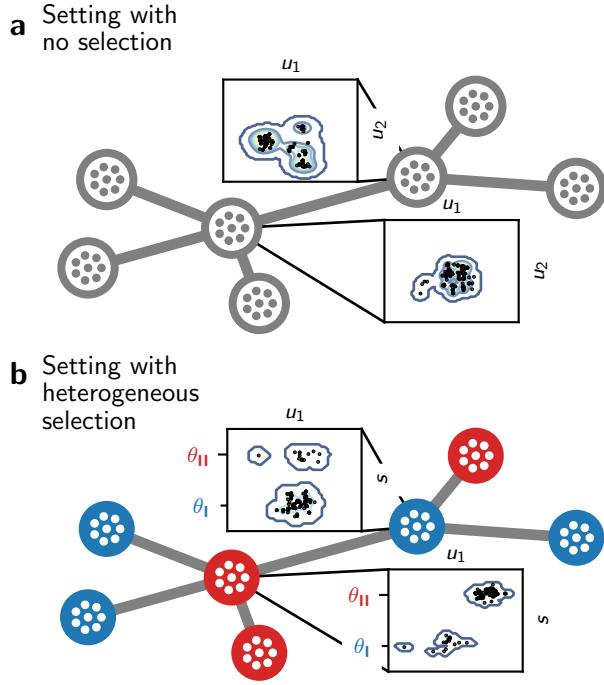
where  $p$  is the selection strength (Mirrahimi and Gandon, 2020). This ensures that the maximum birth rate on  $v_i$  is attained for  $s_k = \Theta_i$ , which results in a differential advantage that acts as an evolutionary stabilising force. In the following we consider two habitat types denoted by  $\bullet$  and  $\bullet$  with symmetric environmental conditions  $\theta_\bullet$  and  $\theta_\bullet$ , so that  $\Theta_i \in \{\theta_\bullet, \theta_\bullet\}$  and  $\theta_\bullet = -\theta_\bullet = \theta$ , where  $\theta$  can be viewed as the habitat heterogeneity (Mirrahimi and Gandon, 2020).

## 2.2.2 Deterministic approximation of the population dynamics under no selection

The model can be formulated as a measure-valued point process ((Bansaye and Méléard, 2015) and Supplementary Note). Under this formalism, we demonstrate in the Supplementary Note how the population size and the trait dynamics show a deterministic behaviour when a stabilising force dampens the stochastic fluctuations. This makes it possible to express the dynamics of the macroscopic properties with deterministic differential equations, connecting emergent patterns to the processes that generate them. In particular, in the setting of no selection, competition stabilizes the population size fluctuations, and the dynamics can be considered deterministic and expressed as

$$\partial_t N_t^{(i)} = N_t^{(i)} \left[ b(1 - m) - \frac{N_t^{(i)}}{K} \right] + mb \sum_{j \neq i} \frac{a_{i,j}}{d_j} N_t^{(j)} \quad (2.3)$$

where  $A = (a_{i,j})_{1 \leq i,j \leq M}$  is the adjacency matrix of the graph  $G$  and  $D = (d_1, d_2, \dots, d_M)$  is a vector containing the degree of each vertex (number of edges incident to the vertex). The first term on the right-hand side corresponds to logistic growth, which accounts for birth and death events of non-migrating individuals. The second term captures the gains due to migrations, which depend on the graph topology. Assuming that all vertices with the same degree have an equivalent position on the graph, corresponding to a mean field approach



**Fig. 2.1:** Graphical representation of the structure of individuals in the eco-evolutionary model. (a) Setting with no selection, where individuals are characterized by a set of neutral traits  $u \in \mathcal{U}$ . The scatter plots represent a projection of the first two components of  $u$  for the individuals present on the designated vertices at time  $t = 1000$ , obtained from one simulation of the IBM. (b) Setting with heterogeneous selection. In this setting, individuals are additionally characterized by adaptive traits  $s \in \mathcal{S}$ . Blue vertices favour the optimal adaptive trait value  $\theta_I$ , while red vertices favour  $\theta_{II}$ . The scatter plots represent a projection of the first component of  $u$  and  $s$  for the individuals present on the designated vertices at time  $t = 1000$ , obtained from one simulation. The majority of individuals are locally well-adapted and have an adaptive trait close to the optimal value, but some maladaptive individuals originating from neighbouring vertices are also present.  $m = 0.05$ .

(see Methods), one can obtain a closed-form solution from Eq. (2.3) (see Eq. (2.12)), which shows that the average population size  $\bar{N}$  scales with  $\langle \sqrt{k} \rangle^2 / \langle k \rangle$ , where  $\langle k \rangle$  is the average vertex degree and  $\langle \sqrt{k} \rangle$  is the average square-rooted vertex degree. The quantity  $\langle \sqrt{k} \rangle^2 / \langle k \rangle$ , denoted as  $h_d$ , relates to the homogeneity in vertex degree of the graph and can therefore be viewed as a measure negatively associated with heterogeneity in connectivity. Simulations of the IBM illustrate that  $h_d$  can explain differences in population size for complex graph topologies with varying migration regimes (Fig. 2.2a for graphs with  $M = 7$  vertices and Fig. S2.6a for  $M = 9$ ). This analytical result is connected to theoretical work on reaction diffusion processes (Colizza et al., 2007) and highlights that irregular graphs (graphs whose vertices do not have the same degree) result in unbalanced migration fluxes that affect the ecological balance between births and deaths. Highly connected vertices present an oversaturated carrying capacity ( $N^{(i)} > bK$ , see Methods), increasing local competition

and lowering total population size compared with regular graphs (Fig. 2.2a). Because populations with small sizes experience more drift ((Bürger, 2000) and Fig. S2.7), this result indicates that graph topology affects neutral differentiation not only through population isolation, but also by affecting population dynamics.

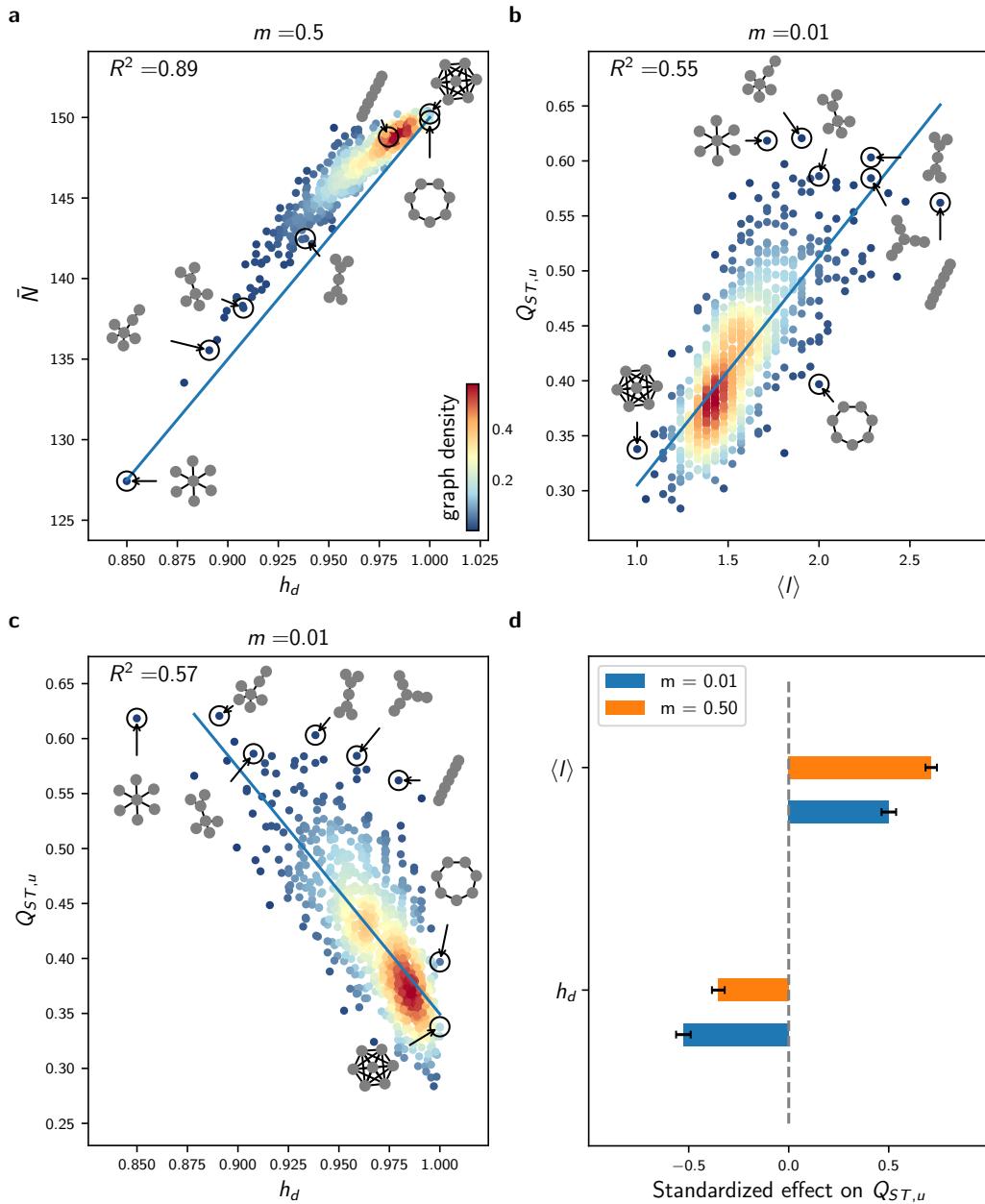
Nonetheless, the stochasticity of the processes at the individual level can propagate to the population level and substantially affect the macroscopic properties. In particular, neutral differentiation emerges from the stochastic fluctuations of the populations' neutral trait distribution. These fluctuations complicate an analytical underpinning of the dynamics, and in this case simulations of the IBM offer a straightforward approach to evaluate the level of neutral differentiation.

### 2.2.3 Effect of graph topology on neutral differentiation under no selection

We study a setting with no selection and investigate the effect of the graph topology on neutral differentiation. When migration is limited, individuals' traits are coherent on each vertex but stochastic drift at the population level generates neutral differentiation between the vertices. Migration attenuates neutral differentiation because it has a correlative effect on local trait distributions. Following (Economou and Keitt, 2007; Carrara et al., 2012; Thompson et al., 2017), we expect that the intensity of the correlative effect depends on the average path length of the graph  $\langle l \rangle$ , defined as the average shortest path between all pairs of vertices (Bounova and Weck, 2012). For a constant number of vertices,  $\langle l \rangle$  is strictly related to the mean betweenness centrality and quantifies the graph connectivity (Bounova and Weck, 2012). High  $\langle l \rangle$  implies low connectivity and a greater isolation of populations, and hence we expect that graphs with high  $\langle l \rangle$  are associated with high differentiation levels. We consider various graphs with an identical number of vertices and run simulations of the IBM to obtain the neutral differentiation level  $Q_{ST,u}$  attained after a time long enough to discard transient dynamics (see Methods). We then interpret the discrepancies in  $Q_{ST,u}$  across the simulations by relating them to the underlying graph topologies.

We observe strong differences in  $Q_{ST,u}$  across graphs for varying  $m$ , and find that  $\langle l \rangle$  explains at least 55% of the variation in  $Q_{ST,u}$  across all graphs with  $M = 7$  vertices for (Fig. 2.2b). Nonetheless, some specific graphs, such as the star graph, present higher levels of  $Q_{ST,u}$  than expected by their average path length. To explain this discrepancy, we explore the effect of homogeneity in vertex degree  $h_d$ , as we showed in Eq. (2.12) that it decreases population size, which should in turn increase  $Q_{ST,u}$  by intensifying stochastic drift. We find that  $h_d$  explains 57% of the variation for low  $m$  (Fig. 2.2c). However, the fit remains similar after correcting for differences in population size (see Table S2.1), indicating that irregular graphs structurally amplify the isolation of populations. Unbalanced migration fluxes lead central vertices to host more individuals than allowed by their carrying capacity. This causes increased competition that results in a higher death rate, so that migrants have a lower probability of further spreading their trait. Highly connected vertices therefore behave as bottlenecks, increasing the isolation of peripheral vertices and consequently amplifying  $Q_{ST,u}$ .

We then evaluate the concurrent effect of  $\langle l \rangle$  and  $h_d$  on  $Q_{ST,u}$  with a multivariate regression model that we fit independently for low and high migration regimes (Fig. 2.2d). The multivariate regression model explains at least 70% of the variation in  $Q_{ST,u}$  for the migration regimes considered and for graphs with  $M = 7$  vertices (see Table S2.2 for details). Moreover, we find that  $\langle l \rangle$  and  $h_d$  have akin contributions to neutral differentiation for low  $m$ , but the effect of  $\langle l \rangle$  increases for higher migration regimes while the effect of  $h_d$  decreases. To ensure that these conclusions can be generalized to larger graphs, we conduct the same analysis on a subset of graphs with  $M = 9$  vertices and find congruent results (Fig. S2.6). In the absence of selection and with competitive interactions, graphs with a high average path length  $\langle l \rangle$  and low homogeneity in vertex degree  $h_d$ , or similarly graphs with low connectivity and high heterogeneity in connectivity, show high levels of neutral differentiation.



**Fig. 2.2: Effect of  $\langle l \rangle$  and  $h_d$  on average population size  $\bar{N}$  and neutral differentiation  $Q_{ST,u}$  in the setting with no selection.** (a) Response of  $\bar{N}$  to homogeneity in degree  $h_d = \langle \sqrt{k} \rangle^2 / \langle k \rangle$  for all undirected connected graphs with  $M = 7$  vertices and  $m = 0.5$ . (b) Response of  $Q_{ST,u}$  to average path length  $\langle l \rangle$  for similar simulations obtained with  $m = 0.01$ . (c) Response of  $Q_{ST,u}$  to homogeneity in degree  $h_d$  for the same data. In (a), (b) and (c), each dot represents average results from 5 replicate simulations of the IBM, the colour scale corresponds to the proportion of the graphs with similar  $x$  and  $y$  axis values (graph density), and the blue line corresponds to a linear fit. (d) Standardized effect of  $h_d$  and  $\langle l \rangle$  on  $Q_{ST,u}$ , obtained from multivariate regression models independently fitted on similar data obtained for  $m = 0.01$  and  $m = 0.5$ . The contributions of  $\langle l \rangle$  and  $h_d$  to  $Q_{ST,u}$  are alike for low migration regimes. Error bars show 95% confidence intervals. Analogous results on graphs with  $M = 9$  vertices are presented in Fig. S2.6 and all regression details can be found in Table S2.2.

## 2.2.4 Deterministic approximation of the population dynamics and adaptation under heterogeneous selection

We next consider heterogeneous selection and investigate the response of adaptive differentiation to the spatial distribution of habitat types, denoted as the  $\Theta$ -spatial distribution. Adaptive differentiation emerges from local adaptation, but migration destabilizes adaptation as a result of the influx of maladaptive migrants. We expect that higher connectivity between vertices of similar habitat type increases the level of adaptive differentiation, because it increases the proportion of well-adapted migrants. Local adaptation can be investigated by approximating the stochastic dynamics of the trait distribution with a deterministic partial differential equation (PDE). We demonstrate under mean field assumption how the deterministic approximation can be reduced to an equivalent two-habitat model. We analyse the reduced model with the theory of adaptive dynamics (Meszéna et al., 1997; Mirrahimi and Gandon, 2020) and find a critical migration threshold  $m^*$  that determines local adaptation.  $m^*$  depends on a quantity coined the habitat assortativity  $r_\Theta$ , and we demonstrate with numerical simulations that  $r_\Theta$  determines the overall adaptive differentiation level  $Q_{ST,s}$  reached at steady state in the deterministic approximation.

Heterogeneous selection, captured by the dependence of the birth rate on  $\Theta_i$ , generates a stabilising force that dampens the stochastic fluctuations of the adaptive trait distribution. The dynamics of the adaptive trait distribution consequently shows a deterministic behavior and we demonstrate in the Supplementary Note and Figs. S2.8 and S2.9 that the number of individuals on  $v_i$  with traits  $s \in \Omega \subset \mathcal{S}$  can be approximated by the quantity  $\int_\Omega n^{(i)}(s)ds$ , where  $n^{(i)}$  is a continuous function solution of the PDE

$$\begin{aligned} \partial_t n_t^{(i)}(s) &= n_t^{(i)}(s) \left[ b^{(i)}(s)(1-m) - \frac{1}{K} \int_S n_t^{(i)}(\mathbf{s})d\mathbf{s} \right] \\ &\quad + m \sum_{j \neq i} b_j(s) \frac{a_{i,j}}{d_j} n_t^{(j)}(s) \\ &\quad + \frac{1}{2} \mu \sigma_\mu^2 \Delta_s \left[ b^{(i)}(s) n_t^{(i)}(s) \right] \end{aligned} \tag{2.4}$$

Equation (2.4) is similar to Eq. (2.3), except that it incorporates an additional term corresponding to mutation processes and that the birth rate is trait dependent. We show how Eq. (2.4) can be reduced to an equivalent two-habitat model under mean field assumption. The mean field approach differs slightly from the setting with no selection because vertices are labelled with  $\Theta_i$ . Here we assume that vertices with similar habitat types have an equivalent position on the graph (see Fig. S2.10 for a graphical representation), so that all vertices with habitat type  $\bullet$  are characterized by the identical adaptive trait distribution that we denote by  $\bar{n}^\bullet$ , and are associated with the birth rate  $b^\bullet(s) = b(1 - p(s - \theta_\bullet)^2)$ . Let  $P(\bullet, \bullet)$  denote the proportion of edges connecting a vertex  $v_i$  of type  $\bullet$  to a vertex  $v_j$  of type  $\bullet$ , and

let  $P(\bullet)$  denote the proportion of vertices  $v_i$  of type  $\bullet$ . By further assuming that habitats are homogeneously distributed on the graph so that  $P(\bullet) = P(\bullet) = \frac{1}{2}$ , Eq. (2.4) transforms into

$$\begin{aligned}\partial_t \bar{n}_t^\bullet(s) &= \bar{n}_t^\bullet(s) \left[ b^\bullet(s)(1-m) - \frac{1}{K} \int_S \bar{n}_t^\bullet(s) ds \right] + \frac{1}{2} \mu \sigma_\mu^2 (\Delta_s b^\bullet \bar{n}_t^\bullet)(s) \\ &\quad + \frac{m}{2} [(1-r_\Theta)b^\bullet(s)\bar{n}_t^\bullet(s) + (1+r_\Theta)b^\bullet(s)\bar{n}_t^\bullet(t)]\end{aligned}\quad (2.5)$$

(see Methods), where we define

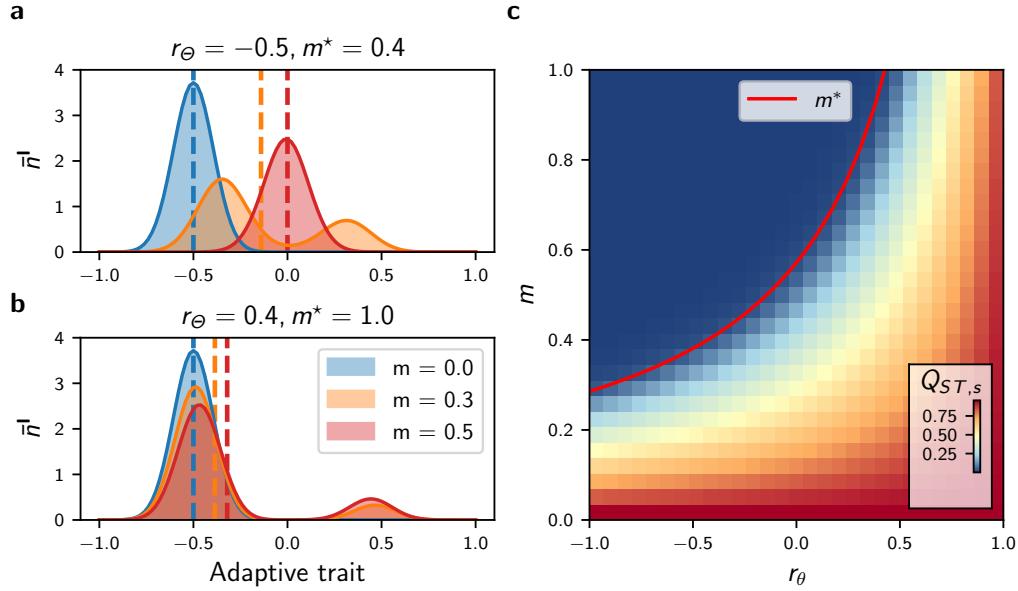
$$r_\Theta = 2(P(\bullet, \bullet) - P(\bullet, \bullet)) \quad (2.6)$$

as the habitat assortativity of the graph, which ranges from  $-1$  to  $1$ . When  $r_\Theta = -1$ , all edges connect dissimilar habitat types (disassortative graph), while as  $r_\Theta$  tends towards  $1$  the graph is composed of two clusters of vertices with identical habitat types (assortative graph). Eq. (2.5) can be analysed with the theory of adaptive dynamics (Meszéna et al., 1997; Débarre et al., 2013; Mirrahimi and Gandon, 2020), a mathematical framework that provides analytical insights by assuming a trait substitution process. Following this assumption, the mutation term in Eq. (2.5) is omitted and the phenotypic distribution results in a collection of discrete individual types that are gradually replaced by others until evolutionary stability is reached (see Methods and (Meszéna et al., 1997; Débarre et al., 2013; Mirrahimi and Gandon, 2020) for details). By applying the theory of adaptive dynamics, we find a critical migration rate  $m^*$

$$m^* = \frac{1}{(1-r_\Theta)} \frac{4p\theta^2}{(1+3p\theta^2)} \quad (2.7)$$

so that when  $m > m^*$ , a single type of individual exists with adaptive trait  $s^* = (\theta_\bullet + \theta_\bullet)/2 = 0$  in the steady state (see Methods for the derivation of Eq. (2.7)). In this case, adaptive differentiation  $Q_{ST,s}$  is nil and the average population size is given by  $\bar{N} = bK(1-p\theta)^2$ . In contrast, when  $m = 0$  and/or  $r_\Theta = 1$ , all individuals are locally well-adapted with trait  $\Theta_i$  on  $v_i$ , and it follows that the average population size is higher and equal to  $\bar{N} = bK$ , while adaptive differentiation is maximal and equal to  $Q_{ST,s} = \text{Var}(\Theta)/(\text{Var}(\Theta)+0) = 1$ . When  $0 < m < m^*$ , the coexistence of two types of individuals on each vertex  $v_i$  is predicted but the calculation of the trait values is more subtle. To understand the effect of  $m$  and  $r_\Theta$  on the local trait distributions and on  $Q_{ST,s}$ , we therefore leave behind the adaptive dynamics framework and numerically solve Eq. (2.5) by including the mutation term. When  $0 < m < m^*$ , the local trait distributions are bimodal with peaks corresponding to the two types of individuals predicted by the adaptive dynamics. The highest peak corresponds to the well-adapted individuals, whose adaptation is destabilized by the influx of maladaptive migrants (Fig. 2.3a). This phenomenon is dampened as  $r_\Theta$  increases, since the proportion of maladaptive migrants is reduced in assortative graphs (Fig. 2.3b). As a consequence, the habitat assortativity  $r_\Theta$  increases the differentiation  $Q_{ST,s}$  when  $0 < m < m^*$  (Fig. 2.3c). The simulations further confirm that the adaptive dynamics prediction given by Eq. (2.7) is still valid when the continuous accumulation of mutations is considered, so that for  $m > m^*$  the local trait distributions obtained from Eq. (2.5) are unimodal and  $Q_{ST,s}$  vanishes (Fig. 2.3a,c). Our analysis of the mean field deterministic approximation Eq. (2.5)

therefore demonstrates that assortative graphs present high levels of adaptive differentiation  $Q_{ST,s}$ . On the other hand, the analysis shows that  $Q_{ST,s}$  rapidly declines with increasing  $m$  on disassortative graphs, until  $Q_{ST,s}$  vanishes when  $m > m^*$ .



**Fig. 2.3:** Effect of habitat assortativity  $r_\Theta$  and migration  $m$  on the local adaptive trait distribution  $\bar{n}^*$  and on the adaptive differentiation level  $Q_{ST,s}$  under the mean field, deterministic approximation Eq. (2.5). (a) Effect of  $m$  and  $r_\Theta$  on  $\bar{n}^*$ . Migration induces the apparition of maladaptive individuals (centred around  $\theta_0 = 0.5$ ), which destabilize local adaptation by displacing the mean value of the well-adapted individuals (centred around  $\theta_0 = -0.5$ ). Together with the decrease in local adaptation, migration causes a displacement of the mean value of the local trait distribution (represented by the vertical dashed lines), which decreases local population size and adaptive differentiation  $Q_{ST,s}$ . (b) Similar data for higher  $r_\Theta$ . Increasing  $r_\Theta$  increases population size and  $Q_{ST,s}$ . (c) Effect of  $r_\Theta$  on  $Q_{ST,s}$ . The red line indicates the critical migration threshold  $m^*$  predicted by Eq. (2.7);  $Q_{ST,s}$  vanishes when  $m > m^*$ .

## 2.2.5 Effect of graph topology on adaptive differentiation under heterogeneous selection

To generalize the conclusions drawn from the mean field deterministic approximation Eq. (2.5), we generate different  $\Theta$ -spatial distributions for varying graph topology, and compare outputs of the IBM simulations with those of Eq. (2.5) (see Methods for the details of the simulations). For each combination of  $\Theta$ -spatial distribution and graph, we compute the habitat assortativity  $r_\Theta$ , since  $r_\Theta$  can be generalized from Eq. (2.6) to any graph topology following the original definition of (Newman, 2003) as

$$r_\Theta = \frac{\text{Cov}(\Theta_x, \Theta_\wedge)}{\sigma_{\Theta_x} \sigma_{\Theta_\wedge}} \quad (2.8)$$

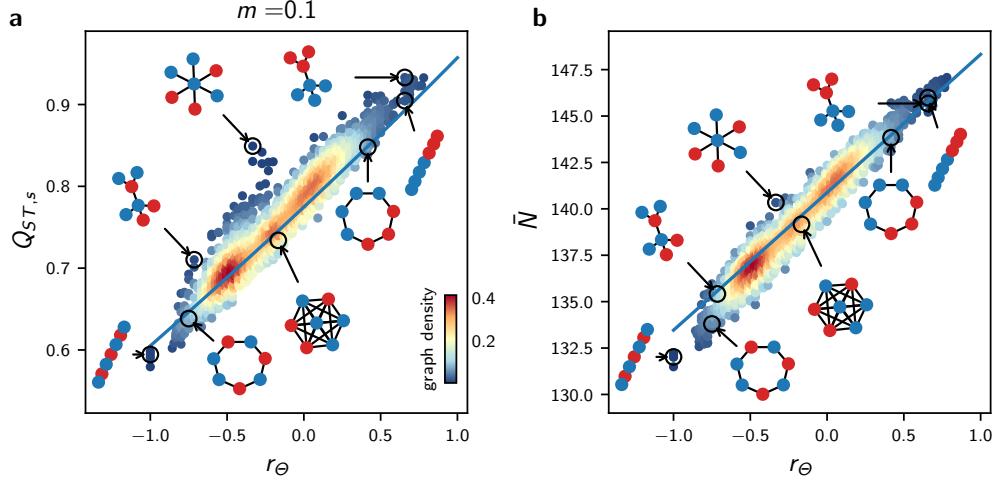
where  $\Theta_{\times}$  and  $\Theta_{\wedge}$  denote the sets of habitats found at the toe and tip of each directed vertex of graph  $V$ , and  $\langle \Theta_{\times} \rangle$ ,  $\langle \Theta_{\wedge} \rangle$  and  $\sigma_{\Theta_{\times}}, \sigma_{\Theta_{\wedge}}$  denote their respective means and standard deviations (see Supplementary Note). The mean field deterministic approximation Eq. (2.5) is in very good agreement with the IBM simulations for general graph ensembles at low migration regimes, and captures the response of  $\bar{N}$  and  $Q_{ST,s}$  to  $r_{\Theta}$  (Fig. 2.4). Nonetheless, under high migration regimes, higher levels of  $Q_{ST,s}$  are observed in the stochastic simulations compared with the mean field deterministic approximation (Fig. S2.11). We hypothesize that this reinforcement is generated by stochastic drift, which must become the main driver of differentiation when local adaptation is lost for  $m > m^*$ , and perform a multivariate regression analysis to investigate the additional effect of  $\langle l \rangle$  and  $h_d$  on  $Q_{ST,s}$ . As expected, the analysis highlights that the effect of  $\langle l \rangle$  and  $h_d$  are substantial and complement the effect of  $r_{\Theta}$  for high  $m$  (Fig. 2.5c for graphs with  $M = 7$  vertices and Fig. S2.12a for  $M = 9$ ), further explaining the discrepancies observed (see Table S2.3).

We extend our analyses to realistic landscapes with a continuum of habitat types by running simulations on graphs obtained from real spatial habitat datasets and by considering mean annual temperature as a proxy for habitat type (see Fig. S2.13 and Table S2.4). We also consider simulations accounting for trait-dependent competition to test whether our results hold under more complex ecological processes (see Supplementary Note for the implementation details and Table S2.5 for the results). The simulations are congruent and show that the effects of  $r_{\Theta}$ ,  $h_d$  and  $\langle l \rangle$  are similar under these alternative settings, underlining the robustness of these metrics and the generality of our conclusions. Taken together, these results indicate that under sufficiently strong selection and sufficiently high habitat heterogeneity, adaptive differentiation  $Q_{ST,s}$  is mainly driven by habitat assortativity  $r_{\Theta}$ . Nonetheless, local adaptation is lost in disassortative graphs when  $m > m^*$ , such that  $\langle l \rangle$  and  $h_d$  become complementary determinants of  $Q_{ST,s}$  for high migration regimes.

## 2.2.6 Effect of habitat assortativity on neutral differentiation under heterogeneous selection

We finally consider a setting with heterogeneous selection where individuals carry both neutral and adaptive traits. With distinct habitat types, selection promotes neutral differentiation by reducing the birth rate of maladaptive migrants, reinforcing the isolation of local populations. We have shown above that adaptive differentiation  $Q_{ST,s}$  is driven by habitat assortativity  $r_{\Theta}$ , so we expect  $r_{\Theta}$ , together with the topological metrics found in the setting with no selection, to influence the level of neutral differentiation  $Q_{ST,u}$ . We first investigate how the response of  $Q_{ST,u}$  to migration compares between the setting with no selection and the setting with heterogeneous selection for graphs with an identical topology. We then examine how the response compares between graphs with an identical topology but different  $r_{\Theta}$ . We finally consider simulations on different graphs with varying  $r_{\Theta}$  to assess the concurrent effect of  $\langle l \rangle$ ,  $h_d$  and  $r_{\Theta}$  on  $Q_{ST,u}$ .

Migration has a fitness cost because maladaptive migrants present lower fitness. Under an equivalent migration regime, migrants therefore have a lower probability of reproduction, increasing the populations' isolation compared with a setting without selection. Simulations with varying  $m$  on the complete graph confirm that selection in heterogeneous habitats

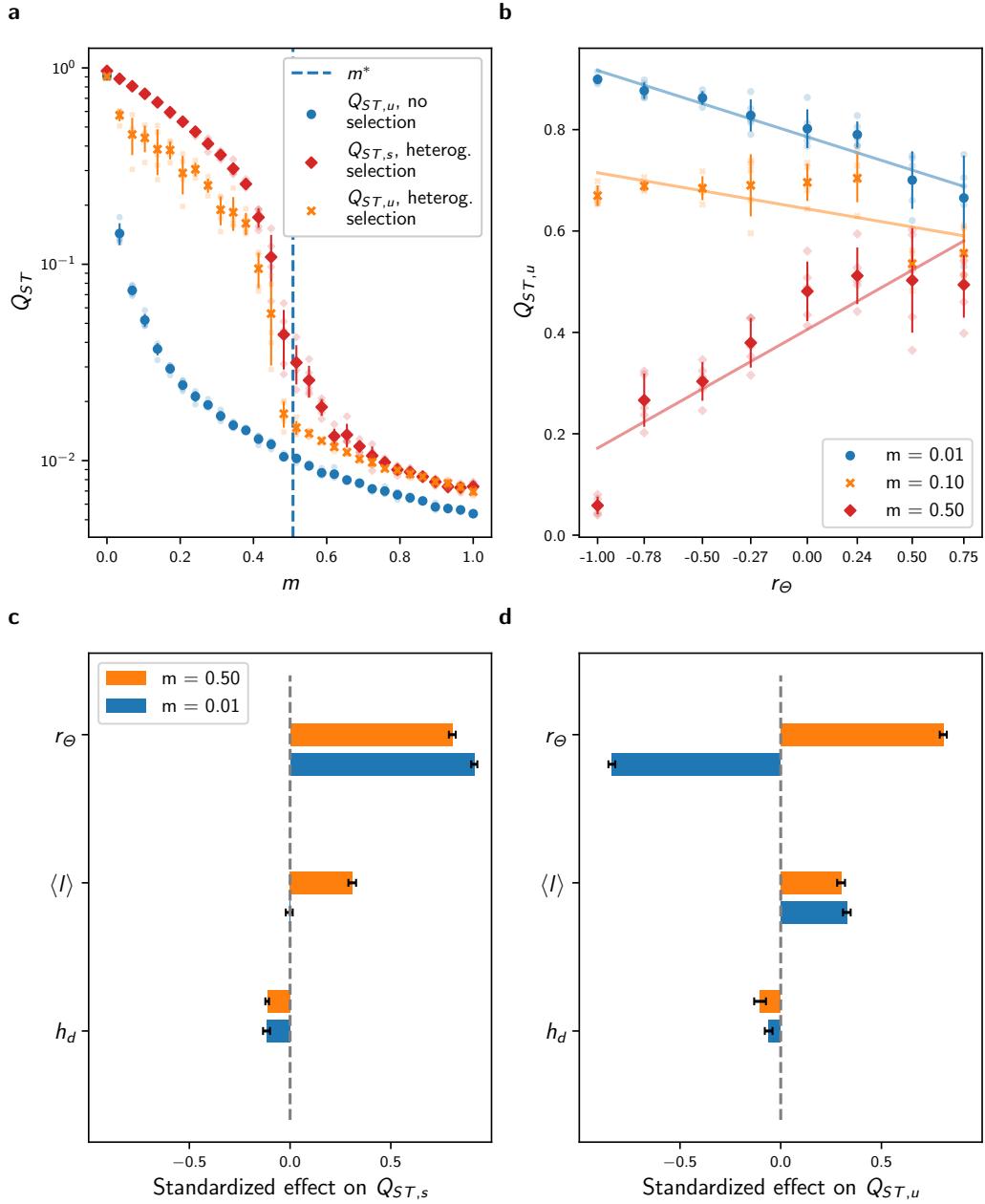


**Fig. 2.4:** Effect of habitat heterogeneity  $r_\Theta$  on  $Q_{ST,s}$  and average population size  $\bar{N}$  for general graph ensembles. (a) Effect of  $r_\Theta$  on  $Q_{ST,s}$  for all undirected connected graphs with  $M = 7$  vertices and varying  $r_\Theta$ , for  $m = 0.1$ . (b) Effect of  $r_\Theta$  on average population size  $\bar{N}$  for the same simulations. In (a) and (b), each dot represents average results from 5 replicate simulations of the IBM, the colour scale corresponds to the proportion of the graphs with similar  $x$  and  $y$  axis values (graph density), and the blue lines correspond to results obtained from the mean field approximation Eq. (2.5). Insights from Eq. (2.5) are congruent with the IBM simulations for complex habitat connectivity patterns at low  $m$ . Similar results with  $m = 0.5$  are presented in Fig. S2.11.

reinforces  $Q_{ST,u}$  compared with a setting without selection (Fig. 2.5a). Nonetheless, previous results show that adaptive differentiation  $Q_{ST,s}$  vanishes on a disassortative graph when  $m > m^*$ , implying that individuals become equally fit in all habitats. In this case, the isolation effect of heterogeneous selection is lost and  $Q_{ST,u}$  reaches a similar level as in the setting with no selection for  $m > m^*$  (Fig. 2.5a), although  $Q_{ST,u}$  is slightly higher in the setting with heterogeneous selection due to a lower population size ( $\bar{N} = bK(1 - p\theta)$  vs.  $\bar{N} = bK$ , see section above and Methods). This suggests that  $r_\Theta$  reinforces  $Q_{ST,u}$ , as assortative graphs sustain higher levels of adaptive differentiation (Figs. 2.3 and 2.4). Simulations on the path graph with varying  $\Theta$ -spatial distribution support this conclusion for high migration regimes, but show the opposite relationship under low migration regimes, where the habitat assortativity  $r_\Theta$  decreases  $Q_{ST,u}$  (Fig. 2.5b). Assortative graphs are composed of large clusters of vertices with similar habitats, within which migrants can circulate without fitness losses. Local neutral trait distributions become more correlated within these clusters, resulting in a decline in  $Q_{ST,u}$  for assortative graphs compared with disassortative graphs. Figure 2.5b therefore highlights the ambivalent effect of  $r_\Theta$  on  $Q_{ST,u}$ .  $r_\Theta$  reinforces  $Q_{ST,u}$  by favouring adaptive differentiation, but also decreases  $Q_{ST,u}$  by decreasing population isolation within clusters of vertices with the same habitat type.

We compare the effect of  $r_\Theta$  on  $Q_{ST,u}$  to the effect of the topology metrics  $\langle l \rangle$  and  $h_d$  found in the setting with no selection using a multivariate regression analysis on simulation results obtained for different graphs with varying  $\Theta$ -spatial distribution (Fig. 2.5d for graphs with  $M = 7$  vertices and Fig. S2.12b for  $M = 9$ ). The multivariate model explains the

discrepancies in  $Q_{ST,u}$  across the simulations for low and high migration regimes (see Table S2.3 for details), and we find that  $r_\Theta$ ,  $\langle l \rangle$  and  $h_d$  contribute similarly to neutral differentiation. Hence, the effects of  $r_\Theta$  and the topology metrics  $\langle l \rangle$  and  $h_d$  add up under heterogeneous selection. A change in sign of the standardized effect of  $r_\Theta$  on  $Q_{ST,s}$  for low and high migration regimes verifies that the ambivalent effect of  $r_\Theta$  on  $Q_{ST,u}$  found on the path graph holds for general graph ensembles. Simulations with trait-dependent competition and simulations on realistic graphs with a continuum of habitat types equally confirm the ambivalent effect of  $r_\Theta$  and further support the complementary effect of  $\langle l \rangle$  and  $h_d$  on  $Q_{ST,u}$  (see Fig. S2.13).  $\langle l \rangle$  and  $h_d$  therefore drive neutral differentiation with and without heterogeneous selection.  $r_\Theta$  becomes an additional determinant of neutral differentiation under heterogeneous selection. In contrast to the non-ambivalent, positive effect of habitat assortativity on adaptive differentiation,  $r_\Theta$  can amplify or depress neutral differentiation depending on the migration regime considered.



**Fig. 2.5: Effect of  $r_\Theta$ ,  $\langle l \rangle$  and  $h_d$  on  $Q_{ST,s}$  and  $Q_{ST,u}$  in the setting with heterogeneous selection.** (a) Comparison of the response of  $Q_{ST,u}$  to migration with the response of  $Q_{ST,u}$  in the setting with no selection for the complete graph. The dashed vertical blue line corresponds to the critical migration regime  $m^*$  predicted by Eq. (2.7). Heterogeneous selection increases  $Q_{ST,u}$  when  $m < m^*$ , but local adaptation is lost when  $m > m^*$ , and in this case  $Q_{ST,u}$  reaches similar levels as  $Q_{ST,u}$  in the setting with no selection. (b) Response of  $Q_{ST,u}$  to  $r_\Theta$  and migration for the path graph.  $r_\Theta$  correlates positively with  $Q_{ST,u}$  for high  $m$ , but correlates negatively for low  $m$ . In (a–b), each plain dot represents average results from 5 replicate simulations, the bars represent one standard deviation, and each fade dot represents a single replicate value. (c–d) Standardized effect of  $h_d$ ,  $\langle l \rangle$  and  $r_\Theta$  on  $Q_{ST,s}$  and  $Q_{ST,u}$  obtained from a multivariate regression model independently fitted for low and high migration regimes on average results from 5 replicate simulations of the IBM on all undirected connected graphs with  $M = 7$  vertices and varying  $r_\Theta$  (see Methods). The ambivalence of the effect of  $r_\Theta$  on  $Q_{ST,u}$  found for the path graph holds for general graph ensembles and adds up to that of  $\langle l \rangle$  and  $h_d$ . Error bars show 95% confidence intervals. Analogous results on graphs with  $M = 9$  vertices are presented in Fig. S2.12 and all regression details can be found in Table S2.3.

## 2.2 Results

## 2.3 Discussion

Using analytical tools and simulations, we have built upon a graph representation of landscapes and a stochastic individual-based model to investigate how landscape features drive phenotypic differentiation. Our study is based on a first principles modelling approach (Champagnat et al., 2006) describing the stochastic dynamics of individuals and capturing the interplay between population dynamics, phenotypic evolution and spatial dynamics in heterogeneous habitats. In contrast to metacommunity models (Holland and Hastings, 2008; Gilarranz and Bascompte, 2012; Mari et al., 2014; Gravel et al., 2016; Carrara et al., 2012; Thompson et al., 2017; Suzuki and Economo, 2021) and evolutionary metacommunity models (Economo and Keitt, 2007; Economo and Keitt, 2010), we have focused on differentiation at the population level. Quantitative genetics and population genetics studies have investigated the effect of topology on differentiation under the assumption of non-overlapping generations, constant population sizes and regular spatial structures (Bürger, 2000; Kimura and Weiss, 1964; Lande, 1991; Nagylaki, 1994; Yeaman and Otto, 2011). Generalising beyond these assumptions, our modelling framework accounts for population dynamics and includes competition and frequency-dependent selection. The systematic investigation of the effect of topology on differentiation over general graph ensembles and under different ecological settings shows that average path length  $\langle l \rangle$ , homogeneity in vertex degree  $h_d$  and habitat assortativity  $r_\Theta$  contribute equally to differentiation. These results support correlative studies that have associated population differentiation (Manel et al., 2003; McRae and Beier, 2007) and species richness (Liu et al., 2018; Dias et al., 2014; Rahbek and Graves, 2001; Kreft and Jetz, 2007; Davies et al., 2007; Veech and Crist, 2007; Guégan et al., 1998; Stein et al., 2014) with a variety of metrics used as surrogates for connectivity, connectivity heterogeneity and habitat heterogeneity. To further our understanding of the origin of spatial biodiversity patterns, the contribution of landscape properties to discrepancies in population differentiation could be investigated at large scales by (i) using techniques to project real landscapes on graphs (see Fig. S2.13a–b); (ii) characterising the landscape features with  $\langle l \rangle$ ,  $h_d$  and  $r_\Theta$ ; and (iii) relating the obtained metrics maps to observation data. More generally, the proposed eco-evolutionary model on spatial graphs could be combined with approximate bayesian computation to estimate ecological, spatial and evolutionary processes of real populations from observation data, similarly to (Lepers et al., 2021). This approach might improve current inferential techniques based on models that do not account for competition nor heterogeneous selection (see e.g. (Petkova et al., 2015)). Overall, our results point to topology metrics that can connect spatial biodiversity patterns to the generating eco-evolutionary and spatial processes.

In the absence of selection, neutral differentiation is more pronounced on graphs with a high average path length  $\langle l \rangle$ , but is also negatively associated with homogeneity in degree  $h_d$  (Fig. 2.2c–d).  $\langle l \rangle$  generalizes the concept of dimensionality in (Kimura and Weiss, 1964; Lande, 1991; Nagylaki, 1994), where it is shown that differentiation is lower for two-dimensional grid graphs compared with path graphs.  $\langle l \rangle$  also closely relates to the concept of resistance distance shown theoretically and empirically to drive genetic

differentiation (McRae, 2006; McRae and Beier, 2007). At the species level, a similar effect of  $\langle l \rangle$  on  $\beta$ -diversity (pairwise differences in species composition) has been reported with the graph metacommunity model of (Carrara et al., 2012) and with the graph eco-evolutionary metacommunity model of (Econo and Keitt, 2007). Accounting for population dynamics and specifically including competition processes, we have shown that not only  $\langle l \rangle$  but also  $h_d$  affects neutral phenotypic differentiation (Fig. 2.2c,d). Our model realistically assumes that population growth is limited by the local carrying capacity. The latter becomes saturated on highly connected vertices in irregular graphs, an effect that has been experimentally documented in microcosm experiments (Altermatt and Fronhofer, 2018). As a consequence, central vertices behave as bottlenecks and amplify the isolation of peripheral vertices (Orsini et al., 2013). The role of  $h_d$  cannot be captured with classical metapopulation and quantitative genetics models or with models of evolutionary dynamics in graphs, as they assume constant population size. This behaviour should be prevalent in patchy landscapes where interspecific competition is high because of limiting resources. Our study highlights that heterogeneity in connectivity can reinforce differentiation patterns through the creation of unbalanced migration fluxes which affect ecological equilibrium.

Habitat assortativity  $r_\Theta$  is a useful indicator for assessing how the spatial distribution of habitat types modulates local adaptation and adaptive differentiation in complex landscapes (Richardson et al., 2014). While adaptation has been extensively studied along environmental gradients (Slatkin, 1973; Slatkin, 1978; Kirkpatrick and Barton, 1997; Polechová and Barton, 2015; Polechová, 2018; AndradeRestrepo et al., 2019; Doeblei and Dieckmann, 2003), landscapes can be patchy and it is unrealistic to assume regularity (Dale and Fortin, 2010). Our model of heterogeneous selection on spatial graphs extends the two-habitat setting investigated in (Meszéna et al., 1997; Yeaman and Otto, 2011; Débarre et al., 2013; Mirrahimi and Gandon, 2020) and captures irregularity in connectivity between distinct habitats (Dale and Fortin, 2010). Similarly to the aforementioned studies, we have found a critical migration regime  $m^*$  that dictates the possibility of adaptation. Equation (2.7) indicates that  $m^*$  increases with increasing selection strength  $p$  and with increasing environmental heterogeneity  $\theta$ , the latter playing a similar role as the slope of the environmental gradient in (Slatkin, 1973; Slatkin, 1978; Polechová and Barton, 2015; Polechová, 2018). Local adaptation would consequently be sustained under higher migration regimes following an increase in these parameters. Additionally, the critical migration regime  $m^*$  in Eq. (2.7) involves the habitat assortativity  $r_\Theta$ , which must be regarded as a measure of habitat spatial auto-correlation based on the dispersal range of a species (Richardson et al., 2014). Our results indicate that for general habitat distributions,  $r_\Theta$  is the main determinant of adaptive differentiation under sufficiently strong selection  $p$  and high habitat heterogeneity  $\theta$ , irrespective of the graph topology (Fig. 2.5c, Fig. S2.12a and Fig. S2.13). As  $p$  decreases, however, the effect of stochastic drift on  $Q_{ST,s}$  should increase, and in this case the topology metrics  $\langle l \rangle$  and  $h_d$  should become the most important determinants of  $Q_{ST,s}$ . Our results predict that in landscapes with heterogeneous habitats and where selection is strong, populations structured over assortative habitats are larger, support higher adaptive differentiation, and can be locally well-adapted even in the case where migration rates are high.

Spatial eco-evolutionary feedbacks in heterogeneous habitats can critically affect differentiation (Richardson et al., 2014). While most eco-evolutionary studies have investigated

diversification by considering a unique adaptive trait (Doebeli and Dieckmann, 2003; Kirkpatrick and Barton, 1997; Polechová and Barton, 2015; Polechová, 2018), distinguishing between neutral and adaptive processes is crucial (Holderegger et al., 2006) and our work underlines the distinct responses of neutral and adaptive differentiation to landscape features (Fig. 2.5c vs. Fig. 2.5d). Our study builds upon recent mathematical models that consider the co-evolution of neutral and adaptive traits (Billiard et al., 2015; Anceschi et al., 2019) and extends those works to a spatial context. Our work provides an analytical framework to the concept of isolation by environment (IBE) (Orsini et al., 2013), which has been suggested to be one of the most important mechanisms governing differentiation in nature (Wang and Bradburd, 2014). Heterogeneous selection leads to more isolation by modifying the fitness of migrants (Polechová, 2018), which further reduces gene flow (Richardson et al., 2014) and therefore affects the level of neutral differentiation (Fig. 2.5a) (Garant et al., 2007). Our work proposes a mechanism by which habitat assortativity, relative to the migration regime, controls the direction of the effect of habitat heterogeneity on differentiation (Fig. 2.5d). Patchy, heterogeneous habitats can promote neutral differentiation as a result of selection that reduces effective migration (Stein et al., 2014). Nonetheless, adaptive differentiation decreases substantially when migration is high relative to the critical migration regime  $m^*$ . In this case, neutral differentiation should be higher in landscapes with more aggregated habitats (Richardson et al., 2014). Our study suggests that habitat assortativity must be considered for a complete understanding of differentiation in complex environments (Stein et al., 2014).

In conclusion, we have established how differentiation can emerge at the population level from eco-evolutionary feedbacks in complex landscapes by using an analytical description of micro-evolutionary processes explicitly accounting for spatial dynamics over graphs. Our study formalizes how differentiation emerges from the interplay between spatial dynamics, the co-evolution of neutral and adaptive traits, and landscape properties. Connectivity and habitat assortativity emerge as core determinants of differentiation in spatial graphs. These results resonate with empirical findings and previous theoretical works. Our study further stresses that habitat assortativity can depress or foster neutral differentiation depending on the migration regime. Additionally, our work highlights that heterogeneity in connectivity is an equally strong determinant of differentiation because highly connected habitats behave as bottlenecks, increasing the isolation of peripheral habitats. The present approach offers a promising framework for studying complex adaptive systems, as it can elucidate how macroscopic properties emerge from microscopic processes acting upon agents structured over complex spatio-evolutionary structures.

## 2.4 Methods

### 2.4.1 Mean field approximation

In the setting with no selection, the mean field approach involves the assumption that all vertices having the same degree are equivalent. For this, let  $P(k, k')$  denote the proportion of edges that map a vertex with degree  $k$  to a vertex with degree  $k'$ , and consider the average

population size  $\bar{N}_t^{(k)}$  in each vertex with degree  $k$  at time  $t$ . An individual has probability  $P(k, k')/k'$  to migrate from a vertex with degree  $k'$  to a vertex with degree  $k$ . Viewing  $a_{i,j}/d_j$  as the probability that an individual on  $v_i$  chosen for migration moves to  $v_j$ , Eq. (2.3) then transforms into

$$\partial_t \bar{N}_t^{(k)} = \bar{N}_t^{(k)} \left[ b(1-m) - \frac{\bar{N}_t^{(k)}}{K} \right] + mbk \sum_{k' \in V} \frac{P(k, k')}{k'} \bar{N}_t^{(k')} \quad (2.9)$$

Assuming uncorrelated graphs for which  $P(k, k')/k' = P(k')^k/\langle k \rangle$ , where  $\langle k \rangle$  denotes the average degree of the graph (Colizza et al., 2007), yields

$$\partial_t \bar{N}_t^{(k)} = \bar{N}_t^{(k)} \left[ b(1-m) - \frac{\bar{N}_t^{(k)}}{K} \right] + mb \frac{k}{\langle k \rangle} \bar{N}_t \quad (2.10)$$

where

$$\bar{N}_t = \sum_k P(k) \bar{N}_t^{(k)}. \quad (2.11)$$

When solving for the stationary state and setting  $m = 1$ , one obtains  $\bar{N}^{(k)} = \sqrt{bK \frac{k}{\langle k \rangle} \bar{N}}$  from Eq. (2.10). Combining this with Eq. (2.11) yields

$$\bar{N} = bK \langle \sqrt{k} \rangle^2 / \langle k \rangle \quad (2.12)$$

In the setting with heterogeneous selection, the mean field approach involves the assumption that all vertices with a similar habitat are equivalent. In this case, an individual from a vertex of habitat type  $\bullet$  has the probability  $P(\bullet, \bullet)/P(\bullet)$  of migrating to a vertex of type  $\bullet$ , and therefore Eq. (2.4) transforms into

$$\begin{aligned} \partial_t \bar{n}_t^\bullet(s) &= \bar{n}_t^\bullet(s) \left[ b^\bullet(s)(1-m) - \frac{1}{K} \int_S \bar{n}_t^\bullet(\mathbf{s}) d\mathbf{s} \right] + \frac{1}{2} \mu \sigma_\mu^2 \Delta_s [b^\bullet(s) \bar{n}_t^\bullet(s)] \\ &\quad + m \sum_{i \in \{\bullet, \bullet\}} b_i(s) \frac{P(\bullet, i)}{P(i)} \bar{n}_t^i(s) \end{aligned} \quad (2.13)$$

Considering that  $P(\bullet) = P(\bullet) = \frac{1}{2}$  (habitats are equally distributed),  $P(\bullet, \bullet) + P(\bullet, \bullet) = P(\bullet)$  (sum of conditional expectations), and  $r_\Theta = 2(P(\bullet, \bullet) - P(\bullet, \bullet))$  (Eq. (2.6)), one obtains

$$P(\bullet, \bullet) = \frac{1}{4}(1 - r_\Theta) \quad \text{and} \quad P(\bullet, \bullet) = \frac{1}{4}(1 + r_\Theta) \quad (2.14)$$

Combining Eq. (2.14) with Eq. (2.13) yields Eq. (2.5). We show in the Supplementary Note how one can derive Eq. (2.6) from the general definition of assortativity given in Eq. (2.8) and initially introduced in (Newman, 2003).

## 2.4.2 Adaptive dynamics on graphs

The adaptive dynamics theory considers a monomorphic population that evolves following a trait substitution process (Meszéna et al., 1997). Accordingly, the trait  $s$  of the monomorphic metapopulation evolves gradually along the direction given by its fitness gradient, until

it reaches a singular strategy  $s^*$  for which the fitness gradient vanishes. By omitting the mutation term, Eq. (2.6) can be written in the matrix form

$$\partial_t \bar{\mathbf{n}}_t(s) = M(s, \bar{\mathbf{N}}_t) \bar{\mathbf{n}}_t(s) \quad (2.15)$$

where  $\bar{\mathbf{n}}_t = (\bar{n}_t^\bullet, \bar{n}_t^\bullet)$  and  $\bar{\mathbf{N}}_t = (\bar{N}_t^\bullet, \bar{N}_t^\bullet)$  are the vectors containing the population densities and the population size on each habitat type, and

$$M(s, \bar{\mathbf{N}}) = \begin{bmatrix} \tau^\bullet(s, \bar{N}^\bullet) & \frac{m}{2}(1 - r_\Theta)b^\bullet(s) \\ \frac{m}{2}(1 - r_\Theta)b^\bullet(s) & \tau^\bullet(s, \bar{N}^\bullet) \end{bmatrix} \quad (2.16)$$

is the so-called projection matrix (Meszéna et al., 1997), with  $\tau^\bullet(s, \bar{N}^\bullet) = b^\bullet(s)(1 + \frac{m}{2}(r_\Theta - 1)) - \bar{N}^\bullet/K$ . The overall fitness of individuals with trait  $s$  is the leading eigenvalue of  $M$ , which we denote with  $\lambda(s, \bar{\mathbf{N}})$ . We obtain the singular strategy  $s^*$  by setting the fitness gradient  $\frac{\partial \lambda}{\partial s}(s, \bar{\mathbf{N}}) = 0$ , from which we further obtain the demographic equilibrium  $\bar{\mathbf{N}}^{s^*}$ . Because of symmetries, we must have  $\bar{N}^{\bullet, s^*} = \bar{N}^{\bullet, s^*}$  and  $s^* = \frac{\theta_\bullet + \theta_\bullet}{2} = 0$ , such that  $\bar{N}^{\bullet, s^*} = \bar{N}^{\bullet, s^*} = bK(1 - p\theta^2)$ .  $s^*$  is said to be evolutionary stable if no mutants can invade, i.e. if  $s^*$  locally maximizes the fitness of a mutant with trait  $y$  in the resident population with trait  $s^*$ , given by  $\lambda(y, \bar{\mathbf{N}}^{s^*})$  (see (Meszéna et al., 1997) for details). One can show that  $\left[ \frac{\partial \lambda}{\partial y}(y, \bar{\mathbf{N}}^{s^*}) \right]_{y=s^*} = 0$  and the condition for evolutionary stability becomes  $\left[ \frac{\partial^2 \lambda}{\partial y^2}(y, \bar{\mathbf{N}}^{s^*}) \right]_{y=s^*} < 0$ . We compute and simplify this inequality through computer algebra (see Mathematica notebook provided in the simulation code), which leads to Eq. (2.7).

### 2.4.3 Numerical simulations

The model was implemented in a multi-purpose Julia package called `EvoId.jl`, available at <https://github.com/vboussange/EvoId.jl>. For each result presented,  $b = 1$ , local carrying capacity  $K = 150$ , selection strength  $p = 1$ , mutation rate  $\mu = 0.1$ , mutation range  $\sigma_\mu = 5 \cdot 10^{-2}$ , and total time span  $t = 1000$ . This parameter choice made it possible to discard transient dynamics while obtaining results in a reasonable computational time (see Fig. S2.14). In settings (1) and (2), we ran simulations on all of the 853 undirected connected graphs with  $M = 7$  vertices and on 1126 of the 261,080 undirected connected graphs with  $M = 9$  vertices, listed at <http://oeis.org/A001349>. Graphs with  $M = 9$  vertices were selected with a stratified sampling method: we randomly sampled without replacement a maximum of 50 graphs for each class of graphs with an equal number of vertices. For the setting with heterogeneous selection, we generated the labeled graphs by randomly generating  $\Theta$ -spatial distributions, and by using a stratified sampling strategy to select without replacement at most 3 and 2  $\Theta$ -spatial distributions corresponding to the quartiles of the  $r_\Theta$  values obtained, respectively for graphs with  $M = 7$  and  $M = 9$  vertices. This sampling strategy allowed to obtain a uniform distribution of the topology metrics investigated in the study, and therefore permitted to correctly represent the population of graphs to investigate their effect on differentiation. We then computed  $Q_{ST,u}$  and  $Q_{ST,s}$ , which we further averaged over the last time steps and across the replicates. Since the dynamics of  $Q_{ST,u}$  is characterized by large quadratic variations, we simulated individuals

with  $d = 300$  neutral traits, where each trait can independently be affected by mutations.  $Q_{ST,u}$  values presented were then obtained from the average  $Q_{ST,u}$  for each trait. This reduced the variance of the numerical simulations and is also biologically meaningful because populations are characterized by many traits, most of which are neutral (Holderegger et al., 2006). As initial conditions,  $MK$  individuals were homogeneously distributed over all of the vertices, with traits centred on 0 and with standard deviation  $\sigma_\mu$ . Graph metrics used for the meta-analysis were calculated using the **LightGraphs.jl** library (Bromberger and Contributors, 2017). We numerically solved the PDEs with a finite difference scheme using **DifferentialEquations.jl** (Rackauckas and Nie, 2017), ensuring that the domain was large enough to avoid border effects.

#### 2.4.4 Statistics and reproducibility

Statistical analyses were conducted in Julia using **StatsKit.jl**. All simulations can be exactly reproduced from the code available at <https://github.com/vboussange/differentiation-in-spatial-graphs>.

### 2.5 Data availability

The data underlying our figures is available at <https://github.com/vboussange/differentiation-in-spatial-graphs>.

### 2.6 Code availability

The simulation code is available at <https://github.com/vboussange/differentiation-in-spatial-graphs>.

### 2.7 Acknowledgements

We thank Thomas Poulet, Sylvian Billiard, Sepideh Mirrahimi, Heike Lischke, Joshua Payne, Conor Waldock, Yaquan Chang, Flora Desmet, Benjamin Flück and Alexander Skeels for helpful discussions and comments on the manuscript. L.P. and V.B. were supported by the Swiss National Science Foundation grant (No 310030\_188550). We thank two anonymous reviewers for constructive comments and valuable suggestions on a previous version of this article.

## 2.A Supplementary Note

### 2.A.1 Mathematical construction of the model

The model is a measure-valued point process (Bansaye and Méléard, 2015), so that individuals are represented as dirac functions  $\delta_{x_k^{(i)}}$ , where  $x_k^{(i)} \in \mathcal{X}$  corresponds to the traits' value of individual  $k$  located on vertex  $v_i$ . Under this formalism, the population on  $v_i$  is represented as a sum of dirac functions  $\nu_t^{(i)} = \sum_k^{N^{(i)}} \delta_{x_k^{(i)}}$ , where  $N^{(i)}$  is the local population size. It follows that the time variation of the process can be described by the so-called infinitesimal generator  $L$ , defined for all real valued functions  $\phi$  as

$$L\phi(\nu_t^{(i)}) = \partial_t \mathbb{E} [\phi(\nu_t^{(i)})] \quad (\text{S2.17})$$

(see (Linke, 2015) for an introduction to infinitesimal generators). Equation (S2.17) provides the expected time variation at time  $t$  of e.g. the population size by choosing  $\phi(\nu_t^{(i)}) = \int_{\mathcal{X}} \nu_t^{(i)}(dx)$ . Recall that we use  $b^{(i)}$  to denote the birth rate on vertex  $v_i$ ,  $d$  for the death rate,  $\mu$  for the mutation probability,  $m$  for the migration probability,  $\mathcal{M}(x, y) = \frac{1}{\sqrt{2\pi}\sigma_\mu} \exp\left(-\frac{\|x-y\|^2}{2\sigma_\mu^2}\right)$  for the mutation kernel,  $K$  for the local carrying capacity,  $A = (a_{i,j})_{1 \leq i,j \leq M}$  for the adjacency matrix of the graph  $G$ , and  $D = (d_1, d_2, \dots, d_M)$  for the vector containing the degree of each vertex. In order to explicitly write the generator  $L$ , let us recall that five events of different natures can alter the number of individuals with trait  $x$  on vertex  $v_i$ :

- an individual on  $v_i$  with trait  $x$  can give birth to an offspring that does not experience mutations nor migration, at rate  $(1 - \mu)(1 - m)b^{(i)}(x)$ ,
- an individual on  $v_i$  with trait  $y$  can give birth to an offspring with mutated trait  $x$  that does not experience migration, at rate  $\mu(1 - m)\mathcal{M}(x, y)b^{(i)}(y)$ ,
- an individual on  $v_i$  with trait  $x$  can die, at rate  $d(N^{(i)}) = \frac{N^{(i)}}{K} = \frac{1}{K} \int_{\mathcal{X}} \nu_t^{(i)}(dx)$ ,
- an individual on  $v_j$  with trait  $x$  can give birth to an offspring that does not experience mutations and migrates to  $v_i$ , at rate  $\frac{a_{i,j}}{d_j}(1 - \mu)m b^{(j)}(x)$ ,
- an individual on  $v_j$  with trait  $y$  can give birth to an offspring with mutated trait  $x$  that migrates to  $v_i$ , at rate  $\frac{a_{i,j}}{d_j}\mu m \mathcal{M}(x, y)b^{(j)}(x)$ .

Summing over all all individuals and all vertices yields

$$\begin{aligned}
L\phi(\nu_t^{(i)}) &= \int_{\mathcal{X}} \left\{ b^{(i)}(\mathbf{x})(1-\mu)(1-m)(\phi(\nu_t^{(i)} + \delta_{\mathbf{x}}) - \phi(\nu_t^{(i)})) \right\} \nu_t^{(i)}(d\mathbf{x}) \\
&\quad \text{births w/o mutations, w/o migrations} \\
&+ \int_{\mathcal{X}} \left\{ \mu(1-m) \int_{\mathcal{X}} b^{(i)}(y)(\phi(\nu_t^{(i)} + \delta_z) - \phi(\nu_t^{(i)})) \mathcal{M}(\mathbf{x}, y) dy \right\} \nu_t^{(i)}(d\mathbf{x}) \\
&\quad \text{births w/ mutations, w/o migrations} \\
&+ \iint_{\mathcal{X}} \left\{ \frac{1}{K} (\phi(\nu_t^{(i)} - \delta_{\mathbf{x}}) - \phi(\nu_t^{(i)})) \nu_t^{(i)}(dy) \nu_t^{(i)}(dx) \right\} \\
&\quad \text{deaths} \\
&+ \sum_{j \neq i} \frac{a_{i,j}}{d_j} \int_{\mathcal{X}} \mu m \left\{ \int_{\mathcal{X}} b^{(j)}(y)(\phi(\nu^{(j)} + \delta_{\mathbf{x}}) - \phi(\nu^{(j)})) \mathcal{M}(\mathbf{x}, y) dy \right\} \nu_t^{(j)}(d\mathbf{x}) \\
&\quad \text{migrations w/ mutations} \\
&+ \sum_{j \neq i} \frac{a_{i,j}}{d_j} \int_{\mathcal{X}} \left\{ b^{(j)}(\mathbf{x})(1-\mu)m(\phi(\nu^{(j)} + \delta_{\mathbf{x}}) - \phi(\nu^{(j)})) \right\} \nu_t^{(j)}(d\mathbf{x}). \quad (\text{S2.18}) \\
&\quad \text{migrations w/o mutations}
\end{aligned}$$

Taking expectations in Eq. (S2.18), one can obtain an equation for the mean trajectory of the quantity of interest,  $\mathbb{E} [\phi(\nu_t^{(i)})]$ . Nonetheless, Eq. (S2.18) involves an integral with respect to  $\nu_t^{(i)}(dx)\nu_t^{(i)}(dy)$ , making it impossible to obtain an explicit solution. It is therefore unclear whether one can gain insight into the stochastic dynamics from Eq. (S2.18) without simplifying assumptions. We refer to (Champagnat et al., 2006) for a detailed discussion on the topic.

## 2.A.2 Deterministic approximation

One strategy to overcome the difficulties encountered above is to assimilate the process to its mean trajectory, assuming that  $\mathbb{E} [\nu_t^{(i)}] \approx \nu_t^{(i)}$  and further approximating  $\nu_t^{(i)}$  with a continuous deterministic function  $n_t^{(i)}$ . Such strategy inherently neglects the stochasticity of the process, which is reasonable provided that a force dampens the stochastic fluctuations of the quantity of interest.

### Setting with no selection

Consider a setting with no selection and recall that in this setting where  $x \equiv u \in \mathcal{X} = \mathcal{U}$  we define

$$b^{(i)}(x) \equiv b \quad (\text{S2.19})$$

By applying the strategy mentioned above and choosing  $\phi(n_t^{(i)}) = \int_{\mathcal{X}} n_t^{(i)}(x) dx$ , Eq. (S2.18) transforms into the deterministic approximation of the population size dynamics given in the main-text by

$$\partial_t N_t^{(i)} = N_t^{(i)} \left[ b(1-m) - \frac{N_t^{(i)}}{K} \right] + mb \sum_{j \neq i} \frac{a_{i,j}}{d_j} N_t^{(j)}. \quad (\text{S2.20})$$

Competition stabilises the population size dynamics, which behaves deterministically. This is supported by Fig. S2.15a, which shows how Eq. (S2.20) accurately describes the population size for varying migration regimes. Nonetheless, stochastic fluctuations drive the dynamics of the neutral trait distribution. Attempting to characterise the neutral trait distribution with the same strategy, this time setting  $\phi(n_t^{(i)}) = n_t^{(i)}(u)$ , yields

$$\begin{aligned} \partial_t n_t^{(i)}(u) &= n_t^{(i)}(u) \left[ b(1-m)(1-\mu) - \frac{1}{K} \int_{\mathcal{U}} n_t^{(i)}(\mathbf{u}) d\mathbf{u} \right] \\ &\quad + (1-m)\mu b \int_{\mathcal{U}} n_t^{(i)}(\mathbf{u}) \mathcal{M}(u, \mathbf{u}) d\mathbf{u} \\ &\quad + m\mu b \sum_{j \neq i} \frac{a_{i,j}}{d_j} \int_{\mathcal{U}} n_t^{(j)}(u) \mathcal{M}(u, \mathbf{u}) d\mathbf{u} \\ &\quad + m(1-\mu)b \sum_{j \neq i} \frac{a_{i,j}}{d_j} b n_t^{(j)}(u). \end{aligned} \quad (\text{S2.21})$$

Solving for Eq. (S2.21), one can show that the variance of  $n_t^{(i)}$  continuously grows in time (see Fig. S2.15) and tends to infinity as time goes to infinity, which is an unrealistic behaviour considering finite populations. Intuitively, this reflects the fact that no stabilising force acts on the neutral trait distribution, such that random fluctuations play a major role in driving the dynamics of the stochastic process. Figure S2.15 shows how IBM trajectories significantly differ from Eq. (S2.21), and Fig. S2.16 illustrates how diversity metrics obtained from Eq. (S2.21) do not match those obtained from simulations of the IBM.

### Setting with heterogeneous selection

In contrast to the neutral trait dynamics, the adaptive distribution can successfully be approximated by a deterministic description because selection pressure acts as a stabilising force and stabilises the populations' adaptive trait, dampening the stochastic fluctuations. Consider the setting with heterogeneous selection and recall that in this setting where  $x \equiv (s, u) \in \mathcal{X} = \mathcal{S} \times \mathcal{U}$  we define

$$b^{(i)}(x) \equiv b(1 - p(s - \theta_i)^2). \quad (\text{S2.22})$$

By applying the same strategy as above to characterise the adaptive trait distribution  $n_t^{(i)}(s)$  by choosing  $\phi(n_t^{(i)}) = n_t^{(i)}(s) \equiv \int_{\mathcal{U}} n_t^{(i)}(u, s) du$ , Eq. (S2.18) transforms into

$$\begin{aligned} \partial_t n_t^{(i)}(s) &= n_t^{(i)}(s) \left[ b^{(i)}(s)(1-m)(1-\mu) - \frac{1}{K} \int_{\mathcal{S}} n_t^{(i)}(\mathbf{s}) d\mathbf{s} \right] \\ &\quad + (1-m)\mu \int_{\mathcal{S}} b^{(i)}(\mathbf{s}) n_t^{(i)}(\mathbf{s}) \mathcal{M}(\mathbf{s}, s) d\mathbf{s} \\ &\quad + m\mu \sum_{j \neq i} \frac{a_{i,j}}{d_j} \int_{\mathbb{R}} b^{(j)}(\mathbf{s}) n_t^{(j)}(s) \mathcal{M}(\mathbf{s}, s) ds \\ &\quad + m(1-\mu) \sum_{j \neq i} \frac{a_{i,j}}{d_j} b^{(j)}(s) n_t^{(j)}(s). \end{aligned} \quad (\text{S2.23})$$

Assuming that the variance of the mutation kernel is small, one can use a diffusion approximation for the mutation term (Kimura, 1965; Débarre et al., 2013; Mirrahimi and Gandon, 2020)

$$\int_{\mathcal{S}} b^{(i)}(\mathbf{s}) n_t^{(i)}(\mathbf{s}) \mathcal{M}(\mathbf{s}, s) d\mathbf{s} = b^{(i)}(s, t) n_t^{(i)}(s) + \frac{1}{2} \sigma_\mu^2 \Delta_s(b^{(i)} n_t^{(i)})(s). \quad (\text{S2.24})$$

Neglecting the terms in  $m\mu$ , we obtain

$$\begin{aligned} \partial_t n_t^{(i)}(s) &= n_t^{(i)}(s) \left[ b^{(i)}(s, t)(1-m-\mu) - \frac{1}{K} \int_{\mathcal{S}} n_t^{(i)}(\mathbf{s}) d\mathbf{s} \right] \\ &\quad + \mu \left[ b^{(i)}(s, t) n_t^{(i)}(s) + \frac{1}{2} \sigma_\mu^2 \Delta_s(b^{(i)} n_t^{(i)})(s) \right] \\ &\quad + m \sum_{j \neq i} b^{(j)}(s, t) n_t^{(j)}(s) a_{i,j} \end{aligned} \quad (\text{S2.25})$$

which, after rearranging terms, yields the elegant deterministic approximation of the adaptive trait dynamics

$$\begin{aligned} \partial_t n_t^{(i)}(s) &= n_t^{(i)}(s) \left[ b^{(i)}(s)(1-m) - \frac{1}{K} \int_{\mathcal{S}} n_t^{(i)}(\mathbf{s}) d\mathbf{s} \right] \\ &\quad + m \sum_{j \neq i} b^{(j)}(s) \frac{a_{i,j}}{d_j} n_t^{(j)}(s) \\ &\quad + \frac{1}{2} \mu \sigma_\mu^2 \Delta_s \left[ b^{(i)}(s) n_t^{(i)}(s) \right]. \end{aligned} \quad (\text{S2.26})$$

Setting  $m = 0$  (Mirrahimi and Gandon, 2020) shows that Eq. (S2.26) admits a stationary solution that is Gaussian, with variance  $\sqrt{\mu\sigma_\mu^2}/\sqrt{p}$ . Therefore, the variance of the adaptive trait distribution stabilises to a finite value. Intuitively, this reflects the fact that the random fluctuations of the adaptive trait distribution are damped by the stabilising force of selection. Provided that the selection strength  $p$  is large enough, Eq. (S2.26) is a good approximation of the adaptive trait distribution obtained from the stochastic process. Figure S2.8 shows how IBM trajectories are similar to the ones obtained from Eq. (S2.21), and Fig. S2.9 illustrates how diversity

metrics obtained from Eq. (S2.21) match those obtained from simulations of the IBM.

### 2.A.3 Trait-dependent competition

To test whether the effects of the metrics hold under more complex ecological processes, we designed an extra experiment considering heterogeneous selection and adaptive trait-dependent competition, where the death rate of individuals on  $v_i$  with traits  $x_k^{(i)} = (u_k^{(i)}, s_k^{(i)}) \in \mathcal{U} \times \mathcal{S}$  is given by

$$d(x_k^{(i)}, \nu^{(i)}) = \frac{1}{K} \int_{\mathcal{S}} \exp\left(-\frac{(s_k^{(i)} - \mathbf{s})^2}{2\sigma_\alpha^2}\right) \nu^{(i)}(\mathbf{s}) \quad (\text{S2.27})$$

where  $\sigma_\alpha$  is the competition bandwidth. This competition kernel tends to increase the population size, as it decreases the overall competition. The adaptive dynamics theory predicts that when  $m = 0$ , competition promotes two distinct types of individuals at either side of the adaptive trait optimum for a competition bandwidth  $\sigma_\alpha < 1/\sqrt{2p}$ , while a single type is observed when  $\sigma_\alpha > 1/\sqrt{2p}$  (Doebeli, 2011). We performed simulations in both cases for graphs with  $M = 7$  vertices and show results of the multivariate regression analyses in Table S2.5. The analyses demonstrate that the trends reported in the main manuscript remain unchanged in both cases.

### 2.A.4 Derivation of the habitat assortativity metric $r_\Theta$ in binary environments

We demonstrate here how the habitat assortativity  $r_\Theta$  relates to the conditional probability of habitats being connected, and we show how  $r_\Theta$  simplifies under mean field assumption.

Following the original definition of (Newman, 2003), habitat assortativity  $r_\Theta$  is defined as the Pearson correlation of environmental conditions  $\theta$  at either ends of the vertices  $V$  of graph  $G$ , that is

$$r_\Theta = \frac{\text{Cov}(\Theta_x, \Theta_\wedge)}{\sqrt{\text{Var}(\Theta_x)\text{Var}(\Theta_\wedge)}} = \frac{\langle \Theta_x \Theta_\wedge \rangle - \langle \Theta_x \rangle \langle \Theta_\wedge \rangle}{\sqrt{(\langle \Theta_x^2 \rangle - \langle \Theta_x \rangle^2)(\langle \Theta_\wedge^2 \rangle - \langle \Theta_\wedge \rangle^2)}} \quad (\text{S2.28})$$

where  $\Theta_x$  and  $\Theta_\wedge$  denote the sets of environmental conditions found at the toe and tip of each directed vertex of graph  $V$ , and  $\langle \Theta_x \rangle$  and  $\langle \Theta_\wedge \rangle$  denote their respective mean values.

Let  $P(\bullet, \bullet)$  be the proportion of edges that connect a vertex of habitat type  $\bullet$  to a vertex of habitat type  $\bullet$ . One can also view  $P(\bullet, \bullet)$  as the conditional probability that

a vertex of type  $\bullet$  is connected to a vertex of type  $\circ$ . Let  $P(\circ)$  denote the proportion of vertices that are of type  $\circ$ . First observe that for undirected graphs, one has  $\langle \Theta_{\times} \rangle = \langle \Theta_{\wedge} \rangle$  and  $\langle \Theta_{\times}^2 \rangle = \langle \Theta_{\wedge}^2 \rangle$ . Assuming that habitats are symmetric and binary, it follows that  $\theta_{\bullet|} = -\theta_{\bullet|}$ . Then

$$\begin{aligned}\langle \Theta_{\times} \Theta_{\wedge} \rangle &= P(\bullet, \bullet) \theta_{\bullet|}^2 + P(\bullet, \circ) \theta_{\bullet|}^2 + [P(\circ, \bullet) + P(\bullet, \circ)] \theta_{\bullet|} \theta_{\bullet|} \\ &= \theta_{\bullet|}^2 (P(\bullet, \bullet) + P(\bullet, \circ) - [P(\circ, \bullet) + P(\bullet, \circ)]),\end{aligned}\tag{S2.29}$$

$$\begin{aligned}\langle \Theta_{\times} \rangle &= P(\bullet) \theta_{\bullet|} + P(\circ) \theta_{\bullet|} \\ &= \theta_{\bullet|} [P(\circ) - P(\bullet)],\end{aligned}\tag{S2.30}$$

$$\begin{aligned}\langle \Theta_{\times}^2 \rangle &= P(\bullet) \theta_{\bullet|}^2 + P(\circ) \theta_{\bullet|}^2 \\ &= \theta_{\bullet|}^2 [P(\bullet) + P(\circ)] \\ &= \theta_{\bullet|}^2.\end{aligned}\tag{S2.31}$$

Combining Eq. (S2.29), Eq. (S2.30) and Eq. (S2.31) with Eq. (S2.28) one gets

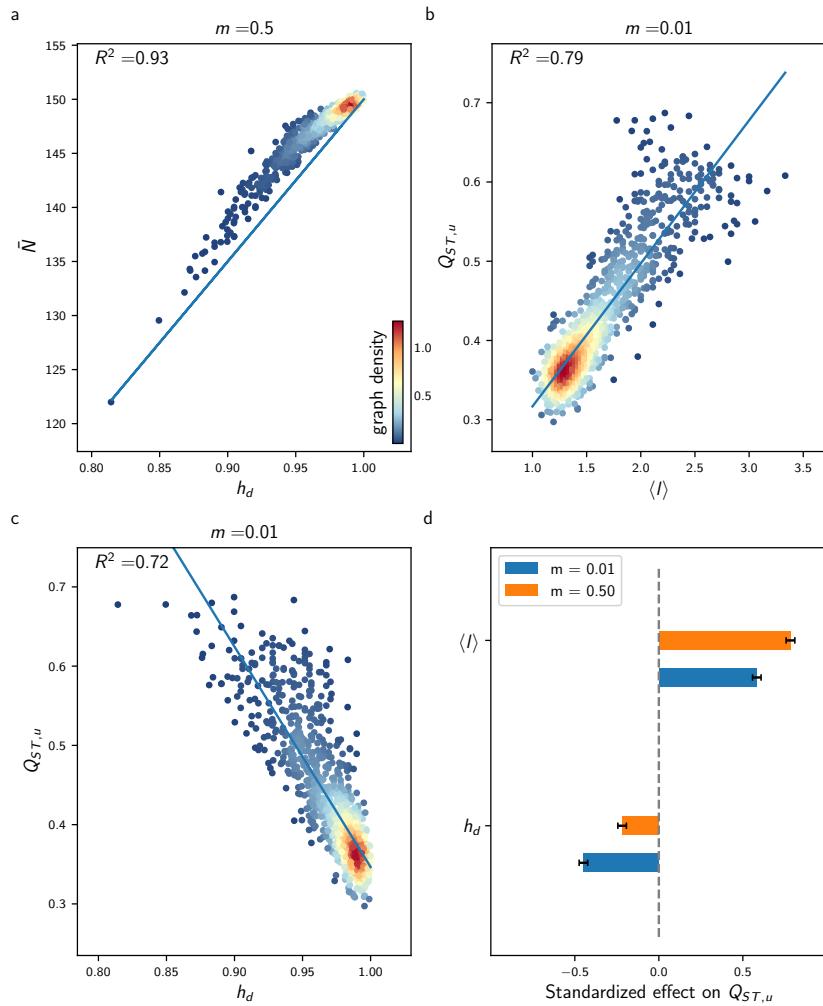
$$\begin{aligned}r_{\Theta} &= \frac{\langle \Theta_{\times} \Theta_{\wedge} \rangle - \langle \Theta_{\times} \rangle \langle \Theta_{\wedge} \rangle}{\langle \Theta_{\times}^2 \rangle - \langle \Theta_{\times} \rangle^2} \\ &= \frac{P(\bullet, \bullet) + P(\bullet, \circ) - [P(\circ, \bullet) + P(\bullet, \circ)] - (P(\circ) - P(\bullet))^2}{P(\bullet) + P(\circ) - (P(\circ) - P(\bullet))^2} \\ &= \frac{P(\bullet, \bullet) + P(\bullet, \circ) - [P(\circ, \bullet) + P(\bullet, \circ)] - (P(\circ) - P(\bullet))^2}{1 - (P(\circ) - P(\bullet))^2}.\end{aligned}\tag{S2.32}$$

Assuming that habitats are homogeneously distributed, we have  $P(\bullet) = P(\circ) = \frac{1}{2}$  and thus we obtain

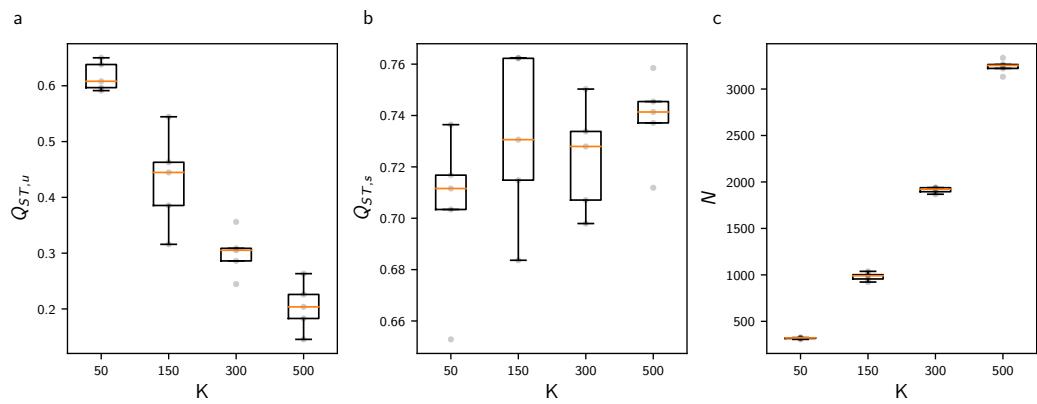
$$r_{\Theta} = P(\bullet, \bullet) + P(\bullet, \circ) - [P(\circ, \bullet) + P(\bullet, \circ)].\tag{S2.33}$$

The mean field approximation involves the assumption that all vertices with similar habitats are equivalent in terms of their connections with other habitats, so that  $P(\bullet, \bullet) = P(\bullet, \circ)$  and  $P(\bullet, \bullet) = P(\circ, \bullet)$ , which yields  $r_{\Theta} = 2(P(\bullet|, \bullet|) - P(\bullet|, \circ|))$ .

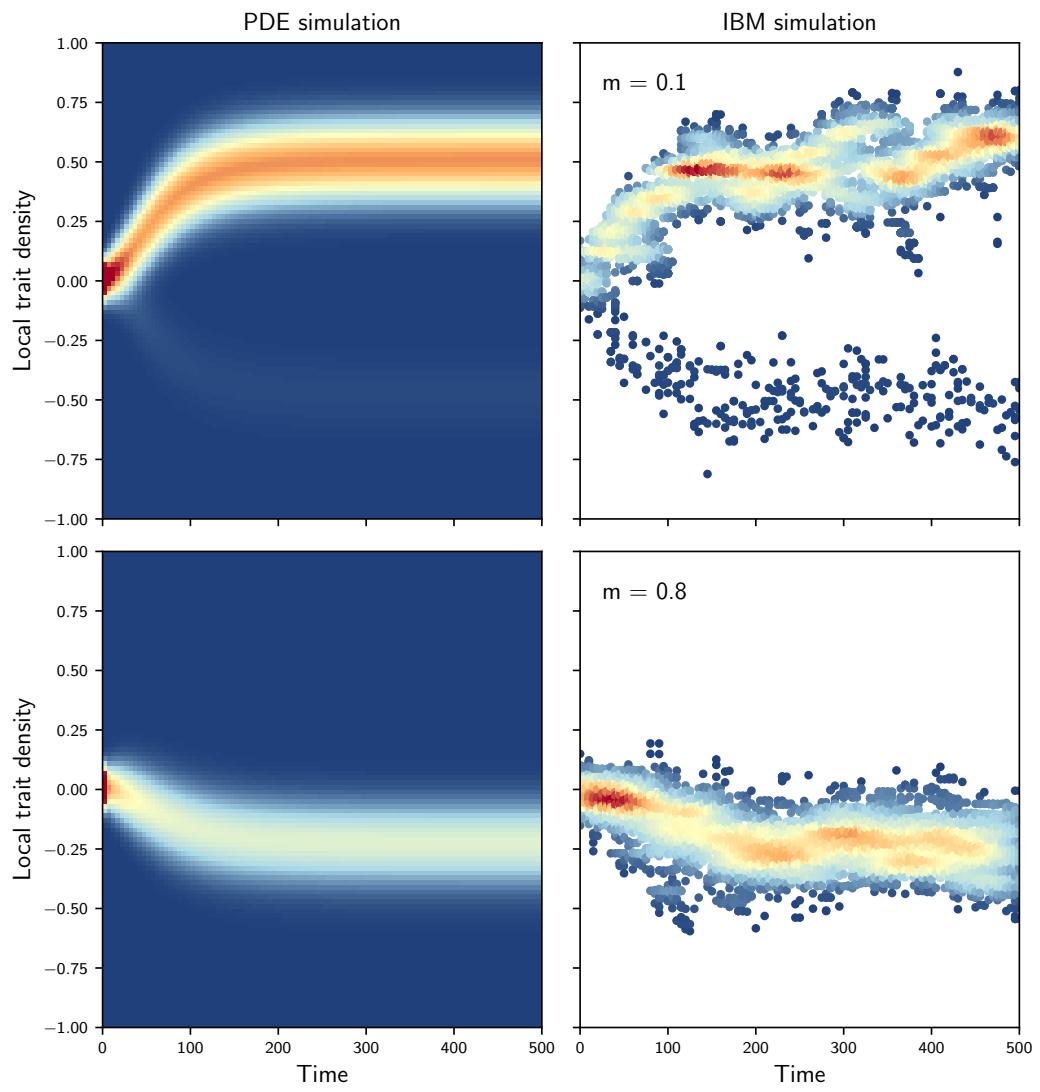
## 2.B Supplementary Figures



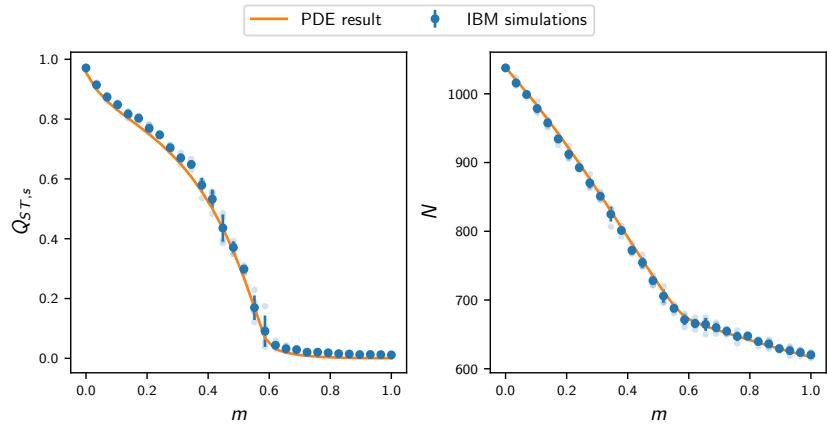
**Fig. S2.6:** Effect of  $\langle I \rangle$  and  $h_d$  on average population size  $\bar{N}$  and neutral differentiation  $Q_{ST,u}$  under the setting with no selection, analogous to Fig. 2.2 but for 1126 of the 261,080 undirected connected graphs with  $M = 9$  vertices.



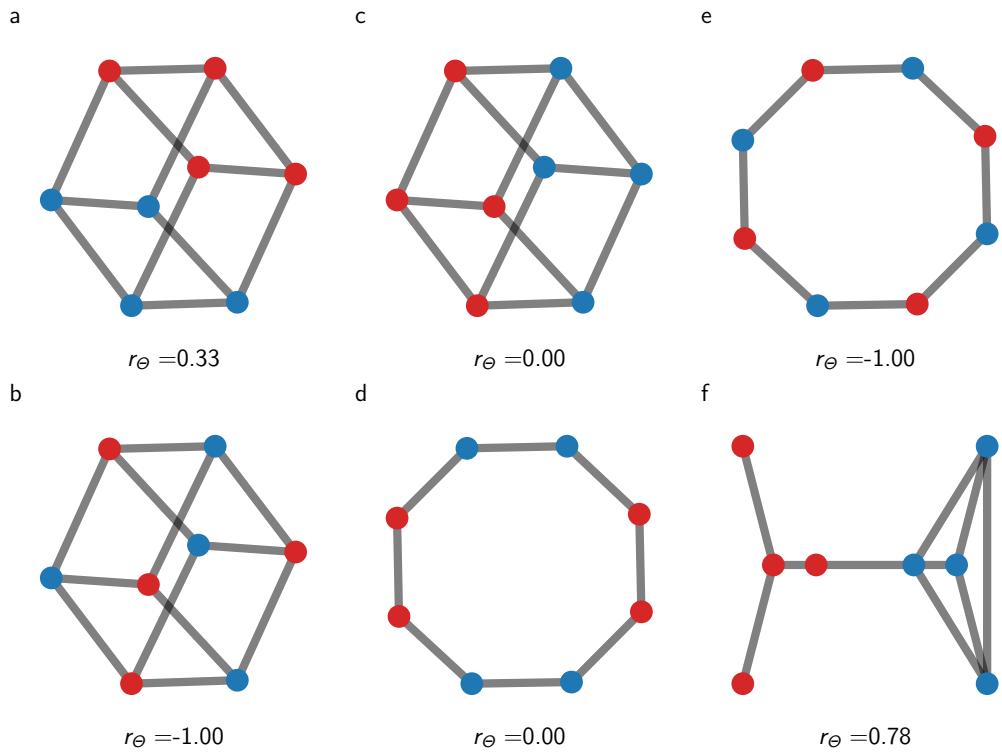
**Fig. S2.7:** Effect of the carrying capacity  $K$  on  $Q_{ST,u}$ ,  $Q_{ST,s}$  and metapopulation size  $N$  for the line graph with  $M = 7$  vertices for  $m = 0.1$ . Decreasing  $K$  increases  $Q_{ST,u}$  as it favours drift, but it does not influence  $Q_{ST,s}$ . Each boxplot is based on 5 replicate simulations of the IBM, and fade dots represent single values for each replicate.



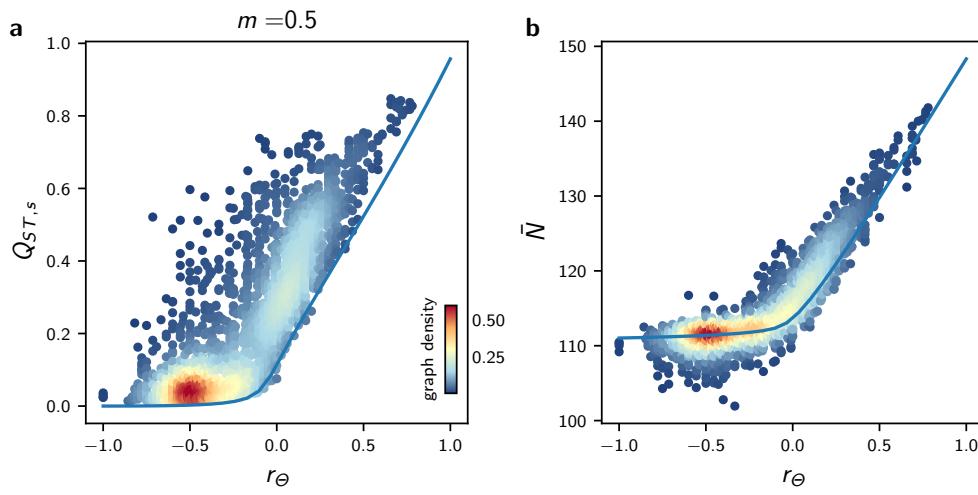
**Fig. S2.8:** Comparison of the adaptive trait density on one vertex obtained from Eq. (S2.26) (left) and from the IBM simulations (right) in the setting with heterogeneous selection, for the star graph with  $M = 7$  vertices. The densities obtained from Eq. (S2.26) and from the IBM are qualitatively similar.



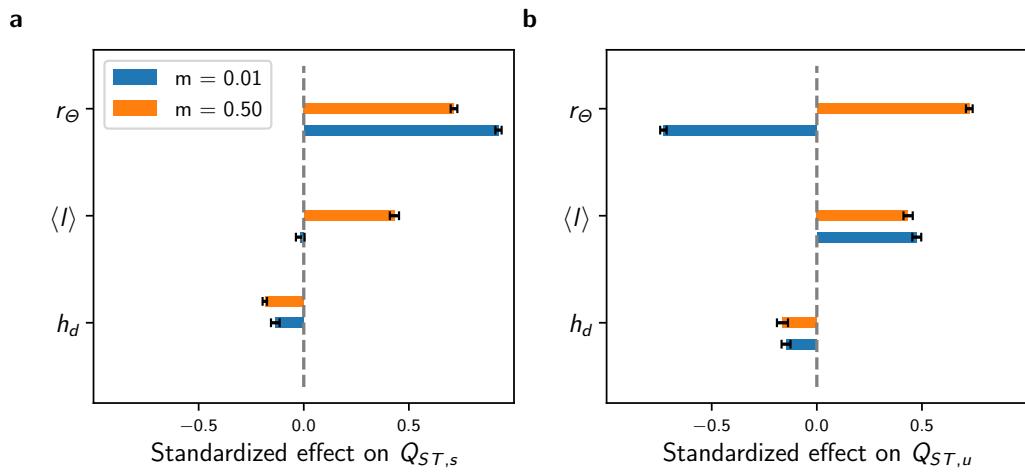
**Fig. S2.9:** Comparison of  $Q_{ST,s}$  and  $N$  obtained from the deterministic approximation Eq. (S2.26) and from IBM simulations in the setting with heterogeneous selection, on the star graph with  $M = 7$  vertices.  $Q_{ST,s}$  and population size obtained from Eq. (S2.26) closely match the IBM simulations. Each plain dot represents average results from 5 replicate simulations of the IBM, bars represent one standard deviation, and each fade dot represents a single replicate value.



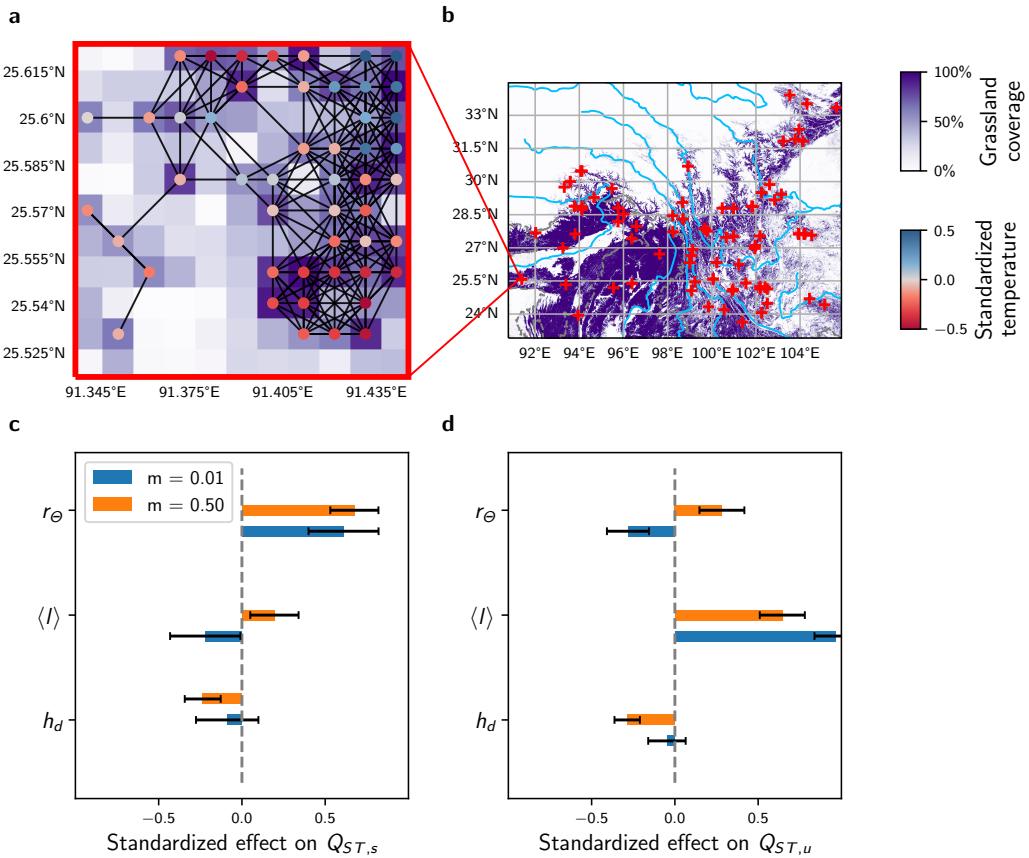
**Fig. S2.10:** Graphs with spatial distribution of habitat types corresponding to different habitat assortativity  $r_\Theta$ . Graphs (ad) can be described exactly with a mean field approach, as blue and red vertices have an equivalent position on the graph.



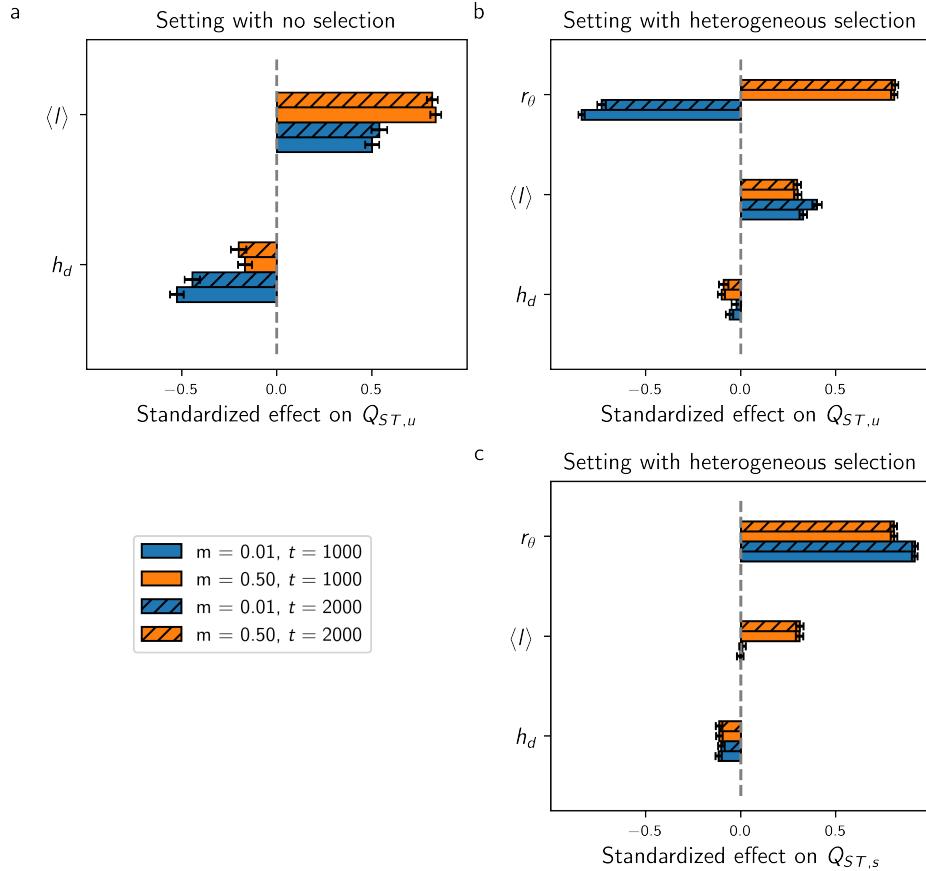
**Fig. S2.11:** Effects of habitat heterogeneity  $r_\Theta$  on  $Q_{ST,s}$  and average population size  $\bar{N}$  for all undirected connected graphs with  $M = 7$  vertices and varying  $r_\Theta$ , obtained for similar simulations to those in Fig. 2.4 with  $m = 0.5$ . In (a) and (b), each dot represents average results from 5 replicate simulations of the IBM, the colour scale corresponds to the proportion of the graph with similar  $x$  and  $y$  axis values (graph density), and the blue lines correspond to results obtained from the mean field, deterministic approximation Eq. (2.5). Deviations from the mean field, deterministic approximation Eq. (2.5) can be explained by differences in  $\langle l \rangle$  and  $h_d$  between the graphs.



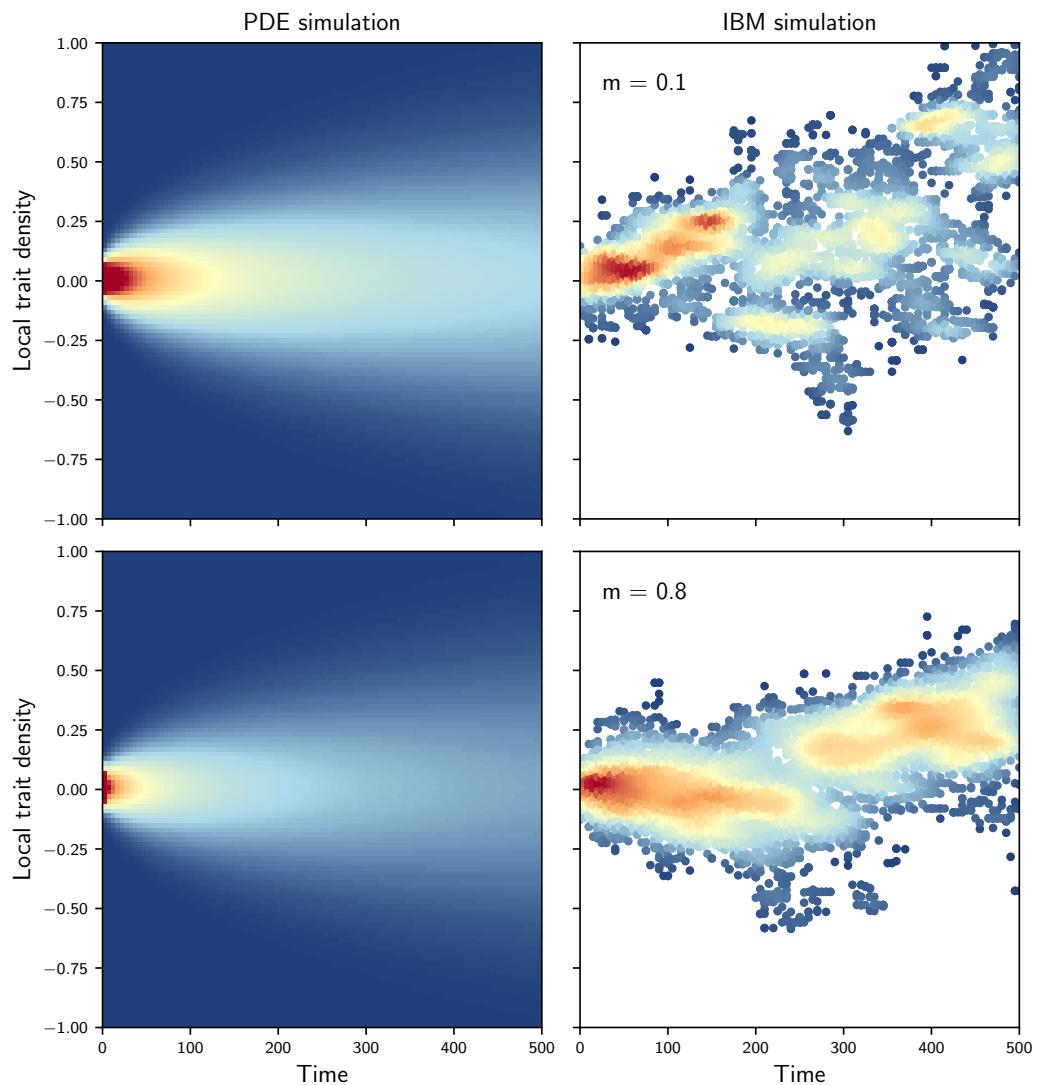
**Fig. S2.12:** Standardized effects of  $h_d$ ,  $\langle l \rangle$  and  $r_\Theta$  on  $Q_{ST,s}$  and  $Q_{ST,u}$  obtained from multivariate regression models independently fitted for low and high migration regimes on average results from 5 replicate simulations of the IBM, analogous Fig. 2.5c-d but for 1126 of the 261,080 undirected connected graphs with  $M = 9$  vertices and varying  $r_\Theta$  (see Methods for details). Error bars show 95% confidence intervals.



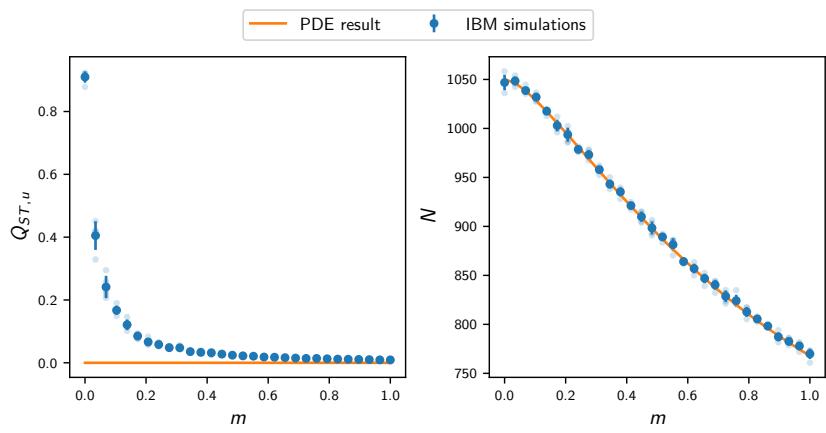
**Fig. S2.13: Simulations on graphs with  $M = 49$  vertices obtained from real spatial habitat datasets, in the setting with heterogeneous selection.** The region from where graphs are obtained is centered on the Hengduan Mountains in Southwest China, one of the most species-rich temperate mountain biota globally (Ding et al., 2020). (a) Graphical representation of a geographical area of size  $0.11^\circ \times 0.11^\circ$ . To create the graph, we considered biological populations living in grasslands, and used the dataset provided in (Jung et al., 2020) containing global grassland coverage at  $0.01^\circ$  resolution. We assigned a vertex to a geographical area of size  $0.01^\circ \times 0.01^\circ$  if its grassland coverage was above a threshold arbitrarily set to 50%. We further assumed that two vertices were connected if their euclidean distance was below a certain dispersal range, which we let vary from 1 to 2.5 km. Local annual average temperature was considered as the value that captures environmental conditions at each vertex. Temperature data was obtained from the CHELSA dataset (Karger et al., 2017). (b) Grassland coverage for the considered region. Blue lines correspond to rivers and dashed grey lines correspond to country borders. Red crosses indicate the locations of the 83 graphs sampled for the simulations used in (c-d). (c-d) Standardized effects of  $h_d$ ,  $\langle l \rangle$  and  $r_\Theta$  on  $Q_{ST,s}$  and  $Q_{ST,u}$  obtained from multivariate regression models independently fitted for low and high migration regimes to average results from 5 replicate simulations of the IBM on the 83 graphs which location is illustrated in (c) (see Table S2.4 for simulation details). Error bars show 95% confidence intervals.



**Fig. S2.14:** Standardized effects of  $h_d$ ,  $\langle I \rangle$  and  $r_\Theta$  on  $Q_{ST,u}$  in the setting with no selection and in the setting with heterogeneous selection for the time horizons  $t = 1000$  and  $t = 2000$ , obtained from multivariate regression models independently fitted for low and high migration regimes to average results from 5 replicate simulations of the IBM on all undirected connected graphs with  $M = 7$  vertices and varying  $r_\Theta$  (see Methods for details). (a–c) illustrate that the effects of the topology metrics on  $Q_{ST,u}$  and  $Q_{ST,s}$  remain constant for  $t > 1000$  in both the settings without selection and with heterogeneous selection. Error bars show 95% confidence intervals.



**Fig. S2.15:** Comparison of the neutral trait density on one vertex obtained from Eq. (S2.21) (left) and from the IBM simulations (right) in the setting with no selection, for the chain graph. The densities obtained from Eq. (S2.21) and from the IBM are dissimilar.



**Fig. S2.16:** Comparison of results obtained from the deterministic approximations Eqs. (S2.20) and (S2.21) and from IBM simulations in the setting with no selection, on the star graph with  $M = 7$  vertices. While Eq. (S2.20) can capture population size, Eq. (S2.21) is not able to capture  $Q_{ST,u}$ . Each plain dot represents average results from 5 replicate simulations, bars represent one standard deviation, and each fade dot represents a single replicate value.

## 2.C Supplementary Tables

	$Q_{ST,u}$				$Q_{ST,u} - bN$	
$m$	0.01	0.50	0.01	0.50	0.01	0.50
(Intercept)	0.000 (0.023)	-0.000 (0.017)	-0.000 (0.023)	-0.000 (0.025)	-0.000 (0.023)	-0.000 (0.028)
$\langle l \rangle$	0.739*** (0.023)	0.872*** (0.017)				
$h_d$			-0.753*** (0.023)	-0.674*** (0.025)	-0.753*** (0.023)	-0.143*** (0.028)
Number of sim.	853	853	853	853	853	853
$R^2$	0.546	0.760	0.567	0.454	0.567	0.030

**Tab. S2.1:** Linear regression model coefficients for the effect of topology metrics on  $Q_{ST,u}$  in the setting with no selection, based on all graphs with  $M = 7$  vertices. \*\*\*  $P < 0.001$

	$M = 7$		$M = 9$	
	$Q_{ST,u}$			
$m$	0.01	0.50	0.01	0.50
(Intercept)	-0.000 (0.017)	-0.000 (0.013)	0.000 (0.009)	-0.000 (0.010)
$h_d$	-0.527*** (0.019)	-0.352*** (0.014)	-0.449*** (0.013)	-0.218*** (0.013)
$\langle l \rangle$	0.500*** (0.019)	0.712*** (0.014)	0.583*** (0.013)	0.784*** (0.013)
Number of sim.	853	853	1,126	1,126
$R^2$	0.766	0.858	0.899	0.896

**Tab. S2.2:** Multivariate linear regression model coefficients for the effect of topology metrics on  $Q_{ST,u}$  in the setting with no selection. \*\*\*  $P < 0.001$

	$M = 7$				$M = 9$			
	$Q_{ST,s}$		$Q_{ST,u}$		$Q_{ST,s}$		$Q_{ST,u}$	
$m$	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50
(Intercept)	-0.000 (0.008)	-0.000 (0.009)	-0.000 (0.009)	-0.000 (0.009)	0.000 (0.008)	0.000 (0.008)	0.000 (0.008)	0.000 (0.008)
$h_d$	-0.117*** (0.009)	-0.114*** (0.010)	-0.060*** (0.010)	-0.102*** (0.010)	-0.135*** (0.010)	-0.185*** (0.011)	-0.146*** (0.011)	-0.164*** (0.011)
$\langle l \rangle$	-0.004 (0.009)	0.308*** (0.010)	0.328*** (0.010)	0.300*** (0.010)	-0.017 (0.010)	0.431*** (0.011)	0.475*** (0.011)	0.434*** (0.011)
$r_\Theta$	0.914*** (0.008)	0.805*** (0.009)	-0.838*** (0.009)	0.807*** (0.009)	0.926*** (0.008)	0.715*** (0.008)	-0.730*** (0.008)	0.725*** (0.008)
Number of sim.	2,548	2,548	2,548	2,548	2,250	2,250	2,250	2,250
$R^2$	0.845	0.808	0.808	0.799	0.870	0.853	0.862	0.851

**Tab. S2.3:** Multivariate linear regression model coefficients for the effect of the topology metrics on  $Q_{ST,u}$  and  $Q_{ST,s}$  in the setting with heterogeneous selection. \*\*\*  $P < 0.001$

	$Q_{ST,s}$		$Q_{ST,u}$	
$m$	0.1	0.50	0.1	0.50
(Intercept)	-0.000 (0.093)	-0.000 (0.064)	0.000 (0.056)	-0.000 (0.059)
$h_d$	-0.088 (0.094)	-0.235*** (0.065)	-0.048 (0.057)	-0.286*** (0.060)
$\langle l \rangle$	-0.220* (0.106)	0.195** (0.073)	0.965*** (0.064)	0.645*** (0.068)
$r_\Theta$	0.610*** (0.106)	0.675*** (0.073)	-0.282*** (0.063)	0.282*** (0.068)
Number of sim.	83	83	83	83
$R^2$	0.313	0.675	0.752	0.717

**Tab. S2.4:** Multivariate linear regression model coefficients for the effect of topology metrics on  $Q_{ST,u}$  and  $Q_{ST,s}$  on real graphs with  $M = 49$  vertices in the setting with heterogeneous selection. \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$

	$\sigma_a = 0.5 < 1/\sqrt{2p}$				$\sigma_a = 1 > 1/\sqrt{2p}$			
	$Q_{ST,s}$		$Q_{ST,u}$		$Q_{ST,s}$		$Q_{ST,u}$	
$m$	0.05	0.50	0.05	0.50	0.05	0.50	0.05	0.50
(Intercept)	0.000 (0.005)	-0.000 (0.010)	-0.000 (0.011)	-0.000 (0.010)	0.000 (0.004)	-0.000 (0.008)	0.000 (0.012)	-0.000 (0.007)
$h_d$	-0.228*** (0.006)	-0.118*** (0.011)	-0.171*** (0.012)	-0.169*** (0.012)	-0.166*** (0.004)	-0.128*** (0.009)	-0.178*** (0.013)	-0.139*** (0.008)
$\langle l \rangle$	0.084*** (0.006)	0.373*** (0.011)	0.461*** (0.012)	0.573*** (0.012)	0.002 (0.004)	0.296*** (0.009)	0.483*** (0.013)	0.286*** (0.008)
$r_\Theta$	0.922*** (0.005)	0.741*** (0.010)	-0.657*** (0.011)	0.508*** (0.010)	0.967*** (0.004)	0.816*** (0.008)	-0.585*** (0.012)	0.837*** (0.007)
Number of sim.	2,548	2,548	2,548	2,548	2,548	2,548	2,548	2,548
$R^2$	0.934	0.768	0.716	0.732	0.962	0.828	0.659	0.861

**Tab. S2.5:** Multivariate linear regression model coefficients for the effect of topology metrics on  $Q_{ST,u}$  and  $Q_{ST,s}$  in the setting of trait-dependent competition and heterogeneous selection (Section 2.A.3), based on all graphs with  $M = 7$  vertices. \*\*\*  $P < 0.001$



# 3

## Mini-batching ecological data to improve ecosystem models with machine learning

by Victor Boussange<sup>1,2</sup>, Pau Vilimelis Aceituno<sup>3</sup>, and Loïc Pellissier<sup>1,2</sup>

<sup>1</sup> Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland

<sup>2</sup> Landscape Ecology, Institute of Terrestrial Ecosystems, Department of Environmental System Science, ETH Zürich, CH-8092 Zürich, Switzerland

<sup>3</sup> Institute of Neuroinformatics, ETH Zürich and University of Zürich, Zürich, Switzerland

bioRxiv:2022.07.25.501365

In review at PLOS Computational Biology

**”** *Les données pertinentes détiennent les réponses.*

— French anagram

Ecosystems are involved in global biogeochemical cycles that regulate climate and provide essential services to human societies. Mechanistic models are required to describe ecosystem dynamics and anticipate their response to anthropogenic pressure, but their adoption has been limited in practice because of issues with parameter identification and because of model inaccuracies. While observations could be used to directly estimate parameters and improve models, model nonlinearities as well as shallow, incomplete and noisy datasets complicate this process. Here, we propose a machine learning (ML) framework relying on a mini-batch method combined with automatic differentiation and state-of-the-art optimizers. By splitting the data into mini-batches with a short time horizon, we show both analytically and numerically that the mini-batch method regularizes the learning problem. When combined with the proposed numerical implementation, the resulting ML framework can efficiently learn the parameters of complex dynamical models and is a workhorse for model selection. We evaluate the performance of the ML framework in recovering the dynamics of a simulated food-web. We show that it can efficiently learn from noisy, incomplete and independent time series, accurately estimating the model parameters and providing reliable short-term forecasts. We further show that the ML framework can provide statistical support for the true generating model among several candidates. In summary, the proposed ML framework can efficiently learn from data and elucidate mechanistic pathways to improve our understanding and predictions of ecosystem dynamics.

### 3.1 Introduction

Ecosystems are complex systems involving many interacting functional entities which together play a major role in regulating global biogeochemical cycles (Bonan, 2008) and delivering essential services to humans (Kremen, 2005). Ecosystems currently face intense disruption from anthropogenic pressure, through pollution and land use (Doney, 2010; Ellis, 2011), and from climate change (Midgley and Hannah, 2019). In order to anticipate the responses of ecosystems to these disruptions, models that can extrapolate ecological dynamics beyond observations are required (Boyd, 2012). A major challenge is that the processes driving ecological dynamics are nonlinear, resulting in complex responses and feedbacks (Scheffer et al., 2001). Nonlinearity greatly affects the capacity of modelling approaches that do not incorporate specific biological knowledge to reliably project current trends into the future (Barnosky et al., 2012). For instance, while methods based on statistical descriptions (Deneu et al., 2021) and nonparametric methods (Ye et al., 2015; Ye and Sugihara, 2016; Deyle et al., 2016) have adequate interpolation capabilities, they are ill-suited for extrapolating beyond observed trends (Barnosky et al., 2012; Urban et al., 2016). In contrast, mechanistic ecosystem models integrate constraints on the expected dynamics by explicitly modelling interactions, feedback loops and dependencies between ecosystem components (Geary et al., 2020). While this should ensure a more robust forecast under large disruptions (Norberg et al., 2012), ecosystem models suffer in practice from parametrization issues, i.e. inaccuracies in the mathematical formulation of the processes and issues in identifying the correct parameter values (DeAngelis and Yurek, 2015). These drawbacks have limited their broad adoption (Urban et al., 2016). Learning the parametrization of ecosystem models from observation data by blending specific biological knowledge and ML methods could improve our representation of ecosystem processes and help us to anticipate ecosystem responses to global changes.

The parametrization of ecosystem models can be indirectly learnt from observations by calibrating the parameters from the data collectively using inference methods. These methods proceed by maximizing the posterior probability of the parameters given the observations, but their success is subject to a number of issues, some of which specifically relating to ecosystem model properties. Among these issues, the exploration of the posterior landscape demands repeated model simulations, but ecosystem models are usually associated with a high computational cost that limits the number of possible runs (Fisher et al., 2018). Additionally, the complexity of processes requires a large number of parameters. Due to the curse of dimensionality (Boyd, 2012), this complicates the exploration of the posterior distribution and can further leave many parameters poorly constrained (Gutenkunst et al., 2007). The limited availability of observations, which are usually composed of multiple partial short-term time series (Scheiter et al., 2013; Schartau et al., 2017), accentuates the lack of parameter constraints. Moreover, ecosystem dynamics can be strongly dependent on the initial conditions (ICs) and show a chaotic behaviour (Hastings et al., 1993; Huisman and Weissing, 1999; Benincà et al., 2008), or can be associated with a large panel of dynamics depending on the parameter values. In such case, small perturbations of the ICs or parameter values lead to large divergences in the model outcomes, causing numerical problems in finding the most probable parameters (DeAngelis and Yurek, 2015). Last but not least, in

contrast to fields such as climate and weather modelling, the derivation of fundamental processes regulating ecosystems is far from being established (Gehlen et al., 2015; Schartau et al., 2017; Purves et al., 2013), resulting in inaccurate mechanistic pathways and uncertain mathematical formulations (Gentleman et al., 2003), limiting extrapolation to unseen data. To summarize, the parametrization of ecosystem models requires inference methods that are robust despite the models' complexity, the limited observation data, and the inaccurate description of ecological processes.

A variety of data assimilation and ML methods are increasingly being used to parametrize ecosystem models. Bayesian inference with Markov Chain Monte Carlo methods, used in (Lignell et al., 2013; Higgins et al., 2010; Xu et al., 2006; Fiechter et al., 2013; Rosenbaum et al., 2019), offer the advantage of quantifying uncertainties by inferring the full posterior probability distribution of the unknown parameters. This is achieved by a global exploration of the parameter space, which makes Bayesian methods computationally expensive and particularly prone to the curse of dimensionality (Gosh et al., 2021). Simulated annealing (Matear, 1995), genetic algorithms (Ward et al., 2010), and sequential methods such as extended Kalman filtering and ensemble Kalman methods (Bertino et al., 2003; Doron et al., 2013; Gharamti et al., 2017) have been used as alternatives, but are similarly subject to the curse of dimensionality and demand a large number of model evaluations. Variational methods rely on the model adjoint, i.e. the model sensitivity to the parameters, to explore more efficiently the parameter space, iteratively updating the parameter estimates using the gradient of the posterior landscape. Such methods therefore demand less evaluation (Schartau et al., 2017), which explains their wide adoption in the field of artificial intelligence to train highly parametrized neural networks (up to the order of  $10^8$  parameters (Vaswani et al., 2017)) and their use in calibrating marine ecosystem models (Fennel et al., 2001; Spitz et al., 1998; Xiao and Friedrichs, 2014; Pelc et al., 2012) (see (Schartau et al., 2017) for a review) and terrestrial ecosystem models (Zhu and Zhuang, 2015; DeLong et al., 2014; Curtsdotter et al., 2019). However, as the complex dynamics of ecosystem models tend to be associated with rugged posterior landscapes, variational methods are prone to converging to local minima, making variational methods very sensitive to the choice of initial model parameters (Gábor and Banga, 2015; Schartau et al., 2017). Ecosystem models are specified as differential equations that depend not only on parameters but also on ICs. The state-dependency of ecosystems means that neglecting the estimation of initial ICs might compromise the correct fitting of the parameters and the forecast skill (Lignell et al., 2013). However, few of the aforementioned studies have addressed the problem of IC estimation (but see (Pelc et al., 2012)). Finally, the numerical implementation of variational methods is also challenging, as the model adjoint is difficult to obtain and maintain as the model is modified (Lawson et al., 1995; Pelc et al., 2012; Gharamti et al., 2017). Novel methods for model parametrization are emerging, thanks to advances in the field of artificial intelligence (Willard et al., 2020; Kashinath et al., 2021; Alber et al., 2019; Peng et al., 2021), providing new opportunities to better address these issues.

Here, we propose a ML framework relying on a mini-batch method inspired by multiple shooting methods (Pisarenko and Sornette, 2004) and on automatic differentiation and state-of-the-art optimizers to efficiently learn the parametrization of ecosystem models from observation data. The mini-batch method divides the training problem into mini-batches

with a short time horizon. We show analytically how this learning strategy regularizes the ill-behaviour of the loss function arising from the strong nonlinearities of ecosystem models. We implement the mini-batch method in the software ecosystem SciML (Rackauckas et al., 2020a), which provides advanced optimizers and allows the automatic generation of efficient and accurate model adjoints, leading to excellent performance. The resulting ML framework makes it possible to efficiently combine the information contained in short, independent time series, and is a workhorse for performing model selection and improving model accuracy. We evaluate the performance of the ML framework in recovering the chaotic dynamics of simulated food webs. We assume a perfect-model setting and test the capacity of the ML framework to recover the true parameters and provide forecasts based on noisy and incomplete observations, and we explore its efficiency in combining the information from multiple time series. Additionally, we investigate whether the ML framework can recover the most appropriate model structure among candidate models. By blending biological knowledge and ML methods, the proposed ML framework is interpretable and data-efficient, and it facilitates mechanism discovery. The proposed approach thus shows promise in improving our ability to understand and forecast ecosystem dynamics.

## 3.2 Machine learning framework for ecosystem models

### 3.2.1 Ecosystem model parametrization as a learning problem

#### Ecosystem models

Ecosystem models generally consist of a system of ordinary differential equations (ODEs) of the form

$$\begin{aligned}\dot{x}(t) &= f(t, x(t), p) \\ x(0) &= x_0 \\ y(t) &= h(x(t)) + \epsilon(t)\end{aligned}\tag{3.1}$$

where  $x(t) \in \mathbb{R}^m$  is a vector of state variables that might represent species abundance, resources availability or functional group biomass,  $y(t) \in \mathbb{R}^d$  is a vector of observables that contains a subset or aggregates of the state variables, and  $p \in \mathbb{R}^q$  is the model parameter vector.  $h$  is a function that maps the state variables to the observables, and we assume that the observables are contaminated with a white noise  $\epsilon$  of Gaussian type, with zero mean and variance–covariance matrix  $\Sigma_y$ . Denoting by  $\theta = (x_0, p)$  the vector containing the ICs and the parameters, the model may be viewed as a map  $\mathcal{M}$  parametrized by time  $t$  that takes the parameters  $\theta$  to the state variables  $x$

$$\begin{aligned}\mathcal{M}(t, \theta) &= x(t) \\ &= \int_0^t f(s, x(s), p) ds + x_0\end{aligned}\tag{3.2}$$

## Inverse modelling

Taking expectations over the noise realizations yields  $\mathbb{E}[y(t)] = h(\mathcal{M}(t, \theta))$ , and it follows that the conditional likelihood of each observation  $y_k \equiv y(t_k)$ , given the parameters  $\theta$  and the model  $\mathcal{M}$  denoted by  $p(y_k|\theta, \mathcal{M})$ , follows the distribution of the residuals  $\epsilon_k \equiv \epsilon(t_k) = y(t_k) - h(\mathcal{M}(t_k, \theta))$ , which corresponds to the multivariate normal distribution  $\mathcal{N}_{0, \Sigma_y}$ . Following a Bayesian approach, the parametrization of the ecosystem model can be performed on the basis of the parameter and model posterior probability  $p(\theta, \mathcal{M}|\mathbf{y}_{1:K})$ , i.e. the conditional probability density of the parameter values  $\theta$  and the model  $\mathcal{M}$  given the data, given by

$$p(\theta, \mathcal{M}|\mathbf{y}_{1:K}) \propto p(\mathbf{y}_{1:K}|\theta, \mathcal{M})p(\theta, \mathcal{M}) \quad (3.3)$$

where  $\mathbf{y}_{1:K} = (y_1, \dots, y_K)$ ,  $p(\mathbf{y}_{1:K}|\theta, \mathcal{M})$  is the product of the conditional likelihood of each observation  $y_k$

$$\begin{aligned} p(\mathbf{y}_{1:K}|\theta, \mathcal{M}) &= \prod_{i=1}^K p(y_i|\theta, \mathcal{M}) \\ &= \prod_{k=1}^K \frac{1}{\sqrt{(2\pi)^d |\Sigma_y|}} \exp\left(-\frac{1}{2} \epsilon_k^T \Sigma_y^{-1} \epsilon_k\right) \end{aligned} \quad (3.4)$$

and  $p(\theta, \mathcal{M})$  is the prior distribution of the model and its associated parameter values. The model  $\mathcal{M}$  is included in the probabilistic quantities in order to accommodate multiple candidate models (see Section 3.3.3).

A variational method to obtain a Bayesian estimate of  $\theta$  involves maximizing  $p(\theta, \mathcal{M}|\mathbf{y}_{1:k})$  to obtain the maximum a posteriori (MAP) estimator (Bocquet et al., 2019), which is equivalent to a maximum likelihood approach under a uniform prior distribution of the parameters, i.e. when no prior information on the parameter values is used (Schartau et al., 2017). Observing that maximizing  $p(\theta, \mathcal{M}|\mathbf{y}_{1:K})$  is equivalent to minimizing  $-\log p(\theta|\mathbf{y}_{1:K}, \mathcal{M})$  and assuming a normal prior distribution of the parameters  $\mathcal{N}_{p_b, \Sigma_p}$ , one can obtain the MAP  $\hat{\theta}$

$$\hat{\theta} = \arg \min_{\theta} L_{\mathcal{M}}(\theta) \quad (3.5)$$

where

$$L_{\mathcal{M}}(\theta) = \frac{1}{2} \left[ \sum_{k=1}^{K-1} \|y_k - h(\mathcal{M}(t_k, \theta))\|_{\Sigma_y}^2 + \|p - p_b\|_{\Sigma_p}^2 \right] \quad (3.6)$$

(Schneider et al., 2017; Raue et al., 2009) and where we use the notation  $\|y\|_{\Sigma}^2 = y \Sigma^{-1} y^T$ . Eq. (3.6) is similar to a traditional least squares function commonly used in regression, where the second summand is the analogue of a regularization term for the weights and biases of e.g. a neural network.

Gradient-based optimizers can then be used to efficiently obtain  $\hat{\theta}$  in Eq. (3.5), iteratively updating the parameter vector  $\theta_m$  given the gradient of the loss function, denoted by  $\nabla_{\theta} L_{\mathcal{M}}$ ,

to navigate the surface defined by  $L_{\mathcal{M}}$  with the aim to find the global minimum where  $\nabla_{\theta}L_{\mathcal{M}}(\hat{\theta}) = 0$ . As an example, the plain vanilla gradient descent algorithm is given by

$$\theta_{m+1} = \theta_m - \gamma \nabla_{\theta}L_{\mathcal{M}}(\theta_m) \quad (3.7)$$

where  $\gamma$  is the learning rate. Other gradient-based algorithms, such as the ADAM optimizer used in the section below, employ more advanced updating strategies to avoid convergence to local minima but stay in the spirit of Eq. (3.7).

### Information indigestion

A naive minimization of  $L_{\mathcal{M}}(\theta)$  with gradient-based methods is likely to fail, as its associated surface cannot be navigated properly. As illustrated in Fig. 3.1A, the loss surface associated with models characterized by complex dynamics consists of multiple local minima, which cause problems of convergence in efforts to reach the global minimum (depicted by the orange curve in Fig. 3.1B). Furthermore, in a neighbourhood of the global minimum, the gradient of the loss function is very large (a "ravine" with almost vertical walls), leading the optimizer to overshoot the true parameter values (depicted by the green curve in Fig. 3.1B). In Section 3.A we show in a general setting that these problematic features arise from the dynamical properties of ecosystem models: when the dynamics are chaotic or exhibit a limit cycle (as is often the case for ecological dynamics (Bjørnstad and Grenfell, 2001; Hastings et al., 1993; Huisman and Weissing, 1999; Benincà et al., 2008)), the dynamical trajectories exhibit high sensitivity to the model parameters and ICs. This means that a small modification of the parameters or ICs leads to large divergences over time. The prevalence of large deviations causes discontinuities, appearing as many sub-optimal local minima on the loss surface. Moreover, the true minimum can only be found in a narrow ravine that becomes narrower as the number of data points increases. Such surfaces are hardly navigable with gradient descent methods, but since the behaviour of  $L_{\mathcal{M}}(\theta)$  critically depends on the time horizon, we reformulate Eq. (3.6) in the following section by splitting the time series into mini-batches with a short time horizon.

## 3.2.2 ML framework for ecosystem models

### Description of the mini-batch method

We propose a mini-batch method that splits the data into mini-batches with a short time horizon. Under perturbed parameters and ICs, chaotic or limit cycle dynamics only diverge after some characteristic simulation time; by splitting the time series into small mini-batches, discontinuities that cause the poor navigability of the landscape are therefore avoided. By averaging the associated losses during the training, the mini-batch method regularizes the loss function and makes it possible to combine the information contained in independent time series.

Consider the availability of  $S$  independent time series, where each time series  $s \in \{1, \dots, S\}$  contains  $K^{(s)}$  observations  $\mathbf{y}_{1:K^{(s)}}^{(s)}$ . To improve the ill-behaviour of Eq. (3.6), we

split each time series into  $M^{(s)}$  mini-batches, each of which defines a loss denoted by  $L_{\mathcal{M}}^{(s,m)}$ . Averaging the losses  $L_{\mathcal{M}}^{(s,m)}$  leads to a reformulation of the loss function in Eq. (3.6), yielding

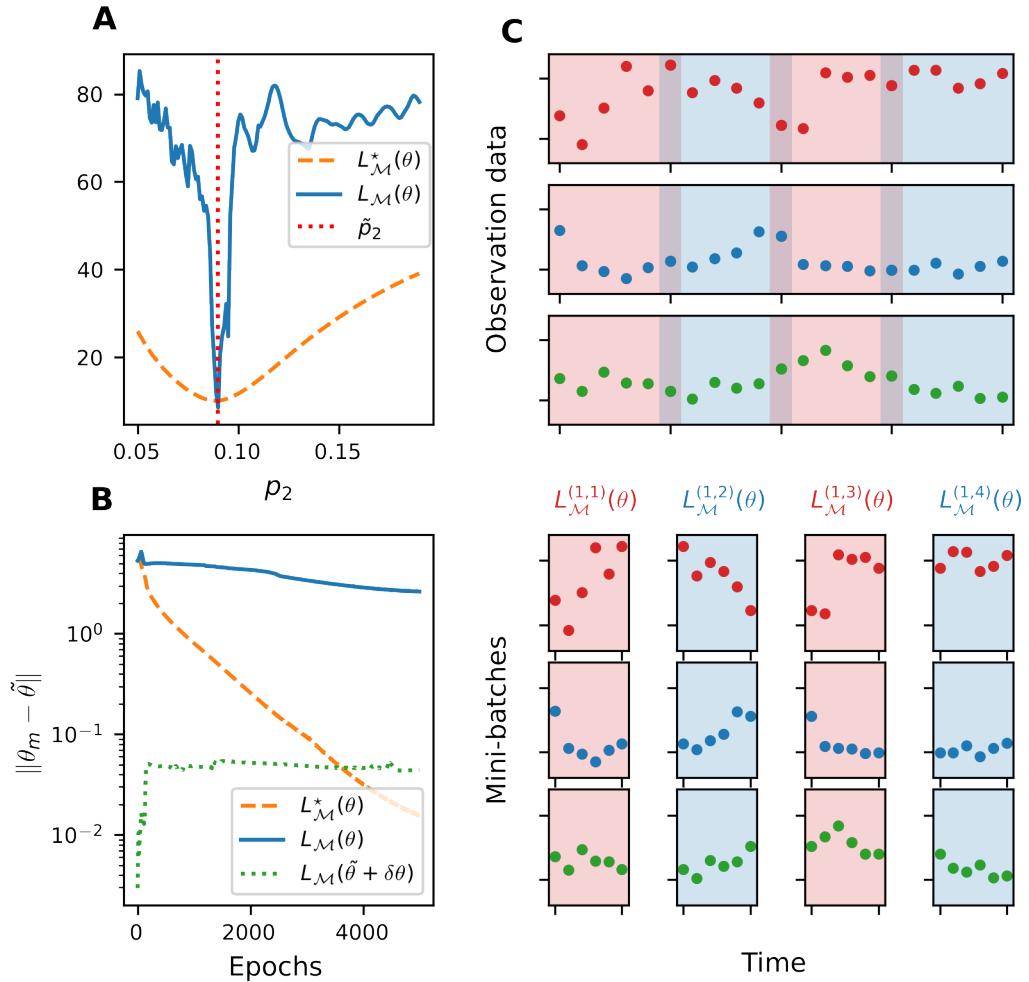
$$\begin{aligned} L_{\mathcal{M}}^*(\theta) &= \frac{1}{S} \sum_{s=1}^S \frac{1}{M^{(s)}} \sum_{m=0}^{M^{(s)}-1} L_{\mathcal{M}}^{(s,m)}(\theta) \\ &= \frac{1}{S} \sum_{s=1}^S \frac{1}{M^{(s)}} \sum_{m=0}^{M^{(s)}-1} \left[ \frac{1}{K^{(s,m)}} \sum_{k=1}^{K^{(s,m)}} \|y_{k+mK^{(s)}/M^{(s)}}^{(s)} - h(\mathcal{M}(t_{k+mK^{(s)}}, p, x_0^{(s,m)}))\|_{\Sigma_y} \right. \\ &\quad \left. + \|y_{mK^{(s)}/M^{(s)}}^{(s)} - h(x_0^{(s,m)})\|_{\Sigma_{x_0}} + \|p_b - p\|_{\Sigma_p} \right] \end{aligned} \quad (3.8)$$

where  $\theta = (p, x_0^{(1,1)}, \dots, x_0^{(1,M_1)}, x_0^{(2,1)}, \dots, x_0^{(S,M_S)})$  consists of the augmented parameter vector containing  $\sum_{s=1}^S [M^{(s)} - 1]$  additional ICs to be estimated and denoted by  $x_0^{(s,m)}$ ,

$$K^{(s,m)} = \begin{cases} K^{(s)}/M^{(s)} + 1 & : m < M \\ K^{(s)}/M^{(s)} & : m = M \end{cases}$$

indicates the number of points to include in the  $m$ th batch, and  $p_b$  designates the prior knowledge of the parameter values. We refer to Fig. 3.1C for a graphical representation of Eq. (3.8). In contrast to Eq. (3.6), the ICs for each batch are constrained by an extra term with weight  $\Sigma_{x_0}^{-1}$ , which is needed in practice for better convergence. The ICs  $x_0^{(s,m)}$  are nuisance parameters that augment the dimensionality of the parameter space as they are inferred for each batch. Nonetheless, the efficiency of gradient-based methods, together with the suggested numerical implementation detailed in the section below, largely circumvent the additional cost associated with the augmented dimensionality.  $K^{(s,m)}$  expresses the overlap between each batch of data for  $m < M^{(s)}$  and ensures that all the available information is utilized to constrain the parameter vector  $p$ .

By smoothing the ruggedness of the loss surface (orange curve in Fig. 3.1A), the proposed mini-batch method yields an improved navigability (orange curve in Fig. 3.1B). We show analytically in Section 3.A that using  $L_{\mathcal{M}}^*$  yields a more navigable loss surface than if  $L_{\mathcal{M}}(\theta)$  is used in Eq. (3.6).



**Fig. 3.1:** Illustration of the mini-batch method. **A** Characteristic features of the naive loss function  $L_M(\theta)$  and the mini-batch loss function  $L_M^*(\theta)$ . The blue and orange lines correspond to cross sections of  $L_M(\theta)$  and  $L_M^*(\theta)$ , respectively, obtained from the ecosystem model presented in the section Simulated food-web model as a case study. While the cross section of  $L_M(\theta)$  presents many local minima and a very large gradient in the neighbourhood of the true parameter, which renders the navigability of the loss surface difficult, the cross section of  $L_M^*(\theta)$  is smooth and shows a single minimum, illustrating the regularization induced by the mini-batch method. **B** Convergence of gradient descent algorithms applied to  $L_M$  and  $L_M^*$ . The blue, orange and green lines correspond to the loss function evaluated against the epochs (number of parameter updates) using  $L_M$ ,  $L_M$  starting from a parameter value close to the true parameters, and  $L_M^*$ , respectively. The ill-behaviour of  $L_M$  leads to the convergence to a local minimum, while  $L_M^*$  is associated with a smooth convergence to the true parameters. **C** Graphical representation of the proposed mini-batch method. To improve the navigability of the posterior landscape, the algorithm splits the time series into mini-batches with short time horizons (blue and red portions of the time series). Since mini-batches are treated independently, the method naturally extends to independent time series.

## Numerical implementation of the ML framework

The choice of the optimizer and the correct calculation of the model sensitivity to the parameters and ICs, upon which the gradient of the loss function  $\nabla L_{\mathcal{M}}^*(\theta)$  depends, play an essential role in the success of the minimization of the loss function and the subsequent correct estimation of the MAP. To accelerate the learning process and make it more robust, we propose to combine the mini-batch method with modern optimizers and automatic differentiation. Building upon the software ecosystem SciML (Rackauckas et al., 2020a), we use **DifferentialEquations.jl** for the forward integration of the ecosystem model, as it provides highly efficient ODE solvers (Rackauckas and Nie, 2017). **DifferentialEquations.jl** is additionally compatible with automatic differentiation and includes an extensive set of sensitivity analysis methods (Ma et al., 2021), enabling the automatic generation of the model sensitivity to the parameters and ICs and guaranteeing their accuracy. This automatic generation greatly reduces the effort and potential errors associated with the adjoint code construction, enabling continuous development of the models. The accuracy is also an essential feature, as the model sensitivities are critically involved in the minimization of Eq. (3.6) and their inaccuracies can compromise the convergence of the gradient-based optimizers (Gholamnejad et al., 2019). The interoperability of the SciML ecosystem further makes it possible to benefit from the tooling of the deep learning library **Flux.jl** and the nonlinear optimization library **Optim.jl** (K Mogensen and N Riseth, 2018), providing state-of-the-art optimizers that are computationally efficient and well suited for highly parameterized models (Ruder, 2016). We use the adaptive, momentum-based Adam optimizer (Kingma and Ba, 2014) to converge in the basin of attraction of the true parameters, which we substitute with the limited memory BroydenFletcherGoldfarbShanno optimizer (L-BFGS) (Liu and Nocedal, 1989) for the final training epochs to ensure faster and more accurate convergence.

The reformulation of the learning problem in Eq. (3.8), together with the numerical implementation suggested above, define the proposed ML framework, which we benchmark with a concrete case scenario in the next section.

### 3.3 Simulated food-web model as a case study

We evaluate the performance of the ML framework by considering a food-web ecosystem composed of three functional compartments including a resource, consumers and predators. We use a reference model to generate the observation data and first assume a perfect-model setting, evaluating the performance of the ML framework in parametrizing the reference model for different noise levels, with incomplete observations, and with an increasing number of independent time series. Second, we relax the perfect-model assumption by considering two plausible candidate models capturing contrasting hypotheses regarding the ecological processes, and test whether the ML framework can provide support for the true generating model by combining it with information-based model selection.

### 3.3.1 Three-compartment food-web ecosystem

We use a reference food-web model investigated in (Hastings and Powell, 1991; McCann and Yodzis, 1994b; McCann and Yodzis, 1994a; Klebanoff and Hastings, 1994) where a resource  $R$  is eaten by consumers  $C$ , which in turn are fed upon by predators  $P$  (model  $\mathcal{M}_1$  in Fig. 3.2). We further consider an "omnivory variant" of the reference model introduced in (McCann and Hastings, 1997), where predators are omnivorous and can feed upon the resource with a determined strength  $\omega$  (model  $\mathcal{M}_2$  in Fig. 3.2). These models generate fluctuations that resemble the behaviour of observed ecological time series (Bjørnstad and Grenfell, 2001), they produce chaotic dynamics that are notoriously challenging to forecast for a wide range of realistic parameters (Post et al., 2000), and they have been used as benchmarks for proposed ecosystem forecasting methods (see (Perretti et al., 2013; Deyle et al., 2016; Ye and Sugihara, 2016)).

After nondimensionalization, the three-compartment model and the omnivory variant comprise a total of six and nine parameters, respectively: the mass-specific metabolic rate of consumers and predators  $x_C$  and  $x_P$ , the ingestion rate per unit metabolic rate of consumers and predators  $y_C$  and  $y_P$  (decomposed into  $y_{PC}$ ,  $y_{PR}$  for the omnivory variant), the half saturation densities for the type II functional responses of the consumers and predators  $R_0$  (decomposed into  $R_0$ ,  $R_{02}$  for the omnivory variant) and  $C_0$ , and the omnivory strength  $\omega$  for the omnivory variant (see Section 3.B for the ODE details). Time is nondimensionalized by the resource growth rate and set to the biologically realistic value of 100% biomass increase per day, so that one unit of time corresponds to one day. The parameters in the simulations are set to the biologically realistic values proposed by (McCann and Yodzis, 1994a; McCann and Hastings, 1997), which additionally ensure that the dynamics of the system are chaotic or show oscillations (see Section 3.B for details).

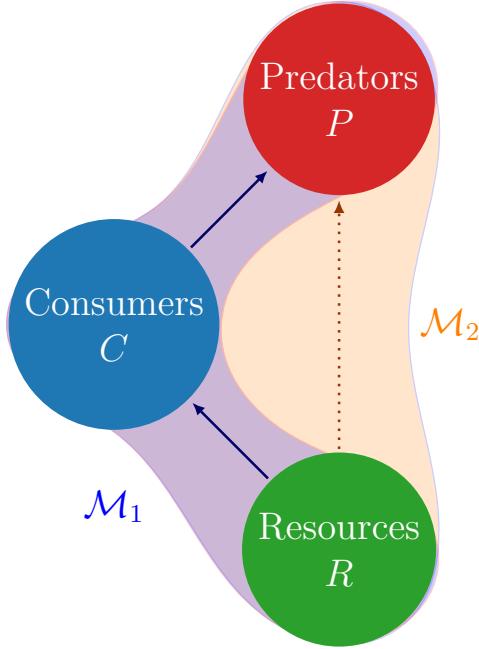
We generate the observation data by sampling the simulated ecosystem dynamics and by contaminating the samples with noise. The noise variance–covariance matrix  $\Sigma_y$  is set to be diagonal, with entries that are proportional to the sample variances of the observables

$$\text{diag } \Sigma_y = r^2[\text{Var}(\tilde{y}_1), \dots, \text{Var}(\tilde{y}_d)] \quad (3.9)$$

where  $r$  indicates the noise level and  $\tilde{y}$  corresponds to the noiseless data generated with the true parameter values  $\tilde{\theta}$ . We sample the simulated dynamics after a long burn-in time ( $t > 500$ ) to ensure that transient dynamics are not observed. A visual representation of the generated data is displayed in Fig. 3.2B. We assume a uniform distribution of the parameter priors, and randomly draw initial parameter estimates from a uniform distribution so that the initial parameter estimates follow  $\mathcal{U}(0, 2\tilde{p})$ .

We consider two different settings: one where all compartment abundances are observable, i.e. the observing system map  $h$  is the identity, and one where only predator and consumer abundances are available, i.e. discarding the resource abundance data. For both settings, structural identifiability is tested with the Julia library **StructuralIdentifiability.jl** (Dong et al., 2021) and verified globally. This means that in theory, the unique observation of predator and consumer abundances carries the information required for a complete characterization of all the model parameters.

For the meta-parameters of the ML framework, we set  $\Sigma_{x_0} = \frac{M^{(1)}}{K^{(1)}} \Sigma_y$ , we use Adam with  $\gamma_m = \mathbb{1}_{[0,2000]}(m)10^{-1} + \mathbb{1}_{[0,2000]}(m)10^{-2} + \mathbb{1}_{[0,2000]}(m)10^{-3}$ ,  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$  for the first 6000 epochs, where  $\gamma_m$  corresponds to the learning rate of the  $m$ th epoch, and we use L-BFGS for the last 200 epochs.



**Fig. 3.2: Reference food web systems considered.** In Section 3.3.2, the blue model  $\mathcal{M}_1$  from (Hastings and Powell, 1991) is considered, where a resource is eaten by consumers, which are themselves eaten by predators. In Section 3.3.3, the orange model  $\mathcal{M}_2$ , corresponding to the omnivory variant introduced in (McCann and Hastings, 1997), is also considered.

### 3.3.2 Parameter learning in a perfect-model setting

We generate time series from the reference food-web model (blue model in Fig. 3.2) under varying noise levels and for both settings with total and partial observations, sampling from the model simulations every four days (see Fig. 3.1C for an illustration of the generated data). We then apply the ML method to the generated data, with a focus on evaluating its performance in recovering the true parameters  $\tilde{p}$  and on its forecast skill.

We evaluate the performance in recovering the true parameters with two metrics, namely the coefficient of determination between the true parameters  $\tilde{x}_P$  and the estimated parameters  $\hat{x}_P$ , denoted by  $R^2_{x_P}$ , and the relative parameter error for the ensemble of training simulations, denoted by  $|\hat{p} - \tilde{p}|/\tilde{p}$  and calculated as the median relative parameter error across the six estimated parameters. To evaluate the out-of-sample forecast skill, we simulate the model beyond the training time span by using the estimated ICs of the last batch of each independent time series. Further, we quantify the forecast skill, denoted by  $\rho^2$ , by

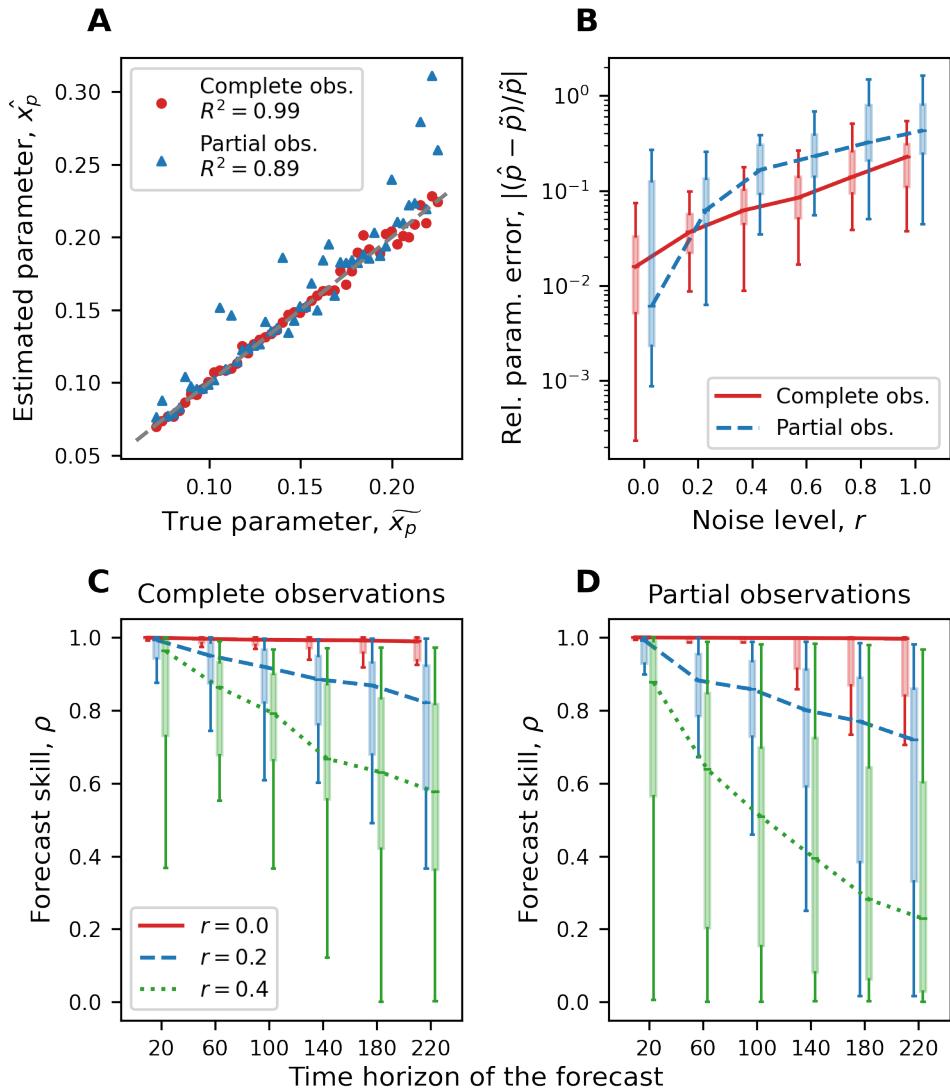
computing the mean squared correlation across all the independent time series between the prey abundance generated with the true parameter  $\tilde{\theta}$  and the predicted prey abundance. We obtain summary statistics of the metrics by varying the critical parameter value  $x_P$ , generating a total of 50 simulations for each noise level and setting considered. While only  $x_P$  is varied, all the parameters together with the ICs are collectively fitted.

### Robustness of the ML framework against noise and incomplete observations

We set the number of time series to  $S = 1$ , the time series length to  $K = 80$ , and the number of mini-batches to  $M = 8$ . We investigate the ML framework performance against observational noise and under the setting with complete or partial observations.

In the complete observation setting, the ML framework can very accurately recover the true parameter values under a moderate observational noise, with mean  $|(\hat{p} - \tilde{p})/\tilde{p}| = 6\%$  and  $R_{x_P}^2 = 0.99$  for 20% observational noise ( $r = 0.2$ ; see Fig. 3.3A, red dots). In the partial observation setting, fair results are also obtained with  $|(\hat{p} - \tilde{p})/\tilde{p}| = 16\%$  and  $R_{x_P}^2 = 0.89$  (Fig. 3.3A, blue triangles). On top of being accurate, the ML framework further shows a very short inference time, i.e. 36 seconds in the complete observation setting and 34 seconds in the partial observation setting (see Table S3.1 for details). To investigate systematically how the ML framework accommodates different levels of observational noise, we vary the noise level from  $r = 0$  to  $r = 1$  and calculate  $|(\hat{p} - \tilde{p})/\tilde{p}|$ . Results reported in Fig. 3.3B show that the ML framework handles observational noise well and the response of the logarithm of  $|(\hat{p} - \tilde{p})/\tilde{p}|$  to  $r$  only differs by a constant factor between the complete and partial observation settings.

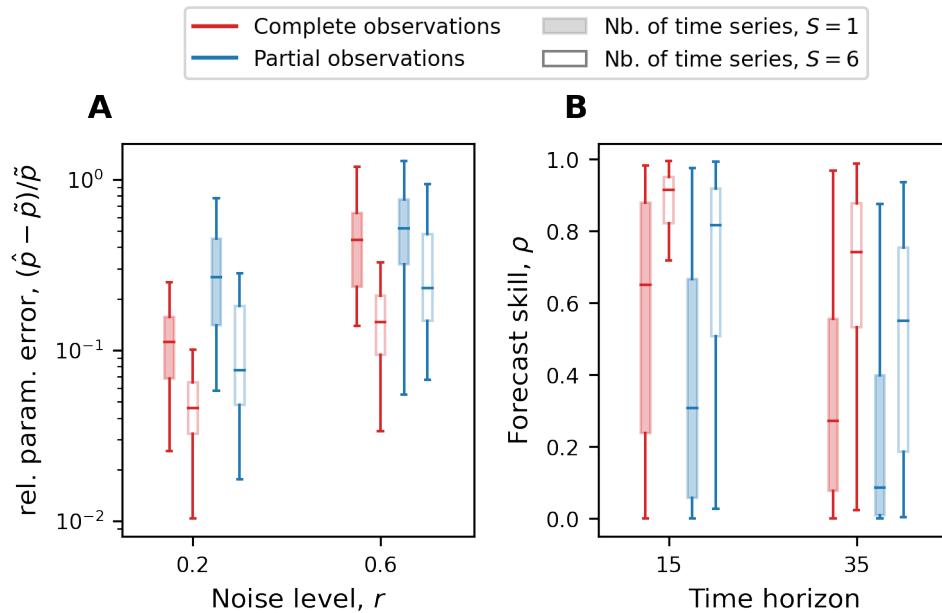
Given that the food-web dynamics are chaotic (Section 3.B), excellent performance in parameter estimation might not be sufficient to provide accurate forecasts. We therefore also test the ML algorithm by evaluating how the forecast skill  $\rho^2$  is affected by the noise level under complete or partial observations. Results for the complete and partial observation settings reported in Fig. 3.3C show good forecast skill under moderate observational noise, where  $\rho^2$  linearly decreases with the time horizon considered.



**Fig. 3.3:** Performance of the proposed ML framework for varying noise levels, under the complete and the partial observation setting. A True parameter  $\tilde{x}_P$  against estimated parameter  $\hat{x}_P$  for  $r = 0.2$  under the complete and the partial observation setting. Although the parameter estimation is more accurate when complete observations of abundance are available, parameters show a fair fit when estimated with partial observations. B Relative parameter error  $|(\hat{p} - \tilde{p})/\tilde{p}|$  for varying noise levels under the setting of complete or partial observations. B supports the above observation for varying noise levels. C Forecast skill  $\rho^2$  of the trained model under the complete observation setting. D Analogous data under the partial observation setting. In A–D, the batch size is set to  $m = 6$ .

## Capability of the ML framework to harness multiple time series

We further investigate the ability of the ML framework to process and combine information from independent datasets. We reduce the time horizon of the observation data by setting the time series length to  $K = 12$ , generate two datasets comprising  $S = 1$  and  $S = 6$  independent time series, respectively, – obtained from independent ICs – and set the number of batches for each time series  $s$  to  $M^{(s)} = 2$ . In both the complete and partial observation settings, we find that the relative parameter error  $|(\hat{p} - \tilde{p})/\tilde{p}|$  is consistently lower in the simulations with a larger number of time series (Fig. 3.4A). The forecast skill is also consistently improved as more independent time series are processed (Fig. 3.4B), and the forecast skill for long-term predictions considerably increases. These results confirm the robustness of the ML framework against noise and partial observations, and show that the ML framework can efficiently harness the information from disparate observation datasets.



**Fig. 3.4:** Performance of the ML framework in processing and combining the information of multiple independent data sets. **A** Relative parameter error  $|\hat{p} - \tilde{p}|/\tilde{p}$  for different numbers of time series and levels of noise, under the complete and the partial observation setting. **B** Forecast skill for different numbers of time series and time horizons of the forecasts, under the complete and the partial observation setting.  $r = 0.1$ . In **A–B**,  $|\hat{p} - \tilde{p}|/\tilde{p}$  decreases while  $\rho^2$  increases as the number of time series processed increases, demonstrating the capacity of the ML framework to process and combine the information from independent time series. In **A–B**, each box plot corresponds to 100 independent simulations where  $\tilde{x}_P$  varies  $\tilde{x}_P \in [0.071, 0.225]$ , and the batch size is set to  $m = 6$ .

### 3.3.3 Elucidating mechanistic pathways

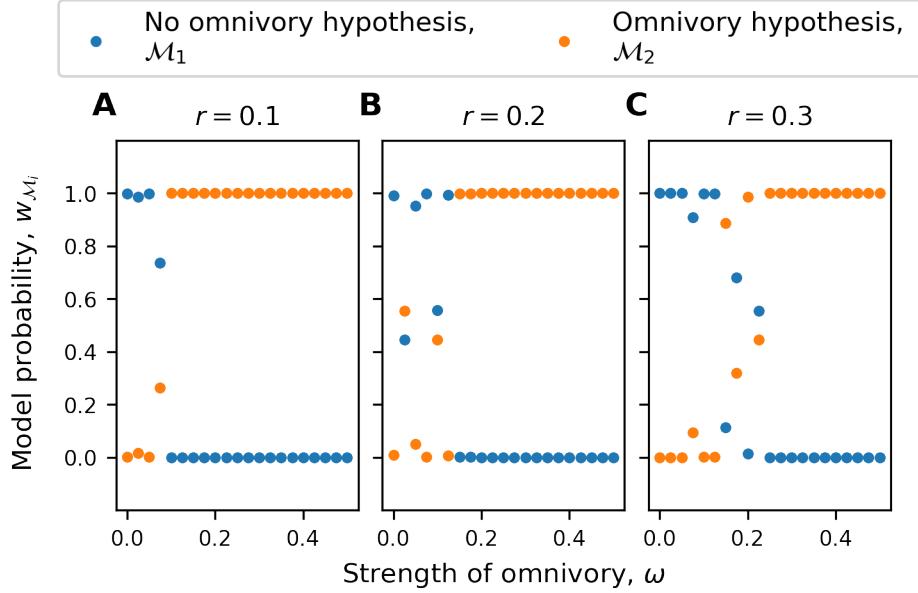
Finally, we relax the perfect-model assumption and investigate whether the ML framework can provide statistical support for the true generating model among several candidates with information-based model selection. Specifically, we investigate whether the ML framework can detect omnivory from single observations of time series. We generate multiple observation datasets from the omnivory variant model  $\mathcal{M}_2$  for different omnivory strengths  $\omega$  and noise levels  $r$ . We consider both the standard model  $\mathcal{M}_1$  and the omnivory variant model  $\mathcal{M}_2$  as two plausible candidate models (see Fig. 3.2 for a graphical illustration of the models). We use the Akaike information criterion (AIC) to select the model with the strongest support in relation to the data (Mangan et al., 2017). In the specific case of our framework, we calculate the AIC as

$$\text{AIC}_{\mathcal{M}_i} = -2 \ln(p(\hat{\theta}, M_i | y_{1:K})) + 2k_{\mathcal{M}_i} \quad (3.10)$$

where  $p(\hat{\theta}, M_i | y_{1:K})$  corresponds to the maximum value of the likelihood of the model  $\mathcal{M}_i$  given the data, and  $k_{\mathcal{M}_i}$  is the number of parameters in the model  $\mathcal{M}_i$ . The AIC ranks the most probable models by penalizing complexity to balance information loss and parsimony, where candidate models with the lowest scores are ranked as the most likely. We consider the Akaike weights  $w_{\mathcal{M}_i} = \frac{\exp(-\Delta\text{AIC}_{\mathcal{M}_i}/2)}{\sum_j \exp(-\Delta\text{AIC}_{\mathcal{M}_j}/2)}$ , where  $\Delta\text{AIC}_{\mathcal{M}_i} = \text{AIC}_{\mathcal{M}_i} - \min_j \text{AIC}_{\mathcal{M}_j}$ , which can be directly interpreted as the probability that  $\mathcal{M}_i$  is the most appropriate model given the data (see (Burnham and Anderson, 2002)). We expect that the Akaike weights provide support for the generating model  $\mathcal{M}_2$  only across values where  $\omega > 0$ , as  $\mathcal{M}_1$  is equivalent to  $\mathcal{M}_2$  when  $\omega = 0$  and  $\mathcal{M}_2$  is penalized by its three additional parameters.

In the complete observation setting, for moderate observational noise ( $r = 0.1$ ) we find that  $\mathcal{M}_1$  is given strong support for  $\omega < 0.07$  ( $w_{\mathcal{M}_1} > 98\%$ ) and that  $\mathcal{M}_2$  is favored for  $\omega > 0.08$  ( $w_{\mathcal{M}_2} > 99\%$ ), providing overall strong support for the true model over a large range of  $\omega$  values (Fig. 3.5A). As the observational noise increases the support strength naturally decreases, leading to an increased range of  $\omega$  values where the simplest model  $\mathcal{M}_1$  is favored or where no model is given strong support (Fig. 3.5B-C). On the other hand, in the partial observation setting, the lack of data prevents the correct estimation of the omnivory variant model parameters, leading model  $\mathcal{M}_1$  to be supported for an even larger range of  $\omega$  values (Fig. S3.6).

Overall, the ML framework provides statistical support for the model embedding the most appropriate hypotheses given the available data. With appropriate data, the proposed ML framework can therefore elucidate mechanistic pathways and infer ecological processes by utilizing information-based model selection.



**Fig. 3.5:** Performance of the ML framework in supporting the predator omnivory hypothesis in a food web. A–C Hypothesis testing for levels of noise  $r = 0.1, 0.2, 0.3$ . Blue dots correspond to  $w_{\mathcal{M}_1}$ , the Akaike weights of the simple food-web model, and orange dots correspond to  $w_{\mathcal{M}_2} = 1 - w_{\mathcal{M}_1}$ , the Akaike weights of the omnivory model, which can be interpreted as model probabilities. A–C indicate that the ML framework can detect omnivory, as the omnivory model  $\mathcal{M}_2$  is given strong support ( $w_{\mathcal{M}_2} > 99\%$ ) for most of the  $w$  range investigated.

## 3.4 Discussion

We propose a ML framework combining a mini-batch method inspired by multiple shooting methods (Pisarenko and Sornette, 2004) with automatic differentiation (Rackauckas et al., 2020a) and state-of-the-art variational optimizers (Kingma and Ba, 2014) to efficiently and accurately parametrize complex dynamical models. We show formally that splitting the data into mini-batches with a short time horizon regularizes the loss function associated with dynamical models characterized by complex dynamics, such as chaotic dynamics and limit cycles (Section 3.A). We demonstrate numerically that this reformulation ensures the success of gradient-based optimizers to parametrize ecosystem models (Figs. 3.1, 3.3 and 3.4). This mini-batch method is also relevant beyond variational methods and applies to any inferential method navigating the posterior landscape, such as evolutionary algorithms (Wilke et al., 2001; Rodriguez-Fernandez et al., 2006) or Markov Chain Monte Carlo methods (Lignell et al., 2013; Higgins et al., 2010; Xu et al., 2006; Fiechter et al., 2013; Rosenbaum et al., 2019). The proposed approach is particularly relevant for the parametrization of ecosystem models incorporating realistic ecological and adaptive mechanisms (Urban et al., 2016), which are generally associated with strong nonlinearities due to the complexity of processes linking interacting ecological compartments (Bjørnstad and Grenfell, 2001; Hastings et al., 1993; Huisman and Weissing, 1999; Benincà et al., 2008). It further integrates the practical constraints of available ecological datasets (Dornelas et al., 2018), accommodating

incomplete, noisy, shallow and independent observation data. Overall, the ML framework successfully blends ML methods with mechanistic ecosystem models to learn from ecological time series, and it could therefore improve our quantitative understanding of ecosystem dynamics and help to anticipate their responses to global changes (Urban et al., 2016).

Our work contributes to the ongoing effort to better assimilate observational data into mechanistic models (Schartau et al., 2017; Raissi et al., 2019a; Kashinath et al., 2021), with a specific focus on the parametrization of ecosystem models with strong nonlinearities. Recently, (Yazdani et al., 2020) proposed an alternative framework dubbed "systems biology informed deep learning", where a neural network is fitted to the data and the additional mechanistic model constraints are integrated. This alternative framework extends previous colocation methods (Ramsay et al., 2007; Cao et al., 2008) and has the advantage of being able to parametrize stochastic models. As it requires the selection of a neural network architecture and a "goodness of fit" parameter, it nevertheless imposes an additional layer of complexity, which might negatively affect the model parametrization (Yazdani et al., 2020). In contrast, the ML framework proposed here trains the model directly against data, using automatic differentiation and sensitivity analysis in order to apply variational optimizers directly to the model simulations. This makes it possible to bypass the use of neural networks, rendering the parametrization process simpler and more amenable to model selection (Ramsay et al., 2007).

By integrating the practical constraints imposed by ecological datasets, the ML framework can learn from short time series with partial and noisy observations (Figs. 3.3 and 3.4). Local ecosystem surveys, such as marine trawling surveys or local terrestrial surveys (see Pinsky et al., 2013; Dornelas et al., 2018; Burrows et al., 2019 and references therein), provide time series that are generally shallow in time but composed of many replicates (Hsieh et al., 2008; Clark et al., 2015), in part due to the practical difficulties of long-term monitoring (Ye and Sugihara, 2016). Our results show that the inclusion of multiple independent time series in the training dataset reduces the error in the parameter estimates and increases the forecast skill (Fig. 3.4). This indicates that the proposed ML framework could, in practice, efficiently harness the information available in current ecological datasets. Instead of directly comparing simulated and observed data, matching time-averaged statistics between observations and simulations (e.g. means and covariances) could further yield an improved assimilation of observations from diverse data sources, such as global observations of productivity from satellites and local surveys, as proposed for climate models (Schneider et al., 2017). Overall, the proposed ML framework accommodates the specificities of current ecological time series and can improve the assimilation of ecological data into mechanistic ecosystem models.

Our work can help elucidate mechanistic pathways by contrasting hypotheses embedded in model variants. Using information-criterion-based model selection, we demonstrate with a case study that the ML framework is able to provide statistical support for the true generating model among two different candidates (Fig. 3.5). Importantly, the ML framework can perform model selection on complex models, incorporating key mechanisms such as trait–species interactions, evolutionary potential and responses to environmental conditions, which have been shown to be important in mediating ecosystem dynamics and must be refined in models to improve predictive accuracy (Urban et al., 2016). The

ML framework can therefore lead to the improvement of current ecosystem models and knowledge, which is crucially needed given that key ecological processes are only partially described in most ecosystem models (Schartau et al., 2017). AIC can also be used to ensure the interpretability of the model parameters, and should be preferred to estimating the parameter uncertainty through e.g. the Cramer Rao inequality (Burnham and Anderson, 2002): by favouring models with less complexity, model selection techniques disqualify uninformative parameters to ensure interpretability (Burnham and Anderson, 2002). This has the extra benefit of reducing the dimensionality of the parameter space, hence improving the estimation of other parameters. Following recent novel approaches to investigate ecological hypotheses (Curtsdotter et al., 2019), our method contributes to the development of a process understanding of ecosystem functions and provides a path forward to better link ecological theory and data.

The proposed approach still presents a set of limitations, which might hamper its success under specific situations. First, while the use of mini-batches smooths the loss surface and ensures better convergence, it also flattens the loss surface around the true parameter value, which consequently deteriorates the precision of the inferred parameters because the loss function takes similar values in an extended neighbourhood of the true parameters. To circumvent this issue, iterative training can be performed, where the learning is initiated by a short batch length  $K^{(s)}$  to identify the region with the most probable parameters, and in subsequent iterations the batch length is increased to improve the precision of the inference. Iterative training could also improve the lack of statistical support obtained in hypothesis testing experiments (see simulations in Figs. 3.5 and S3.6 where none of the models is given statistical support), as it would increase differences in likelihood for parameter values around the neighbourhood of the true parameter values. Second, our results highlight that the data might not provide enough constraints for a correct parametrization (Fig. 3.4, partial observation setting and  $S = 1$ ). Pre-experimental analyses with simulated synthetic data might therefore be required to design the sampling protocol and campaign to ensure an adequate sampling effort (Banks et al., 2017; Laubmeier et al., 2018). Third, while in Eq. (3.8) it is assumed that the parameter values are the same across the time series, strong regional variability might also be observed among the spatially replicated data, causing parameter values to vary across the replicates. The knowledge of this variability could motivate partial pooling (Beaumont, 2010) or parametrization of the parameters in terms of environmental conditions (Pahlow et al., 2008), to account for the independence of the parameter values across the replicates. Finally, while the proposed ML framework greatly improves convergence, it could still be that – even with a large amount of data – poor initial parameter estimates, a large number of free parameters, or high noise levels prevent convergence to the true minimum. Performing multiple runs with varying initial parameter estimates can ensure that the maximum a priori estimate is reliable. If this is not the case, stochasticity could further be introduced within the ML framework to prevent the convergence to local minima, where only a subset of mini-batches are fitted at each epoch (Bottou, 2012).

## 3.5 Conclusion

We proposed a ML framework based on a mini-batch method combined with automatic differentiation and state-of-the-art optimizers to estimate the parameters and improve the forecast skill of complex ecosystem models from observation data. The ML framework was benchmarked with a realistic ecosystem model characterized by strong nonlinearities and delivered excellent performance, accommodating the practical constraints imposed by the quality and availability of ecological datasets. Our experiments have further illustrated the ability of the ML framework to discriminate between several candidate models, enabling the testing of ecological theories against data and the improvement of current mechanistic models. Given the increasing number of ecological datasets following the development of monitoring technologies such as environmental DNA (Ruppert et al., 2019), remote sensing (Jetz et al., 2019), bioacoustics (Aide et al., 2013), and citizen observations (GBIF: The Global Biodiversity Information Facility, 2022a), the proposed ML framework opens up new opportunities for the quantitative investigation of current ecosystem functions (Curtsdotter et al., 2019) and the prediction of ecosystem responses to increasing disruptions (Urban et al., 2016).

## 3.6 Code availability

The ML framework is implemented in the multi-purpose Julia package **MiniBatchInference.jl** available at <https://github.com/vboussange/MiniBatchInference.jl>, and the simulation code is available at <https://github.com/vboussange/mini-batching-ecological-data>.

## 3.7 Acknowledgements

L.P. and V.B. were supported by the SNF grant 310030E\_205556. P.V.A. was supported by an ETH postdoctoral fellowship.

## 3.A Supplementary Information

We show that the loss function  $L_{\mathcal{M}}(\theta)$  in the main manuscript in Eq. (3.6) is ill-behaved for models with complex dynamics when the time horizon is large. We proceed by first analysing the dynamics of models with complex dynamics showing chaotic behaviour or limit cycles, and approximate the divergence of perturbed dynamical trajectories. We then show that the divergence in dynamics translates into a loss function whose surface is rugged in most of the parameter space, and that the gradient of the loss function around the true parameters becomes exponentially steeper with time. We conclude by formally discussing how the proposed mini-batch method regularizes the ill-behaviour of the loss function.

### 3.A.1 Dynamics under perturbations

#### Perturbed initial conditions

Consider the trajectory of the state variables

$$\begin{aligned} x(t) &= \mathcal{M}(t, p, x_0) \\ &= \int_0^t f(s, x(s), p) ds + x_0 \end{aligned} \quad (\text{S3.11})$$

and consider the perturbed trajectory

$$x_{\delta x_0}(t) = \mathcal{M}(t, p, x_0 + \delta x_0) \quad (\text{S3.12})$$

whose initial conditions (ICs)  $x_0$  are perturbed by  $\delta x_0$ . Assuming that the system is chaotic and that  $\delta x_0$  is small, the distance between the perturbed trajectory and the original one grows as

$$\|x(t) - x_{\delta x_0}(t)\| \sim e^{\lambda t} \delta x_0 \quad (\text{S3.13})$$

where  $\lambda$  is the largest Lyapunov exponent of the system (Strogatz, 2018). After enough time, the trajectories diverge so much that they effectively become independent samples of the phasespace: the trajectories forget their ICs, and ergodic theory ensures that the positions of  $x(t)$  and  $x_{\delta x_0}(t)$  are better described by a random variable  $\mathcal{R}$  with probability density given by the density of orbits in the chaotic attractor, the so-called invariant measure of the chaotic attractor (Jost, 2005). The distance between  $x(t)$  and  $x_{\delta x_0}(t)$  can therefore be described as

$$\|x(t) - x_{\delta x_0}(t)\| \sim \|\mathcal{R}_1 - \mathcal{R}_2\| \quad (\text{S3.14})$$

Considering the observation function  $h$ , it follows that

$$\begin{aligned} \|h(x(t)) - h(x_{\delta x_0}(t))\| &\sim \left. \frac{\partial h}{\partial x} \right|_{x(t)} e^{\lambda t} \delta x_0 \text{ for } t \ll \frac{1}{\lambda} \\ \|h(x(t)) - h(x_{\delta x_0}(t))\| &\sim \|h(\mathcal{R}_1) - h(\mathcal{R}_2)\| \text{ for } t \gg \frac{1}{\lambda} \end{aligned} \quad (\text{S3.15})$$

(see Fig. S3.7 for an illustration of the divergence behaviour over time).

#### Perturbed model parameters

Consider now a trajectory  $x_{\delta p}$  with a small perturbation of the parameters  $\delta p$ . From Eqs. (S3.11) and (S3.12), it follows that

$$\begin{aligned} \dot{x}_{\delta p}(t) &= f(t, x_{\delta p}(t), p + \delta p) \\ &\sim f(t, x_{\delta p}(t), p) + \frac{\partial f(t, x_{\delta p}(t), p)}{\partial p} \delta p \\ &\sim f(t, x(t), p) + \frac{\partial f(t, x_{\delta p}(t), p)}{\partial p} \delta p + \frac{\partial f(t, x(t), p)}{\partial x(t)} \frac{\partial x(t)}{\partial p} \delta p \end{aligned} \quad (\text{S3.16})$$

which is dominated by the first term under small values of  $\delta p$ , and is thus subject to chaotic dynamics. Similar to a perturbation of the ICs, the small perturbation  $\delta p$  generates a divergence in the dynamical trajectories that grows exponentially until they become uncorrelated. For small  $\delta p$ , the distance between the true and deviated trajectory can therefore be approximated as

$$\|x(t) - x_{\delta p}(t)\| \sim \int_0^t e^{\lambda(t-s)} \|x(s) - x_{\delta p}(s)\| ds \sim e^{\lambda t} g(\delta p) \quad (\text{S3.17})$$

where  $g(\delta p) = \left\| \frac{\partial f(t, x_{\delta p}(t), p)}{\partial p} + \frac{\partial f(t, x(t), p)}{\partial x(t)} \frac{\partial x(t)}{\partial p} \right\|$  gives the scale of the divergence between the two trajectories as a function of  $\delta p$ . Similar to a perturbation of the ICs, the difference in trajectories grows to the point where the trajectories become effectively independent after a long time. Hence, it follows that

$$\begin{aligned} \|h(x(t)) - h(x_{\delta p}(t))\| &\sim \left. \frac{\partial h}{\partial x} \right|_{x(t)} e^{\lambda t} g(\delta p) \text{ for } t \ll \frac{1}{\lambda} \\ \|h(x(t)) - h(x_{\delta p}(t))\| &\sim \|h(\mathcal{R}_1) - h(\mathcal{R}_2)\| \text{ for } t \gg \frac{1}{\lambda} \end{aligned} \quad (\text{S3.18})$$

(see Fig. S3.8 for an illustration of the divergence behaviour over time).

In the following section, we call the first divergence regime the informative divergence regime, where the loss grows with the distance to the true parameters, and we call the second divergence regime the mixed divergence regime, where the loss is dominated by the random-like behaviour. Given that  $\delta p$  and  $\delta x_0$  behave similarly, we employ  $\theta$  and  $\delta\theta$  to encompass both perturbations of parameters and ICs, and denote by  $g(\delta\theta)$  the function that gives the scale of the divergence in trajectories for both  $g(\delta p)$  and  $\|\delta x_0\|$ .

### Transition in the parameter space between the informative and the mixed regime

For a fixed time horizon  $t$ , and depending on the shape of the chaotic attractor, the magnitude of the perturbation determines the divergence regime. If the perturbation is small, the trajectories will be aligned, but for large perturbations they will effectively become two independent trajectories.

The transition between the two regimes can be studied by noting that the informative divergence should remain in the same order of magnitude as the mixed divergence. The reason is that the expected value of the squared divergence between two trajectories  $x(t)$  and  $x_{\delta\theta}(t)$  for large  $t$  is

$$\mathbb{E} [\|h(\mathcal{R}_1) - h(\mathcal{R}_2)\|^2] = 2\text{Var}[h(\mathcal{R})] \lesssim \max |h(\mathcal{R}) - \mathbb{E}[h(\mathcal{R})]|^2 \quad (\text{S3.19})$$

meaning that the expected value of the squared divergence in the mixed regime is in the same order of magnitude as the maximum distance within the phasespace. On the other hand, the divergence of any two trajectories in the chaotic attractor cannot be larger than the maximum distance between two points in the attractor, which is itself bounded through the triangle inequality as  $2 \max |h(\mathcal{R}) - \mathbb{E}[h(\mathcal{R})]|$ .

Since the growth of the informative regime has to remain in the same order of magnitude as the mixed regime, at the regime transition we must have

$$e^{\lambda t} g(\delta\theta) \sim \mathbb{E} [\|h(\mathcal{R})\|] \quad (\text{S3.20})$$

Equation (S3.20) implies that, for a given time horizon  $t$ , the magnitude of the critical perturbation  $\delta\theta^*$  associated with the regime transition satisfies

$$\|\delta\theta^*\| \sim e^{-\lambda t} \quad (\text{S3.21})$$

### Limit cycles

While in the section above a chaotic system was assumed to provide an approximation for the divergence of the trajectories, a similar approximation applies for systems characterized by limit cycles. Considering a system  $x(t)$  with a limit cycle characterized by the phase  $\omega t$  with frequency  $\omega$ , i.e.

$$x(t) = f(\phi(t)) = f(\omega t) \mod 2\pi \quad (\text{S3.22})$$

a perturbation of the parameters  $\delta p$  might lead to a perturbed frequency  $\delta\omega$ , further leading to a difference in phases

$$\delta\phi(t) = \phi(t) - \phi_{\delta\omega}(t) = \delta\omega t \mod 2\pi \quad (\text{S3.23})$$

For  $t \lesssim \frac{1}{\delta\omega}$ ,  $\delta\phi(t)$  grows linearly with  $\delta\omega$ , but once  $t \gg \frac{1}{\delta\omega}$ , the change of phase  $\delta\phi(t)$  is affected by the modulo operation. As this operation is nonlinear, a small random perturbation  $\delta\omega$  results in a random uniform phase over the interval  $[0, 2\phi]$ . For a large time horizon,  $x_{\delta\omega}$  is thus uniformly spread over the circular line given by the phasespace of the dynamical system. Hence, the approximation in Eq. (S3.18) applies for cyclic dynamics, except that the initial divergence is linear rather than exponential, and that in contrast to Eq. (S3.15), a change affecting the initial position will not grow over time.

### 3.A.2 Consequences for the shape of the loss surface

The approximation of the divergence of trajectories in Eqs. (S3.15) and (S3.18), together with the transition boundary determined by Eq. (S3.21), can be used to characterize the surface associated with the loss function.

Omitting the term corresponding to the priors and the variance–covariance matrix  $\Sigma_y$  for simplicity, the loss function presented in the main text is expressed as

$$\begin{aligned} L_{\mathcal{M}}(\theta) &= \frac{1}{K} \sum_{k=1}^K \|y_k - h(\mathcal{M}(t_k, \theta))\|^2 \\ &= \frac{1}{K} \sum_{k=1}^K \|h(\tilde{x}(t_k)) + \epsilon(t_k) - h(\mathcal{M}(t_k, \theta))\|^2 \end{aligned} \quad (\text{S3.24})$$

where the parameter vector  $\theta$  is decomposed into the model parameter vector  $p$  and the ICs  $x_0$ ,  $y_k$  correspond to the observations,  $\tilde{x}$  corresponds to the true trajectory and  $\epsilon(t)$  is the

observational noise. As the noise is independent of the dynamics, it is uncorrelated with  $h(\tilde{x}(t_k)) - \mathcal{M}(t_k, \theta)$ , meaning that the loss can be split in expectation

$$\mathbb{E}[L_{\mathcal{M}}(\theta)] = \frac{1}{K} \sum_{k=1}^K \|h(\tilde{x}(t_k)) - h(\mathcal{M}(t_k, \theta))\|^2 + \text{Var}[\epsilon] \quad (\text{S3.25})$$

where the noise term ( $\text{Var}[\epsilon]$ ) is independent from the parameters  $\theta$ . Assuming that  $\theta = \tilde{\theta} + \delta\theta$ , where  $\tilde{\theta}$  correspond to the true parameters and ICs, every term in the sum corresponds to a squared distance between the true trajectory  $\tilde{x}(t)$  and a perturbed trajectory  $\tilde{x}_{\delta\theta}(t)$ . Using Eqs. (S3.15) and (S3.18) to (S3.20) we obtain the loss function approximation

$$L_{\mathcal{M}}(\theta) \sim \sum_{k=1}^{\min\{K, K^*\}} e^{2\lambda t_k} g^2(\delta\theta) + \sum_{k=\min\{K, K^*\}+1}^K \text{Var}[h(\mathcal{R})] + \text{Var}[\epsilon] \quad (\text{S3.26})$$

where  $K^*$  is the observation index corresponding to the time horizon  $t_{K^*}$ , where the transition between the informative and the mixed regime happens for the perturbation  $\delta\theta$ , obtained from Eq. (S3.21).

The distribution of the observation times and the magnitude of the perturbation  $\delta\theta$  determine whether the loss is dominated by the informative or by the mixed divergence regime. For a fixed perturbation  $\delta\theta$ , assuming that the observations  $y_k$  are uniformly distributed over the time interval  $[0, t]$ , Eq. (S3.21) yields for  $t \ll \frac{\log(|\delta\theta|)}{\lambda}$  that the loss  $L_{\mathcal{M}}$  is dominated by the informative divergence regime, whereas if  $t \gg \frac{\log(|\delta\theta|)}{\lambda}$  the loss is dominated by the mixed divergences. In the region where the loss is dominated by mixed divergences, the loss  $L_{\mathcal{M}}$  has an expected value of order  $\mathcal{O}(\text{Var}[h(\mathcal{R})] + \text{Var}[\epsilon])$  and does not grow monotonically with  $\delta\theta$ . It corresponds to a "random-like" surface populated with local minima, and is consequently characterized by an uninformative gradient preventing local optimizers from converging to the true parameters  $\tilde{\theta}$  (see Fig. 3.1B, orange dashed curve). On the other hand, in the region where the divergence in trajectories belongs to the informative regime, the loss is convex and grows with  $\delta\theta$ . Its associated gradient  $\nabla_{\theta} L_{\mathcal{M}}(\theta) \sim \sum_{k=1}^K e^{2\lambda t_k} \nabla_{\theta} g^2(\delta\theta)$  consequently contains relevant information for the use of variational optimizers, but the loss surface becomes exponentially steeper as the time horizon increases. As shown by the green dotted curve in Fig. 3.1B, this large gradient in the vicinity of the optimal parameters is likely to lead gradient-based optimizers to overshoot and not converge to the true parameters  $\tilde{\theta}$ . Eq. (S3.21) further indicates that the volume of the region in the parameter space where the loss is informative shrinks exponentially as the time horizon increases, implying that for large time horizons, the uninformative region is predominant.

### 3.A.3 Regularizing the loss surface with mini-batches

To prevent the situation with a mixed divergence regime and to decrease the gradient in a vicinity of the true parameters, we reformulate the loss function as the average of loss functions defined over mini-batches of short time horizons. In the following section, we compare the properties of the naive loss function  $L_{\mathcal{M}}$  in Eq. (S3.25) with the mini-batch

loss function  $L_{\mathcal{M}}^*$  proposed in Eq. (3.8) in the main manuscript, and further discuss the limitations of the method in the presence of noise.

Omitting the term corresponding to the priors and ICs and the noise shape  $\Sigma_y$ , and assuming a single time series for simplicity, the mini-batch loss function presented in the main text can be expressed as

$$L_{\mathcal{M}}^*(\theta) = \frac{1}{M} \sum_{m=0}^{M-1} \frac{1}{K} \sum_{k=1}^{K^{(m)}} \|y_{k+mK/M} - h(\mathcal{M}(t_{k+mK/M}, p, x_0^{(m)}))\|^2 \quad (\text{S3.27})$$

where  $M$  is the number of mini-batches,  $x_0^{(m)}$  corresponds to the ICs for mini-batch  $m$  inferred at time  $t_{mK/M}$ , and  $K^{(m)} = \begin{cases} K/M + 1 & : m < M \\ K/M & : m = M \end{cases}$  is the number of data points in the  $m$ th batch. The loss function can be split in expectation as

$$\mathbb{E}[L_{\mathcal{M}}^*(\theta)] = \frac{1}{M} \sum_{m=0}^{M-1} \frac{1}{K} \sum_{k=1}^{K^{(m)}} \|h(\tilde{x}(t_{k+mK/M})) - \mathcal{M}(t_{k+mK/M}, p, x_0^{(m)})\|^2 + \text{Var}[\epsilon] \quad (\text{S3.28})$$

Assuming that  $x_0^{(m)} = \tilde{x}(t_{mK/M}) + \delta x_0^{(m)}$  when  $\delta x_0^{(m)}$  is small, and assuming that the observation times are regularly spaced so that  $t_{k+1} - t_k = \Delta t$ , the time length of simulated trajectories (i.e. the time elapsed between the time when the perturbation is applied and the end time of the simulated trajectory) is divided by the number of mini-batches, in comparison to the time length of the simulated trajectory in Eq. (S3.25) ( $t = t_{K/M+1} - t_0 = K/M\Delta t$  in Eq. (S3.28), in comparison to  $t = K\Delta t$  in Eq. (S3.25)). Further assuming that  $K\Delta t >> \frac{\log(|\delta\theta|)}{\lambda}$  and choosing the number of mini-batches  $M$  so that  $K/M\Delta t << \frac{\log(|\delta\theta|)}{\lambda}$ , we apply the approximation Eq. (S3.26), which leads to

$$\begin{aligned} L_{\mathcal{M}}(\theta) &\sim \text{Var}[\epsilon] + \text{Var}[h(\mathcal{R})] \\ L_{\mathcal{M}}^*(\theta) &\sim \text{Var}[\epsilon] + g^2(\delta\theta) \sum_{m=0}^{M-1} e^{2\lambda(K/M\Delta t)} \end{aligned} \quad (\text{S3.29})$$

While  $L_{\mathcal{M}}$  is dominated by the mixed regime,  $L_{\mathcal{M}}^*$  is dominated by the informative regime because the simulation time remains small, permitting the successful use of variational optimizers.

The number of mini-batches  $M$  should be determined by considering the dynamical behaviour of the system and the level of noise in the observation, because a large number of mini-batches  $M$  smooths out the loss surface but also entails more sensitivity to the level of noise. Indeed, in  $L_{\mathcal{M}}(\theta)^*$  the relative effect of the second term corresponding to the observational noise  $\epsilon$  increases when the number of mini-batches  $M$  increases. The value of  $M$  should therefore be chosen wisely to balance the benefits of mini-batches, i.e. widening the region of the parameter space where  $L_{\mathcal{M}}$  is well behaved and reducing the overshooting problem, and their cons, i.e. their tendency to increase the importance of noise.

## 3.B Three-compartment food-web models

### 3.B.1 Reference food-web model

We used the three-species chaotic food-web model from (Hastings and Powell, 1991), formulated as

$$\begin{aligned}\frac{d}{dt}R &= R(1 - R) - x_C y_C \frac{CR}{R + R_0} \\ \frac{d}{dt}C &= x_C C \left[ -1 + x_C \frac{R}{R + R_0} \right] - x_P y_P \frac{PC}{C + C_0} \\ \frac{d}{dt}P &= x_P P \left[ -1 + y_P \frac{C}{C + C_0} \right],\end{aligned}\tag{S3.30}$$

with the biologically realistic parameter values  $x_C = 0.4$ ,  $0.071 \leq x_P \leq 0.225$ ,  $y_C = 2.01$ ,  $y_P = 5$ ,  $R_0 = 0.16129$ , and  $C_0 = 0.5$  (McCann and Yodzis, 1994a). The dynamics of the system are chaotic for this set of parameter values.

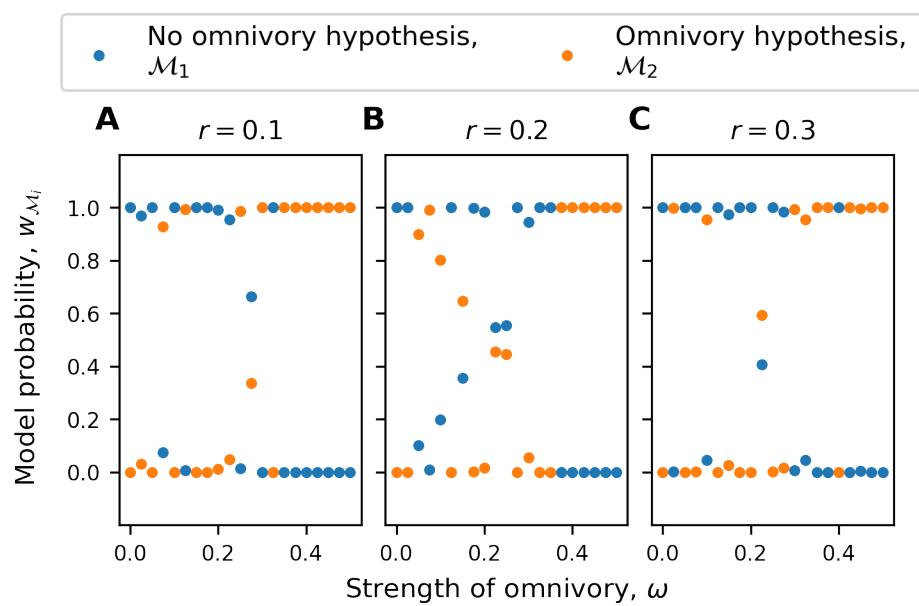
### 3.B.2 Omnivory variant food-web model

We used the three-species food-web model from (McCann and Hastings, 1997), formulated as

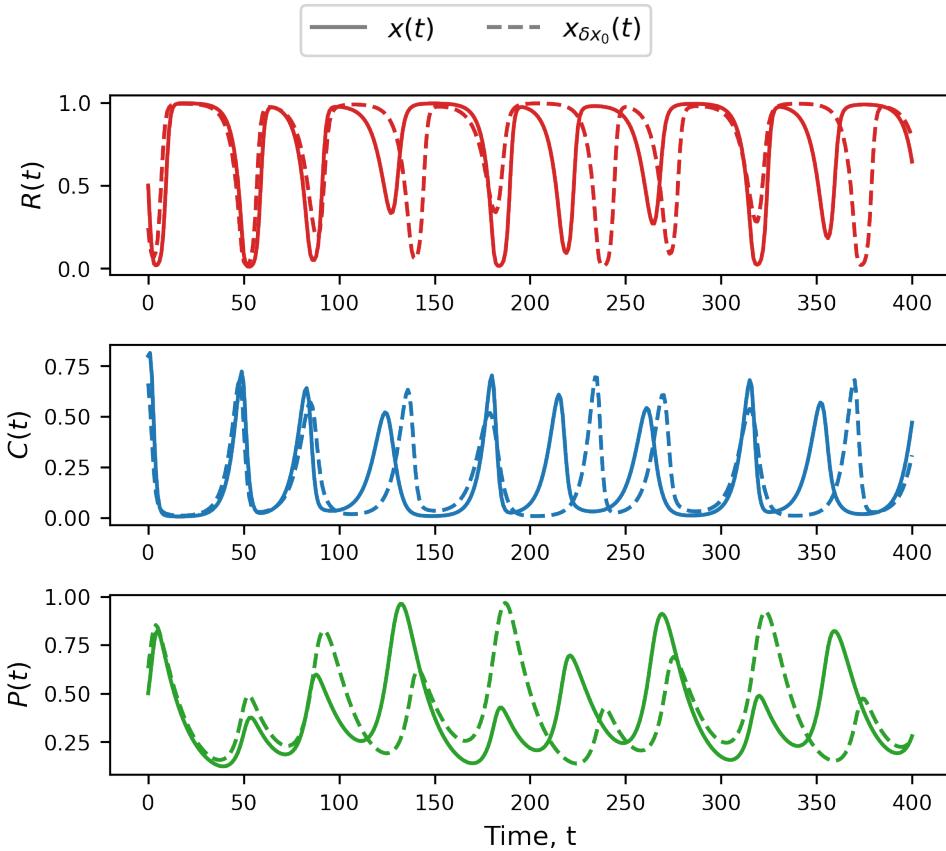
$$\begin{aligned}\frac{d}{dt}R &= R(1 - R) - x_C x_C \frac{CR}{R + R_0} - \omega x_P y_{PR} \frac{PR}{R_{02} + (1 - \omega)C + \omega R} \\ \frac{d}{dt}C &= x_C C \left[ -1 + x_C \frac{R}{R + R_0} \right] - (1 - \omega)x_P y_{PC} \frac{PC}{\omega R + (1 - \omega)C + C_0} \\ \frac{d}{dt}P &= x_P P \left[ -1 + (1 - \omega)x_P y_{PC} \frac{C}{\omega R + (1 - \omega)C + C_0} \right. \\ &\quad \left. + \omega x_P y_{PR} \frac{R}{\omega R + (1 - \omega)C + R_{02}} \right]\end{aligned}\tag{S3.31}$$

with the biologically realistic parameter values  $x_C = 0.4$ ,  $x_P = 0.08$ ,  $y_C = 2.009$ ,  $y_{PR} = 2$ ,  $y_{PC} = 5$ ,  $R_0 = 0.16129$ ,  $C_0 = 0.5$ , and  $0 \leq \omega \leq 0.5$ . For this set of parameter values, the dynamics of the system are chaotic for  $\omega \lesssim 0.20$ , consist of a limit cycle for  $0.20 \lesssim \omega \lesssim 0.35$ , and consist of damped oscillations for  $0.35 \lesssim \omega$ .

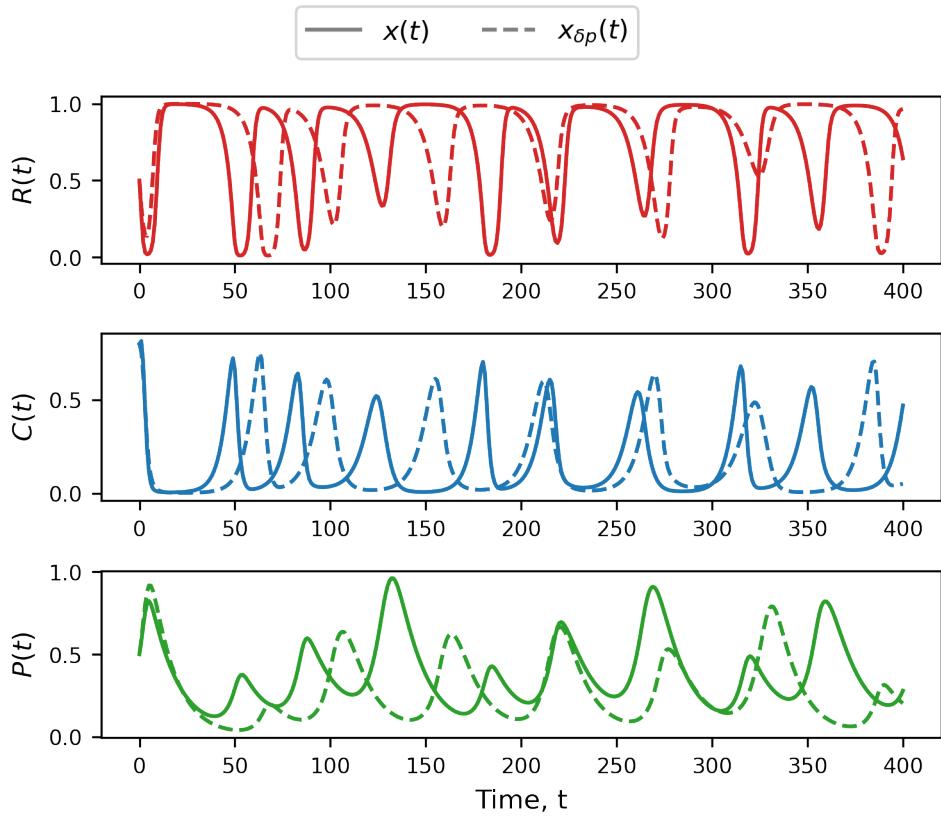
## 3.C Supplementary Figures



**Fig. S3.6:** Performance of the ML framework in supporting the predator omnivory hypothesis in a food web for the partial observation setting. In A, B and C where  $r = 0.1, 0.2, 0.3$ , the lack of data prevents the correct estimation of the omnivory variant model parameters, leading model  $M_1$  to be supported for a wider range of  $\omega$  in contrast to the complete observation setting (Fig. 3.5A).



**Fig. S3.7:** Divergence between the trajectory  $x(t)$  and a perturbed trajectory  $x_{\delta x_0}(t)$ , obtained from the reference food-web model from Hastings and Powell, 1991 and detailed in Section 3.B. For  $t \lesssim 100$ ,  $x(t)$  and  $x_{\delta x_0}(t)$  are correlated and the divergence regime is informative, but for  $t \gtrsim 100$  the trajectories become essentially uncorrelated, corresponding to the mixed divergence regime.



**Fig. S3.8:** Divergence between the trajectory  $x(t)$  and a perturbed trajectory  $x_{\delta p}(t)$ , obtained from the reference food-web model from Hastings and Powell, 1991 and detailed in Section 3.B. For  $t \lesssim 40$ ,  $x(t)$  and  $x_{\delta p}(t)$  are correlated and the divergence regime is informative, but for  $t \gtrsim 40$  the trajectories become essentially uncorrelated, corresponding to the mixed divergence regime.

### 3.D Supplementary Tables

Setting	Median simulation time	Mean simulation time	Std. simulation time
Complete observations	35.9977436	39.5980174	20.5789433
Partial observations	34.1293998	39.1896534	21.8191709

**Tab. S3.1:** Simulation time for the complete and partial observation settings.



# 4

## Analogous processes to ecological interactions and dispersal shape the dynamics of economic activities

by Victor Boussange<sup>1,2</sup>, Didier Sornette, Heike Lischke<sup>3</sup>, and Loïc Pellissier<sup>1,2</sup>

<sup>1</sup> Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland

<sup>2</sup> Landscape Ecology, Institute of Terrestrial Ecosystems, Department of Environmental System Science, ETH Zürich, CH-8092 Zürich, Switzerland

<sup>3</sup> Dynamic Macroecology, Land Change Science, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

**”**

*Thus, dare! your heritage, your acquisition,  
Stories, lessons from the mouths of your fathers,  
Laws and customs, names of the ancient Gods,  
Forget them with hardness to raise your eyes  
like newborns, on divine nature.*

— Friedrich Hölderlin (1770-1843)  
The Death of Empedocles (1st version)

The fundamental processes of ecological interactions, dispersal and mutations shape the dynamics of biological systems, and analogous eco-evolutionary processes acting upon firms have been proposed to explain economic change. This hypothesis is compelling because it explains economic change through endogenous mechanisms, but whether it can characterize economic change at the country level is unclear. Here, we use an inverse modelling technique, together with 59 years of economic data covering 77 countries, to test whether the collective dynamics of economic activities at the country level can be characterized by eco-evolutionary processes. We estimate the statistical support of population dynamic models coupling the dynamics of economic activities by accounting, respectively, for positive and negative interactions between themselves, their spatial dispersal, and their transformations into other economic activities. We compare the models' statistical support against a logistic growth model taken as a null model, and find that the null model is rejected in 45 countries against the alternative models. In particular, we find a strong support for the models capturing positive interactions between economic activities, and spatial dispersal across countries. These results suggest that processes akin to those displayed in ecosystems play a significant role on the dynamics of economic systems at the national level. The strength-of-evidence obtained for each model varies across countries, and might underlie differences in proximities between countries, or specific institutional contexts and historical contingencies. Overall, our study provides a new quantitative, biologically inspired framework to study the forces shaping economic change.

## 4.1 Introduction

The fields of evolutionary biology and economics have mutually exchanged ideas for the past two centuries (Dopfer and Potts, 2007), and analogies between processes in biology and economics have been increasingly used during the 20th century to explain economic change (Ruth, 1996). A pioneer was Schumpeter, who notoriously proposed that economic dynamics is driven by innovations that transform the economy – so called periods of "creative destruction" (Schumpeter, 1934)– similar to the punctuated equilibrium changes observed throughout the development of life on Earth (Gould and Eldredge, 1972). The analogy with biology was further developed within the field of evolutionary economics (Hodgson, 2019), largely promoted by the seminal work of Nelson, 1985. The premise of evolutionary economics considers habits, customs, and organizational routines as "replicators", i.e. atomic units playing the role of genes and defining as a whole the cohesive identity of an economic entity (Hodgson, 2019). Under this framework, firm business strategies determine how they transform commodities and knowledge into new knowledge, new technologies or other industrial products with more added value. The replicators define the fitness of an economic entity within a given economic context, which grows in terms of assets and human capital and survive within an ecosystem of entities (Hodgson, 2002), continuously adapting and experiencing evolutionary processes (Veblen, 1898). While biological analogies have resulted in useful insights on the plausible drivers of economic growth (Dopfer and Potts, 2007), the proposed pathways have mostly been qualitatively investigated. Computational tools have recently allowed to test biological hypotheses against data to gain a quantitative understanding of the ecological and evolutionary processes shaping the dynamics of ecological systems (Pontarp et al., 2019; Boussange et al., 2022; Skeels et al., 2022). These tools could leverage the qualitative insights gained from biological analogies and provide a quantitative framework to investigate the forces shaping economic change.

Interactions between biological organisms, dispersal (movement of individuals across space), and mutations, represent fundamental processes that drive the dynamics of ecosystems (Vellend, 2010b), and similar eco-evolutionary processes may shape the dynamic of economic systems. Analogously to biological organisms, economic entities interact in a mutualistic or competitive fashion (Pistorius and Utterback, 1997). While biological organisms engage in negative interactions when e.g. competing for similar resources (Grime, 1973), economic entities have negative effects on each other when relying on workforce with similar knowledge or attracting similar potential investors (Wernerfelt, 1989). Positive interactions are observed between biological organisms with e.g. the accumulation of nutrients, the provision of shade, or the protection from herbivores (Callaway et al., 2002). In parallel, positive effects between economic entities appear when they are connected through supply chains (Ozman, 2009; Saavedra et al., 2009), or when they benefit from positive agglomeration externalities (Van Der Panne, 2004), such as knowledge spillovers (Caragliu and Nijkamp, 2016) or the attraction of elite workers (Cohendet et al., 2018). Dispersal processes play a major role in the development of ecosystems, where e.g. colonization initiate ecological succession with the settlement and growth of new species (Leibold et al., 2004). Similarly, socio-economic processes contribute to diffuse knowledge and organizational routines across space. Those processes involve, e.g., international business expansions

(Andersen, 1993; Zahra et al., 2000; Luo and Tung, 2007), labor mobility (Boschma et al., 2008), or the diffusion of innovations through social networks (Rogers, 2003; Keller, 2004; Bahar et al., 2014). Finally, evolutionary processes allow the adaptation of organisms to changing environmental conditions (Bell, 2017), while in economic systems, variations in organizational routines allow economic entities to adapt to economic contexts (Cordes, 2006) and transform into new economic entities (Freeman, 2002; Hodgson and Knudsen, 2004; Aldrich et al., 2008). While processes acting upon economic activities, including their interactions, their spatial dispersal, and their transformations into other economic activities, are regularly documented in evolutionary economics, the significance of these processes on long-term economic change has seldom been quantified.

Focusing on endogenous forces and borrowing concepts and methods from biology, a number of modelling approaches have broken with the traditions of standard economic modelling (Lawson, 2013) to study the processes driving economic patterns (Tacchella et al., 2018). For instance, Saavedra et al., 2009 investigated the effect of cooperation within firms with a model of consumer-resource interactions. Scholl et al., 2021 developed a theory of market ecology to interpret market phenomena and predict market behavior, where financial trading strategies are the analogous to biological species. Lotka-Volterra models were used in the prediction of technology evolution (Zhang et al., 2018), to evaluate product and firm competition (Modis, 1997; Saavedra et al., 2014), to understand the drivers of market share dynamics (Farmer and Lo, 1999; Michalakelis et al., 2011; Marasco et al., 2016; Gatabazi et al., 2019) and to estimate the value of firms (Cauwels and Sornette, 2012). Applegate and Lampert, 2021 investigated an ecological model of competition-colonization dynamics to understand the distribution of firm sizes, and Suweis et al., 2015 used a population dynamic model to evaluate the effect of international trade on global food security. The availability of fine-grained datasets of economic activities, together with dimensionality reduction techniques, have provided insights into the endogenous processes shaping economic development (Mealy et al., 2019; Hidalgo, 2021). In particular, the detailed global trade data compiled by the United Nations statistical Division and cleaned in Hidalgo, 2021 consists in a time series of 59 years of economic activity that can be combined with inverse modelling techniques to learn about the processes influencing the long-term dynamics of economic systems.

Here, we investigate whether eco-evolutionary processes can quantitatively explain economic growth using an inverse modelling technique together with 59 years of data on economic activity. Relating the temporal evolution of the capital of economic activities to the temporal evolution of the biomass of functional groups within an ecosystem, we quantitatively estimate the effect of eco-evolutionary processes on the long-term development of national economies. Functional groups, as the aggregation of species sharing similar characteristics and having similar functions within the ecosystem, are the required elemental units to model ecosystem dynamics. Analogously, as the aggregation of firms with similar output nature (Applegate and Lampert, 2021), we consider 9 economic activities to form the elemental units of economic systems, and conduct our investigation by modelling their temporal dynamics. We consider a null model implementing the fundamental processes of self-replication and self-limitation ( $\mathcal{M}_{null}$ ), where no couplings between economic activities are captured, which we contrast to alternative dynamic models coupling their dynamics,

including negative and positive ecological interactions between themselves ( $\mathcal{M}_{\alpha^-}$ ,  $\mathcal{M}_{\alpha^+}$ ), their spatial dispersal ( $\mathcal{M}_\delta$ ), and their transformations into other economic activities ( $\mathcal{M}_\mu$ ). We use the mini-batching machine learning technique of Boussange et al., 2022 to estimate the maximum likelihood of each model for 100 countries, and use a model selection technique to evaluate the statistical support of each alternative model, based on a data set of national exports from 1962 to 2020, taken as a proxy for the temporal development of the capital of economic activities. We first show that the machine learning method, together with the model selection procedure, can provide support for the generating processes in a controlled experiment. Further, applying the model selection procedure to the empirical data, we find evidences for eco-evolutionary processes, and observe differences in model support across countries, which we relate to socio-economic drivers. By adopting a biological perspective on economic dynamics, our study sets the basis for an understanding of the endogenous forces determining economic growth, and proposes a complementary perspective to mainstream economic theory on the drivers of economic dynamics.

## 4.2 Methods

### 4.2.1 Eco-evolutionary model to characterize the dynamics of economic activities

We derive a general population dynamic model where the dynamics of national economic activities is driven by self-replication, self-limitation, interactions between themselves, their spatial dispersal across countries, and their transformations into other economic activities within the country. We present the general model in the following, which we further split into alternative sub-models to test the support of each process.

In the general population dynamic model that includes all the processes investigated, we consider that the rate of change of the size of an economic activity  $i$  in the country  $c$ , denoted as  $n_i^{(c)}$ , is determined as follows

$$\begin{aligned} \frac{d}{dt} n_i^{(c)}(t) = & r_i^{(c)} n_i^{(c)}(t) \left( 1 - b_i^{(c)} n_i^{(c)}(t) + \sum_{j \neq i}^{N^{(c)}} \alpha_{i,j}^{(c)} n_j^{(c)}(t) \right) + \sum_l^M \delta_i^{(l,c)} \left( n_i^{(l)}(t) - n_i^{(c)}(t) \right) \\ & + \sum_j^{N^{(c)}} \mu_{j,i}^{(c)} \left( n_j^{(c)}(t) - n_i^{(c)}(t) \right), \end{aligned} \quad (4.1)$$

where  $N^{(c)}$  is the number of economic activities considered in country  $c$  and  $M$  is the number of countries considered. In Eq. (4.1), the first summand corresponds to a Lotka-Volterra model for ecological communities (see, e.g., Bunin, 2017; Scheffer and Nes, 2006; Case, 1990), where  $r_i^{(c)}$  is the growth rate, involved in self-replication, and  $b_i^{(c)}$  is the self-interaction coefficient, involved in self-limitation.  $b_i^{(c)}$  can also be interpreted as the inverse of the country's carrying capacity for the activity. The individual dynamics of economic

activities are altered by interactions, where  $\alpha_{i,j}^{(c)}$  captures the interaction between activity  $i$  and activity  $j$  at location  $c$ . The interaction between  $i$  and  $j$  is mutualistic if  $\alpha_{i,j}^{(c)} > 0$ , and competitive when  $\alpha_{i,j}^{(c)} < 0$ , altering positively or negatively the rate of change of the activity  $i$  by the presence of other activities. The second summand accounts for spatial dispersal across locations (see, e.g., Tilman, 1994), where  $\delta_i^{(l,c)}$  is the rate of spatial dispersal for activity  $i$  between locations  $l$  and  $c$ . The last summand corresponds to economic activity transformations within country  $c$ , where  $\mu_{j,i}^{(c)}$  is the rate of transfer dictating how fast activity  $j$  transforms into activity  $i$ . In the sense of the quasi-species model (Eigen et al., 1988), this term can be interpreted as accounting for evolutionary processes, and can contribute to the development of an activity in the presence of activity  $j$ . A graphical representation of the model is given in Fig. 4.1.

Equation (4.1) involves a high number of independent parameters to fit (i.e.,  $N^{(c)}(1+M+2N^{(c)})$  parameters for each country), which can be reduced for the sake of parsimony under mean field assumptions. Specifically, we assume that all economic activities interact similarly, so that  $\alpha_{i,j}^{(c)} = \alpha^{(c)}$ . We assume that activity transformations are symmetric and occur at similar rates for all activities, so that  $\mu_{i,j}^{(c)} = \mu^{(c)}$ , and we assume that spatial dispersal occur at similar rates for all activities and all countries, so that  $\delta_i^{(l,c)} = \delta^{(c)}$ . These mean field assumptions reduce the number of parameters to  $2N + 3$ . In summary, the parameters  $r_i^{(c)}$  and  $b_i^{(c)}$  are assumed activity- and country-dependent, and  $\alpha^{(c)}$ ,  $\mu^{(c)}$ , and  $\delta^{(c)}$  are assumed country-dependent. We further assume that the values of these parameters are determined by the activity's characteristics and the country's institutional system (comprising e.g. the taxation regime, the system of innovation, the legal system, the intellectual property right and the socio-cultural background of the society) and resources (comprising e.g. labor force, knowledge capital, agricultural resources, mineral resources and energy resources), but that they do not change through time.

To investigate how the data supports each of the process embedded in Eq. (4.1), we further decompose the model into five different sub-models, as follows

$$\mathcal{M}_{\text{null}} : \frac{d}{dt} n_i^{(c)}(t) = r_i^{(c)} n_i^{(c)}(t)(1 - b_i^{(c)} n_i^{(c)}(t)) \quad (4.2)$$

$$\mathcal{M}_{\alpha^+} : \frac{d}{dt} n_i^{(c)}(t) = r_i^{(c)} n_i^{(c)}(t) \left( 1 - b_i^{(c)} n_i^{(c)}(t) + \alpha^{(c)} \sum_{j \neq i} n_j^{(c)}(t) \right), \quad \alpha > 0 \quad (4.3)$$

$$\mathcal{M}_{\alpha^-} : \frac{d}{dt} n_i^{(c)}(t) = r_i^{(c)} n_i^{(c)}(t) \left( 1 - b_i^{(c)} n_i^{(c)}(t) + \alpha^{(c)} \sum_{j \neq i} n_j^{(c)}(t) \right), \quad \alpha < 0 \quad (4.4)$$

$$\mathcal{M}_\delta : \frac{d}{dt} n_i^{(c)}(t) = r_i^{(c)} n_i^{(c)}(t)(1 - b_i^{(c)} n_i^{(c)}(t)) + \delta^{(c)} \left( \overline{n_i^{(c)}}(t) - n_i^{(c)}(t) \right) \quad (4.5)$$

$$\mathcal{M}_\mu : \frac{d}{dt} n_i^{(c)}(t) = r_i^{(c)} n_i^{(c)}(t)(1 - b_i^{(c)} n_i^{(c)}(t)) + \mu^{(c)} \sum_j^{N^{(c)}} \left( n_j^{(c)}(t) - n_i^{(c)}(t) \right) \quad (4.6)$$

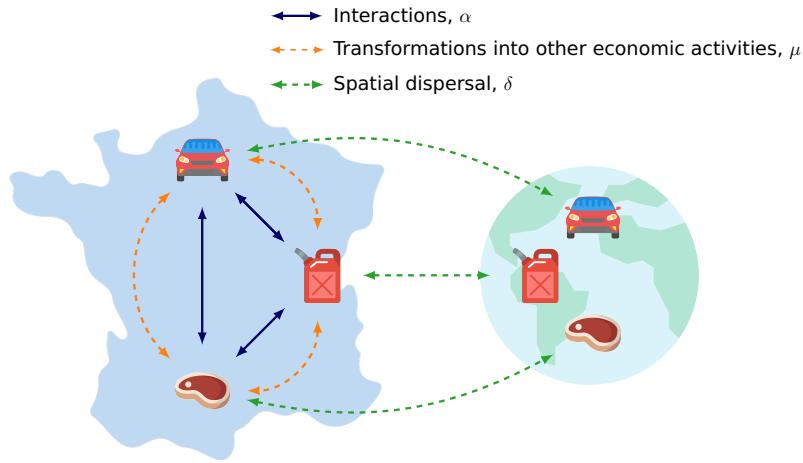
where

$$\overline{n_i^{(c)}} = \frac{1}{M-1} \sum_{l \neq c}^M n_i^{(l)} \quad (4.7)$$

accounts for the capital of activity  $i$  at the global level, excluding country  $c$ . The term  $\overline{n_i^{(c)}}$  appearing in Eqs. (4.6) and (4.7) arises from the mean field assumption used for spatial dispersal, where

$$\begin{aligned}
\sum_l^M \delta_i^{(l,c)} \left( n_i^{(l)}(t) - \overline{n_i^{(c)}}(t) \right) &= \delta^{(c)} \sum_l^M \left( n_i^{(l)}(t) - n_i^{(c)}(t) \right) \\
&= \delta^{(c)} \sum_{l \neq c}^M \left( n_i^{(l)}(t) - n_i^{(c)}(t) \right) \\
&= \delta^{(c)} \left( \sum_{l \neq c}^M n_i^{(l)}(t) - (M-1)n_i^{(c)} \right) \\
&\propto \delta^{(c)} \left( \overline{n_i^{(c)}} - n_i^{(c)} \right).
\end{aligned} \tag{4.8}$$

The simplest model  $\mathcal{M}_{null}$  embeds self-replication and self-limitation without any further coupling force acting upon economic activities, and is considered as the null model. The alternative models  $\mathcal{M}_{\alpha+}$ ,  $\mathcal{M}_{\alpha-}$ ,  $\mathcal{M}_\delta$ , and  $\mathcal{M}_\mu$  capture self-replication and self-limitation together with one of the additional eco-evolutionary process investigated. Because we find inconsistent maximum likelihood estimations in the controlled experiment detailed in the Results for models incorporating combinations of eco-evolutionary processes, we do not test more complex models. For  $\mathcal{M}_\delta$ , in Eq. (4.5), we assume that spatial dispersal do not affect the dynamics of  $n_i^{(c)}$ . As such, the dynamics of  $n_i^{(c)}$  does not depend on  $n_i^{(c)}$ , and we use linear interpolations of the data to model  $\overline{n_i^{(c)}}$ . Dividing by two the number of state variables to simulate in each country, this assumption considerably improves the computational complexity of the maximum likelihood estimation procedure. In the following, we designate by  $\mathcal{M}(t, \theta^{(c)})$  the vector of economic activity capitals predicted by model  $\mathcal{M}$  at time  $t$  for country  $c$  with the parameter vector  $\theta^{(c)}$ , that includes the growth rates  $r_i^{(c)}$ , the self limitation rates  $b_i^{(c)}$ , the specific model parameters  $\alpha^{(c)}$ ,  $\mu^{(c)}$ ,  $\delta^{(c)}$ , and the initial condition vector  $n^{(c)}(t_0)$  for country  $c$ .



**Fig. 4.1:** General population dynamic model used to characterize the dynamics of economic activities, capturing interactions between themselves, their spatial dispersal and their transformations into other economic activities. For each country  $c$ , each one of the 9 economic activities considered is assumed to be driven by replication and self limitation processes, characterized respectively by a growth rate  $r_i^{(c)}$  and a self-limitation coefficient  $b_i^{(c)}$ , which values are assumed to be country-dependent and determined by the institutional context and the resources available (in terms of human capital, knowledge capital and natural resources). In addition, interactions between activities within each country are determined by  $\alpha^{(c)}$  and can be mutualistic ( $\alpha^{(c)} > 0$ ) or competitive ( $\alpha^{(c)} < 0$ ). Due to spatial dispersal, the capital of an economic activity can experience positive (resp. negative) fluxes of capital when it is lower (resp. higher) than the average global capital for the economic activity at time  $t$ , arising at rate  $\delta^{(c)}$ . Due to transformations into other economic activities, the capital of an economic activity can experience positive (resp. negative) fluxes of capital when it is lower (resp. higher) than the capital of other economic sectors within the country at time  $t$ , arising at rate  $\mu^{(c)}$ . The parameters  $r_i^{(c)}$ ,  $b_i^{(c)}$ ,  $\alpha^{(c)}$ ,  $\delta^{(c)}$  and  $\mu^{(c)}$  that best fit the empirical data are estimated in each country for each sub-model in Eqs. (4.2) to (4.6), to investigate whether the proposed processes have a significant effect on the temporal dynamics of economic activities. This diagram has been designed using resources from Flaticon.com.

## 4.2.2 Empirical data and model likelihood

We consider time series of global trade data as a proxy for the evolution of the capital of economic activities through time. Export data is synthetic of the capital of an economic activity within a country because the more it is exporting, the more it is competing in the global market, and therefore the larger it is in terms of capital (Tacchella et al., 2018). We specifically use the database compiled by the United Nations Statistical Division COMTRADE, categorized in the Standard International Trade Classification (SITC, revision 2) at the 1 digit level, covering 11 categories of economic activities in 249 countries from 1962 to 2020. Because of their low quality (Hidalgo, 2021), we discard the "Unspecified" and "Services" category, resulting in 9 economic categories to model (see Fig. 4.3 for details). As shown in the Results, this large number of data points contains the information necessary to recover the plausible generating processes. Export values for activity  $i$  in country  $c$ , denoted by  $X_i^{(c)}(t)$ , are discounted by the national population at time  $t$ , denoted by  $P^{(c)}(t)$ , in order to compare the economic variables across time and across countries. This leads to expressing the observation data for the capital of activity  $i$  in country  $c$  as  $y_i^{(c)}(t) = X_i^{(c)}(t)/P^{(c)}(t)$ . In each country, we only consider economic activities which have sustained a significant size relative to the world trends for at least 4 years (revealed comparative advantage  $> 1$ , see Hidalgo, 2021). We assume that the observation data is contaminated by a white noise with log-normal distribution  $\epsilon$  with zero mean and a variance-covariance matrix  $\Sigma = \sigma^2 I$  where  $\sigma$  is the noise level, which is a reasonable error model for population dynamics (Schartau et al., 2017), so that  $y_i^{(c)}(t) = n_i^{(c)}(t) \exp(\epsilon_i^{(c)}(t))$ . This leads to expressing the likelihood of model  $\mathcal{M}$  in country  $c$ , denoted by  $\mathcal{L}(\theta_{\mathcal{M}}^{(c)} | \mathbf{y}^{(c)}, \mathcal{M})$ , as

$$\mathcal{L}(\theta_{\mathcal{M}}^{(c)} | \mathbf{y}^{(c)}, \mathcal{M}) = \prod_{j=1}^{T^{(c)}} \frac{1}{\sqrt{(2\pi)^{N^{(c)}} |\Sigma| y_j^{(c)}(t_j)}} \exp\left(-\frac{1}{2} [\tilde{d}^{(c)}(t_j)]^{T^{(c)}} \Sigma^{-1} [\tilde{d}^{(c)}(t_j)]\right) \quad (4.9)$$

where  $\theta_{\mathcal{M}}^{(c)}$  refers to the parameters and the initial conditions for model  $\mathcal{M}$ ,  $\mathbf{y}^{(c)} = (y^{(c)}(t_1), \dots, y^{(c)}(t_T))$  designate the  $T^{(c)}$  time points of economic data available for country  $c$ ,  $y^{(c)}(t) = (y_1(t), \dots, y_N^{(c)}(t))$  is the vector of economic activity capital at time  $t$ , and  $\tilde{d}^{(c)}(t_j) = \ln(y^{(c)}(t_j)) - \ln(\mathcal{M}(t_j, \theta_{\mathcal{M}}^{(c)}))$ . We detail in the next section how do we obtain the maximum likelihood estimate  $\hat{\theta}_{\mathcal{M}}^{(c)}$  that maximizes Eq. (4.9) given  $\mathbf{y}^{(c)}$  for each model  $\mathcal{M}$  and country  $c$ , further used for selecting the most probable model given the data (Burnham and Anderson, 2002). In the following, we drop the indices  $(c)$  for clarification.

## 4.2.3 Maximum likelihood estimation

In order to obtain the maximum likelihood of each model given the data, we employ the machine learning framework detailed in Boussange et al., 2022, which is based on a variational optimization method and a learning strategy coined "mini-batching". The likelihood of model  $\mathcal{M}_i$  is maximized by training  $\mathcal{M}_i$  against mini-batches of data comprising

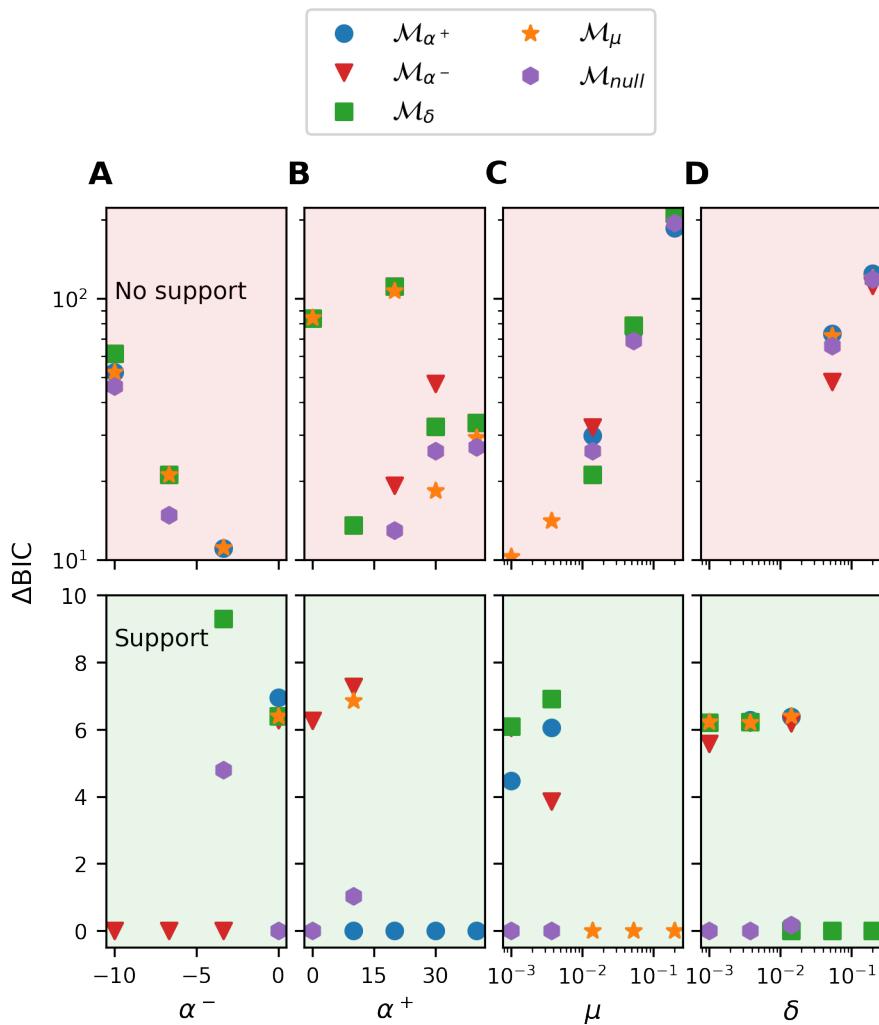
only  $K < T^{(c)}$  data points of the full time series, where the parameters and the initial conditions for each mini-batch are estimated. The mini-batching strategy ensures the convergence towards the maximum likelihood estimate, provided that the choice of  $K$  is appropriate given the data and the model investigated (Boussange et al., 2022). A large  $K$  might induce convergence towards a local minimum, while a low  $K$  might flatten the likelihood landscape, where all models would be provided equal support. In the controlled experiments in the Results, we find that  $K = 20$  (models are trained against batches of 20 years) ensures a good convergence, while reasonably discriminating the models (Fig. 4.2). We therefore use  $K = 20$  throughout all the experiments detailed in the manuscript, and discard countries where data is available for less than  $K = 20$  years. The library **MiniBatchInference.jl** (Boussange et al., 2022) is employed for the numerical implementation, and we use the gradient descent optimizer ADAM (Kingma and Ba, 2014) during the first 800 epochs to converge in the basin of attraction of the maximum likelihood estimate, which we substitute for the Broyden-Fletcher-Goldfarb-Shanno optimizer (BFGS) (Fletcher, 2013) for the final 800 training epochs to ensure faster and more accurate convergence. As a cross-checking procedure, for each country and each model we perform 5 optimization runs with different initial parameter values, where  $r_i, b_i, \alpha, \mu, \delta$  are drawn from the uniform random distribution  $\mathcal{U}_{[0.05, 0.15]}, \mathcal{U}_{[0.5, 1.5]}, \mathcal{U}_{[0.5, 1.5]}, \mathcal{U}_{[0.0005, 0.0015]}$  and  $\mathcal{U}_{[0.0005, 0.0015]}$  respectively. We then take the best run among the 5 optimization runs, making sure that the likelihood estimates of each run are similar (see Table S4.2 for full details of the results).

## 4.3 Results

### 4.3.1 Validation with synthetic data

We first investigate in a controlled experiment whether the inverse modelling technique of Boussange et al., 2022, together with the proposed sub-models in Eqs. (4.2) to (4.6), can detect signatures of eco-evolutionary processes in a controlled experiment. We proceed by generating multiple synthetic datasets from the models  $\mathcal{M}_{\alpha^-}, \mathcal{M}_{\alpha^+}, \mathcal{M}_\delta$  and  $\mathcal{M}_\mu$  with realistic  $r_i$  and  $b_i$  parameters and with different values for the parameters  $\alpha, \mu$ , and  $\delta$  and with different values for the noise level  $\sigma$ . We consider  $\mathcal{M}_{null}, \mathcal{M}_{\alpha^-}, \mathcal{M}_{\alpha^+}, \mathcal{M}_\delta$  and  $\mathcal{M}_\mu$  as equally plausible candidate models for each generated datasets, and apply the maximum likelihood estimation method detailed in the Methods to obtain the maximum likelihood  $\mathcal{L}(\hat{\theta}, \mathcal{M}|\mathbf{y})$  of each model for each generated dataset. We then use the Bayesian information criterion (BIC) to select the model with the strongest strength-of-evidence in relation to the data (Mangan et al., 2017), calculating the BIC for model  $\mathcal{M}_i$  as  $BIC_{\mathcal{M}_i} = -2 \ln(\mathcal{L}(\hat{\theta}, \mathcal{M}_i|\mathbf{y})) + k_{\mathcal{M}} \ln(NT)$  where  $N$  is the number of activities in country  $c$ ,  $T$  is the number of time points in the time series considered ( $NT$  therefore being the number of data points for country  $c$ ) and  $k_{\mathcal{M}}$  is the number of parameters in the model  $\mathcal{M}$  for country  $c$ . The BIC ranks the most probable models by penalizing complexity to balance information loss and parsimony, where candidate models with the lowest scores are ranked as the most likely (Mangan et al., 2017). We consider the relative BIC score  $\Delta BIC_{\mathcal{M}_i}$ , which allows a strength-

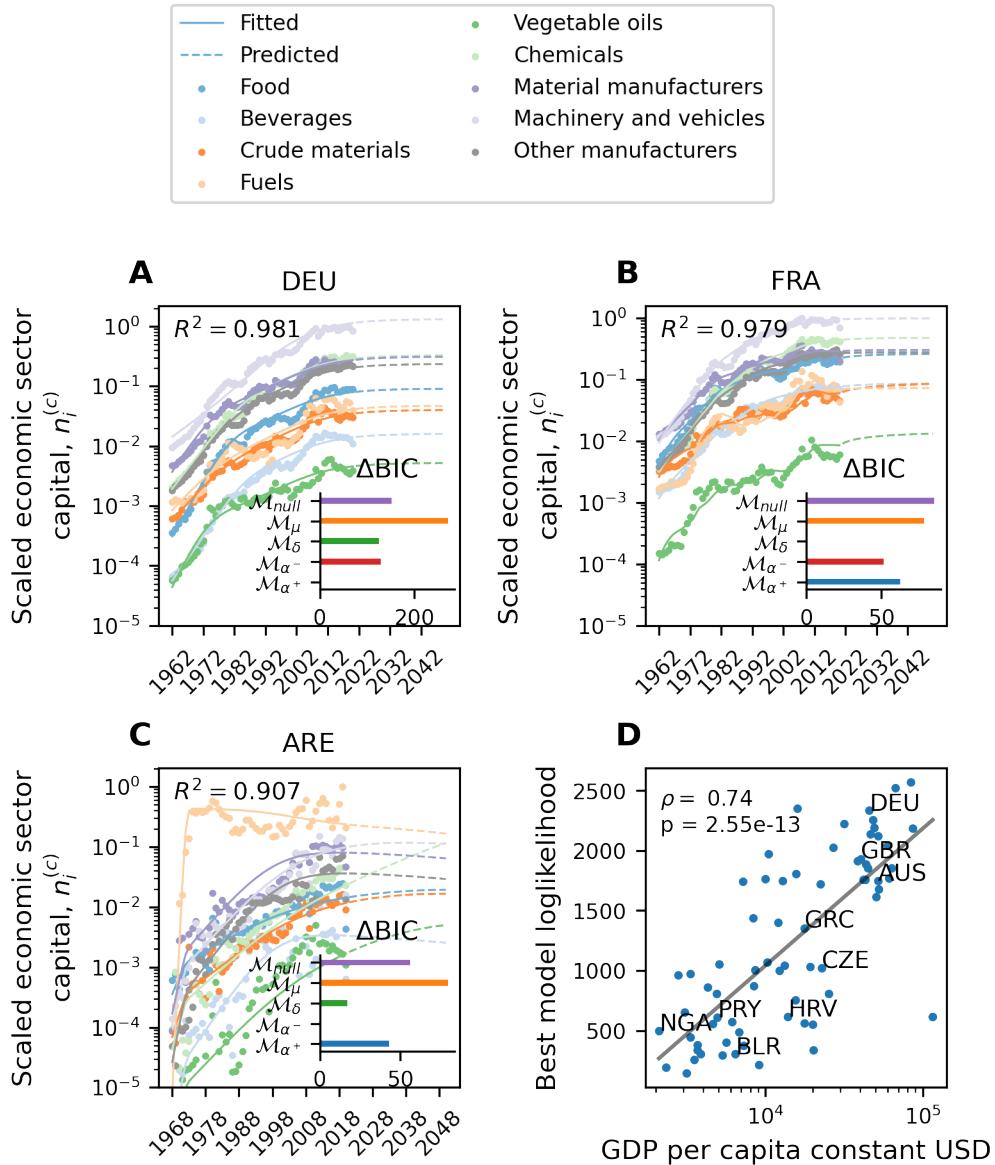
of-evidence comparison across models and is calculated as  $\Delta\text{BIC}_{\mathcal{M}_i} = \text{BIC}_{\mathcal{M}_i} - \min_j \text{BIC}_{\mathcal{M}_j}$ . We expect the  $\Delta\text{BIC}$  scores to only provide support for the true generating model when the process considered has a significant effect on the observed dynamics. Under realistic observational noise ( $\sigma = 0.2$ ), we find overall strong support for the true models (i.e., when  $\mathcal{M}_i$  is the true generating model,  $\Delta\text{BIC}_{\mathcal{M}_i} = 0$  and  $\Delta\text{BIC}_{\mathcal{M}_{j,j \neq i}} > 10$ , see Fig. 4.2A-C) when the values of the parameters  $\alpha$ ,  $\mu$  and  $\delta$  exceed a certain threshold. Moreover, the more extreme the value of  $\alpha$ ,  $\mu$ , and  $\delta$ , the more support is given to the true generating model (Fig. 4.2A-C). In contrast, when the signature of the underlying process on the dynamics is not sufficiently strong, the null model  $\mathcal{M}_{null}$  is given the most strength-of-evidence ( $\Delta\text{BIC}_{\mathcal{M}_{null}} = 0$ ). These results hold for varying noise levels (see Fig. S4.6 for  $\sigma = 0.3$ ), indicating that the models together with the maximum likelihood estimation method are well adapted to investigate the influence of eco-evolutionary processes with the dataset considered. Based upon this experiment and classical model selection criterion (Burnham and Anderson, 2002), we decide to accept the hypothesis that the null model is the best model given the data when  $\Delta\text{BIC}_{\mathcal{M}_{null}} \leq 10$ . In addition, when  $\text{BIC}_{\mathcal{M}_i} - \text{BIC}_{\mathcal{M}_{null}} < -10$  we conclude that  $\mathcal{M}_i$  is supported against  $\mathcal{M}_{null}$ , and we conclude that  $\mathcal{M}_i$  is the best model given the data if  $\Delta\text{BIC}_{\mathcal{M}_i} = 0$  and if for all other model  $\mathcal{M}_{j,j \neq i}$ , we have that  $\Delta\text{BIC}_{\mathcal{M}_{j,j \neq i}} > 10$ . Overall, the proposed eco-evolutionary model together with the maximum likelihood estimation leads to a good discrimination ability when combined with the BIC-based model selection procedure, and can provide strength-of-evidence for eco-evolutionary processes that may shape the dynamics of economic systems.



**Fig. 4.2: Validation of the model selection procedure.** In A and B we generate data with models  $\mathcal{M}_{\alpha^-}$  and  $\mathcal{M}_{\alpha^+}$  with varying  $\alpha$ , in C we generate data with model  $\mathcal{M}_\delta$  with varying  $\delta$  and in D we generate data with model  $\mathcal{M}_\mu$  with varying  $\mu$ . We then contaminate the generated data with realistic noise ( $\sigma = 0.2$ ), and seek for the maximum likelihood of all models  $\mathcal{M}_{\text{null}}$ ,  $\mathcal{M}_{\alpha^-}$ ,  $\mathcal{M}_{\alpha^+}$ ,  $\mathcal{M}_\delta$  and  $\mathcal{M}_\mu$ . The maximum likelihood of each model is then used to compute difference in BIC for each model,  $\Delta\text{BIC}$ , which comparison provides a strength-of-evidence for each model. A-D show that the model selection procedure is valid, since only the true models are placed in the "Strong support" category when no other model are placed in the "No support" category.

### 4.3.2 Discrepancies in dynamical regime across countries

We apply the maximum likelihood estimation method detailed in the Methods to obtain the maximum likelihood of each model for the world's richest 100 countries (highest GDP per capita as of 2020, see list in Table S4.1). We first investigate the quality of best fit, and further detail the results for each model in the following section. In the following, we refer to countries by their ISO 3166-1 reference code, which associated plain names are given in Table S4.1. Details of the numerical simulations are provided in Table S4.2. Over the 100 countries investigated, only 78 countries have sufficiently long time series to be included in the analyses (countries with at least 20 years of data, see Methods), while 77 countries show consistent maximum likelihood estimates (best model coefficient of determination  $R^2 > 0$ ). Among the 77 countries, we find that the best models have a good fit to the data, and capture well the long term growth of economic activities (median explained variance  $R^2 = 0.940$ , see Fig. 4.3 and Figs. S4.7 to S4.10 for graphical illustrations of the fit). While this demonstrates the relevance of the proposed models, we observe discrepancies across countries in how well the best models perform (standard deviation of explained variance  $\text{std}(R^2) = 0.074$ , see Fig. 4.3D for a graphical illustration of the variance in terms of the model loglikelihoods). This discrepancy is explained by the number of data points available in each country, as it differs among countries and since more data points yield a higher model loglikelihood ( $\beta = 0.928 \pm 0.046$ ,  $p < 0.001$ , see Table S4.3). We additionally observe a positive effect of the GDP on the residuals ( $\beta = 0.693 \pm 0.187$ ,  $p < 0.001$ , see Table S4.3 and Fig. 4.3D), indicating that rich economies have a more deterministic behavior than poorer ones. The residuals are further negatively associated with the number of economic activities ( $\beta = -0.230 \pm 0.071$ ,  $p < 0.01$ , see Table S4.3), indicating that the models proposed might be less appropriate for qualifying the dynamics of diversified economies.



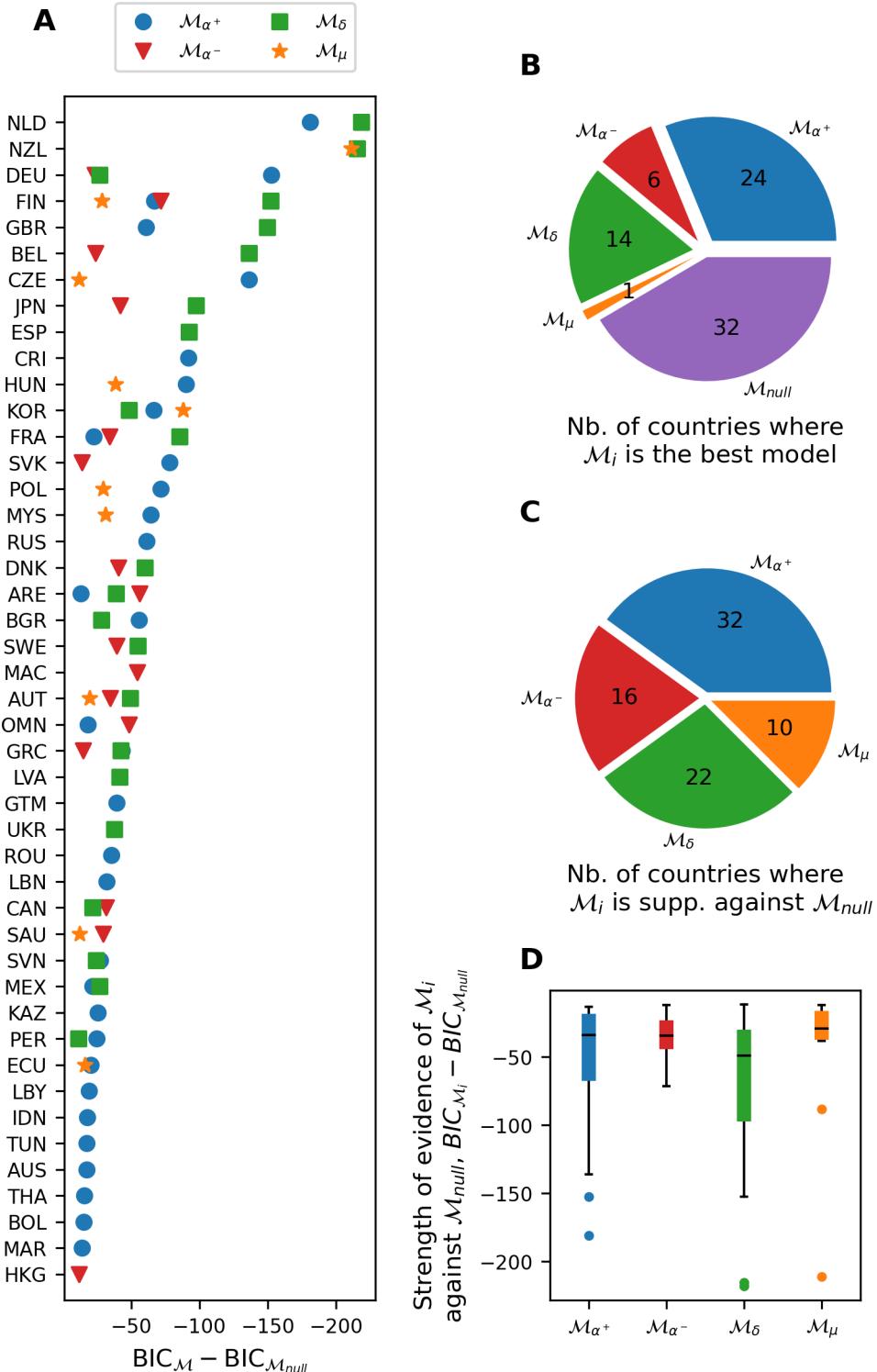
**Fig. 4.3: Best model fits and associated predictions for DEU, FRA and ARE, and effect of GDP on best model loglikelihoods.** A-C Best model fits and associated predictions for DEU, FRA and ARE. Scatter dots represent the empirical data, plain lines correspond to the best model fits, and dotted lines correspond to the best model predictions. The inset plots provide the  $\Delta BIC$  for each country. D Effect of GDP on model loglikelihood. The plain line corresponds to a linear fit.

### 4.3.3 Evidence for eco-evolutionary processes

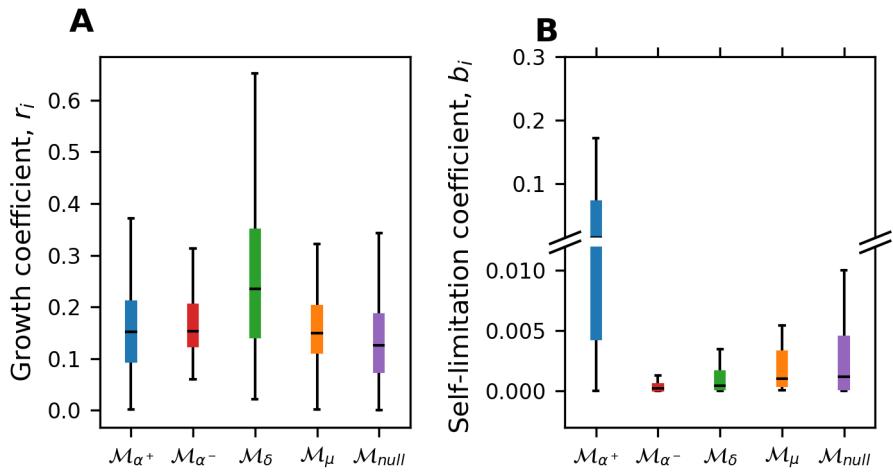
Applying the model selection procedure detailed in the controlled experiment, we investigate in details the effect of eco-evolutionary processes on the dynamics of the countries investigated. We find that the null model is rejected in 45 of the 77 countries, in favor of alternative models (Fig. 4.4B). Among the alternative models,  $\mathcal{M}_{\alpha+}$ , capturing positive interactions between economic activities, is the most frequently supported (Fig. 4.4B,C), and is given considerably more support against  $\mathcal{M}_{null}$  than  $\mathcal{M}_{\alpha-}$  and  $\mathcal{M}_\mu$  (Fig. 4.4A,D). The model with spatial dispersal,  $\mathcal{M}_\delta$ , is the second most supported (Fig. 4.4B,C) and is given the most support against  $\mathcal{M}_{null}$  (Fig. 4.4A,D). The models capturing negative interactions,  $\mathcal{M}_{\alpha-}$ , and capturing economic activity transformations,  $\mathcal{M}_\mu$ , are the least supported and given the least strength-of-evidence (Fig. 4.4B,D), but are still ranked as most supported models in some countries (Fig. 4.4A).

In contrast to  $\mathcal{M}_{\alpha-}$ ,  $\mathcal{M}_\delta$ ,  $\mathcal{M}_\mu$  and  $\mathcal{M}_{\alpha+}$ ,  $\mathcal{M}_\delta$ ,  $\mathcal{M}_\mu$ , which implement complementary processes,  $\mathcal{M}_{\alpha-}$  and  $\mathcal{M}_{\alpha+}$  are structurally antagonistic, but are nonetheless equally supported in some countries (see, e.g., FIN and FRA in Fig. 4.4A). This ambiguity highlights that the success of each model arises from their ability to capture different features of the empirical data. For instance,  $\mathcal{M}_{\alpha-}$  is the only model capturing capital decrease (see, e.g., ARE in Figs. 4.3 and S4.7) or very low capital growth (see, e.g., HRV in Fig. S4.7).  $\mathcal{M}_{\alpha+}$  captures well periods of exponential economic growth followed by a growth slow down, and correlation patterns in growth between activities (see, e.g., DEU and HUN in Figs. 4.3 and S4.8). We find that this dynamics is associated with an increase in the self limitation coefficients  $b_i$ , that compensates the beneficial effect of positive interactions (Fig. 4.5B). In contrast, the model  $\mathcal{M}_{null}$  tends to overestimate the capital size for such dynamics.  $\mathcal{M}_\delta$  is the only model reproducing the oscillations observed in the empirical data (see e.g. FRA in Fig. 4.3 and BEL, GBR in Fig. S4.9), where oscillations arise from the term  $\overline{n^{(c)}}$  through mismatches between the local capital size and the global capital size of a given economic activity.  $\mathcal{M}_\delta$  tends to overestimate the growth rate coefficients  $r_i$  against other models (Fig. 4.5A), due to a source-sink dynamics. Indeed, the dispersal term acts as capital sink for most economic activities in most countries, where only a handful of under-developed activities, relative to global trends in the world economy, benefit from source dynamics (see, e.g., KOR in Fig. S4.11C). Characteristic trends of  $\mathcal{M}_\mu$  are difficult to observe given the low number of cases where it is given statistical support.

Overall, positive interactions and spatial dispersal stand out as the most supported processes, both in term of the number of countries where they are given support, and in term of the strength-of-evidence against  $\mathcal{M}_{null}$ .



**Fig. 4.4:** Statistical support or the population dynamic models  $\mathcal{M}_{\alpha^-}$ ,  $\mathcal{M}_{\alpha^+}$ ,  $\mathcal{M}_\delta$  and  $\mathcal{M}_\mu$ , based on Bayesian information criterion f. A Strength-of-evidence comparison for  $\mathcal{M}_{\alpha^-}$ ,  $\mathcal{M}_{\alpha^+}$ ,  $\mathcal{M}_\delta$  and  $\mathcal{M}_\mu$  against  $\mathcal{M}_{\text{null}}$ , for countries where  $\mathcal{M}_{\text{null}}$  is rejected, sorted in decreasing order. B Number of countries where models  $\mathcal{M}_{\text{null}}$ ,  $\mathcal{M}_{\alpha^-}$ ,  $\mathcal{M}_{\alpha^+}$ ,  $\mathcal{M}_\delta$ , and  $\mathcal{M}_\mu$  are ranked as best models. C Number of countries where models  $\mathcal{M}_{\alpha^-}$ ,  $\mathcal{M}_{\alpha^+}$ ,  $\mathcal{M}_\delta$  and  $\mathcal{M}_\mu$  are accepted against  $\mathcal{M}_{\text{null}}$ . D Global strength-of-evidence comparison for  $\mathcal{M}_{\alpha^-}$ ,  $\mathcal{M}_{\alpha^+}$ ,  $\mathcal{M}_\delta$ , and  $\mathcal{M}_\mu$  against  $\mathcal{M}_{\text{null}}$ . A-D show that  $\mathcal{M}_{\alpha^+}$  and  $\mathcal{M}_\delta$  stand out as the most supported models, both in term of number of countries where they are given support, and in term of strength-of-evidence.



**Fig. 4.5:** Differences in growth rates  $r_i$  and coefficients of self-limitation  $b_i$  between  $M_{\alpha^-}$ ,  $M_{\alpha^+}$ ,  $M_\delta$ ,  $M_\mu$ , and  $M_{null}$ . In A-B, we aggregate the parameters  $r_i$  and  $b_i$  across all economic activities and all countries where models are supported. A shows that  $M_\delta$  overestimates  $r_i$  against the other models. This is because for a dispersal rate  $\delta > 0$ , most economic activities experience a negative capital flux, and consequently need to grow more to compensate for the losses. B shows that  $M_{\alpha^+}$  tends to overestimate  $b_i$  relative to the other models, which negative effect on growth dynamics compensates the beneficial effect of positive interactions.

## 4.4 Discussion

Our results indicate that dynamic models embedding eco-evolutionary processes, commonly used to characterize the dynamics of biological populations (Bunin, 2017; Scheffer and Nes, 2006; Case, 1990; Tilman, 1994; Eigen et al., 1988), can appropriately characterize the dynamics of economic activities (Fig. 4.3). Together with state of the art computational methods and 59 years of economic data, the models provide a framework to assess the effect of eco-evolutionary processes on the long-term development of national economies. Over the 77 countries investigated, the null model – embedding replication and self limitation processes only – was rejected 45 times in favor of alternative eco-evolutionary models (Fig. 4.4B), and each of the alternative model was supported against the null model in some countries. Hence, eco-evolutionary models, that couple the growth of economic activities, better characterize their dynamics. This result suggests that economic change carries the effect of the inter-dependency between economic activities. The models capturing positive interactions between economic activities ( $\mathcal{M}_{\alpha+}$ ) and spatial dispersal ( $\mathcal{M}_\delta$ ) were overall given more strength-of-evidence, providing evidences that these processes are central in the development of economic systems. Together, our study suggests that biologically inspired dynamic models are well suited to characterize economic dynamics.

The statistical support obtained for the model with positive interactions between economic activities and the associated parameter value estimates highlight that economic activities sustain the development of each other, supporting the hypothesis that economic diversity promotes economic growth (Saviotti et al., 2020).  $\mathcal{M}_{\alpha+}$  was the most frequently supported model (Fig. 4.4B-C), outlining that positive interactions may be ubiquitous in economic systems. We also found that  $\mathcal{M}_{\alpha+}$  was associated with high self-limitation coefficients (Fig. 4.5B). This finding highlights that the growth of economic activities is driven by positive interactions rather than by their endogenous growth. Positive interactions may arise from a variety of sub-processes acting upon firms, including beneficial business interactions through supply chains (Ozman, 2009; Saavedra et al., 2009) and agglomeration externalities (Van Der Panne, 2004). Together, our results suggest that the effect of positive interactions acting upon firms scales up and at the national level and result in economic activities sustaining each other through positive feedbacks.

The dynamics of economic activities seem also to be greatly affected by spatial dispersal (Fig. 4.4B,D), which may originate from socio-economic processes operating at lower organizational levels and involved in the spatial diffusion of organizational routines (Hodgson and Knudsen, 2004) and knowledge (Rogers, 2003; Keller, 2004; Bahar et al., 2014) across countries. Candidates socio-economic processes underlying the diffusion of organizational routines are international business expansions (Andersen, 1993; Zahra et al., 2000; Luo and Tung, 2007), labor mobility (Boschma et al., 2008), and the diffusion of innovation through social networks (Rogers, 2003; Keller, 2004). However, we found that  $\mathcal{M}_\delta$  was less frequently supported against  $\mathcal{M}_{null}$  than  $\mathcal{M}_{\alpha+}$  (Fig. 4.4C). To explain this discrepancy, we hypothesize that transfers of knowledge and routines are blocked by barriers due to differences in cognitive, organizational, social, institutional or geographic proximity between

countries (Boschma, 2005; Caragliu and Nijkamp, 2016; Bahar et al., 2014). This hypothesis should be investigated in future work. Overall, our results indicate that socio-economic processes involved in the transfer of knowledge and organizational routines have an important effect on economic change at the national level.

The strong discrepancies in model supports observed across countries may indicate strong idiosyncratic processes, that can relate to the countries' institutional context (Acemoglu et al., 2005) or to historical contingencies (Silverberg and Verspagen, 2005). We found that developing economies have a distinct, less predictable dynamical behavior, as the likelihood of the best models showed a negative relationship with country GDP (Fig. 4.3D and Table S4.3). In line with our findings, (Cristelli et al., 2015) also reports a difference in dynamical regime between mature and developing economies, where the GDP dynamics of developing economies is found less predictable than that of rich countries. Those lower model fits indicate that singular processes may importantly influence the dynamics of developing countries. Also, while we found a statistical advantage for  $\mathcal{M}_{\alpha^+}$  and  $\mathcal{M}_\delta$  over the alternative models (Fig. 4.4B-D), the support and strength-of-evidence for each model strongly varied across countries (Fig. 4.4A), without associations with macroeconomic characteristics such as, e.g., country GDP per capita. This may be due to the noise in the data, or to the specific assumptions considered. For instance, while we assumed equal strength of positive interactions between pairs of economic sectors (Eq. (4.4)), it could be that interaction strength is pair specific, implying that the likelihood estimated for the positive interaction hypothesis is underrated in countries with many economic activities. Yet, these discrepancies could also originate from the idiosyncrasy of the country histories and characteristics. Together, our results point towards country and economic activity-specific idiosyncrasies in the nature and the role of eco-evolutionary processes in economic systems.

For the sake of parsimony, the population dynamic models proposed in Eqs. (4.2) to (4.6) rely on a set of assumptions, which could be relaxed to further detail eco-evolutionary processes acting upon economic activities. The models investigated rely on the mean field assumptions that economic activities interact with equal strengths, and proceed to spatial dispersal across countries and capital transfer between each other with equal rates. Nonetheless, similarly to biological organisms (Bascompte et al., 2003), economic entities interact through organized networks (Hidalgo et al., 2007; Bustos et al., 2012; Saavedra et al., 2009; Schweitzer et al., 2009; Giuliani, 2007). For instance, Hidalgo et al., 2007 suggests that economic activities are related to one another through a network of relatedness, which could be integrated in  $\mathcal{M}_{\alpha^-}$ ,  $\mathcal{M}_{\alpha^+}$  and  $\mathcal{M}_\mu$ , weighting the interactions and transfers by the relatedness of two activities. Likewise, spatial dispersal are likely to depend on proximity metrics between countries (Boschma, 2005; Caragliu and Nijkamp, 2016; Bahar et al., 2014). The consideration of generalized variants of the models investigated, such as in Eq. (4.1), will be required to assess the importance of economic activity relatedness and country proximity on economic dynamics. Additionally, considering the variation of the parameter values through time may be required, as the strength and directionality of interactions, spatial dispersal and economic activity transformations may temporally change throughout the development of economic activities (Pistorius and Utterback, 1997). Exploring more complex eco-evolutionary models will demand, however, richer time series, in order to extract the extra information needed to constrain the additional parameters.

In conclusion, our results foster a biologically inspired approach to understand the mechanisms shaping the endogenous dynamics of economic systems. Using an inverse modelling approach relying on alternative population dynamic models and temporal economic data, our study demonstrates that positive interactions between economic activities and spatial dispersal considerably influence the dynamics of economic activities at the national, and may be fundamental drivers of economic change. The quantitative paradigm to study economic systems has mainly relied on models inspired from mechanics, which assume a world in equilibrium (Sornette, 2014). Evolutionary biology, focusing on nonlinear dynamical processes and emergence, seems to be a more appropriate paradigm to study the fundamental forces shaping the dynamics of economic systems. It can provide quantitative models to test general hypotheses on organizational principles, and suggest new ones. We call this promising research venue *econobiology*.

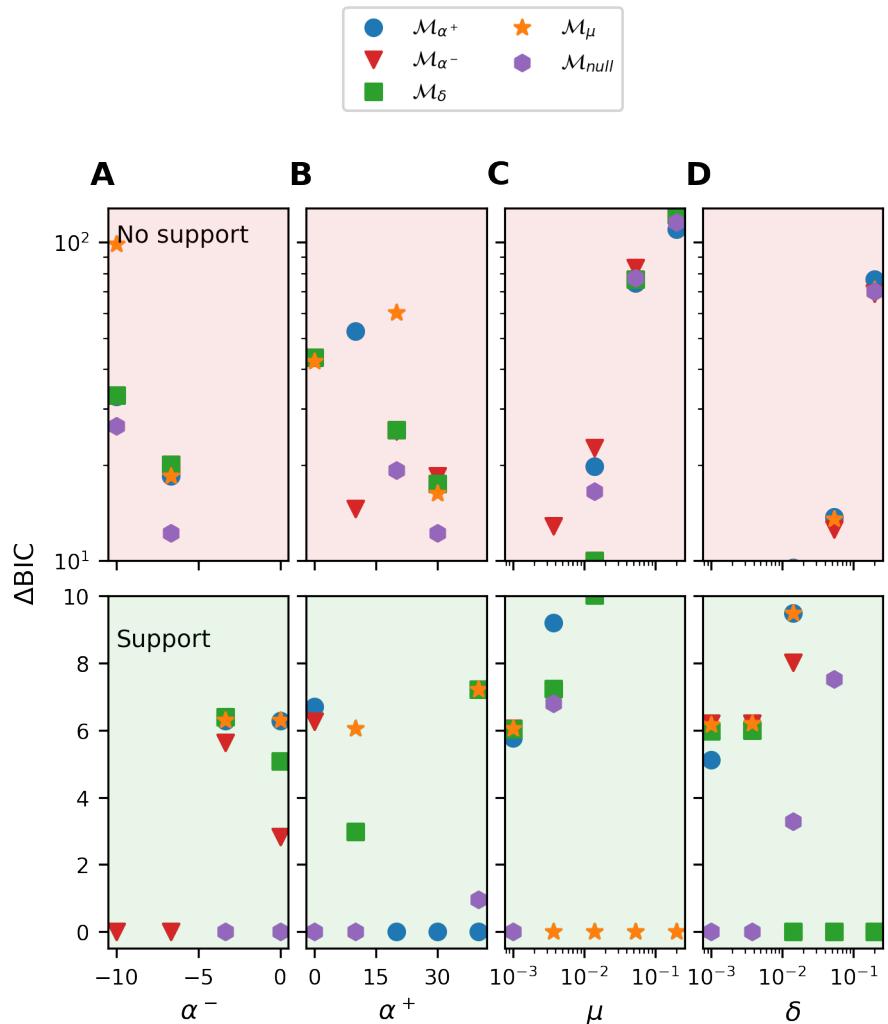
## 4.5 Code availability

The simulation code is available on request.

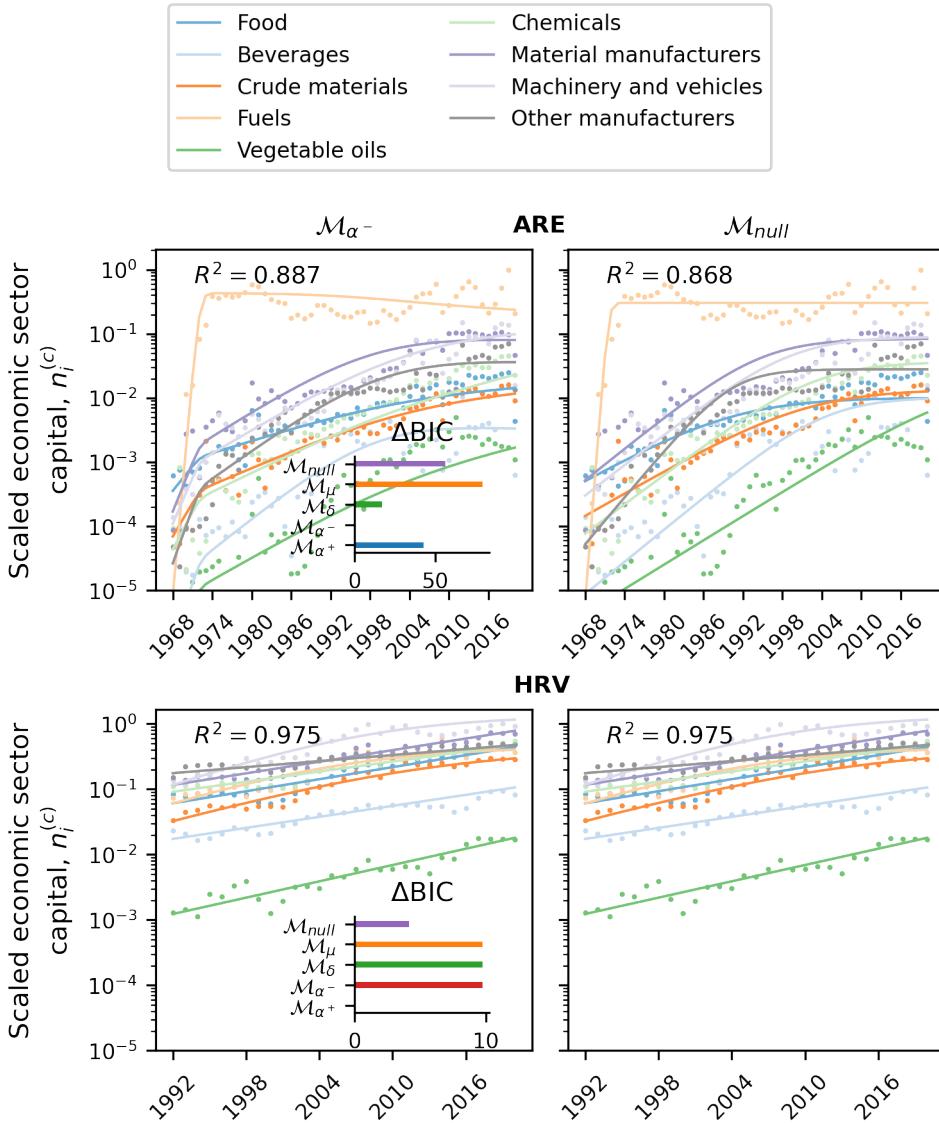
## 4.6 Acknowledgements

L.P. and V.B. were supported by the SNF grant 310030E\_205556.

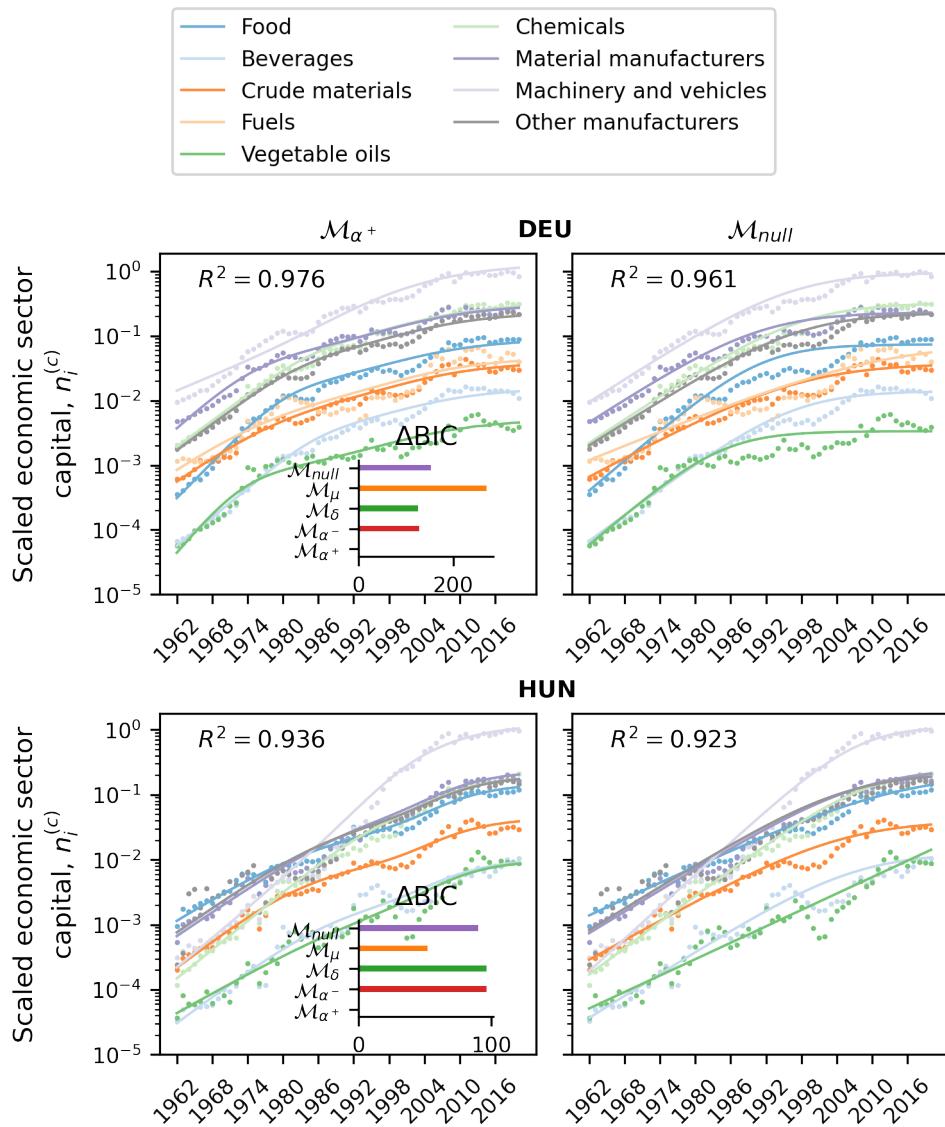
## 4.A Supplementary Figures



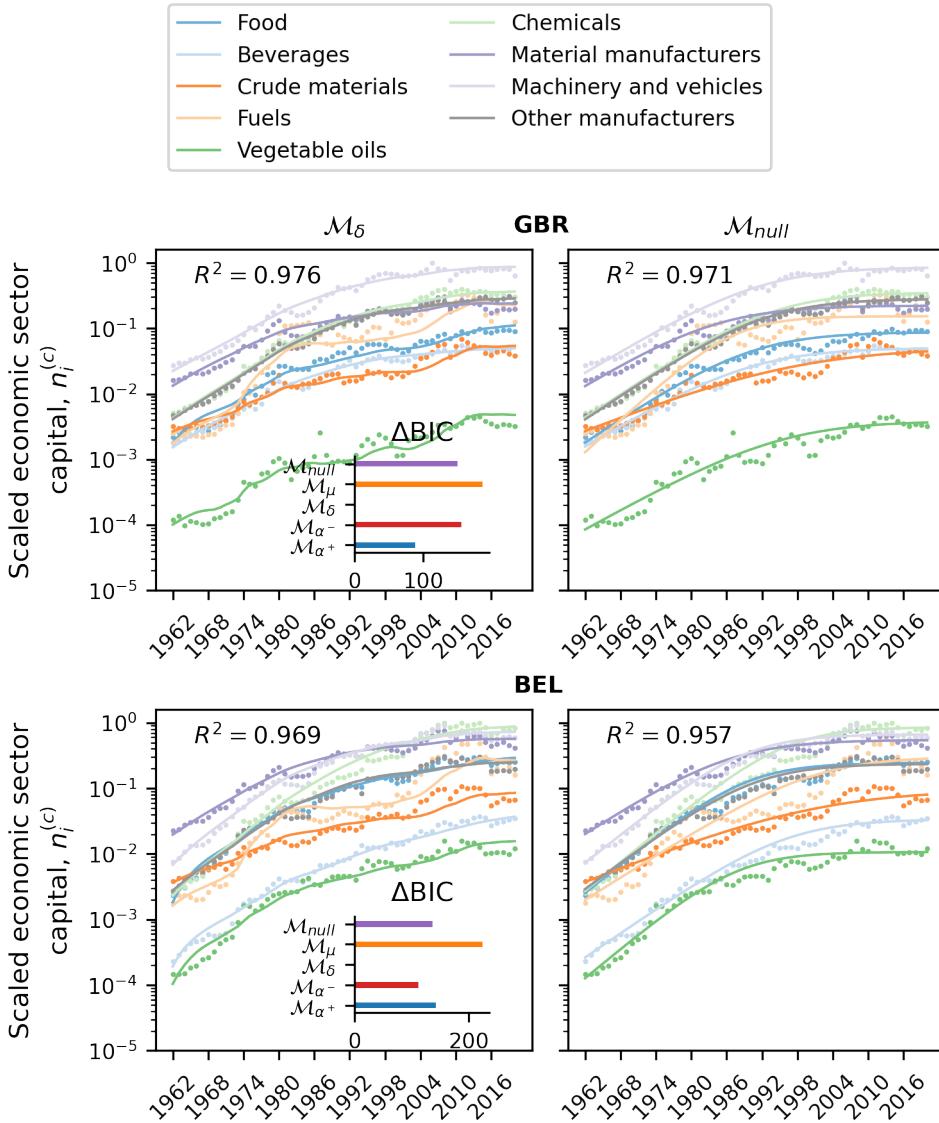
**Fig. S4.6:** Validation of the model selection procedure. The figure is analogous to Fig. 4.2 but with  $\sigma = 0.3$ .



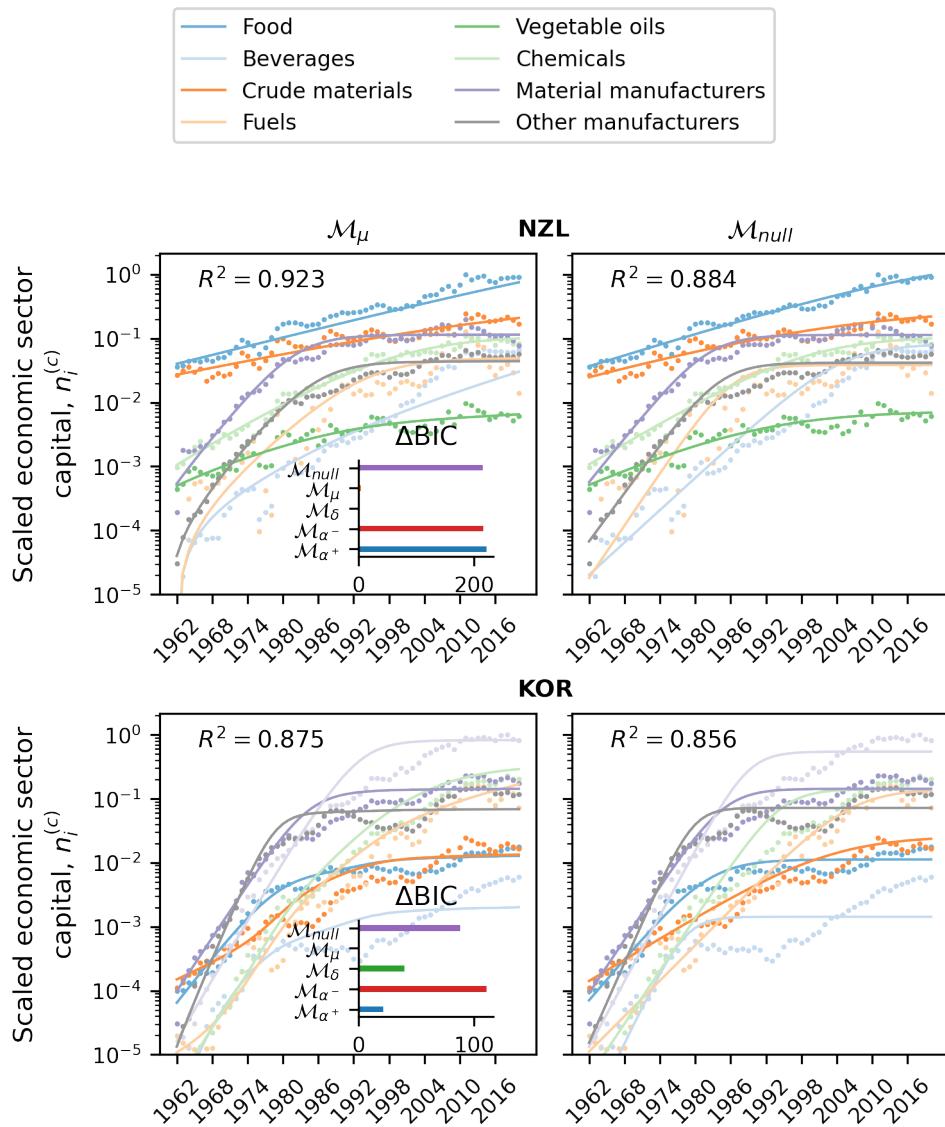
**Fig. S4.7:** Fits of model  $\mathcal{M}_{\alpha^-}$  for ARE and HRV, where  $\mathcal{M}_{\alpha^-}$  is given support against  $\mathcal{M}_{null}$ .



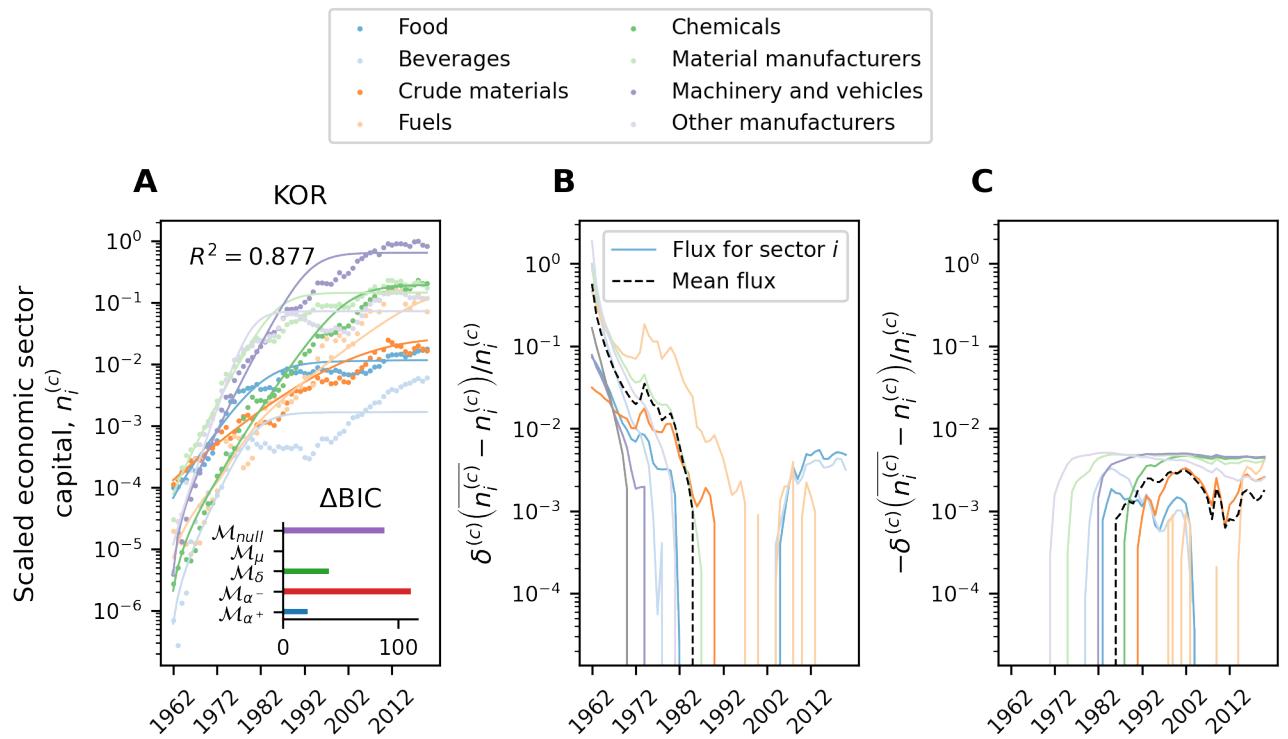
**Fig. S4.8:** Fits of model  $\mathcal{M}_{\alpha^+}$  for DEU and HUN, where  $\mathcal{M}_{\alpha^+}$  is given support against  $\mathcal{M}_{\text{null}}$ .



**Fig. S4.9:** Fits of model  $\mathcal{M}_\delta$  for GBR and BEL, where  $\mathcal{M}_\delta$  is given support against  $\mathcal{M}_{null}$ .



**Fig. S4.10:** Fits of model  $\mathcal{M}_\mu$  for NZL and KOR, where  $\mathcal{M}_\mu$  is given support against  $\mathcal{M}_{null}$ .



**Fig. S4.11:** Contributions of fluxes due to spatial dispersal on growth dynamics. A shows the growth dynamics for KOR, together with the best model fit. B-C show the positive and negative fluxes, respectively, due to spatial dispersal for each economic activity, relative to the capital size of economic activity. The contributions of spatial dispersal to growth decreases over time.

## 4.B Supplementary Tables

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta BIC$
ARE	$\mathcal{M}_{\alpha^+}$	1.2527837	0.0443378	0.8985771	2194.9947804	-331.4965062
ARE	$\mathcal{M}_{\alpha^-}$	1.1323544	0.0001987	0.9070505	2209.3385519	-374.6811568
ARE	$\mathcal{M}_\delta$	1.1547997	0.0018924	0.9037398	2204.4213038	-357.3571273
ARE	$\mathcal{M}_\mu$	1.3237895	0.0096762	0.8908284	2175.1338748	-295.0532417
ARE	$\mathcal{M}_{null}$	1.3210473	0.0185026	0.8945233	2184.4961044	-318.3014521
ARG	$\mathcal{M}_{\alpha^+}$	0.5362716	0.0000000	0.9126885	873.2783542	-479.0901860
ARG	$\mathcal{M}_{\alpha^-}$	0.5260177	0.0000000	0.9148312	877.0708569	-486.6684522
ARG	$\mathcal{M}_\delta$	0.5347459	0.0882225	0.9127092	873.5096285	-479.1622984
ARG	$\mathcal{M}_\mu$	0.5362716	0.0000000	0.9126884	873.2790114	-479.0898098
ARG	$\mathcal{M}_{null}$	0.5362716	0.0000000	0.9126885	873.2797678	-484.8104788
AUS	$\mathcal{M}_{\alpha^+}$	0.2493686	0.0259488	0.9701803	1746.6372258	-1152.0686387
AUS	$\mathcal{M}_{\alpha^-}$	0.2639190	0.0000105	0.9687099	1734.9134560	-1128.5804718
AUS	$\mathcal{M}_\delta$	0.3009456	0.0013734	0.9620972	1683.0297057	-1035.0200199
AUS	$\mathcal{M}_\mu$	0.2639179	0.0000000	0.9687119	1734.9252830	-1128.6123545
AUS	$\mathcal{M}_{null}$	0.2639178	0.0000001	0.9687114	1734.9245624	-1134.7946683
AUT	$\mathcal{M}_{\alpha^+}$	0.2068390	0.0001061	0.9818722	2225.2530696	-1414.8850719
AUT	$\mathcal{M}_{\alpha^-}$	0.1927354	0.0000609	0.9831719	2245.3364190	-1455.7289956
AUT	$\mathcal{M}_\delta$	0.1858071	0.0000049	0.9836237	2253.3349649	-1470.6709565
AUT	$\mathcal{M}_\mu$	0.1967636	0.0000009	0.9827050	2238.1320372	-1440.7057723
AUT	$\mathcal{M}_{null}$	0.2068390	0.0000001	0.9818725	2225.2566207	-1421.2047025
BEL	$\mathcal{M}_{\alpha^+}$	0.1723450	0.0000000	0.9732590	1784.0358726	-1495.7959922
BEL	$\mathcal{M}_{\alpha^-}$	0.1648341	0.0000001	0.9746902	1798.8747676	-1525.9943707
BEL	$\mathcal{M}_\delta$	0.1324942	0.2790846	0.9793798	1854.1829608	-1638.4943651
BEL	$\mathcal{M}_\mu$	0.1968834	0.0000000	0.9689876	1743.5335080	-1414.4390435
BEL	$\mathcal{M}_{null}$	0.1723450	0.0000000	0.9732588	1784.0363416	-1502.0995394
BGR	$\mathcal{M}_{\alpha^+}$	0.7781416	0.0479922	0.8288290	1764.4641596	-649.6618052
BGR	$\mathcal{M}_{\alpha^-}$	0.8838256	0.0000000	0.8083396	1732.3309835	-587.5907358
BGR	$\mathcal{M}_\delta$	0.8266382	0.0009712	0.8200675	1750.0324119	-622.2562137
BGR	$\mathcal{M}_\mu$	0.8597119	0.0000000	0.8124862	1737.9191192	-599.5986287
BGR	$\mathcal{M}_{null}$	0.8838256	0.0000922	0.8083396	1732.3319709	-593.8988785
BHR	$\mathcal{M}_{\alpha^+}$	1.6567106	0.0171776	0.8702070	1089.2193585	55.6414398
BHR	$\mathcal{M}_{\alpha^-}$	1.5961899	0.0523684	0.8790107	1092.1370550	33.5162361
BHR	$\mathcal{M}_\delta$	1.6725460	0.0380687	0.8704729	1089.9680617	54.9955669
BHR	$\mathcal{M}_\mu$	1.6588562	0.0042337	0.8710776	1090.6511513	53.5213442
BHR	$\mathcal{M}_{null}$	1.6091897	0.0288135	0.8782411	1100.0620639	29.7611037
BLR	$\mathcal{M}_{\alpha^+}$	0.2314568	0.0000000	0.7546141	302.7120092	-265.0157361
BLR	$\mathcal{M}_{\alpha^-}$	0.2310920	0.0000000	0.7559656	303.1611263	-265.8441406
BLR	$\mathcal{M}_\delta$	0.2312815	0.0000001	0.7550474	302.8040002	-265.2808168

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta BIC$
BLR	$\mathcal{M}_\mu$	0.2314569	0.0000000	0.7546081	302.7103969	-265.0120688
BLR	$\mathcal{M}_{null}$	0.2314568	0.0000000	0.7546151	302.7104437	-270.0269676
BOL	$\mathcal{M}_{\alpha^+}$	0.3539741	0.5305266	0.9268044	146.3653765	-107.0340735
BOL	$\mathcal{M}_{\alpha^-}$	0.4496981	0.0000247	0.9048095	135.9258779	-87.5909551
BOL	$\mathcal{M}_\delta$	0.4209790	0.0000005	0.9104235	138.0751496	-92.0891975
BOL	$\mathcal{M}_\mu$	0.4205466	0.0000066	0.9104306	138.0818576	-92.0950317
BOL	$\mathcal{M}_{null}$	0.4496985	0.0000003	0.9048066	135.9237089	-91.8927180
BRA	$\mathcal{M}_{\alpha^+}$	0.3172497	0.0000001	0.9440110	488.0802898	-388.2378022
BRA	$\mathcal{M}_{\alpha^-}$	0.3249407	0.0000000	0.9424996	485.4392675	-383.3634446
BRA	$\mathcal{M}_\delta$	0.3249407	0.0058828	0.9424995	485.4379807	-383.3630892
BRA	$\mathcal{M}_\mu$	0.3206030	0.0000000	0.9433234	486.8511429	-386.0040451
BRA	$\mathcal{M}_{null}$	0.3249407	0.0000000	0.9424990	485.4350185	-388.5710346
CAN	$\mathcal{M}_{\alpha^+}$	0.1420790	0.0000000	0.9813800	1740.6230285	-1592.8111504
CAN	$\mathcal{M}_{\alpha^-}$	0.1348187	0.0000000	0.9826218	1759.3484402	-1630.7054690
CAN	$\mathcal{M}_\delta$	0.1345303	0.0002367	0.9823122	1755.3549921	-1621.0102338
CAN	$\mathcal{M}_\mu$	0.1420790	0.0000000	0.9813802	1740.6331982	-1592.8186471
CAN	$\mathcal{M}_{null}$	0.1420790	0.0000000	0.9813800	1740.6275036	-1599.1212257
CHE	$\mathcal{M}_{\alpha^+}$	0.2379502	0.0002883	0.9792924	2188.2874195	-1290.4488236
CHE	$\mathcal{M}_{\alpha^-}$	0.2381246	0.0725138	0.9792943	2188.2300435	-1290.5016663
CHE	$\mathcal{M}_\delta$	0.2686683	0.0343412	0.9764831	2154.3432229	-1220.6077504
CHE	$\mathcal{M}_\mu$	0.2331059	0.0096259	0.9796038	2192.4685594	-1298.7697313
CHE	$\mathcal{M}_{null}$	0.2381246	0.1101355	0.9792943	2188.2241859	-1296.8080908
CHL	$\mathcal{M}_{\alpha^+}$	0.5874845	0.0004616	0.9580845	1049.6437073	-446.2645519
CHL	$\mathcal{M}_{\alpha^-}$	0.6179149	0.0000022	0.9565743	1043.9289736	-435.4684853
CHL	$\mathcal{M}_\delta$	0.6179150	0.0000543	0.9565743	1043.9278574	-435.4687821
CHL	$\mathcal{M}_\mu$	0.6179149	0.0000583	0.9565745	1043.9242279	-435.4695969
CHL	$\mathcal{M}_{null}$	0.6179149	0.0000344	0.9565743	1043.9283281	-441.1885246
COL	$\mathcal{M}_{\alpha^+}$	0.4490345	0.0000000	0.7782291	295.7157945	-221.5722013
COL	$\mathcal{M}_{\alpha^-}$	0.4690859	0.0000000	0.7690281	293.0967240	-216.6127614
COL	$\mathcal{M}_\delta$	0.4690859	0.0000000	0.7690293	293.0990791	-216.6134084
COL	$\mathcal{M}_\mu$	0.4690859	0.0000000	0.7690286	293.0969254	-216.6130390
COL	$\mathcal{M}_{null}$	0.4690859	0.0000000	0.7690282	293.0970732	-221.4168595
CRI	$\mathcal{M}_{\alpha^+}$	2.1618680	0.1028749	0.8450607	1399.4320760	-59.4886382
CRI	$\mathcal{M}_{\alpha^-}$	3.0070923	0.0161348	0.7967607	1355.2447368	38.1976757
CRI	$\mathcal{M}_\delta$	3.2498039	0.0035771	0.7495929	1312.1428722	113.3310469
CRI	$\mathcal{M}_\mu$	2.6857576	0.0535064	0.7957142	1349.6977712	40.0465823
CRI	$\mathcal{M}_{null}$	3.0070923	0.0000000	0.7967602	1355.2396475	32.3123594
CUB	$\mathcal{M}_{\alpha^+}$	0.9908768	0.0000001	0.7215276	478.2930704	-171.2727897
CUB	$\mathcal{M}_{\alpha^-}$	0.8902225	0.0000001	0.7426569	485.8364570	-185.7130367
CUB	$\mathcal{M}_\delta$	0.9908767	0.1727308	0.7215307	478.2948060	-171.2747793
CUB	$\mathcal{M}_\mu$	0.9585458	0.0000020	0.7316088	481.3950009	-178.0205603
CUB	$\mathcal{M}_{null}$	0.9908762	0.0000003	0.7215403	478.3330403	-176.4906120
CZE	$\mathcal{M}_{\alpha^+}$	0.0580083	0.0054248	0.9869966	1018.8726966	-778.2751708

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta BIC$
CZE	$\mathcal{M}_{\alpha^-}$	0.0904964	0.0069555	0.9776120	947.2785419	-636.4698424
CZE	$\mathcal{M}_\delta$	0.0898238	0.0017040	0.9775473	945.9773418	-635.7157255
CZE	$\mathcal{M}_\mu$	0.0886011	0.0000000	0.9790624	955.8358104	-653.9506041
CZE	$\mathcal{M}_{null}$	0.0904964	0.0000000	0.9776120	947.2785787	-642.0340782
DEU	$\mathcal{M}_{\alpha^+}$	0.1477448	0.2011817	0.9809597	2336.7255710	-1555.4873746
DEU	$\mathcal{M}_{\alpha^-}$	0.1947722	0.0000414	0.9759068	2272.0947772	-1426.2671483
DEU	$\mathcal{M}_\delta$	0.1853032	0.0018003	0.9760610	2275.9987490	-1429.7922153
DEU	$\mathcal{M}_\mu$	0.2447523	0.1317466	0.9688203	2202.4488643	-1284.7143596
DEU	$\mathcal{M}_{null}$	0.2020155	0.0000000	0.9745613	2257.6743661	-1402.7410593
DNK	$\mathcal{M}_{\alpha^+}$	0.1199929	0.0000000	0.9750129	1737.6862271	-1681.0788644
DNK	$\mathcal{M}_{\alpha^-}$	0.1114577	0.0012059	0.9770632	1761.0406979	-1728.0820104
DNK	$\mathcal{M}_\delta$	0.1067551	0.0075976	0.9778510	1770.9719428	-1747.2704361
DNK	$\mathcal{M}_\mu$	0.1199929	0.0001201	0.9750130	1737.6879470	-1681.0794476
DNK	$\mathcal{M}_{null}$	0.1199929	0.0000000	0.9750128	1737.6829698	-1687.3839457
DOM	$\mathcal{M}_{\alpha^+}$	0.6961046	0.0003898	0.9155814	373.7708776	-261.0889773
DOM	$\mathcal{M}_{\alpha^-}$	0.7001650	0.0000000	0.9152206	373.2591147	-260.3086751
DOM	$\mathcal{M}_\delta$	0.7001650	0.0000001	0.9152197	373.2484394	-260.3065475
DOM	$\mathcal{M}_\mu$	0.7001650	0.0000000	0.9152199	373.2546911	-260.3071059
DOM	$\mathcal{M}_{null}$	0.7001650	0.0000000	0.9152205	373.2634442	-265.5179004
DZA	$\mathcal{M}_{\alpha^+}$	1.2993726	0.0001052	0.9607808	442.7559336	-90.4793137
DZA	$\mathcal{M}_{\alpha^-}$	1.2997836	0.0000901	0.9607793	442.6671504	-90.4747253
DZA	$\mathcal{M}_\delta$	1.2996518	0.0001250	0.9607617	442.6369237	-90.4198633
DZA	$\mathcal{M}_\mu$	1.2996155	0.0000682	0.9607943	442.7400231	-90.5214100
DZA	$\mathcal{M}_{null}$	1.2996035	0.0000559	0.9607710	442.4031491	-95.2526666
ECU	$\mathcal{M}_{\alpha^+}$	1.5337628	0.0523325	0.9297865	401.2774868	-67.1390744
ECU	$\mathcal{M}_{\alpha^-}$	1.6887306	0.0000045	0.9209043	384.0521122	-47.8421360
ECU	$\mathcal{M}_\delta$	1.5963688	0.0266030	0.9251918	390.1548236	-56.8705370
ECU	$\mathcal{M}_\mu$	1.5439082	0.0000015	0.9279190	392.9889600	-62.8867317
ECU	$\mathcal{M}_{null}$	1.7791235	0.0279990	0.9178951	382.7367272	-46.8807035
ESP	$\mathcal{M}_{\alpha^+}$	0.1856080	0.0000028	0.9422837	1975.7328569	-1434.0717545
ESP	$\mathcal{M}_{\alpha^-}$	0.1855373	0.0076910	0.9424047	1976.2639184	-1435.2236056
ESP	$\mathcal{M}_\delta$	0.1548412	0.0000287	0.9517866	2025.4954451	-1532.8383743
ESP	$\mathcal{M}_\mu$	0.1856080	0.0000000	0.9422840	1975.7359458	-1434.0745080
ESP	$\mathcal{M}_{null}$	0.1856081	0.0000000	0.9422835	1975.7284854	-1440.3779295
FIN	$\mathcal{M}_{\alpha^+}$	0.3280622	0.0797648	0.9728173	2152.1380163	-1126.4399534
FIN	$\mathcal{M}_{\alpha^-}$	0.3360629	0.0000399	0.9730371	2153.3108657	-1130.8958714
FIN	$\mathcal{M}_\delta$	0.2788228	0.0981958	0.9767311	2193.0232499	-1211.7880069
FIN	$\mathcal{M}_\mu$	0.3539563	0.0007086	0.9708473	2130.2102048	-1088.0271204
FIN	$\mathcal{M}_{null}$	0.3797915	0.0000000	0.9689312	2113.4406174	-1059.3878652
FRA	$\mathcal{M}_{\alpha^+}$	0.1373927	0.0299971	0.9764265	1880.1804428	-1599.3274779
FRA	$\mathcal{M}_{\alpha^-}$	0.1386926	0.0000073	0.9769100	1886.0075800	-1610.7067588
FRA	$\mathcal{M}_\delta$	0.1198313	0.0474908	0.9789739	1914.1483197	-1662.1121244
FRA	$\mathcal{M}_\mu$	0.1421161	0.1610016	0.9757525	1872.4761689	-1583.8529740

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta BIC$
FRA	$\mathcal{M}_{null}$	0.1466084	0.0000000	0.9751495	1865.9199173	-1576.6741035
GBR	$\mathcal{M}_{\alpha^+}$	0.1669051	0.1074419	0.9775211	1884.4171476	-1484.4759333
GBR	$\mathcal{M}_{\alpha^-}$	0.1938987	0.0000000	0.9746001	1850.8502461	-1417.4039255
GBR	$\mathcal{M}_\delta$	0.1413371	0.0012231	0.9808639	1931.1521839	-1572.8654147
GBR	$\mathcal{M}_\mu$	0.1999852	0.0000000	0.9731472	1835.8110921	-1386.8661426
GBR	$\mathcal{M}_{null}$	0.1903921	0.0000000	0.9745839	1850.8009895	-1423.3628352
GHA	$\mathcal{M}_{\alpha^+}$	0.3490916	0.0000916	0.7666365	195.2026830	-249.8849106
GHA	$\mathcal{M}_{\alpha^-}$	0.3531896	0.0000000	0.7657012	194.8203537	-249.3969268
GHA	$\mathcal{M}_\delta$	0.3076765	0.0053948	0.7871236	200.1576962	-261.0949413
GHA	$\mathcal{M}_\mu$	0.3531896	0.0000000	0.7657012	194.8204784	-249.3969083
GHA	$\mathcal{M}_{null}$	0.3531896	0.0000000	0.7657012	194.8205084	-254.2009194
GRC	$\mathcal{M}_{\alpha^+}$	0.4896514	0.0161565	0.8280443	1354.6952066	-793.6899983
GRC	$\mathcal{M}_{\alpha^-}$	0.5290440	0.0000000	0.8177233	1337.4385948	-765.2452269
GRC	$\mathcal{M}_\delta$	0.4874272	0.0273553	0.8276619	1353.0776742	-792.6062212
GRC	$\mathcal{M}_\mu$	0.5490437	0.0000000	0.8073630	1324.7477342	-738.2677202
GRC	$\mathcal{M}_{null}$	0.5412571	0.0000001	0.8096391	1327.2211290	-750.2581772
GTM	$\mathcal{M}_{\alpha^+}$	1.3642600	0.0080365	0.9021763	558.2459476	-159.0748496
GTM	$\mathcal{M}_{\alpha^-}$	1.5874596	0.0000000	0.8823431	535.2421777	-114.0311039
GTM	$\mathcal{M}_\delta$	1.5874598	0.0000000	0.8823458	535.2459962	-114.0367881
GTM	$\mathcal{M}_\mu$	1.6555121	0.0000000	0.8757806	528.0917573	-100.7875649
GTM	$\mathcal{M}_{null}$	1.5874596	0.0000000	0.8823431	535.2423513	-119.5283970
HKG	$\mathcal{M}_{\alpha^+}$	0.2509271	0.0014428	0.9787234	2126.7218635	-1260.9135186
HKG	$\mathcal{M}_{\alpha^-}$	0.2445060	0.0000013	0.9794208	2135.6830150	-1279.2081293
HKG	$\mathcal{M}_\delta$	0.2517358	0.1089110	0.9785874	2124.9567552	-1257.4134379
HKG	$\mathcal{M}_\mu$	0.2509189	0.2308280	0.9787245	2126.7325395	-1260.9402382
HKG	$\mathcal{M}_{null}$	0.2509189	0.0014568	0.9787245	2126.7360861	-1267.2474677
HRV	$\mathcal{M}_{\alpha^+}$	0.0659912	0.0001262	0.9787833	618.7632652	-772.8094871
HRV	$\mathcal{M}_{\alpha^-}$	0.0696456	0.0000000	0.9780032	613.7143315	-763.0604576
HRV	$\mathcal{M}_\delta$	0.0696456	0.0000271	0.9780032	613.7143075	-763.0603906
HRV	$\mathcal{M}_\mu$	0.0696456	0.0000000	0.9780032	613.7143096	-763.0604842
HRV	$\mathcal{M}_{null}$	0.0696456	0.0000345	0.9780032	613.7143409	-768.6588048
HUN	$\mathcal{M}_{\alpha^+}$	0.2514814	0.0000000	0.9582531	2350.3148004	-1119.3586592
HUN	$\mathcal{M}_{\alpha^-}$	0.3220814	0.0000345	0.9491220	2303.4750542	-1022.8291131
HUN	$\mathcal{M}_\delta$	0.3220814	0.0000002	0.9491218	2303.4782840	-1022.8276039
HUN	$\mathcal{M}_\mu$	0.2834268	0.0000000	0.9535596	2324.8152044	-1067.3653028
HUN	$\mathcal{M}_{null}$	0.3220814	0.0000000	0.9491220	2303.4784333	-1029.0201300
IDN	$\mathcal{M}_{\alpha^+}$	0.5192693	0.0000000	0.9146493	305.0747698	-197.7229477
IDN	$\mathcal{M}_{\alpha^-}$	0.6211057	0.0000000	0.8971687	292.8324976	-174.9917751
IDN	$\mathcal{M}_\delta$	0.5703114	0.0003749	0.9049279	298.3149198	-184.5632250
IDN	$\mathcal{M}_\mu$	0.6211057	0.0000000	0.8971704	292.8323470	-174.9937641
IDN	$\mathcal{M}_{null}$	0.6211057	0.0000000	0.8971695	292.8317590	-179.7967128
IRL	$\mathcal{M}_{\alpha^+}$	0.2088275	0.0143105	0.9772782	2569.1155517	-1359.9304403
IRL	$\mathcal{M}_{\alpha^-}$	0.2073867	0.0000036	0.9774790	2571.5504173	-1364.8037711

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta BIC$
IRL	$\mathcal{M}_\delta$	0.2090907	0.0200323	0.9772362	2568.5462746	-1358.9156430
IRL	$\mathcal{M}_\mu$	0.2088306	0.0352422	0.9772776	2569.1090642	-1359.9153338
IRL	$\mathcal{M}_{null}$	0.2088274	0.0144979	0.9772784	2569.1199148	-1366.2417945
ISR	$\mathcal{M}_{\alpha^+}$	0.4160179	0.0062564	0.9709201	1893.2428702	-1010.4705652
ISR	$\mathcal{M}_{\alpha^-}$	0.4349423	0.0000136	0.9701674	1885.8473780	-996.4417954
ISR	$\mathcal{M}_\delta$	0.4339349	0.0008507	0.9698446	1882.2493491	-990.5338945
ISR	$\mathcal{M}_\mu$	0.4349423	0.0000322	0.9701674	1885.8468109	-996.4418756
ISR	$\mathcal{M}_{null}$	0.4349423	0.0000111	0.9701674	1885.8485980	-1002.7501047
JOR	$\mathcal{M}_{\alpha^+}$	1.3140447	0.0000227	0.8158654	859.9004778	-213.5884682
JOR	$\mathcal{M}_{\alpha^-}$	1.3883468	0.0000158	0.8112079	854.4614037	-204.7457457
JOR	$\mathcal{M}_\delta$	1.3140447	0.0010034	0.8158620	859.9018193	-213.5819671
JOR	$\mathcal{M}_\mu$	1.3044454	0.0040242	0.8169297	860.5974454	-215.6406373
JOR	$\mathcal{M}_{null}$	1.3140446	0.0000000	0.8158654	859.9040348	-219.4577945
JPN	$\mathcal{M}_{\alpha^+}$	0.0814135	0.0000000	0.9672378	794.7905570	-850.5394247
JPN	$\mathcal{M}_{\alpha^-}$	0.0683921	0.0113106	0.9730056	818.4672910	-897.7889660
JPN	$\mathcal{M}_\delta$	0.0513784	0.0000323	0.9785380	846.1711542	-953.7491905
JPN	$\mathcal{M}_\mu$	0.0814137	0.0000001	0.9672373	794.7938193	-850.5360642
JPN	$\mathcal{M}_{null}$	0.0814135	0.0000002	0.9672371	794.7868478	-856.0313369
KAZ	$\mathcal{M}_{\alpha^+}$	0.1865516	0.0000476	0.8295751	214.6552620	-174.1895588
KAZ	$\mathcal{M}_{\alpha^-}$	0.2496131	0.0000000	0.7623985	199.8289770	-144.2825364
KAZ	$\mathcal{M}_\delta$	0.2390774	0.0000000	0.7721660	201.6233048	-148.0605220
KAZ	$\mathcal{M}_\mu$	0.2362297	0.0000001	0.7769984	202.2588602	-149.9899724
KAZ	$\mathcal{M}_{null}$	0.2496130	0.0000001	0.7624011	199.8317347	-148.7833403
KOR	$\mathcal{M}_{\alpha^+}$	0.6920979	0.0649634	0.9155851	2216.2272046	-646.0307164
KOR	$\mathcal{M}_{\alpha^-}$	0.8387203	0.0000073	0.8985858	2168.9233801	-556.4976910
KOR	$\mathcal{M}_\delta$	0.6985206	0.0000000	0.9123594	2202.6982338	-627.7305826
KOR	$\mathcal{M}_\mu$	0.6614961	0.2464106	0.9192266	2224.9307739	-667.5495675
KOR	$\mathcal{M}_{null}$	0.7946315	0.0000000	0.9020007	2176.7551440	-579.4032059
LBN	$\mathcal{M}_{\alpha^+}$	1.3016604	0.0240624	0.8243273	807.8076783	-327.5331892
LBN	$\mathcal{M}_{\alpha^-}$	1.4386036	0.0000001	0.8073042	782.3615142	-282.3978581
LBN	$\mathcal{M}_\delta$	1.4380388	0.0005367	0.8117306	789.9437981	-293.7384027
LBN	$\mathcal{M}_\mu$	1.4320212	0.0000001	0.8135593	792.1142739	-298.5017188
LBN	$\mathcal{M}_{null}$	1.4416777	0.0000000	0.8100418	788.0379102	-295.5709325
LBY	$\mathcal{M}_{\alpha^+}$	2.2780600	0.0000128	0.9267840	378.4283856	-11.0745573
LBY	$\mathcal{M}_{\alpha^-}$	2.8894278	0.0000473	0.9079340	364.6708243	16.8746157
LBY	$\mathcal{M}_\delta$	3.5313144	0.0002249	0.8853192	349.5494084	43.6715402
LBY	$\mathcal{M}_\mu$	3.4722522	0.0000224	0.8868652	350.6939375	42.0157078
LBY	$\mathcal{M}_{null}$	2.7553149	0.0000477	0.9107669	365.7889379	8.2576748
LKA	$\mathcal{M}_{\alpha^+}$	1.5941187	0.0003071	0.8115942	336.9975384	-92.8885083
LKA	$\mathcal{M}_{\alpha^-}$	1.6362543	0.0000000	0.8102917	337.0514707	-91.6277204
LKA	$\mathcal{M}_\delta$	1.6322208	0.0022002	0.8084860	335.3889986	-89.8940407
LKA	$\mathcal{M}_\mu$	1.6362543	0.0025755	0.8102917	337.0514702	-91.6277204
LKA	$\mathcal{M}_{null}$	1.6362543	0.0000000	0.8102917	337.0514708	-96.8372065

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta BIC$
LTU	$\mathcal{M}_{\alpha^+}$	0.2792254	0.0000068	0.9197774	550.3188221	-359.2088757
LTU	$\mathcal{M}_{\alpha^-}$	0.2792263	0.0011184	0.9197673	550.3111851	-359.1746619
LTU	$\mathcal{M}_\delta$	0.2539895	0.0000002	0.9241161	555.7579934	-374.2208641
LTU	$\mathcal{M}_\mu$	0.2554806	0.0000000	0.9240155	556.4417023	-373.8632704
LTU	$\mathcal{M}_{null}$	0.2792264	0.0000000	0.9197682	550.3064726	-364.7762372
LUX	$\mathcal{M}_{\alpha^+}$	0.1687301	0.0027144	0.9756622	606.0315078	-387.5202145
LUX	$\mathcal{M}_{\alpha^-}$	0.1556444	0.0423426	0.9776990	616.1184100	-405.6120311
LUX	$\mathcal{M}_\delta$	0.1556444	0.0037217	0.9776974	616.1053905	-405.5967384
LUX	$\mathcal{M}_\mu$	0.1556444	0.0514556	0.9776986	616.1152846	-405.6079128
LUX	$\mathcal{M}_{null}$	0.1556444	0.0000000	0.9776985	616.1172377	-410.9401680
IVA	$\mathcal{M}_{\alpha^+}$	0.2953602	0.0000000	0.9317351	543.8952197	-366.6813027
IVA	$\mathcal{M}_{\alpha^-}$	0.2877696	0.0000698	0.9322366	544.6567294	-368.6722851
IVA	$\mathcal{M}_\delta$	0.2280066	0.0002703	0.9426518	563.8282263	-413.7302090
IVA	$\mathcal{M}_\mu$	0.2682572	0.0000000	0.9354481	550.4990299	-381.7817245
IVA	$\mathcal{M}_{null}$	0.2953602	0.0000000	0.9317370	543.8963214	-372.2873909
MAC	$\mathcal{M}_{\alpha^+}$	1.1294488	0.0004599	0.9315991	1531.4525516	-289.1433884
MAC	$\mathcal{M}_{\alpha^-}$	0.9767320	0.0000041	0.9406784	1567.2754683	-349.5263422
MAC	$\mathcal{M}_\delta$	1.1294497	0.0051249	0.9315974	1531.4577652	-289.1325538
MAC	$\mathcal{M}_\mu$	1.1294485	0.0005140	0.9315978	1531.4587640	-289.1354956
MAC	$\mathcal{M}_{null}$	1.1294486	0.0005148	0.9315990	1531.4584828	-295.1922572
MAR	$\mathcal{M}_{\alpha^+}$	0.8537848	0.0234547	0.9349568	650.3604985	-264.3866710
MAR	$\mathcal{M}_{\alpha^-}$	0.9295200	0.0000018	0.9296139	640.4149824	-245.1246477
MAR	$\mathcal{M}_\delta$	0.9016048	0.0040674	0.9319661	643.9018432	-253.4178828
MAR	$\mathcal{M}_\mu$	0.9173997	0.0000204	0.9310205	642.7052567	-250.0499976
MAR	$\mathcal{M}_{null}$	0.9295200	0.0000035	0.9296140	640.4143693	-250.6218776
MEX	$\mathcal{M}_{\alpha^+}$	0.4508260	0.0000993	0.9151172	1435.2035699	-631.2102997
MEX	$\mathcal{M}_{\alpha^-}$	0.4876843	0.0000001	0.9084882	1420.6198791	-603.6884578
MEX	$\mathcal{M}_\delta$	0.4392520	0.0149692	0.9163000	1437.2548352	-636.3463466
MEX	$\mathcal{M}_\mu$	0.4866618	0.0092329	0.9082319	1420.0192358	-602.6647468
MEX	$\mathcal{M}_{null}$	0.4876843	0.0000001	0.9084879	1420.6159465	-609.5898333
MYS	$\mathcal{M}_{\alpha^+}$	0.2895046	0.0785445	0.9539404	1973.5958108	-1124.8476489
MYS	$\mathcal{M}_{\alpha^-}$	0.3399726	0.0000000	0.9474011	1937.0418339	-1054.3528091
MYS	$\mathcal{M}_\delta$	0.3399726	0.0000000	0.9474014	1937.0487616	-1054.3553106
MYS	$\mathcal{M}_\mu$	0.3169605	0.0000000	0.9509908	1955.6429441	-1091.8872434
MYS	$\mathcal{M}_{null}$	0.3399726	0.0000000	0.9474012	1937.0402581	-1060.6285880
NGA	$\mathcal{M}_{\alpha^+}$	2.7990310	0.0000239	0.9395854	495.7796031	11.7048325
NGA	$\mathcal{M}_{\alpha^-}$	2.7992006	0.0000615	0.9395758	495.6501030	11.7242375
NGA	$\mathcal{M}_\delta$	2.6256522	0.0000111	0.9431462	498.4576558	4.2936995
NGA	$\mathcal{M}_\mu$	2.7268460	0.0000217	0.9411712	496.2771356	8.4598284
NGA	$\mathcal{M}_{null}$	2.7990746	0.0000159	0.9395804	495.7132708	6.9110458
NLD	$\mathcal{M}_{\alpha^+}$	0.1466304	0.0000001	0.9582107	1657.4935262	-1562.8725186
NLD	$\mathcal{M}_{\alpha^-}$	0.2092032	0.0000000	0.9418981	1567.4351623	-1381.9447140
NLD	$\mathcal{M}_\delta$	0.1340786	0.0351018	0.9609515	1678.9038968	-1600.1142643

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta BIC$
NLD	$\mathcal{M}_\mu$	0.2097818	0.0272475	0.9412098	1564.2116085	-1375.4789279
NLD	$\mathcal{M}_{null}$	0.2097818	0.0000000	0.9412096	1564.2134536	-1381.7853394
NOR	$\mathcal{M}_{\alpha^+}$	0.2399445	0.0130421	0.9780959	2521.5074784	-1298.8700700
NOR	$\mathcal{M}_{\alpha^-}$	0.2390888	0.0016231	0.9781834	2522.6463097	-1301.0687029
NOR	$\mathcal{M}_\delta$	0.2400056	0.0128037	0.9780893	2521.4101099	-1298.7067950
NOR	$\mathcal{M}_\mu$	0.2399438	0.0130562	0.9780959	2521.5037000	-1298.8707092
NOR	$\mathcal{M}_{null}$	0.2398829	0.0193799	0.9780996	2521.5638888	-1305.2721046
NZL	$\mathcal{M}_{\alpha^+}$	0.9073281	0.0037365	0.9129915	1654.8440041	-567.6419686
NZL	$\mathcal{M}_{\alpha^-}$	0.8772374	0.0000059	0.9140447	1656.6103203	-573.5852855
NZL	$\mathcal{M}_\delta$	0.4805584	0.0203982	0.9447784	1759.6846130	-789.5124824
NZL	$\mathcal{M}_\mu$	0.4917836	0.0120060	0.9442763	1758.4536220	-785.0947066
NZL	$\mathcal{M}_{null}$	0.9073281	0.0000000	0.9129914	1654.8396788	-573.8317817
OMN	$\mathcal{M}_{\alpha^+}$	0.7889506	0.0396981	0.9139409	1112.8720701	-153.1574333
OMN	$\mathcal{M}_{\alpha^-}$	0.6700664	0.0860673	0.9239838	1120.4887405	-183.5588674
OMN	$\mathcal{M}_\delta$	0.8933917	0.0678704	0.9031544	1102.8136604	-124.2268460
OMN	$\mathcal{M}_\mu$	0.8791464	0.0000000	0.9020953	1095.4922578	-121.5620907
OMN	$\mathcal{M}_{null}$	0.8520861	0.0006775	0.9052430	1100.9144521	-135.0695316
PAN	$\mathcal{M}_{\alpha^+}$	1.1143223	0.0000000	0.8199065	998.7151866	-355.1878253
PAN	$\mathcal{M}_{\alpha^-}$	1.1149436	0.0000000	0.8201169	998.8693793	-355.6869349
PAN	$\mathcal{M}_\delta$	1.1100368	0.0002924	0.8199278	998.2834172	-355.2383815
PAN	$\mathcal{M}_\mu$	1.1149435	0.0000000	0.8201169	998.8720122	-355.6869611
PAN	$\mathcal{M}_{null}$	1.1149436	0.0000000	0.8201166	998.8746400	-361.7432027
PER	$\mathcal{M}_{\alpha^+}$	0.8254032	0.0311618	0.8826033	573.8679392	-274.9069780
PER	$\mathcal{M}_{\alpha^-}$	0.9361490	0.0000000	0.8671464	559.4016095	-244.7267453
PER	$\mathcal{M}_\delta$	0.8882912	0.0000974	0.8760467	567.2656212	-261.6465567
PER	$\mathcal{M}_\mu$	0.9055167	0.0000000	0.8725708	564.1687729	-254.8983509
PER	$\mathcal{M}_{null}$	0.9361490	0.0000000	0.8671464	559.4031274	-250.2239393
POL	$\mathcal{M}_{\alpha^+}$	0.2535426	0.0000000	0.9438326	1807.0469575	-1001.0879697
POL	$\mathcal{M}_{\alpha^-}$	0.3163218	0.0000000	0.9326013	1768.7678021	-923.2501964
POL	$\mathcal{M}_\delta$	0.3162856	0.0000403	0.9326143	1768.7760416	-923.3322541
POL	$\mathcal{M}_\mu$	0.2782767	0.0000000	0.9379993	1785.4443202	-958.8962784
POL	$\mathcal{M}_{null}$	0.3163218	0.0000000	0.9326025	1768.7577185	-929.3142565
PRT	$\mathcal{M}_{\alpha^+}$	0.3321047	0.0000076	0.9133453	1726.5468270	-1093.0105775
PRT	$\mathcal{M}_{\alpha^-}$	0.3453640	0.0127740	0.9112486	1719.7441003	-1079.8850541
PRT	$\mathcal{M}_\delta$	0.3369898	0.0000004	0.9136215	1727.2952235	-1094.7635003
PRT	$\mathcal{M}_\mu$	0.3417620	0.0000000	0.9114441	1720.4004837	-1081.0956205
PRT	$\mathcal{M}_{null}$	0.3453640	0.0000000	0.9112488	1719.7422418	-1086.1944995
PRY	$\mathcal{M}_{\alpha^+}$	0.7911077	0.0000299	0.8628434	618.4255027	-283.9049804
PRY	$\mathcal{M}_{\alpha^-}$	0.8457188	0.0000000	0.8541465	610.2534769	-268.9040302
PRY	$\mathcal{M}_\delta$	0.8457189	0.0000006	0.8541473	610.2555041	-268.9052712
PRY	$\mathcal{M}_\mu$	0.8457188	0.0000003	0.8541469	610.2551282	-268.9046708
PRY	$\mathcal{M}_{null}$	0.8457188	0.0000007	0.8541468	610.2536061	-274.4016218
QAT	$\mathcal{M}_{\alpha^+}$	1.9266798	0.0038751	0.9118724	1618.1838222	-63.5800421

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta BIC$
QAT	$\mathcal{M}_{\alpha^-}$	2.0139147	0.0000255	0.9086920	1608.4377172	-51.8806944
QAT	$\mathcal{M}_\delta$	1.9353840	0.0001128	0.9115033	1615.0203438	-62.2009448
QAT	$\mathcal{M}_\mu$	1.9353839	0.0226246	0.9115036	1615.0306966	-62.2018858
QAT	$\mathcal{M}_{null}$	1.9353843	0.0000104	0.9115025	1615.0092877	-67.9971407
ROU	$\mathcal{M}_{\alpha^+}$	0.8900993	0.0290867	0.8907507	1747.7760599	-447.7210801
ROU	$\mathcal{M}_{\alpha^-}$	1.0376017	0.0000015	0.8796222	1727.7138904	-406.3009748
ROU	$\mathcal{M}_\delta$	1.0376021	0.0000221	0.8796235	1727.7104815	-406.3056340
ROU	$\mathcal{M}_\mu$	1.0008829	0.0125723	0.8786329	1723.9683488	-402.8060015
ROU	$\mathcal{M}_{null}$	1.0376017	0.0000001	0.8796230	1727.7051904	-412.3607662
RUS	$\mathcal{M}_{\alpha^+}$	0.6174841	0.0000000	0.9566539	1070.3858708	-349.4521333
RUS	$\mathcal{M}_{\alpha^-}$	0.8291502	0.0000002	0.9429638	1037.3556155	-282.4815040
RUS	$\mathcal{M}_\delta$	0.8291506	0.0000000	0.9429651	1037.3633751	-282.4866992
RUS	$\mathcal{M}_\mu$	0.8194430	0.0000824	0.9438272	1038.7924091	-286.2030003
RUS	$\mathcal{M}_{null}$	0.8291502	0.0000000	0.9429641	1037.3532266	-287.9799069
SAU	$\mathcal{M}_{\alpha^+}$	1.4447598	0.0000015	0.9319666	318.2772364	-63.7460639
SAU	$\mathcal{M}_{\alpha^-}$	1.1467943	0.0000016	0.9491154	335.8594189	-98.0178968
SAU	$\mathcal{M}_\delta$	1.4317991	0.0000000	0.9331044	319.2716499	-65.7362927
SAU	$\mathcal{M}_\mu$	1.3103942	0.0000004	0.9409868	327.3080046	-80.5301406
SAU	$\mathcal{M}_{null}$	1.4447565	0.0000027	0.9319483	318.1644064	-68.4849842
SGP	$\mathcal{M}_{\alpha^+}$	0.4390113	0.0002142	0.9326026	2040.0506241	-979.1514085
SGP	$\mathcal{M}_{\alpha^-}$	0.4389674	0.1024984	0.9330349	2041.7687561	-982.6841384
SGP	$\mathcal{M}_\delta$	0.4381149	0.0988135	0.9327193	2040.8073272	-980.1034145
SGP	$\mathcal{M}_\mu$	0.4390113	0.0001842	0.9326026	2040.0476396	-979.1513242
SGP	$\mathcal{M}_{null}$	0.4390113	0.1632410	0.9326026	2040.0510546	-985.4594657
SVK	$\mathcal{M}_{\alpha^+}$	0.0897265	0.2508679	0.9841495	1031.5427644	-671.5084754
SVK	$\mathcal{M}_{\alpha^-}$	0.1076807	0.0000152	0.9797558	999.7833418	-607.6491798
SVK	$\mathcal{M}_\delta$	0.1297093	0.0563428	0.9763511	979.6865381	-567.0771140
SVK	$\mathcal{M}_\mu$	0.1164278	0.0012728	0.9782005	991.2554924	-588.3305464
SVK	$\mathcal{M}_{null}$	0.1135388	0.0000000	0.9781674	991.0807478	-593.4997133
SVN	$\mathcal{M}_{\alpha^+}$	0.1106088	0.0196362	0.9794616	807.4033352	-625.2063291
SVN	$\mathcal{M}_{\alpha^-}$	0.1204472	0.0001807	0.9768129	791.5610114	-592.4550281
SVN	$\mathcal{M}_\delta$	0.1060602	0.0012029	0.9792287	805.5613897	-622.1622849
SVN	$\mathcal{M}_\mu$	0.1198190	0.0006017	0.9769728	792.4983162	-594.3233439
SVN	$\mathcal{M}_{null}$	0.1204460	0.0001852	0.9768135	791.5636796	-598.0604505
SWE	$\mathcal{M}_{\alpha^+}$	0.1794411	0.0000246	0.9827111	2093.2525551	-1458.8338866
SWE	$\mathcal{M}_{\alpha^-}$	0.1679885	0.0011001	0.9840927	2115.9414425	-1504.5572179
SWE	$\mathcal{M}_\delta$	0.1604230	0.0619388	0.9845382	2124.0345953	-1520.1547048
SWE	$\mathcal{M}_\mu$	0.1773051	0.0333840	0.9828882	2095.9941251	-1464.4867308
SWE	$\mathcal{M}_{null}$	0.1794411	0.0000000	0.9827112	2093.2494231	-1465.1444556
THA	$\mathcal{M}_{\alpha^+}$	0.3214689	0.0000002	0.9515592	1740.3783079	-881.8617066
THA	$\mathcal{M}_{\alpha^-}$	0.3563113	0.0000219	0.9490399	1729.7252912	-860.2127021
THA	$\mathcal{M}_\delta$	0.3563231	0.0000376	0.9490292	1729.6876773	-860.1227083
THA	$\mathcal{M}_\mu$	0.3563111	0.0000166	0.9490411	1729.7307152	-860.2231110

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta BIC$
THA	$\mathcal{M}_{null}$	0.3563111	0.0000002	0.9490411	1729.7276038	-866.2793417
TKM	$\mathcal{M}_{\alpha^+}$	0.6120721	0.0002175	0.8484762	110.4929157	-52.1194340
TKM	$\mathcal{M}_{\alpha^-}$	0.6125306	0.0002738	0.8486140	110.5452317	-52.1740147
TKM	$\mathcal{M}_\delta$	0.6125441	0.0000282	0.8483613	110.4933587	-52.0739457
TKM	$\mathcal{M}_\mu$	0.6125429	0.0000481	0.8483465	110.4568505	-52.0681256
TKM	$\mathcal{M}_{null}$	0.6125373	0.0000492	0.8483298	110.4627760	-56.1558440
TUN	$\mathcal{M}_{\alpha^+}$	0.6040435	0.0519402	0.8800402	970.7830183	-525.8558531
TUN	$\mathcal{M}_{\alpha^-}$	0.6553858	0.0000039	0.8721401	959.5907734	-502.5129657
TUN	$\mathcal{M}_\delta$	0.6413280	0.0134767	0.8740118	962.3288209	-507.9103306
TUN	$\mathcal{M}_\mu$	0.6486814	0.0000013	0.8731389	960.7741688	-505.3833771
TUN	$\mathcal{M}_{null}$	0.6553859	0.0000000	0.8721396	959.5919379	-508.4144076
TUR	$\mathcal{M}_{\alpha^+}$	0.5613733	0.0036904	0.9221724	1006.5304764	-455.9797216
TUR	$\mathcal{M}_{\alpha^-}$	0.5743223	0.0000163	0.9212163	1004.4008375	-452.2555678
TUR	$\mathcal{M}_\delta$	0.5537710	0.0000113	0.9241992	1010.4413593	-464.0276218
TUR	$\mathcal{M}_\mu$	0.5564595	0.0000000	0.9225319	1006.5641070	-457.3917962
TUR	$\mathcal{M}_{null}$	0.5743230	0.0000163	0.9212166	1004.4010786	-457.9771897
UKR	$\mathcal{M}_{\alpha^+}$	0.2824879	0.0000003	0.8475939	237.3301802	-170.0780074
UKR	$\mathcal{M}_{\alpha^-}$	0.2448255	0.0000002	0.8612119	241.8217522	-181.3100499
UKR	$\mathcal{M}_\delta$	0.1863009	0.0000001	0.8928971	257.3043951	-212.4090202
UKR	$\mathcal{M}_\mu$	0.2496387	0.0000212	0.8602308	242.1071349	-180.4647504
UKR	$\mathcal{M}_{null}$	0.2824878	0.0000003	0.8476001	237.3437052	-174.8704237
URY	$\mathcal{M}_{\alpha^+}$	0.3707057	0.0004125	0.9539609	760.9348292	-446.7383706
URY	$\mathcal{M}_{\alpha^-}$	0.3801091	0.0000909	0.9515936	754.4475193	-433.2002436
URY	$\mathcal{M}_\delta$	0.3801091	0.0739791	0.9515934	754.4467727	-433.1992421
URY	$\mathcal{M}_\mu$	0.4148870	0.0000000	0.9462801	739.8061158	-405.0794658
URY	$\mathcal{M}_{null}$	0.3801091	0.0000489	0.9515934	754.4502665	-438.7974684
USA	$\mathcal{M}_{\alpha^+}$	0.1668963	0.0055876	0.9697761	1854.4974683	-1504.7078129
USA	$\mathcal{M}_{\alpha^-}$	0.1691152	0.0000027	0.9698283	1854.8253141	-1505.6582104
USA	$\mathcal{M}_\delta$	0.1869925	0.0006049	0.9658160	1820.5736973	-1437.1131786
USA	$\mathcal{M}_\mu$	0.1691152	0.0211774	0.9698284	1854.8269747	-1505.6591624
USA	$\mathcal{M}_{null}$	0.1691152	0.0221088	0.9698284	1854.8267976	-1511.9671769
VEN	$\mathcal{M}_{\alpha^+}$	0.7248894	0.0302883	0.9314021	368.7917156	-158.4962726
VEN	$\mathcal{M}_{\alpha^-}$	0.6719343	0.0000000	0.9369305	373.8565502	-168.7471573
VEN	$\mathcal{M}_\delta$	0.7248895	0.0000351	0.9314017	368.7906482	-158.4955756
VEN	$\mathcal{M}_\mu$	0.7248898	0.0371293	0.9314029	368.7920561	-158.4976122
VEN	$\mathcal{M}_{null}$	0.7248892	0.0000001	0.9314027	368.7897007	-163.3013450
VNM	$\mathcal{M}_{\alpha^+}$	2.3967138	0.0012084	0.5717167	960.8749519	-14.6593163
VNM	$\mathcal{M}_{\alpha^-}$	2.4133088	0.0000400	0.5678732	960.0251253	-13.0243703
VNM	$\mathcal{M}_\delta$	2.4132716	0.0000647	0.5679233	959.9523137	-13.0455685
VNM	$\mathcal{M}_\mu$	2.2242733	0.0000015	0.5854469	961.8716340	-20.6221612
VNM	$\mathcal{M}_{null}$	2.4132930	0.0000029	0.5678849	959.9594100	-18.2388068
ZAF	$\mathcal{M}_{\alpha^+}$	0.4738056	0.0014702	0.9471435	1054.7190010	-621.0423136
ZAF	$\mathcal{M}_{\alpha^-}$	0.4936819	0.0000042	0.9466293	1051.1700514	-617.4985641

Country	Model	Loss	Rel. std. loss	R2	Log-likelihood	$\Delta\text{BIC}$
ZAF	$\mathcal{M}_\delta$	0.4936820	0.0001680	0.9466294	1051.1688630	-617.4996371
ZAF	$\mathcal{M}_\mu$	0.4936819	0.0000044	0.9466295	1051.1689693	-617.4999436
ZAF	$\mathcal{M}_{null}$	0.4936819	0.0001973	0.9466293	1051.1698206	-623.4012024

**Tab. S4.2: Results of the numerical simulations.** Numerical values for each model and each country correspond to the optimization run with the minimum loss value. "Rel. std. loss" corresponds to the standard deviation of the loss values obtained for the five independent runs, divided by the mean loss value. Low "Rel. std. loss" indicates that the optimization runs have converged towards the same maximum likelihood.

Country Code	Country Name	Country Code	Country Name
AGO	Angola	LBN	Lebanon
ARE	United Arab Emirates	LBY	Libya
ARG	Argentina	LKA	Sri Lanka
AUS	Australia	LTU	Lithuania
AUT	Austria	LUX	Luxembourg
AZE	Azerbaijan	LVA	Latvia
BEL	Belgium	MAC	Macao SAR, China
BGD	Bangladesh	MAR	Morocco
BGR	Bulgaria	MEX	Mexico
BHR	Bahrain	MMR	Myanmar
BLR	Belarus	MYS	Malaysia
BOL	Bolivia	NGA	Nigeria
BRA	Brazil	NLD	Netherlands
CAN	Canada	NOR	Norway
CHE	Switzerland	NZL	New Zealand
CHL	Chile	OMN	Oman
CHN	China	PAK	Pakistan
COL	Colombia	PAN	Panama
CRI	Costa Rica	PER	Peru
CUB	Cuba	POL	Poland
CZE	Czech Republic	PRI	Puerto Rico
DEU	Germany	PRT	Portugal
DNK	Denmark	PRY	Paraguay
DOM	Dominican Republic	QAT	Qatar
DZA	Algeria	ROU	Romania
ECU	Ecuador	RUS	Russian Federation
EGY	Egypt, Arab Rep.	SAU	Saudi Arabia
ESP	Spain	SDN	Sudan
ETH	Ethiopia	SGP	Singapore
FIN	Finland	SRB	Serbia
FRA	France	SVK	Slovak Republic
GBR	United Kingdom	SVN	Slovenia
GHA	Ghana	SWE	Sweden
GRC	Greece	SYR	Syrian Arab Republic
GTM	Guatemala	THA	Thailand
HKG	Hong Kong SAR, China	TKM	Turkmenistan
HRV	Croatia	TUN	Tunisia
HUN	Hungary	TUR	Turkey
IDN	Indonesia	TZA	Tanzania
IND	India	UKR	Ukraine
IRL	Ireland	URY	Uruguay
IRN	Iran, Islamic Rep.	USA	United States
IRQ	Iraq	UZB	Uzbekistan
ISR	Israel	VEN	Venezuela, RB
JOR	Jordan	VNM	Vietnam
JPN	Japan	WLD	World
KAZ	Kazakhstan	YEM	Yemen, Rep.
KEN	Kenya	ZAF	South Africa
KOR	Korea, Rep.		

**Tab. S4.1:** ISO 3166-1 country codes for each of the 100 country investigated.

	Loglikelihood	Loglikelihood residuals
(Intercept)	0.000 (0.045)	1.495** (0.474)
NT.	0.928*** (0.046)	
N		-0.230** (0.071)
log(GDP)		0.693*** (0.187)
Number of countries	69	69
R <sup>2</sup>	0.861	0.174

**Tab. S4.3:** Linear regression model coefficients for the effect of the number of data points ( $NT$ ) on the best model loglikelihood, and linear regression model coefficients for the effect of GDP and number of economic activities ( $N$ ) on the residuals. \*\*\* $p < 0.001$





# Deep learning approximations for non-local nonlinear PDEs with Neumann boundary conditions

by Victor Boussange<sup>1,2</sup>, Sebastian Becker<sup>3</sup>, Arnulf Jentzen<sup>4,5</sup>,  
Benno Kuckuck<sup>5</sup>, and Loïc Pellissier<sup>1,2</sup>

<sup>1</sup> Unit of Land Change Science, Swiss Federal Research Institute  
for Forest, Snow and Landscape (WSL), Switzerland

<sup>2</sup> Landscape Ecology, Institute of Terrestrial Ecosystems,  
Department of Environmental Systems Science, ETH Zürich,

<sup>3</sup> Risklab, Department of Mathematics, ETH Zürich,

<sup>4</sup> School of Data Science and Shenzhen Research Institute of Big Data,  
The Chinese University of Hong Kong, Shenzhen, China,

<sup>5</sup> Applied Mathematics: Institute for Analysis and Numerics,  
Faculty of Mathematics and Computer Science, University of Münster,

arXiv:2205.03672

Revision requested from Journal of Partial Differential Equations and Applications

**”** *The key to growth is the introduction of higher dimensions of consciousness into our awareness.*

— Lao Tzu (c. 4th century BCE)

Nonlinear partial differential equations (PDEs) are used to model dynamical processes in a large number of scientific fields, ranging from finance to biology. In many applications standard local models are not sufficient to accurately account for certain non-local phenomena such as, e.g., interactions at a distance. In order to properly capture these phenomena non-local nonlinear PDE models are frequently employed in the literature. In this article we propose two numerical methods based on machine learning and on Picard iterations, respectively, to approximately solve non-local nonlinear PDEs. The proposed machine learning-based method is an extended variant of a deep learning-based splitting-up type approximation method previously introduced in the literature and utilizes neural networks to provide approximate solutions on a subset of the spatial domain of the solution. The Picard iterations-based method is an extended variant of the so-called *full history recursive multilevel Picard* approximation scheme previously introduced in the literature and provides an approximate solution for a single point of the domain. Both methods are mesh-free and allow non-local nonlinear PDEs with Neumann boundary conditions to be solved in high dimensions. In the two methods, the numerical difficulties arising due to the dimensionality of the PDEs are avoided by (i) using the correspondence between the expected trajectory of reflected stochastic processes and the solution of PDEs (given by the Feynman–Kac formula) and by (ii) using a plain vanilla Monte Carlo integration to handle the non-local term. We evaluate the performance of the two methods on five different PDEs arising in physics and biology. In all cases, the methods yield good results in up to 10 dimensions with short run times. Our work extends recently developed methods to overcome the curse of dimensionality in solving PDEs.

## 5.1 Introduction

In this article, we derive numerical schemes to approximately solve high-dimensional non-local nonlinear partial differential equations (PDEs) with Neumann boundary conditions. Such PDEs have been used to describe a variety of processes in physics, engineering, finance, and biology, but can generally not be solved analytically, requiring numerical methods to provide approximate solutions. However, traditional numerical methods are for the most part computationally infeasible for high-dimensional problems, calling for the development of novel approximation methods.

The need for solving non-local nonlinear PDEs has been expressed in various fields as they provide a more general description of the dynamical systems than their local counterparts (Kavallaris and Suzuki, 2018; D’Elia et al., 2020; Sunderasan, 2020). In physics and engineering, non-local nonlinear PDEs are found, e.g., in models of Ohmic heating production (Lacey, 1995), in the investigation of the fully turbulent behavior of real flows (Caglioti et al., 1995), in phase field models allowing non-local interactions (Barone et al., 1971; Gajewski and Zacharias, 2003; Coleman, 1994; Hairer and Shen, 2016), or in phase transition models with conservation of mass (Rubinstein and Sternberg, 1992; Stoleriu, 2011); see Kavallaris and Suzuki, 2018 for further references. In finance, non-local PDEs are used, e.g., in jump-diffusion models for the pricing of derivatives where the dynamics of stock prices are described by stochastic processes experiencing large jumps (Merton, 1976; Chan, 1999; Kou, 2002; Abergel and Tachet, 2010; Benth et al., 2001; Sunderasan, 2020; Cruz and Ševčovič, 2020; Cont and Tankov, 2004). Penalty methods for pricing American put options such as in Kou’s jump-diffusion model (Huang et al., 2013; Gan et al., 2020), considering large investors where the agent policy affects the assets prices (Amadori, 2003; Abergel and Tachet, 2010), or considering default risks (Pham, 2009; Henry-Labordère, 2012) can further introduce nonlinear terms in non-local PDEs. In economics, non-local nonlinear PDEs appear, e.g., in evolutionary game theory with the so-called replicator-mutator equation capturing continuous strategy spaces (Oechssler and Riedel, 2001; Kavallaris et al., 2017; Hamel et al., 2020; Alfaro and Carles, 2017; Alfaro and Veruete, 2019) or in growth models where consumption is non-local (Banerjee et al., 2021). In biology, non-local nonlinear PDEs are used, e.g., to model processes determining the interaction and evolution of organisms. Examples include models of morphogenesis and cancer evolution (Lorz et al., 2013; Chen et al., 2020b; Villa et al., 2021), models of gene regulatory networks (Pájaro et al., 2017), population genetics models with the non-local Fisher–Kolmogorov–Petrovsky–Piskunov (Fisher–KPP) equations (FISHER, 1937; Hamel and Nadirashvili, 2001; Bian et al., 2017; Perthame and Génieys, 2007; Berestycki et al., 2009; Houchmandzadeh and Vallade, 2017; Wang et al., 2021), and quantitative genetics models where populations are structured on a phenotypic and/or a geographical space (Burger and Hofbauer, 1994; Génieys et al., 2006; Berestycki et al., 2016; Nordbotten and Stenseth, 2016; Nordbotten et al., 2018; Roques and Bonnefon, 2016; Doebeli and Ispolatov, 2010a; Nordbotten et al., 2020). In such models, Neumann boundary conditions are used, e.g., to model the effect of the borders of the geographical domain on the movement of the organisms.

Real world systems such as those just mentioned may be of considerable complexity and accurately capturing the dynamics of these systems may require models of high dimensionality (Doebeli and Ispolatov, 2010a), leading to complications in obtaining numerical approximations. For example, the number of dimensions of the PDEs may correspond in finance to the number of financial assets (such as stocks, commodities, exchange rates, and interest rates) in the involved portfolio; in evolutionary dynamics, to the dimension of the strategy space; and in biology, to the number of genes modelled (Pájaro et al., 2017) or to the dimension of the geographical or the phenotypic space over which the organisms are structured. Standard approximation methods for PDEs such as finite difference approximation methods, finite element methods, spectral Galerkin approximation methods, and sparse grid approximation methods all suffer from the so called *curse of dimensionality* (Bellman, 2010), meaning that their computational costs increase exponentially in the number of dimensions of the PDE under consideration.

Numerical methods exploiting stochastic representations of the solutions of PDEs can in some cases overcome the curse of dimensionality. Specifically, simple Monte Carlo averages of the associated stochastic processes have been proposed a long time ago to solve high-dimensional linear PDEs, such as, e.g., Black–Scholes and Kolmogorov PDEs (Metropolis and Ulam, 1949; Bauer, 1958). Recently, two novel classes of methods have proved successful in dealing with high-dimensional nonlinear PDEs, namely deep learning-based and full history recursive multilevel Picard approximation methods (in the following we will abbreviate *full history recursive multilevel Picard* by MLP). The explosive success of deep learning in recent years across a wide range of applications (LeCun et al., 2015) has inspired a variety of neural network-based approximation methods for high-dimensional PDEs; see Beck et al., 2020 for an overview. One class of such methods is based on reformulating the PDE as a stochastic learning problem through the Feynman–Kac formula (E et al., 2017; Han et al., 2018; Beck et al., 2021b). In particular, the *deep splitting* scheme introduced in (Beck et al., 2021a) relies on splitting the differential operator into a linear and a nonlinear part and in that sense belongs to the class of splitting-up methods (Cox and Neerven, 2013; Gyöngy and Krylov, 2003; Hochbruck and Ostermann, 2005). The PDE approximation problem is then decomposed along the time axis into a sequence of separate learning problems. The deep splitting approximation scheme has proved capable of computing reasonable approximations to the solutions of nonlinear PDEs in up to 10000 dimensions. On the other hand, the MLP approximation method, introduced in (E et al., 2019a; Hutzenthaler et al., 2020; E et al., 2021b), utilizes the Feynman–Kac formula to reformulate the PDE problem as a fixed point equation. It further reduces the complexity of the numerical approximation of the time integral through a multilevel Monte Carlo approach. However, neither the deep splitting nor the MLP method can, until now, account for non-localness and Neumann boundary conditions.

The goal of this article is to overcome these limitations and thus we generalize the deep splitting method and the MLP approximation method to approximately solve non-local nonlinear PDEs with Neumann boundary conditions. We handle the non-local term by a plain vanilla Monte Carlo integration and address Neumann boundary conditions by constructing reflected stochastic processes. While the MLP method can, in one run, only provide an approximate solution at a single point  $x \in D$  of the spatial domain  $D \subseteq \mathbb{R}^d$

where  $d \in \mathbb{N} = \{1, 2, \dots\}$ , the machine learning-based method can in principle provide an approximate solution on a full subset of the spatial domain  $D$  (however, cf., e.g., (Heinrich, 1998; Heinrich and Sindambiwe, 1999; Grohs and Voigtlaender, 2021) for results on limitations on the performance of such approximation schemes). We use both methods to solve five non-local nonlinear PDEs arising in models from biology and physics and cross-validate the results of the simulations. We manage to solve the non-local nonlinear PDEs with reasonable accuracy in up to 10 dimensions.

For an account of classical numerical methods for solving non-local PDEs, such as finite differences, finite elements, and spectral methods, we refer the reader to the recent survey of D'Elia et al., 2020. Several machine-learning based schemes for solving non-local PDEs can also be found in the literature. In particular, the *physics-informed neural network* and *deep Galerkin* approaches (Raissi et al., 2019b; Sirignano and Spiliopoulos, 2018), based on representing an approximation of the whole solution of the PDE as a neural network and using automatic differentiation to do a least-squares minimization of the residual of the PDE, have been extended to fractional PDEs and other non-local PDEs (Pang et al., 2019; Lu et al., 2021; Guo et al., 2022; Al-Aradi et al., 2019; Yuan et al., 2022). While some of these approaches use classical methods susceptible to the curse of dimensionality for the non-local part (Pang et al., 2019; Lu et al., 2021), mesh-free methods suitable for high-dimensional problems have also been investigated (Guo et al., 2022; Al-Aradi et al., 2019; Yuan et al., 2022).

The literature also contains approaches that are more closely related to the machine learning-based algorithm presented here. Frey & Köck (Frey and Köck, 2021; Frey and Köck, 2022) propose an approximation method for non-local semilinear parabolic PDEs with Dirichlet boundary conditions based on and extending the deep splitting method in (Beck et al., 2021a) and carry out numerical simulations for example PDEs in up to 4 dimensions. Castro, 2021 proposes a numerical scheme for approximately solving non-local nonlinear PDEs based on Huré et al., 2020 and proves convergence results for this scheme. Finally, Gonon and Schwab, 2021 provide theoretical results showing that neural networks with ReLU activation functions have sufficient expressive power to approximate solutions of certain high-dimensional non-local linear PDEs without the curse of dimensionality.

There is a more extensive literature on machine learning-based methods for approximately solving standard PDEs without non-local terms but with various boundary conditions, going back to early works by Lagaris et al., 1998; Lagaris et al., 2000 (see also McFall and Mahan, 2009), which employed a grid-based method based on least-squares minimization of the residual and shallow neural networks to solve low-dimensional ODEs and PDEs with Dirichlet, Neumann, and mixed boundary conditions. More recently, approximation methods for PDEs with Neumann (and other) boundary conditions have been proposed using, e.g., physics-informed neural networks (Lu et al., 2021; Sukumar and Srivastava, 2022; Wang and Perdikaris, 2020), the *deep Ritz* method (based on a variational formulation of certain elliptic PDEs) (E and Yu, 2018; Liao and Ming, 2021; Chen et al., 2020a), or adversarial networks (Zang et al., 2020).

The remainder of this article is organized as follows. Section 5.2 discusses a special case of the proposed machine learning-based method, in order to provide a readily comprehensible exposition of the key ideas of the method. Section 5.3 discusses the general case, which is

flexible enough to cover a larger class of PDEs and to allow more sophisticated optimization methods. Section 5.4 presents our extension of the MLP approximation method to non-local nonlinear PDEs, which we use to obtain reference solutions in Section 5.5. Section 5.5 provides numerical simulations for five concrete examples of (non-local) nonlinear PDEs.

## 5.2 Machine learning-based approximation method in a special case

In this section, we present in Lemma 5.2.2 in Section 5.2.3 below a simplified version of our general machine learning-based algorithm for approximating solutions of non-local nonlinear PDEs with Neumann boundary conditions proposed in Section 5.3 below. This simplified version applies to a smaller class of non-local heat PDEs, specified in Section 5.2.1 below. In Section 5.2.2 we introduce some notation related to the reflection of straight lines on the boundaries of a suitable subset  $D \subseteq \mathbb{R}^d$  where  $d \in \mathbb{N}$ , which will be used to describe time-discrete reflected stochastic processes that are employed in our approximations throughout the rest of the article. The simplified algorithm described in Section 5.2.3 below is limited to using neural networks of a particular architecture that are trained using plain vanilla stochastic gradient descent, whereas the full version proposed in Lemma 5.3.1 in Section 5.3.2 below is formulated in such a way that it encompasses a wide array of neural network architectures and more sophisticated training methods, in particular Adam optimization, minibatches, and batch normalization. Stripping away some of these more intricate aspects of the full algorithm is intended to exhibit more acutely the central ideas in the proposed approximation method.

The simplified algorithm described in this section as well as the more general version proposed in Lemma 5.3.1 in Section 5.3.2 below are based on the deep splitting method introduced in Beck et al., 2021a, which combines operator splitting with a previous deep learning-based approximation method for Kolmogorov PDEs (Beck et al., 2021b); see also Beck et al., 2020, Sections 2 and 3 for an exposition of these methods.

### 5.2.1 Partial differential equations (PDEs) under consideration

Let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ , let  $\mathbb{D} \subseteq \mathbb{R}^d$  be a closed set with sufficiently smooth boundary  $\partial_{\mathbb{D}}$ , let  $\mathbf{n}: \partial_{\mathbb{D}} \rightarrow \mathbb{R}^d$  be an outer unit normal vector field associated to  $\mathbb{D}$ , let  $g \in C(D, \mathbb{R})$ , let  $\nu_x: \mathcal{B}(\mathbb{D}) \rightarrow [0, 1]$ ,  $x \in \mathbb{D}$ , be probability measures, let  $f: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  be measurable, let  $u = (u(t, x))_{(t,x) \in [0,T] \times \mathbb{D}} \in C^{1,2}([0, T] \times \mathbb{D}, \mathbb{R})$  have at most polynomially growing partial derivatives, assume<sup>1</sup> for every  $t \in (0, T]$ ,  $x \in \partial_{\mathbb{D}}$  that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$ , and assume for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$  that  $u(0, x) = g(x)$ ,  $\int_{\mathbb{D}} |f(u(t, x), u(t, \mathbf{x}))| \nu_x(d\mathbf{x}) < \infty$ , and

$$(\frac{\partial}{\partial t} u)(t, x) = (\Delta_x u)(t, x) + \int_{\mathbb{D}} f(u(t, x), u(t, \mathbf{x})) \nu_x(d\mathbf{x}). \quad (5.1)$$

---

<sup>1</sup>Throughout this article we denote by  $\langle \cdot, \cdot \rangle: (\bigcup_{n \in \mathbb{N}} (\mathbb{R}^n \times \mathbb{R}^n)) \rightarrow \mathbb{R}$  and  $\|\cdot\|: (\bigcup_{n \in \mathbb{N}} \mathbb{R}^n) \rightarrow \mathbb{R}$  the functions which satisfy for all  $n \in \mathbb{N}$ ,  $v = (v_1, \dots, v_n)$ ,  $w = (w_1, \dots, w_n) \in \mathbb{R}^n$  that  $\langle v, w \rangle = \sum_{i=1}^n v_i w_i$  and  $\|v\| = \sqrt{\langle v, v \rangle} = [\sum_{i=1}^n |v_i|^2]^{1/2}$ .

Our goal in this section is to approximately calculate under suitable hypotheses the solution  $u: [0, T] \times \mathbb{D} \rightarrow \mathbb{R}$  of the PDE in (5.1).

## 5.2.2 Reflection principle for the simulation of time discrete reflected processes

**Framework 5.2.1** (Reflection principle for the simulation of time discrete reflected processes). Let  $d \in \mathbb{N}$ , let  $\mathbb{D} \subseteq \mathbb{R}^d$  be a closed set with sufficiently smooth boundary  $\partial_{\mathbb{D}}$ , let  $\mathbf{n}: \partial_{\mathbb{D}} \rightarrow \mathbb{R}^d$  be a suitable outer unit normal vector field associated to  $\mathbb{D}$ , let  $\mathbf{c}: (\mathbb{R}^d)^2 \rightarrow \mathbb{R}^d$  satisfy for every  $a, b \in \mathbb{R}^d$  that

$$\mathbf{c}(a, b) = a + [\inf(\{r \in [0, 1]: a + r(b - a) \notin \mathbb{D}\} \cup \{1\})](b - a), \quad (5.2)$$

let  $\mathcal{R}: (\mathbb{R}^d)^2 \rightarrow (\mathbb{R}^d)^2$  satisfy for every  $a, b \in \mathbb{R}^d$  that

$$\mathcal{R}(a, b) = \begin{cases} (a, b) & : \mathbf{c}(a, b) = a \\ (\mathbf{c}(a, b), b - 2\mathbf{n}(\mathbf{c}(a, b))\langle b - \mathbf{c}(a, b), \mathbf{n}(\mathbf{c}(a, b)) \rangle) & : \mathbf{c}(a, b) \notin \{a, b\} \\ (b, b) & : \mathbf{c}(a, b) = b, \end{cases} \quad (5.3)$$

let  $P: (\mathbb{R}^d)^2 \rightarrow \mathbb{R}^d$  satisfy for every  $a, b \in \mathbb{R}^d$  that  $P(a, b) = b$ , let  $\mathcal{R}_n: (\mathbb{R}^d)^2 \rightarrow (\mathbb{R}^d)^2$ ,  $n \in \mathbb{N}_0 = \{0\} \cup \mathbb{N}$ , satisfy for every  $n \in \mathbb{N}_0$ ,  $x, y \in \mathbb{R}^d$  that  $\mathcal{R}_0(x, y) = (x, y)$  and  $\mathcal{R}_{n+1}(x, y) = \mathcal{R}(\mathcal{R}_n(x, y))$ , and let  $R: (\mathbb{R}^d)^2 \rightarrow \mathbb{R}^d$  satisfy for every  $x, y \in \mathbb{R}^d$  that

$$R(x, y) = \lim_{n \rightarrow \infty} P(\mathcal{R}_n(x, y)). \quad (5.4)$$

## 5.2.3 Description of the proposed approximation method in a special case

**Framework 5.2.2** (Special case of the machine learning-based approximation method). Assume Lemma 5.2.1, let  $T, \gamma \in (0, \infty)$ ,  $N, M, K \in \mathbb{N}$ ,  $g \in C^2(\mathbb{R}^d, \mathbb{R})$ ,  $\mathfrak{d}, \mathfrak{h} \in \mathbb{N} \setminus \{1\}$ ,  $t_0, t_1, \dots, t_N \in [0, T]$  satisfy  $\mathfrak{d} = \mathfrak{h}(N + 1)d(d + 1)$  and

$$0 = t_0 < t_1 < \dots < t_N = T, \quad (5.5)$$

let  $\tau_0, \tau_1, \dots, \tau_N \in [0, T]$  satisfy for every  $n \in \{0, 1, \dots, N\}$  that  $\tau_n = T - t_{N-n}$ , let  $f: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  be measurable, let  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in [0, T]})$  be a filtered probability space, let  $\xi^m: \Omega \rightarrow \mathbb{R}^d$ ,  $m \in \mathbb{N}$ , be i.i.d.  $\mathcal{F}_0/\mathcal{B}(\mathbb{R}^d)$ -measurable random variables, let  $W^m: [0, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $m \in \mathbb{N}$ , be i.i.d. standard  $(\mathcal{F}_t)_{t \in [0, T]}$ -Brownian motions, for every  $m \in \mathbb{N}$  let  $\mathcal{Y}^m: \{0, 1, \dots, N\} \times \Omega \rightarrow \mathbb{R}^d$  be the stochastic process which satisfies for every  $n \in \{0, 1, \dots, N-1\}$  that  $\mathcal{Y}_0^m = \xi^m$  and

$$\mathcal{Y}_{n+1}^m = R(\mathcal{Y}_n^m, \mathcal{Y}_n^m + \sqrt{2}(W_{\tau_{n+1}}^m - W_{\tau_n}^m)), \quad (5.6)$$

let  $\mathcal{L}: \mathbb{R}^d \rightarrow \mathbb{R}^d$  satisfy for every  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  that

$$\mathcal{L}(x) = \left( \frac{\exp(x_1)}{\exp(x_1) + 1}, \dots, \frac{\exp(x_d)}{\exp(x_d) + 1} \right), \quad (5.7)$$

for every  $\theta = (\theta_1, \dots, \theta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$ ,  $k, l, v \in \mathbb{N}$  with  $v + l(k + 1) \leq \mathfrak{d}$  let  $A_{k,l}^{\theta,v}: \mathbb{R}^k \rightarrow \mathbb{R}^l$  satisfy for every  $x = (x_1, \dots, x_k) \in \mathbb{R}^k$  that

$$A_{k,l}^{\theta,v}(x) = \left( \theta_{v+kl+1} + \left[ \sum_{i=1}^k x_i \theta_{v+i} \right], \dots, \theta_{v+kl+l} + \left[ \sum_{i=1}^k x_i \theta_{v+(l-1)k+i} \right] \right), \quad (5.8)$$

let  $\mathbb{V}_n: \mathbb{R}^{\mathfrak{d}} \times \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $n \in \{0, 1, \dots, N\}$ , satisfy for every  $n \in \{1, 2, \dots, N\}$ ,  $\theta \in \mathbb{R}^{\mathfrak{d}}$ ,  $x \in \mathbb{R}^d$  that  $\mathbb{V}_0(\theta, x) = g(x)$  and

$$\begin{aligned} \mathbb{V}_n(\theta, x) &= \\ &\left( A_{d,1}^{\theta,(\mathfrak{h}n+\mathfrak{h}-1)d(d+1)} \circ \mathcal{L} \circ A_{d,d}^{\theta,(\mathfrak{h}n+\mathfrak{h}-2)d(d+1)} \circ \dots \circ \mathcal{L} \circ A_{d,d}^{\theta,(\mathfrak{h}n+1)d(d+1)} \circ \mathcal{L} \circ A_{d,d}^{\theta,\mathfrak{h}nd(d+1)} \right)(x), \end{aligned} \quad (5.9)$$

let  $\nu_x: \mathcal{B}(\mathbb{D}) \rightarrow [0, 1]$ ,  $x \in \mathbb{D}$ , be probability measures, for every  $x \in \mathbb{D}$  let  $Z_{x,k}^{n,m}: \Omega \rightarrow \mathbb{D}$ ,  $k, n, m \in \mathbb{N}$ , be i.i.d. random variables which satisfy for every  $A \in \mathcal{B}(\mathbb{D})$  that  $\mathbb{P}(Z_{x,1}^{1,1} \in A) = \nu_x(A)$ , let  $\Theta^n: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^{\mathfrak{d}}$ ,  $n \in \{0, 1, \dots, N\}$ , be stochastic processes, for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$  let  $\phi^{n,m}: \mathbb{R}^{\mathfrak{d}} \times \Omega \rightarrow \mathbb{R}$  satisfy for every  $\theta \in \mathbb{R}^{\mathfrak{d}}$ ,  $\omega \in \Omega$  that

$$\begin{aligned} \phi^{n,m}(\theta, \omega) &= \left[ \mathbb{V}_n(\theta, \mathcal{Y}_{N-n}^m(\omega)) - \mathbb{V}_{n-1}(\Theta_M^{n-1}(\omega), \mathcal{Y}_{N-n+1}^m(\omega)) \right. \\ &\quad \left. - \frac{(t_n - t_{n-1})}{K} \left[ \sum_{k=1}^K f(\mathbb{V}_{n-1}(\Theta_M^{n-1}(\omega), \mathcal{Y}_{N-n+1}^m(\omega)), \mathbb{V}_{n-1}(\Theta_M^{n-1}(\omega), Z_{\mathcal{Y}_{N-n+1}^m(\omega), k}^{n,m}(\omega))) \right]^2 \right], \end{aligned} \quad (5.10)$$

for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$  let  $\Phi^{n,m}: \mathbb{R}^{\mathfrak{d}} \times \Omega \rightarrow \mathbb{R}^{\mathfrak{d}}$  satisfy for every  $\theta \in \mathbb{R}^{\mathfrak{d}}$ ,  $\omega \in \Omega$  that  $\Phi^{n,m}(\theta, \omega) = (\nabla_{\theta} \phi^{n,m})(\theta, \omega)$ , and assume for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$  that

$$\Theta_m^n = \Theta_{m-1}^n - \gamma \Phi^{n,m}(\Theta_{m-1}^n). \quad (5.11)$$

As indicated in Section 5.2.1 above, the algorithm described in Lemma 5.2.2 computes an approximation for a solution of the PDE in Eq. (5.1), i.e., a function  $u \in C^{1,2}([0, T] \times D, \mathbb{R})$  which has at most polynomially growing derivatives, which satisfies for every  $t \in (0, T]$ ,  $x \in \partial_{\mathbb{D}}$  that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$  and which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$  that  $u(0, x) = g(x)$ ,  $\int_{\mathbb{D}} |f(u(t, x), u(t, \mathbf{x}))| \nu_x(d\mathbf{x}) < \infty$ , and

$$\left( \frac{\partial}{\partial t} u \right)(t, x) = (\Delta_x u)(t, x) + \int_{\mathbb{D}} f(u(t, x), u(t, \mathbf{x})) \nu_x(d\mathbf{x}). \quad (5.12)$$

Let us now add some explanatory comments on the objects and notations employed in Lemma 5.2.2 above. The algorithm in Lemma 5.2.2 decomposes the time interval  $[0, T]$  into  $N$  subintervals at the times  $t_0, t_1, t_2, \dots, t_N \in [0, T]$  (cf. Eq. (5.5)). For every  $n \in \{1, 2, \dots, N\}$  we aim to approximate the function  $\mathbb{R}^d \ni x \mapsto u(t_n, x) \in \mathbb{R}$  by a suitable (realization function of a) fully-connected feedforward neural network. Each of these

neural networks is an alternating composition of  $\mathfrak{h} - 1$  affine linear functions from  $\mathbb{R}^d$  to  $\mathbb{R}^d$  (where we think of  $\mathfrak{h} \in \mathbb{N} \setminus \{1\}$  as the *length* or *depth* of the neural network),  $\mathfrak{h} - 1$  instances of a  $d$ -dimensional version of the standard logistic function and finally an affine linear function from  $\mathbb{R}^d$  to  $\mathbb{R}$ . Every such neural network can be specified by means of  $(\mathfrak{h} - 1)(d^2 + d) + d + 1 \leq \mathfrak{h}d(d + 1)$  real parameters and so  $N + 1$  of these neural networks can be specified by a parameter vector of length  $\mathfrak{d} = \mathfrak{h}(N + 1)d(d + 1) \in \mathbb{N}$ . Note that  $\mathcal{L}: \mathbb{R}^d \rightarrow \mathbb{R}^d$  in Lemma 5.2.2 above denotes the  $d$ -dimensional version of the standard logistic function (cf. Eq. (5.7)) and for every  $k, l, v \in \mathbb{N}$ ,  $\theta \in \mathbb{R}^\mathfrak{d}$  with  $v + kl + l \leq \mathfrak{d}$  the function  $A_{k,l}^{\theta,v}: \mathbb{R}^k \rightarrow \mathbb{R}^l$  in Lemma 5.2.2 denotes an affine linear function specified by means of the parameters  $v + 1, v + 2, \dots, v + kl + l$  (cf. Eq. (5.8)). Furthermore, observe that for every  $n \in \{1, 2, \dots, N\}$ ,  $\theta \in \mathbb{R}^\mathfrak{d}$  the function

$$\mathbb{R}^d \ni x \mapsto \mathbb{V}_n(\theta, x) \in \mathbb{R} \quad (5.13)$$

denotes a neural network specified by means of the parameters  $\mathfrak{h}nd(d + 1) + 1, \mathfrak{h}nd(d + 1) + 2, \dots, (\mathfrak{h}n + \mathfrak{h} - 1)d(d + 1) + d + 1$ .

The goal of the optimization algorithm in Lemma 5.2.2 above is to find a suitable parameter vector  $\theta \in \mathbb{R}^\mathfrak{d}$  such that for every  $n \in \{1, 2, \dots, N\}$  the neural network  $\mathbb{R}^d \ni x \mapsto \mathbb{V}_n(\theta, x) \in \mathbb{R}$  is a good approximation for the solution  $\mathbb{R}^d \ni x \mapsto u(t_n, x) \in \mathbb{R}$  to the PDE in Eq. (5.12) at time  $t_n$ . This is done by performing successively for each  $n \in \{1, 2, \dots, N\}$  a plain vanilla stochastic gradient descent (SGD) optimization on a suitable loss function (cf. Eq. (5.11)).

Observe that for every  $n \in \{1, 2, \dots, N\}$  the stochastic process  $\Theta^n: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^\mathfrak{d}$  describes the successive estimates computed by the SGD algorithm for the parameter vector that represents (via  $\mathbb{V}_n: \mathbb{R}^\mathfrak{d} \times \mathbb{R}^d \rightarrow \mathbb{R}$ ) a suitable approximation to the solution  $\mathbb{R}^d \ni x \mapsto u(t_n, x) \in \mathbb{R}$  of the PDE in Eq. (5.12) at time  $t_n$ . Next note that  $M \in \mathbb{N}$  in Lemma 5.2.2 above denotes the number of gradient descent steps taken for each  $n \in \{1, 2, \dots, N\}$  and that  $\gamma \in (0, \infty)$  denotes the learning rate employed in the SGD algorithm. Moreover, observe that for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \{1, 2, \dots, M\}$  the function  $\phi^{n,m}: \mathbb{R}^\mathfrak{d} \times \Omega \rightarrow \mathbb{R}$  denotes the loss function employed in the  $m$ th gradient descent step during the approximation of the solution of the PDE in Eq. (5.12) at time  $t_n$  (cf. Eq. (5.10)). The loss functions employ a family of i.i.d. time-discrete stochastic processes  $\mathcal{Y}^m: \{0, 1, \dots, N\} \times \Omega \rightarrow \mathbb{R}^d$ ,  $m \in \mathbb{N}$ , which we think of as discretizations of suitable reflected Brownian motions (cf. Eq. (5.6)). In addition, for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \{1, 2, \dots, M\}$ ,  $x \in D$  the loss function  $\phi^{n,m}: \mathbb{R}^\mathfrak{d} \times \Omega \rightarrow \mathbb{R}$  employs a family of i.i.d. random variables  $Z_{x,k}^{n,m}: \Omega \rightarrow D$ ,  $k \in \mathbb{N}$ , which are used for the Monte Carlo approximation of the non-local term in the PDE in Eq. (5.12) whose solution we are trying to approximate. The number of samples used in these Monte Carlo approximations is denoted by  $K \in \mathbb{N}$  in Lemma 5.2.2 above.

Finally, for sufficiently large  $N, M, K \in \mathbb{N}$  and sufficiently small  $\gamma \in (0, \infty)$  the algorithm in Lemma 5.2.2 above yields for every  $n \in \{1, 2, \dots, N\}$  a (random) parameter vector  $\Theta_M^n: \Omega \rightarrow \mathbb{R}^\mathfrak{d}$  which represents a function  $\mathbb{R}^d \times \Omega \ni (x, \omega) \mapsto \mathbb{V}_n(\Theta_M^n(\omega), x) \in \mathbb{R}$  that we think of as providing for every  $x \in \mathbb{D}$  a suitable approximation

$$\mathbb{V}_n(\Theta_M^n, x) \approx u(t_n, x). \quad (5.14)$$

## 5.3 Machine learning-based approximation method in the general case

In this section we describe in Lemma 5.3.1 in Section 5.3.2 below the full version of our deep learning-based method for approximating solutions of non-local nonlinear PDEs with Neumann boundary conditions (see Section 5.3.1 for a description of the class of PDEs our approximation method applies to), which generalizes the algorithm introduced in Lemma 5.2.2 in Section 5.2.3 above and which we apply in Section 5.5 below to several examples of non-local nonlinear PDEs.

### 5.3.1 PDEs under consideration

Let  $T \in (0, \infty)$ ,  $d \in \mathbb{N}$ , let  $\mathbb{D} \subseteq \mathbb{R}^d$  be a closed set with sufficiently smooth boundary  $\partial_{\mathbb{D}}$ , let  $\mathbf{n}: \partial_{\mathbb{D}} \rightarrow \mathbb{R}^d$  be an outer unit normal vector field associated to  $\mathbb{D}$ , let  $g: \mathbb{D} \rightarrow \mathbb{R}$ ,  $\mu: \mathbb{D} \rightarrow \mathbb{R}^d$ , and  $\sigma: \mathbb{D} \rightarrow \mathbb{R}^{d \times d}$  be continuous, let  $\nu_x: \mathcal{B}(\mathbb{D}) \rightarrow [0, 1]$ ,  $x \in \mathbb{D}$ , be probability measures, let  $f: [0, T] \times \mathbb{D} \times \mathbb{D} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  be measurable, let  $u = (u(t, x))_{(t,x) \in [0,T] \times \mathbb{D}} \in C^{1,2}([0, T] \times \mathbb{D}, \mathbb{R})$  have at most polynomially growing partial derivatives, assume for every  $t \in [0, T]$ ,  $x \in \partial_{\mathbb{D}}$  that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$ , and assume for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$  that  $u(0, x) = g(x)$ ,  $\int_{\mathbb{D}} |f(t, x, \mathbf{x}, u(t, x), u(t, \mathbf{x}))| \nu_x(d\mathbf{x}) < \infty$ , and

$$\begin{aligned} \left( \frac{\partial}{\partial t} u \right)(t, x) &= \int_{\mathbb{D}} f(t, x, \mathbf{x}, u(t, x), u(t, \mathbf{x})) \nu_x(d\mathbf{x}) \\ &\quad + \langle \mu(x), (\nabla_x u)(t, x) \rangle + \frac{1}{2} \operatorname{Trace}(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t, x)). \end{aligned} \tag{5.15}$$

Our goal is to approximately calculate under suitable hypotheses the solution  $u: [0, T] \times \mathbb{D} \rightarrow \mathbb{R}$  of the PDE in (5.15).

### 5.3.2 Description of the proposed approximation method in the general case

**Framework 5.3.1** (General case of the machine learning-based approximation method). Assume Lemma 5.2.1, let  $T \in (0, \infty)$ ,  $N, \varrho, \mathfrak{d}, \varsigma \in \mathbb{N}$ ,  $(M_n)_{n \in \mathbb{N}_0} \subseteq \mathbb{N}$ ,  $(K_n)_{n \in \mathbb{N}} \subseteq \mathbb{N}$ ,  $(J_m)_{m \in \mathbb{N}} \subseteq \mathbb{N}$ ,  $t_0, t_1, \dots, t_N \in [0, T]$  satisfy

$$0 = t_0 < t_1 < \dots < t_N = T, \tag{5.16}$$

let  $\tau_0, \tau_1, \dots, \tau_N \in [0, T]$  satisfy for every  $n \in \{0, 1, \dots, N\}$  that  $\tau_n = T - t_{N-n}$ , let  $\nu_x: \mathcal{B}(\mathbb{D}) \rightarrow [0, 1]$ ,  $x \in \mathbb{D}$ , be probability measures, for every  $x \in \mathbb{D}$  let  $Z_{x,k}^{n,m,j}: \Omega \rightarrow \mathbb{D}$ ,  $k, n, m, j \in \mathbb{N}$ , be i.i.d. random variables which satisfy for every  $A \in \mathcal{B}(\mathbb{D})$  that  $\mathbb{P}(Z_{x,1}^{1,1,1} \in A) = \nu_x(A)$ , let  $f: [0, T] \times \mathbb{D} \times \mathbb{D} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  be measurable, let  $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in [0, T]})$  be a filtered probability space, for every  $n \in \{1, 2, \dots, N\}$  let  $W^{n,m,j}: [0, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $m, j \in \mathbb{N}$ , be i.i.d. standard  $(\mathcal{F}_t)_{t \in [0, T]}$ -Brownian motions, for every  $n \in \{1, 2, \dots, N\}$  let  $\xi^{n,m,j}: \Omega \rightarrow \mathbb{R}^d$ ,  $m, j \in \mathbb{N}$ , be i.i.d.  $\mathcal{F}_0/\mathcal{B}(\mathbb{R}^d)$ -measurable random variables, let  $H: [0, T]^2 \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a function, for every  $j \in \mathbb{N}$ ,  $s \in \mathbb{R}^\varsigma$ ,  $n \in \{0, 1, \dots, N\}$  let  $\mathbb{V}_n^{j,s}: \mathbb{R}^\mathfrak{d} \times \mathbb{R}^d \rightarrow \mathbb{R}$  be a function, for

every  $n \in \{1, 2, \dots, N\}$ ,  $m, j \in \mathbb{N}$  let  $\mathcal{Y}^{n,m,j}: \{0, 1, \dots, N\} \times \Omega \rightarrow \mathbb{R}^d$  be a stochastic process which satisfies for every  $k \in \{0, 1, \dots, N-1\}$  that  $\mathcal{Y}_0^{n,m,j} = \xi^{n,m,j}$  and

$$\mathcal{Y}_{k+1}^{n,m,j} = H(\tau_{k+1}, \tau_k, \mathcal{Y}_k^{n,m,j}, W_{\tau_{k+1}}^{n,m,j} - W_{\tau_k}^{n,m,j}), \quad (5.17)$$

let  $\Theta^n: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^\vartheta$ ,  $n \in \{0, 1, \dots, N\}$ , be stochastic processes, for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$ ,  $s \in \mathbb{R}^s$  let  $\phi^{n,m,s}: \mathbb{R}^\vartheta \times \Omega \rightarrow \mathbb{R}$  satisfy for every  $\theta \in \mathbb{R}^\vartheta$ ,  $\omega \in \Omega$  that

$$\begin{aligned} \phi^{n,m,s}(\theta, \omega) &= \frac{1}{J_m} \sum_{j=1}^{J_m} \left[ \mathbb{V}_n^{j,s}(\theta, \mathcal{Y}_{N-n}^{n,m,j}(\omega)) - \mathbb{V}_{n-1}^{j,s}(\Theta_{M_{n-1}}^{n-1}(\omega), \mathcal{Y}_{N-n+1}^{n,m,j}(\omega)) \right. \\ &\quad \left. - \frac{(t_n - t_{n-1})}{K_n} \left[ \sum_{k=1}^{K_n} f(t_{n-1}, \mathcal{Y}_{N-n+1}^{n,m,j}(\omega), Z_{\mathcal{Y}_{N-n+1}^{n,m,j}(\omega), k}^{n,m,j}(\omega), \right. \right. \\ &\quad \left. \left. \mathbb{V}_{n-1}^{j,s}(\Theta_{M_{n-1}}^{n-1}(\omega), \mathcal{Y}_{N-n+1}^{n,m,j}(\omega)), \mathbb{V}_{n-1}^{j,s}(\Theta_{M_{n-1}}^{n-1}(\omega), Z_{\mathcal{Y}_{N-n+1}^{n,m,j}(\omega), k}^{n,m,j}(\omega)) \right] \right]^2, \end{aligned} \quad (5.18)$$

for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$ ,  $s \in \mathbb{R}^s$  let  $\Phi^{n,m,s}: \mathbb{R}^\vartheta \times \Omega \rightarrow \mathbb{R}^\vartheta$  satisfy for every  $\omega \in \Omega$ ,  $\theta \in \{\vartheta \in \mathbb{R}^\vartheta: (\mathbb{R}^\vartheta \ni \eta \mapsto \phi^{n,m,s}(\eta, \omega) \in \mathbb{R}) \text{ is differentiable at } \vartheta\}$  that

$$\Phi^{n,m,s}(\theta, \omega) = (\nabla_\theta \phi^{n,m,s})(\theta, \omega), \quad (5.19)$$

let  $\mathcal{S}^n: \mathbb{R}^s \times \mathbb{R}^\vartheta \times (\mathbb{R}^d)^{\{0,1,\dots,N\} \times \mathbb{N}} \rightarrow \mathbb{R}^s$ ,  $n \in \{1, 2, \dots, N\}$ , be functions, for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$  let  $\psi_m^n: \mathbb{R}^\varrho \rightarrow \mathbb{R}^\vartheta$  and  $\Psi_m^n: \mathbb{R}^\varrho \times \mathbb{R}^\vartheta \rightarrow \mathbb{R}^\varrho$  be functions, and for every  $n \in \{1, 2, \dots, N\}$  let  $\mathbb{S}^n: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^s$  and  $\Xi^n: \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^\varrho$  be stochastic processes which satisfy for every  $m \in \mathbb{N}$  that

$$\mathbb{S}_m^n = \mathcal{S}^n(\mathbb{S}_{m-1}^n, \Theta_{m-1}^n, (\mathcal{Y}_k^{n,m,i})_{(k,i) \in \{0,1,\dots,N\} \times \mathbb{N}}), \quad (5.20)$$

$$\Xi_m^n = \Psi_m^n(\Xi_{m-1}^n, \Phi^{n,m,\mathbb{S}_m^n}(\Theta_{m-1}^n)), \quad \text{and} \quad \Theta_m^n = \Theta_{m-1}^n - \psi_m^n(\Xi_m^n). \quad (5.21)$$

In the setting of Lemma 5.3.1 above we think under suitable hypotheses for sufficiently large  $N \in \mathbb{N}$ , sufficiently large  $(M_n)_{n \in \mathbb{N}_0} \subseteq \mathbb{N}$ , sufficiently large  $(K_n)_{n \in \mathbb{N}} \subseteq \mathbb{N}$ , every  $n \in \{0, 1, \dots, N\}$ , and every  $x \in \mathbb{D}$  of  $\mathbb{V}_n^{1,\mathbb{S}_M^n}(\Theta_{M_n}^n, x): \Omega \rightarrow \mathbb{R}$  as a suitable approximation

$$\mathbb{V}_n^{1,\mathbb{S}_M^n}(\Theta_{M_n}^n, x) \approx u(t_n, x) \quad (5.22)$$

of  $u(t_n, x)$  where  $u = (u(t, x))_{(t,x) \in [0,T] \times \mathbb{R}^d} \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$  is a function with at most polynomially growing derivatives which satisfies for every  $t \in (0, T]$ ,  $x \in \partial_{\mathbb{D}}$  that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$  and which satisfies for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$  that  $u(0, x) = g(x)$ ,  $\int_{\mathbb{D}} |f(t, x, \mathbf{x}, u(t, x), u(t, \mathbf{x}))| \nu_x(d\mathbf{x}) < \infty$ , and

$$\begin{aligned} \left( \frac{\partial}{\partial t} u \right)(t, x) &= \int_{\mathbb{D}} f(t, x, \mathbf{x}, u(t, x), u(t, \mathbf{x})) \nu_x(d\mathbf{x}) \\ &\quad + \langle \mu(x), (\nabla_x u)(t, x) \rangle + \frac{1}{2} \operatorname{Trace}(\sigma(x)[\sigma(x)]^*(\operatorname{Hess}_x u)(t, x)) \end{aligned} \quad (5.23)$$

(cf. (5.15)). Compared to the simplified algorithm in Lemma 5.2.2 above, the major new elements introduced in Lemma 5.3.1 are the following:

- (a) The numbers of gradient descent steps taken to compute approximations for the solution of the PDE at the times  $t_n$ ,  $n \in \{1, 2, \dots, N\}$ , are allowed to vary with  $n$ , and so are specified by a sequence  $(M_n)_{n \in \mathbb{N}_0} \subseteq \mathbb{N}$  in Lemma 5.3.1 above.
- (b) The numbers of samples used for the Monte Carlo approximation of the non-local term in the approximation for the solution of the PDE at the times  $t_n$ ,  $n \in \{1, 2, \dots, N\}$ , are allowed to vary with  $n$ , and so are specified by a sequence  $(K_n)_{n \in \mathbb{N}_0} \subseteq \mathbb{N}$  in Lemma 5.3.1 above.
- (c) The approximating functions  $\mathbb{V}_n^{j,s}$ ,  $(j, s, n) \in \mathbb{N} \times \mathbb{R}^s \times \{0, 1, \dots, N\}$ , in Lemma 5.3.1 above are not specified concretely in order to allow for a variety of neural network architectures. For the concrete choice of these functions employed in our numerical simulations, we refer the reader to Section 5.5.
- (d) For every  $m \in \{1, 2, \dots, M\}$  the loss function used in the  $m$ th gradient descent step may be computed using a minibatch of samples instead of just one sample (cf. Eq. (5.18)). The sizes of these minibatches are specified by a sequence  $(J_m)_{m \in \mathbb{N}} \subseteq \mathbb{N}$ .
- (e) Compared to Lemma 5.2.2 above, the more general form of the PDEs considered in this section (cf. Eq. (5.23)) requires more flexibility in the definition of the time-discrete stochastic processes  $\mathcal{Y}^{n,m,j} : \{0, 1, \dots, N\} \times \Omega \rightarrow \mathbb{R}^d$ ,  $(n, m, j) \in \{1, 2, \dots, N\} \times \mathbb{N} \times \mathbb{N}$ , which are specified in Lemma 5.3.1 above in terms of the Brownian motions  $W^{n,m,j} : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $(n, m, j) \in \{1, 2, \dots, N\} \times \mathbb{N} \times \mathbb{N}$ , via a function  $H : [0, T]^2 \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  (cf. Eq. (5.17)). We refer the reader to Eq. (5.44) in Section 5.5.1 below, Eq. (5.46) in Section 5.5.2 below, Eq. (5.48) in Section 5.5.3 below, Eq. (5.50) in Section 5.5.4 below, and Eq. (5.72) in Section 5.5.5 below for concrete choices of  $H$  in the approximation of various example PDEs.
- (f) For every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$  the optimization step in Eq. (5.21) in Lemma 5.3.1 above is specified generically in terms of the functions  $\psi_m^n : \mathbb{R}^\varrho \rightarrow \mathbb{R}^\vartheta$  and  $\Psi_m^n : \mathbb{R}^\varrho \times \mathbb{R}^\vartheta \rightarrow \mathbb{R}^\varrho$  and the random variable  $\Xi_m^n : \Omega \rightarrow \mathbb{R}^\varrho$ . This generic formulation covers a variety of SGD based optimization algorithms such as Adagrad (Duchi et al., 2011), RMSprop, or Adam (Kingma and Ba, 2014). For example, in order to implement the Adam optimization algorithm, for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$  the random variable  $\Xi_m^n$  can be used to hold suitable first and second moment estimates (see Eq. (5.42) and Eq. (5.43) in Section 5.5 below for the concrete specification of these functions implementing the Adam optimization algorithm).
- (g) The processes  $\mathbb{S}^n : \mathbb{N}_0 \times \Omega \rightarrow \mathbb{R}^s$ ,  $n \in \{1, 2, \dots, N\}$ , and functions  $\mathcal{S}^n : \mathbb{R}^s \times \mathbb{R}^\vartheta \times (\mathbb{R}^d)^{\{0,1,\dots,N\} \times \mathbb{N}} \rightarrow \mathbb{R}^s$ ,  $n \in \{1, 2, \dots, N\}$ , in Lemma 5.3.1 above can be used to implement batch normalization; see Ioffe and Szegedy, 2015 for details. Loosely speaking, for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$  the random variable  $\mathbb{S}_m^n : \Omega \rightarrow \mathbb{R}^s$  then holds mean and variance estimates of the outputs of each layer of the approximating neural networks related to the minibatches that are used as inputs to the neural networks in computing the loss function at the corresponding gradient descent step.

## 5.4 Multilevel Picard approximation method for non-local PDEs

In this section we introduce in Lemma 5.4.1 in Section 5.4.1 below our extension of the full history recursive multilevel Picard approximation method for approximating solutions of non-local nonlinear PDEs with Neumann boundary conditions. The MLP method was first introduced in E et al., 2021b; Hutzenthaler et al., 2020 and later extended in a number of directions; see E et al., 2021a and Beck et al., 2020 for recent surveys. We also refer the reader to Becker et al., 2020 and E et al., 2019b for numerical simulations illustrating the performance of MLP methods across a range of example PDE problems.

In Section 5.4.2 below, we will specify five concrete examples of (non-local) nonlinear PDEs and describe how Lemma 5.4.1 can be specialized to compute approximate solutions to these example PDEs. These computations will be used in Section 5.5 to obtain reference values to compare the deep learning-based approximation method proposed in Section 5.3 above against.

### 5.4.1 Description of the proposed approximation method

**Framework 5.4.1** (Multilevel Picard approximation method). *Assume Lemma 5.2.1, let  $c, T \in (0, \infty)$ ,  $\mathfrak{I} = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$ ,  $f \in C([0, T] \times D \times D \times \mathbb{R} \times \mathbb{R}, \mathbb{R})$ ,  $g \in C(D, \mathbb{R})$ ,  $u \in C([0, T] \times \mathbb{D}, \mathbb{R})$ , assume  $u|_{[0, T] \times \mathbb{D}} \in C^{1,2}([0, T] \times \mathbb{D}, \mathbb{R})$ , let  $\nu_x: \mathcal{B}(\mathbb{D}) \rightarrow [0, 1]$ ,  $x \in \mathbb{D}$ , be probability measures, for every  $x \in \mathbb{D}$  let  $Z_x^i: \Omega \rightarrow \mathbb{D}$ ,  $i \in \mathfrak{I}$ , be i.i.d. random variables, assume for every  $A \in \mathcal{B}(\mathbb{D})$ ,  $i \in \mathfrak{I}$  that  $\mathbb{P}(Z_x^i \in A) = \nu_x(A)$ , let  $\phi_r: \mathbb{R} \rightarrow \mathbb{R}$ ,  $r \in [0, \infty]$ , satisfy for every  $r \in [0, \infty]$ ,  $y \in \mathbb{R}$  that*

$$\phi_r(y) = \min\{r, \max\{-r, y\}\}, \quad (5.24)$$

*let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, let  $\mathcal{V}^i: \Omega \rightarrow (0, 1)$ ,  $i \in \mathfrak{I}$ , be independent  $\mathcal{U}_{(0,1)}$ -distributed random variables, let  $V^i: [0, T] \times \Omega \rightarrow [0, T]$ ,  $i \in \mathfrak{I}$ , satisfy for every  $t \in [0, T]$ ,  $i \in \mathfrak{I}$  that*

$$V_t^i = t + (T - t)\mathcal{V}^i, \quad (5.25)$$

*let  $W^i: [0, T] \times \Omega \rightarrow \mathbb{R}^d$ ,  $i \in \mathfrak{I}$ , be independent standard Brownian motions, assume that  $(\mathcal{V}^i)_{i \in \mathfrak{I}}$  and  $(W^i)_{i \in \mathfrak{I}}$  are independent, let  $\mu: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  be globally Lipschitz continuous, for every  $x \in \mathbb{R}^d$ ,  $i \in \mathfrak{I}$ ,  $t \in [0, T]$  let  $X_t^{x,i} = (X_{t,s}^{x,i})_{s \in [t, T]}: [t, T] \times \Omega \rightarrow \mathbb{R}^d$  be a stochastic process with continuous sample paths, let  $(K_{n,l,m})_{n,l,m \in \mathbb{N}_0} \subseteq \mathbb{N}$ , for every  $i \in \mathfrak{I}$ ,*

$n, M \in \mathbb{N}_0$ ,  $r \in [0, \infty]$  let  $U_{n,M,r}^i: [0, T] \times \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^k$  satisfy for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  that

$$\begin{aligned} U_{n,M,r}^i(t, x) = & \left[ \sum_{l=0}^{n-1} \frac{(T-t)}{M^{n-l}} \sum_{m=1}^{M^{n-l}} \frac{1}{K_{n,l,m}} \sum_{k=1}^{K_{n,l,m}} \left[ f\left(V_t^{(i,l,m)}, X_{t,V_t^{(i,l,m)}}^{x,(i,l,m)}, Z_{X_{t,V_t^{(i,l,m)}}^{x,(i,l,m),k}}^{(i,l,m,k)}, \right. \right. \right. \right. \\ & \phi_r\left(U_{l,M,r}^{(i,l,m)}(V_t^{(i,l,m)}, X_{t,V_t^{(i,l,m)}}^{x,(i,l,m)})\right), \phi_r\left(U_{l,M,r}^{(i,l,m)}(V_t^{(i,l,m)}, Z_{X_{t,V_t^{(i,l,m)}}^{x,(i,l,m),k}}^{(i,l,m,k)})\right) \\ & - \mathbb{1}_{\mathbb{N}}(l) f\left(V_t^{(i,l,m)}, X_{t,V_t^{(i,l,m)}}^{x,(i,l,m)}, Z_{X_{t,V_t^{(i,l,m)}}^{x,(i,l,m),k}}^{(i,l,m,k)}, \phi_r\left(U_{\max\{l-1,0\},M,r}^{(i,l,-m)}(V_t^{(i,l,m)}, X_{t,V_t^{(i,l,m)}}^{x,(i,l,m)})\right)\right. \\ & \left. \left. \left. \left. \phi_r\left(U_{\max\{l-1,0\},M,r}^{(i,l,-m)}(V_t^{(i,l,m)}, Z_{X_{t,V_t^{(i,l,m)}}^{x,(i,l,m),k}}^{(i,l,m,k)})\right)\right)\right]\right] + \frac{\mathbb{1}_{\mathbb{N}}(n)}{M^n} \left[ \sum_{m=1}^{M^n} g(X_{t,T}^{x,(i,0,-m)}) \right], \end{aligned} \quad (5.26)$$

assume for every  $t \in [0, T]$ ,  $x \in \partial_D$  that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$ , and assume for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$  that  $\|u(t, x)\| \leq c(1 + \|x\|^c)$ ,  $u(T, x) = g(x)$ , and

$$\begin{aligned} \left(\frac{\partial}{\partial t} u\right)(t, x) + \frac{1}{2} \text{Trace}\left(\sigma(x)[\sigma(x)]^*(\text{Hess}_x u)(t, x)\right) + \langle \mu(x), (\nabla_x u)(t, x) \rangle \\ + \int_D f(t, x, \mathbf{x}, u(t, x), u(t, \mathbf{x})) \nu_x(d\mathbf{x}) = 0. \end{aligned} \quad (5.27)$$

## 5.4.2 Examples for the approximation method

**Example 5.4.2** (Fisher–KPP PDEs with Neumann boundary conditions). In this example we specialize Lemma 5.4.1 to the case of certain Fisher–KPP PDEs with Neumann boundary conditions (cf., e.g., (Bian et al., 2017; Wang et al., 2021)).

Assume Lemma 5.4.1, let  $\epsilon \in (0, \infty)$  satisfy  $\epsilon = \frac{1}{10}$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $D = [-\frac{1}{2}, \frac{1}{2}]^d$ , and  $T \in \{\frac{1}{5}, \frac{1}{2}, 1\}$ , assume for every  $n, l, m \in \mathbb{N}$  that  $K_{n,l,m} = 1$ , assume for every  $t \in [0, T]$ ,  $x, \mathbf{x} \in \mathbb{D}$ ,  $y, \mathbf{y} \in \mathbb{R}$ ,  $v \in \mathbb{R}^d$  that  $g(x) = \exp(-\frac{1}{4}\|x\|^2)$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = \epsilon v$ , and  $f(t, x, \mathbf{x}, y, \mathbf{y}) = y(1 - y)$ , and assume that for every  $x \in \mathbb{R}^d$ ,  $i \in \mathcal{I}$ ,  $t \in [0, T]$ ,  $s \in [t, T]$  it holds  $\mathbb{P}$ -a.s. that

$$X_{t,s}^{x,i} = R\left(x, x + \int_t^s \mu(X_{t,r}^{x,i}) dr + \int_t^s \sigma(X_{t,r}^{x,i}) dW_r^i\right) = R(x, x + \epsilon(W_s^i - W_t^i)). \quad (5.28)$$

The solution  $u: [0, T] \times \mathbb{D} \rightarrow \mathbb{R}$  of the PDE in (5.27) then satisfies that for every  $t \in [0, T]$ ,  $x \in \partial_D$  it holds that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$  and that for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$  it holds that  $u(T, x) = \exp(-\frac{1}{4}\|x\|^2)$  and

$$\left(\frac{\partial}{\partial t} u\right)(t, x) + \frac{\epsilon^2}{2} (\Delta_x u)(t, x) + u(t, x)(1 - u(t, x)) = 0. \quad (5.29)$$

**Example 5.4.3** (Non-local competition PDEs). In this example we specialize Lemma 5.4.1 to the case of certain non-local competition PDEs (cf., e.g., Doebeli and Ispolatov, 2010a; Berestycki et al., 2009; Perthame and Génieys, 2007; Génieys et al., 2006).

Assume Lemma 5.4.1, let  $\mathfrak{s}, \epsilon \in (0, \infty)$  satisfy  $\mathfrak{s} = \epsilon = \frac{1}{10}$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $D = \mathbb{R}^d$ , and  $T \in \{\frac{1}{5}, \frac{1}{2}, 1\}$ , assume for every  $n, l, m \in \mathbb{N}$  that  $K_{n,l,m} = 10$ , assume for

every  $x \in \mathbb{R}^d$ ,  $A \in \mathcal{B}(\mathbb{R}^d)$  that  $\nu_x(A) = \pi^{-d/2} \mathfrak{s}^{-d} \int_A \exp(-\mathfrak{s}^{-2} \|x - \mathbf{x}\|^2) d\mathbf{x}$ , assume for every  $t \in [0, T]$ ,  $v, x, \mathbf{x} \in \mathbb{R}^d$ ,  $y, \mathbf{y} \in \mathbb{R}$  that  $g(x) = \exp(-\frac{1}{4} \|x\|^2)$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = \epsilon v$ , and  $f(t, x, \mathbf{x}, y, \mathbf{y}) = y(1 - \mathbf{y}\pi^{d/2} \mathfrak{s}^d)$ , and assume that for every  $x \in \mathbb{R}^d$ ,  $\mathfrak{i} \in \mathfrak{I}$ ,  $t \in [0, T]$ ,  $s \in [t, T]$  it holds  $\mathbb{P}$ -a.s. that

$$X_{t,s}^{x,\mathfrak{i}} = x + \int_t^s \mu(X_{t,r}^{x,\mathfrak{i}}) dr + \int_t^s \sigma(X_{t,r}^{x,\mathfrak{i}}) dW_r^{\mathfrak{i}} = x + \epsilon(W_s^{\mathfrak{i}} - W_t^{\mathfrak{i}}). \quad (5.30)$$

The solution  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  of the PDE in (5.27) then satisfies that for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  it holds that  $u(T, x) = \exp(-\frac{1}{4} \|x\|^2)$  and

$$(\frac{\partial}{\partial t} u)(t, x) + \frac{\epsilon^2}{2} (\Delta_x u)(t, x) + u(t, x) \left( 1 - \int_{\mathbb{R}^d} u(t, \mathbf{x}) \exp(-\frac{\|\mathbf{x}-\mathbf{x}\|^2}{\mathfrak{s}^2}) d\mathbf{x} \right) = 0. \quad (5.31)$$

**Example 5.4.4** (Non-local sine-Gordon PDEs). In this example we specialize Lemma 5.4.1 to the case of certain non-local sine-Gordon type PDEs (cf, e.g., Hairer and Shen, 2016; Barone et al., 1971; Coleman, 1994).

Assume Lemma 5.4.1, let  $\mathfrak{s}, \epsilon \in (0, \infty)$  satisfy  $\mathfrak{s} = \epsilon = \frac{1}{10}$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $D = \mathbb{R}^d$ , and  $T \in \{1/5, 1/2, 1\}$ , assume for every  $n, l, m \in \mathbb{N}$  that  $K_{n,l,m} = 10$ , assume for every  $x \in \mathbb{R}^d$ ,  $A \in \mathcal{B}(\mathbb{R}^d)$  that  $\nu_x(A) = \pi^{-d/2} \mathfrak{s}^{-d} \int_A \exp(-\mathfrak{s}^{-2} \|x - \mathbf{x}\|^2) d\mathbf{x}$ , assume for every  $t \in [0, T]$ ,  $v, x, \mathbf{x} \in \mathbb{R}^d$ ,  $y, \mathbf{y} \in \mathbb{R}$  that  $g(x) = \exp(-\frac{1}{4} \|x\|^2)$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = \epsilon v$ , and  $f(t, x, \mathbf{x}, y, \mathbf{y}) = \sin(y) - \mathbf{y}\pi^{d/2} \mathfrak{s}^d$ , and assume that for every  $x \in \mathbb{R}^d$ ,  $\mathfrak{i} \in \mathfrak{I}$ ,  $t \in [0, T]$ ,  $s \in [t, T]$  it holds  $\mathbb{P}$ -a.s. that

$$X_{t,s}^{x,\mathfrak{i}} = x + \int_t^s \mu(X_{t,r}^{x,\mathfrak{i}}) dr + \int_t^s \sigma(X_{t,r}^{x,\mathfrak{i}}) dW_r^{\mathfrak{i}} = x + \epsilon(W_s^{\mathfrak{i}} - W_t^{\mathfrak{i}}). \quad (5.32)$$

The solution  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  of the PDE in (5.27) then satisfies that for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  it holds that  $u(T, x) = \exp(-\frac{1}{4} \|x\|^2)$  and

$$(\frac{\partial}{\partial t} u)(t, x) + \frac{\epsilon^2}{2} (\Delta_x u)(t, x) + \sin(u(t, x)) - \int_{\mathbb{R}^d} u(t, \mathbf{x}) \exp(-\frac{\|\mathbf{x}-\mathbf{x}\|^2}{\mathfrak{s}^2}) d\mathbf{x} = 0. \quad (5.33)$$

**Example 5.4.5** (Replicator-mutator PDEs). In this example we specialize Lemma 5.4.1 to the case of certain  $d$ -dimensional replicator-mutator PDEs (cf, e.g., Hamel et al., 2020).

Assume Lemma 5.4.1, let  $\mathfrak{m}_1, \mathfrak{m}_2, \dots, \mathfrak{m}_d, \mathfrak{s}_1, \mathfrak{s}_2, \dots, \mathfrak{s}_d, \mathfrak{u}_1, \mathfrak{u}_2, \dots, \mathfrak{u}_d \in \mathbb{R}$  satisfy for every  $k \in \{1, 2, \dots, d\}$  that  $\mathfrak{m}_k = \frac{1}{10}$ ,  $\mathfrak{s}_k = \frac{1}{20}$ , and  $\mathfrak{u}_k = 0$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $\mathbb{D} = \mathbb{R}^d$ , and  $T \in \{1/5, 1/2, 1\}$ , assume for every  $n, l, m \in \mathbb{N}$  that  $K_{n,l,m} = 10$ , let  $a: \mathbb{R}^d \rightarrow \mathbb{R}$  satisfy for every  $x \in \mathbb{R}^d$  that  $a(x) = -\frac{1}{2} \|x\|^2$ , assume for every  $x \in \mathbb{R}^d$ ,  $A \in \mathcal{B}(\mathbb{R}^d)$  that  $\nu_x(A) = \int_{A \cap [-1/2, 1/2]^d} d\mathbf{x}$ , assume for every  $t \in [0, T]$ ,  $v = (v_1, \dots, v_d)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ ,  $\mathbf{x} \in \mathbb{R}^d$ ,  $y, \mathbf{y} \in \mathbb{R}$  that  $g(x) = (2\pi)^{-d/2} [\prod_{i=1}^d |\mathfrak{s}_i|^{-1/2}] \exp(-\sum_{i=1}^d \frac{(x_i - \mathfrak{u}_i)^2}{2\mathfrak{s}_i})$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = (\mathfrak{m}_1 v_1, \dots, \mathfrak{m}_d v_d)$ , and

$$f(t, x, \mathbf{x}, y, \mathbf{y}) = y(a(x) - \mathbf{y}a(\mathbf{x})), \quad (5.34)$$

and assume that for every  $x \in \mathbb{R}^d$ ,  $i \in \mathfrak{I}$ ,  $t \in [0, T]$ ,  $s \in [t, T]$  it holds  $\mathbb{P}$ -a.s. that

$$X_{t,s}^{x,i} = x + \int_t^s \mu(X_{t,r}^{x,i}) dr + \int_t^s \sigma(X_{t,r}^{x,i}) dW_r^i = x + \sigma(0)(W_s^i - W_t^i). \quad (5.35)$$

The solution  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  of the PDE in (5.27) then satisfies that for every  $t \in [0, T]$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that  $u(T, x) = (2\pi)^{-d/2} [\prod_{i=1}^d |\mathfrak{s}_i|^{-1/2}] \exp(-\sum_{i=1}^d \frac{(x_i - u_i)^2}{2s_i})$  and

$$(\frac{\partial}{\partial t} u)(t, x) + u(t, x) \left( a(x) - \int_{[-1/2, 1/2]^d} u(t, \mathbf{x}) a(\mathbf{x}) d\mathbf{x} \right) + \sum_{i=1}^d \frac{1}{2} |\mathfrak{m}_i|^2 (\frac{\partial^2}{\partial x_i^2} u)(t, x) = 0. \quad (5.36)$$

**Example 5.4.6** (Allen–Cahn PDEs with conservation of mass). In this example we specialize Lemma 5.4.1 to the case of certain Allen–Cahn PDEs with cubic nonlinearity, conservation of mass, and no-flux boundary conditions (cf., e.g., Rubinstein & Sternberg (Rubinstein and Sternberg, 1992)).

Assume Lemma 5.4.1, let  $\epsilon \in (0, \infty)$  satisfy  $\epsilon = \frac{1}{10}$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $D = [-1/2, 1/2]^d$ , and  $T \in \{1/5, 1/2, 1\}$ , assume for every  $n, l, m \in \mathbb{N}$  that  $K_{n,l,m} = 10$ , assume for every  $x \in D$ ,  $A \in \mathcal{B}(D)$  that  $\nu_x(A) = \int_A d\mathbf{x}$ , assume for every  $t \in [0, T]$ ,  $x, \mathbf{x} \in \mathbb{D}$ ,  $y, \mathbf{y} \in \mathbb{R}$ ,  $v \in \mathbb{R}^d$  that  $g(x) = \exp(-\frac{1}{4}\|x\|^2)$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = \epsilon v$ , and  $f(t, x, \mathbf{x}, y, \mathbf{y}) = y - y^3 - (\mathbf{y} - \mathbf{y}^3)$ , and assume that for every  $x \in \mathbb{R}^d$ ,  $i \in \mathfrak{I}$ ,  $t \in [0, T]$ ,  $s \in [t, T]$  it holds  $\mathbb{P}$ -a.s. that

$$X_{t,s}^{x,i} = R \left( x, x + \int_t^s \mu(X_{t,r}^{x,i}) dr + \int_t^s \sigma(X_{t,r}^{x,i}) dW_r^i \right) = R(x, x + \epsilon(W_s^i - W_t^i)). \quad (5.37)$$

The solution  $u: [0, T] \times \mathbb{D} \rightarrow \mathbb{R}$  of the PDE in (5.27) then satisfies that for every  $t \in [0, T]$ ,  $x \in \partial\mathbb{D}$  it holds that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$  and that for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$  it holds that  $u(T, x) = \exp(-\frac{1}{4}\|x\|^2)$  and

$$(\frac{\partial}{\partial t} u)(t, x) + \frac{\epsilon^2}{2} (\Delta_x u)(t, x) + u(t, x) - [u(t, x)]^3 - \int_{[-1/2, 1/2]^d} u(t, \mathbf{x}) - [u(t, \mathbf{x})]^3 d\mathbf{x} = 0. \quad (5.38)$$

## 5.5 Numerical simulations

In this section we illustrate the performance of the machine learning-based approximation method proposed in Lemma 5.3.1 in Section 5.3.2 above by means of numerical simulations for five concrete (non-local) nonlinear PDEs; see Sections 5.5.1 to 5.5.5 below. In each of these numerical simulations we employ the general machine learning-based approximation method proposed in Lemma 5.3.1 with certain 4-layer neural networks and using the Adam optimizer (cf. (5.42) and (5.43) in Lemma 5.5.1 below and Kingma & Ba (Kingma and Ba, 2014)).

More precisely, in each of the numerical simulations in Sections 5.5.1 to 5.5.5 the functions  $\mathbb{V}_n^{j,\mathbf{s}}: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  with  $n \in \{1, 2, \dots, N\}$ ,  $j \in \{1, 2, \dots, 8000\}$ ,  $\mathbf{s} \in \mathbb{R}^s$  are implemented as  $N$  fully-connected feedforward neural networks. These neural networks consist of 4 layers (corresponding to 3 affine linear transformations in the neural networks) where the input layer is  $d$ -dimensional (with  $d$  neurons on the input layer), where the two hidden layers

are both  $(d + 50)$ -dimensional (with  $d + 50$  neurons on each of the two hidden layers), and where the output layer is 1-dimensional (with 1 neuron on the output layer). We refer to Fig. 5.1 for a graphical illustration of the neural network architecture used in the numerical simulations in Sections 5.5.1 to 5.5.5.

As activation functions just in front of the two hidden layers we employ, in Sections 5.5.1 to 5.5.4 below, multidimensional versions of the hyperbolic tangent function

$$\mathbb{R} \ni x \mapsto (e^x + e^{-x})^{-1}(e^x - e^{-x}) \in \mathbb{R}, \quad (5.39)$$

and we employ, in Section 5.5.5 below, multidimensional versions of the ReLU function

$$\mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R}. \quad (5.40)$$

In addition, in Sections 5.5.1, 5.5.2 and 5.5.4 we use the square function  $\mathbb{R} \ni x \mapsto x^2 \in \mathbb{R}$  as activation function just in front of the output layer and in Sections 5.5.3 and 5.5.5 we use the identity function  $\mathbb{R} \ni x \mapsto x \in \mathbb{R}$  as activation function just in front of the output layer. Furthermore, we employ Xavier initialization to initialize all neural network parameters; see Glorot and Bengio, 2010 for details. We did not employ batch normalization in our simulations.

Each of the numerical experiments presented below was performed with the Julia library HighDimPDE.jl on a NVIDIA TITAN RTX GPU with 1350 MHz core clock and 24 GB GDDR6 memory with 7000 MHz clock rate where the underlying system consisted of an AMD EPYC 7742 64-core CPU with 2TB memory running Julia 1.7.2 on Ubuntu 20.04.3.

**Framework 5.5.1.** Assume Lemma 5.3.1, assume  $\mathfrak{d} = (d+50)(d+1)+(d+50)(d+51)+(d+51)$ , let  $\varepsilon, \beta_1, \beta_2 \in \mathbb{R}$ ,  $(\gamma_m)_{m \in \mathbb{N}} \subseteq (0, \infty)$  satisfy  $\varepsilon = 10^{-8}$ ,  $\beta_1 = \frac{9}{10}$ , and  $\beta_2 = \frac{999}{1000}$ , let  $g: D \rightarrow \mathbb{R}$ ,  $\mu: D \rightarrow \mathbb{R}^d$ , and  $\sigma: D \rightarrow \mathbb{R}^{d \times d}$  be continuous, let  $u = (u(t, x))_{(t, x) \in [0, T] \times \mathbb{D}} \in C^{1,2}([0, T] \times \mathbb{D}, \mathbb{R})$  have at most polynomially growing partial derivatives, assume for every  $t \in (0, T]$ ,  $x \in \partial_{\mathbb{D}}$  that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$ , assume for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$ ,  $j \in \mathbb{N}$ ,  $s \in \mathbb{R}^s$  that  $u(0, x) = g(x) = \mathbb{V}_0^{j,s}(\theta, x)$ ,  $\int_{\mathbb{D}} |f(t, x, \mathbf{x}, u(t, x), u(t, \mathbf{x}))| \nu_x(d\mathbf{x}) < \infty$ , and

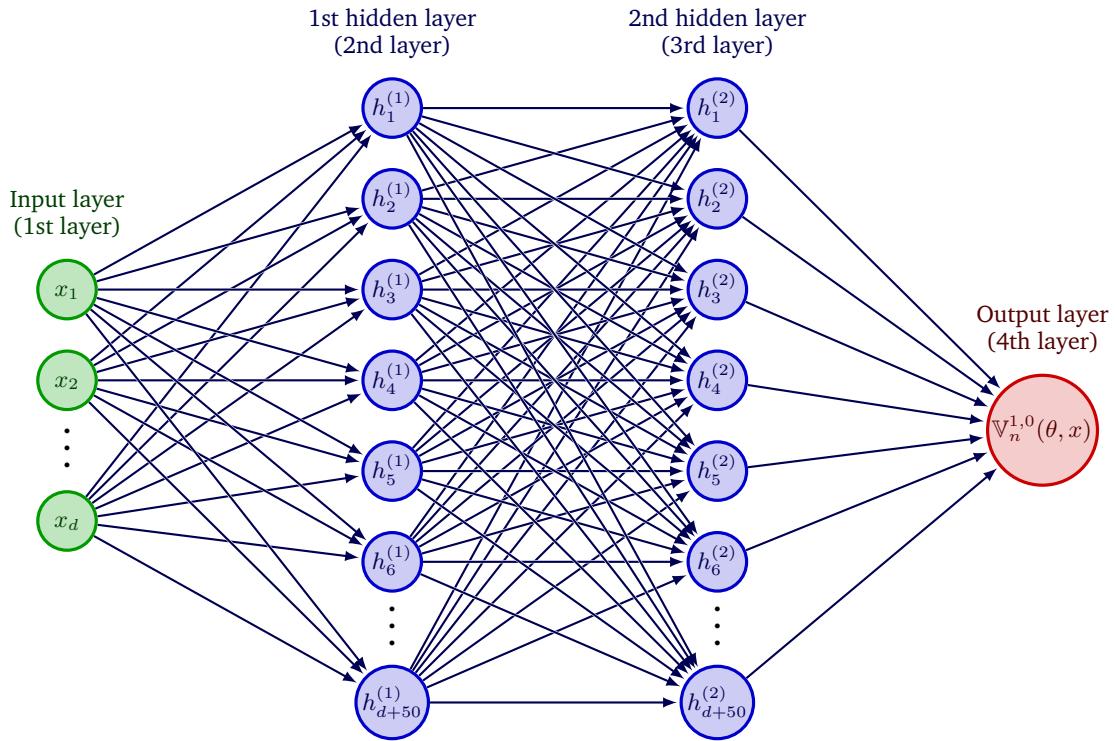
$$\begin{aligned} \left( \frac{\partial}{\partial t} u \right)(t, x) &= \frac{1}{2} \text{Trace}(\sigma(x)[\sigma(x)]^*(\text{Hess}_x u)(t, x)) + \langle \mu(x), (\nabla_x u)(t, x) \rangle \\ &\quad + \int_{\mathbb{D}} f(t, x, \mathbf{x}, u(t, x), u(t, \mathbf{x})) \nu_x(d\mathbf{x}), \end{aligned} \quad (5.41)$$

assume for every  $m \in \mathbb{N}$ ,  $i \in \{0, 1, \dots, N\}$  that  $J_m = 8000$ ,  $t_i = \frac{iT}{N}$ , and  $\varrho = 2\mathfrak{d}$ , and assume for every  $n \in \{1, 2, \dots, N\}$ ,  $m \in \mathbb{N}$ ,  $x = (x_1, \dots, x_{\mathfrak{d}})$ ,  $y = (y_1, \dots, y_{\mathfrak{d}})$ ,  $\eta = (\eta_1, \dots, \eta_{\mathfrak{d}}) \in \mathbb{R}^{\mathfrak{d}}$  that

$$\Xi_0^n(x, y, \eta) = 0, \quad \Psi_m^n(x, y, \eta) = (\beta_1 x + (1 - \beta_1)\eta, \beta_2 y + (1 - \beta_2)((\eta_1)^2, \dots, (\eta_{\mathfrak{d}})^2)), \quad (5.42)$$

and

$$\psi_m^n(x, y) = \left( \left[ \sqrt{\frac{|y_1|}{1 - (\beta_2)^m}} + \varepsilon \right]^{-1} \frac{\gamma_m x_1}{1 - (\beta_1)^m}, \dots, \left[ \sqrt{\frac{|y_{\mathfrak{d}}|}{1 - (\beta_2)^m}} + \varepsilon \right]^{-1} \frac{\gamma_m x_{\mathfrak{d}}}{1 - (\beta_1)^m} \right). \quad (5.43)$$



**Fig. 5.1:** Graphical illustration of the neural network architecture used in the numerical simulations. In Sections 5.5.1 to 5.5.5 we employ neural networks with 4 layers (corresponding to 3 affine linear transformations in the neural networks) with  $d$  neurons on the input layer (corresponding to a  $d$ -dimensional input layer), with  $d + 50$  neurons on the 1st hidden layer (corresponding to a  $(d + 50)$ -dimensional 1st hidden layer), with  $d + 50$  neurons on the 2nd hidden layer (corresponding to a  $(d + 50)$ -dimensional 2nd hidden layer), and with 1 neuron on the output layer (corresponding to a 1-dimensional output layer) in the numerical simulations.

### 5.5.1 Fisher–KPP PDEs with Neumann boundary conditions

In this subsection we use the machine learning-based approximation method in Lemma 5.5.1 to approximately calculate the solutions of certain Fisher–KPP PDEs with Neumann boundary conditions (cf., e.g., Bian et al., 2017 and Wang et al., 2021).

Assume Lemma 5.5.1, let  $\epsilon \in (0, \infty)$  satisfy  $\epsilon = \frac{1}{10}$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $D = [-1/2, 1/2]^d$ ,  $T \in \{1/5, 1/2, 1\}$ ,  $N = 10$ ,  $K_1 = K_2 = \dots = K_N = 1$ , and  $M_1 = M_2 = \dots = M_N = 500$ , assume for every  $n, m, j \in \mathbb{N}$ ,  $\omega \in \Omega$  that  $\xi^{n,m,j}(\omega) = (0, \dots, 0)$ , assume for every  $m \in \mathbb{N}$  that  $\gamma_m = 10^{-2}$ , and assume for every  $s, t \in [0, T]$ ,  $x, \mathbf{x} \in \mathbb{D}$ ,  $y, \mathbf{y} \in \mathbb{R}$ ,  $v \in \mathbb{R}^d$  that  $g(x) = \exp(-\frac{1}{4}\|x\|^2)$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = \epsilon v$ ,  $f(t, x, \mathbf{x}, y, \mathbf{y}) = y(1 - y)$ , and

$$H(t, s, x, v) = R(x, x + \mu(x)(t - s) + \sigma(x)v) = R(x, x + \epsilon v) \quad (5.44)$$

(cf. (5.6) and (5.17)). The solution  $u: [0, T] \times \mathbb{D} \rightarrow \mathbb{R}$  of the PDE in (5.41) then satisfies that for every  $t \in (0, T]$ ,  $x \in \partial_{\mathbb{D}}$  it holds that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$  and that for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$  it holds that  $u(0, x) = \exp(-\frac{1}{4}\|x\|^2)$  and

$$\left(\frac{\partial}{\partial t} u\right)(t, x) = \frac{\epsilon^2}{2} (\Delta_x u)(t, x) + u(t, x)(1 - u(t, x)). \quad (5.45)$$

In (5.45) the function  $u: [0, T] \times D \rightarrow \mathbb{R}$  models the proportion of a particular type of alleles in a biological population spatially structured over  $D$ . For every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  the number  $u(t, x) \in \mathbb{R}$  describes the proportion of individuals with a particular type of alleles located at position  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  at time  $t \in [0, T]$ . In Table 5.1 we use the machine learning-based approximation method in Lemma 5.5.1 to approximately calculate the mean of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the standard deviation of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the relative  $L^1$ -approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the uncorrected sample standard deviation of the approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , and the average runtime in seconds needed for calculating one realization of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$  based on 5 independent realizations (5 independent runs). The reference value, which is used as an approximation for the unknown value  $u(T, (0, \dots, 0))$  of the exact solution of (5.45), has been calculated via the MLP approximation method for non-local nonlinear PDEs in Lemma 5.4.1 (cf. Lemma 5.4.2 and Beck et al., 2019, Remark 3.3).

### 5.5.2 Non-local competition PDEs

In this subsection we use the machine learning-based approximation method in Lemma 5.5.1 to approximately calculate the solutions of certain non-local competition PDEs (cf., e.g., Doebeli and Ispolatov, 2010a; Berestycki et al., 2009; Perthame and Génieys, 2007; Génieys et al., 2006).

Assume Lemma 5.5.1, let  $\mathfrak{s}, \epsilon \in (0, \infty)$  satisfy  $\mathfrak{s} = \epsilon = \frac{1}{10}$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $\mathbb{D} = \mathbb{R}^d$ ,  $T \in \{1/5, 1/2, 1\}$ ,  $N = 10$ ,  $K_1 = K_2 = \dots = K_N = 5$ , and  $M_1 = M_2 = \dots = M_N = 500$ , assume for every  $n, m, j \in \mathbb{N}$ ,  $\omega \in \Omega$  that  $\xi^{n,m,j}(\omega) = (0, \dots, 0)$ , assume for every  $m \in \mathbb{N}$  that  $\gamma_m = 10^{-2}$ , and assume for every  $s, t \in [0, T]$ ,  $v, x, \mathbf{x} \in \mathbb{R}^d$ ,  $y, \mathbf{y} \in \mathbb{R}$ ,  $A \in \mathcal{B}(\mathbb{R}^d)$

$d$	$T$	$N$	Mean of the approx. method	Standard deviation of the approx. method	Reference value	Relative $L^1$ -approx. error	Standard deviation of the error	Average runtime in seconds
1	$1/5$	10	0.9995902	0.0000107	0.9996057	0.0000155	0.0000107	24.887
2	$1/5$	10	0.9991759	0.0000191	0.9991887	0.0000186	0.0000116	26.175
5	$1/5$	10	0.9979572	0.0000388	0.9979693	0.0000303	0.0000235	27.312
10	$1/5$	10	0.9959224	0.0000341	0.9959337	0.0000275	0.0000196	28.972
1	$1/2$	10	0.9992463	0.0000341	0.9992572	0.0000237	0.0000248	26.631
2	$1/2$	10	0.9984982	0.0000287	0.9985442	0.0000460	0.0000287	27.007
5	$1/2$	10	0.9962227	0.0000330	0.9962314	0.0000306	0.0000041	27.632
10	$1/2$	10	0.9925257	0.0001663	0.9921744	0.0003541	0.0001676	28.743
1	1	10	0.9991423	0.0000331	0.9989768	0.0001657	0.0000332	26.601
2	1	10	0.9982349	0.0000782	0.9982498	0.0000605	0.0000430	26.965
5	1	10	0.9956516	0.0000853	0.9957053	0.0000839	0.0000466	27.428
10	1	10	0.9912297	0.0001072	0.9904936	0.0007431	0.0001083	28.521

**Tab. 5.1:** Numerical simulations for the approximation method in Lemma 5.3.1 in the case of the Fisher–KPP PDEs with Neumann boundary conditions in (5.45) in Section 5.5.1.

that  $\nu_x(A) = \pi^{-d/2} \mathfrak{s}^{-d} \int_A \exp(-\mathfrak{s}^{-2} \|x - \mathbf{x}\|^2) d\mathbf{x}$ ,  $g(x) = \exp(-\frac{1}{4} \|x\|^2)$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = \epsilon v$ ,  $f(t, x, \mathbf{x}, y, \mathbf{y}) = y(1 - \mathbf{y}\mathfrak{s}^d \pi^{d/2})$ , and

$$H(t, s, x, v) = x + \mu(x)(t - s) + \sigma(x)v = x + \epsilon v \quad (5.46)$$

(cf. (5.6) and (5.17)). The solution  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  of the PDE in (5.41) then satisfies that for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  it holds that  $u(0, x) = \exp(-\frac{1}{4} \|x\|^2)$  and

$$(\frac{\partial}{\partial t} u)(t, x) = \frac{\epsilon^2}{2} (\Delta_x u)(t, x) + u(t, x) \left( 1 - \int_{\mathbb{R}^d} u(t, \mathbf{x}) \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}\|^2}{\mathfrak{s}^2}\right) d\mathbf{x} \right). \quad (5.47)$$

In (5.47) the function  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  models the evolution of a population characterized by a set of  $d$  biological traits under the combined effects of selection, competition and mutation. For every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  the number  $u(t, x) \in \mathbb{R}$  describes the number of individuals with traits  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  at time  $t \in [0, T]$ . In Table 5.2 we use the machine learning-based approximation method in Lemma 5.5.1 to approximately calculate the mean of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the standard deviation of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the relative  $L^1$ -approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the uncorrected sample standard deviation of the approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , and the average runtime in seconds needed for calculating one realization of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$  based on 5 independent realizations (5 independent runs). The reference value, which is used as an approximation for the unknown value  $u(T, (0, \dots, 0))$  of the exact solution of (5.47), has been calculated via the MLP approximation method for non-local nonlinear PDEs in Lemma 5.4.1 (cf. Lemma 5.4.3 and Beck et al., 2019, Remark 3.3).

$d$	$T$	$N$	Mean of the approx. method	Standard deviation of the approx. method	Reference value	Relative $L^1$ -approx. error	Standard deviation of the error	Average runtime in seconds
1	$1/5$	5	1.1748404	0.0006512	1.1735975	0.0010591	0.0005549	20.571
2	$1/5$	5	1.2114236	0.0008700	1.2096305	0.0014823	0.0007193	25.042
5	$1/5$	5	1.2186650	0.0007070	1.2159038	0.0022709	0.0005814	54.644
10	$1/5$	5	1.2153864	0.0007789	1.2128666	0.0020776	0.0006422	74.331
1	$1/2$	5	1.4755801	0.0032738	1.4694976	0.0041392	0.0022278	20.182
2	$1/2$	5	1.6112576	0.0110426	1.5948898	0.0103067	0.0068414	25.178
5	$1/2$	5	1.6433913	0.0067468	1.6186897	0.0152602	0.0041681	53.618
10	$1/2$	5	1.6323552	0.0053956	1.6090688	0.0144720	0.0033532	73.648
1	1	5	2.0795628	0.0223341	2.0493301	0.0147525	0.0108982	19.836
2	1	5	2.5651031	0.0513671	2.4683060	0.0392160	0.0208107	24.700
5	1	5	2.6977694	0.0381160	2.5606137	0.0535636	0.0148855	52.343
10	1	5	2.6490054	0.0155291	2.5299994	0.0470380	0.0061380	73.186

**Tab. 5.2:** Numerical simulations for the approximation method in Lemma 5.3.1 in the case of the non-local competition PDEs in (5.47) in Section 5.5.2.

### 5.5.3 Non-local sine-Gordon type PDEs

In this subsection we use the machine learning-based approximation method in Lemma 5.5.1 to approximately calculate the solutions of non-local sine-Gordon type PDEs (cf., e.g., Hairer and Shen, 2016; Barone et al., 1971; Coleman, 1994).

Assume Lemma 5.5.1, let  $\varsigma, \epsilon \in (0, \infty)$  satisfy  $\varsigma = \epsilon = \frac{1}{10}$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $\mathbb{D} = \mathbb{R}^d$ ,  $T \in \{1/5, 1/2, 1\}$ ,  $N = 10$ ,  $K_1 = K_2 = \dots = K_N = 5$ , and  $M_1 = M_2 = \dots = M_N = 500$ , assume for every  $n, m, j \in \mathbb{N}$ ,  $\omega \in \Omega$  that  $\xi^{n,m,j}(\omega) = (0, \dots, 0)$ , assume for every  $m \in \mathbb{N}$  that  $\gamma_m = 10^{-3}$ , and assume for every  $s, t \in [0, T]$ ,  $v, x, \mathbf{x} \in \mathbb{R}^d$ ,  $y, \mathbf{y} \in \mathbb{R}$ ,  $A \in \mathcal{B}(\mathbb{R}^d)$  that  $\nu_x(A) = \pi^{-d/2} \varsigma^{-d} \int_A \exp(-\varsigma^{-2} \|x - \mathbf{x}\|^2) d\mathbf{x}$ ,  $g(x) = \exp(-\frac{1}{4} \|x\|^2)$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = \epsilon v$ ,  $f(t, x, \mathbf{x}, y, \mathbf{y}) = \sin(y) - \mathbf{y} \pi^{d/2} \varsigma^d$ , and

$$H(t, s, x, v) = x + \mu(x)(t - s) + \sigma(x)v = x + \epsilon v \quad (5.48)$$

(cf. (5.6) and (5.17)). The solution  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  of the PDE in (5.41) then satisfies that for every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  it holds that  $u(0, x) = \exp(-\frac{1}{4} \|x\|^2)$  and

$$(\frac{\partial}{\partial t} u)(t, x) = \frac{\epsilon^2}{2} (\Delta_x u)(t, x) + \sin(u(t, x)) - \int_{\mathbb{R}^d} u(t, \mathbf{x}) \exp(-\frac{\|\mathbf{x}-x\|^2}{\varsigma^2}) d\mathbf{x}. \quad (5.49)$$

In Table 5.3 we use the machine learning-based approximation method in Lemma 5.5.1 to approximately calculate the mean of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the standard deviation of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the relative  $L^1$ -approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the uncorrected sample standard deviation of the approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , and the average runtime in seconds needed for calculating one realization of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$  based on 5 independent realizations (5 independent runs). The reference value, which is used as an approximation for the unknown value  $u(T, (0, \dots, 0))$

$d$	$T$	$N$	Mean of the approx. method	Standard deviation of the approx. method	Reference value	Relative $L^1$ -approx. error	Standard deviation of the error	Average runtime in seconds
1	$1/5$	10	1.1363013	0.0000101	1.1366512	0.0003079	0.0000089	23.635
2	$1/5$	10	1.1678476	0.0000118	1.1685004	0.0005586	0.0000101	24.788
5	$1/5$	10	1.1731812	0.0000087	1.1740671	0.0007546	0.0000074	24.233
10	$1/5$	10	1.1704700	0.0000063	1.1715686	0.0009377	0.0000054	24.767
1	$1/2$	10	1.3514235	0.0000152	1.3529022	0.0010930	0.0000112	22.622
2	$1/2$	10	1.4393708	0.0000245	1.4423641	0.0020753	0.0000170	23.419
5	$1/2$	10	1.4546282	0.0000816	1.4598476	0.0035754	0.0000559	23.739
10	$1/2$	10	1.4473282	0.0000739	1.4503958	0.0021150	0.0000510	24.222
1	1	10	1.7114614	0.0000309	1.7136091	0.0012533	0.0000180	22.067
2	1	10	1.9019763	0.0000288	1.9062322	0.0022326	0.0000151	22.707
5	1	10	1.9364921	0.0000602	1.9411610	0.0024052	0.0000310	22.899
10	1	10	1.9223347	0.0001494	1.9272222	0.0025360	0.0000775	23.719

**Tab. 5.3:** Numerical simulations for the approximation method in Lemma 5.3.1 in the case of the non-local sine-Gordon PDEs in (5.49) in Section 5.5.3.

of the exact solution of (5.49), has been calculated via the MLP approximation method for non-local nonlinear PDEs in Lemma 5.4.1 (cf. Lemma 5.4.4 and Beck et al., 2019, Remark 3.3).

## 5.5.4 Replicator-mutator PDEs

In this subsection we use the machine learning-based approximation method in Lemma 5.5.1 to approximately calculate the solutions of certain replicator-mutator PDEs describing the dynamics of a phenotype distribution under the combined effects of selection and mutation (cf., e.g., Hamel et al., 2020).

Assume Lemma 5.5.1, let  $\mathcal{D} \subseteq \mathbb{R}^d$ ,  $m_1, m_2, \dots, m_d, s_1, s_2, \dots, s_d, u_1, u_2, \dots, u_d, t \in \mathbb{R}$  satisfy for every  $k \in \{1, 2, \dots, d\}$  that  $m_k = \frac{1}{10}$ ,  $s_k = \frac{1}{20}$ ,  $u_k = 0$ , and  $t = \frac{1}{50}$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $\mathbb{D} = \mathbb{R}^d$ ,  $T \in \{1/10, 1/5, 1/2\}$ ,  $N = 10$ ,  $K_1 = K_2 = \dots = K_N = 5$ , let  $a \in C(\mathbb{R}^d, \mathbb{R})$ ,  $\delta \in C(\mathbb{R}^d, (0, \infty))$  satisfy for every  $x \in \mathbb{R}^d$  that  $a(x) = -\frac{1}{2}\|x\|^2$ , and assume for every  $s, t \in [0, T]$ ,  $v = (v_1, \dots, v_d)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ ,  $\mathbf{x} \in \mathbb{R}^d$ ,  $y, \mathbf{y} \in \mathbb{R}$ ,  $A \in \mathcal{B}(\mathbb{R}^d)$  that  $\nu_x(A) = \int_{A \cap \mathcal{D}} \delta(\mathbf{x}) d\mathbf{x}$ ,  $g(x) = (2\pi)^{-d/2} [\prod_{i=1}^d |s_i|^{-1/2}] \exp(-\sum_{i=1}^d \frac{(x_i - u_i)^2}{2s_i})$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = (m_1 v_1, \dots, m_d v_d)$ ,  $f(t, x, \mathbf{x}, y, \mathbf{y}) = y(a(x) - \mathbf{y}a(\mathbf{x})[\delta(\mathbf{x})]^{-1})$ , and

$$H(t, s, x, v) = x + \mu(x)(t - s) + \sigma(x)v = x + (m_1 v_1, \dots, m_d v_d) \quad (5.50)$$

(cf. (5.6) and (5.17)). The solution  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  of the PDE in (5.41) then satisfies that for every  $t \in [0, T]$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$u(0, x) = (2\pi)^{-d/2} \left[ \prod_{i=1}^d |s_i|^{-1/2} \right] \exp\left(-\sum_{i=1}^d \frac{(x_i - u_i)^2}{2s_i}\right) \quad (5.51)$$

and

$$(\frac{\partial}{\partial t} u)(t, x) = u(t, x) \left( a(x) - \int_{\mathcal{D}} u(t, \mathbf{x}) a(\mathbf{x}) d\mathbf{x} \right) + \sum_{i=1}^d \frac{1}{2} |\mathfrak{m}_i|^2 (\frac{\partial^2}{\partial x_i^2} u)(t, x). \quad (5.52)$$

In (5.52) the function  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  models the evolution of the phenotype distribution of a population composed of a set of  $d$  biological traits under the combined effects of selection and mutation. For every  $t \in [0, T]$ ,  $x \in \mathbb{R}^d$  the number  $u(t, x) \in \mathbb{R}$  describes the number of individuals with traits  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  at time  $t \in [0, T]$ . The function  $a$  models a quadratic Malthusian fitness function.

In Table 5.4 we use the machine learning-based method in Lemma 5.5.1 to approximately solve the PDE in Eq. (5.52) above in the case  $\mathcal{D} = \mathbb{R}^d$ . More precisely, we assume for every  $n, m, j \in \mathbb{N}$  that  $\xi^{n,m,j} = 0$ ,  $\gamma_m = 1/100$ ,  $M_n = 1000$  and we assume for every  $\mathbf{x} \in \mathbb{R}^d$  that  $\delta(\mathbf{x}) = (2\pi)^{-d/2} t^{-d} \exp(-\frac{\|\mathbf{x}\|^2}{2t^2})$  to approximately calculate the mean of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the standard deviation of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the relative  $L^1$ -approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the uncorrected sample standard deviation of the approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , and the average runtime in seconds needed for calculating one realization of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$  based on 5 independent realizations (5 independent runs). The value  $u(T, (0, \dots, 0))$  of the exact solution of (5.52) has been calculated by means of Lemma 5.5.2 below.

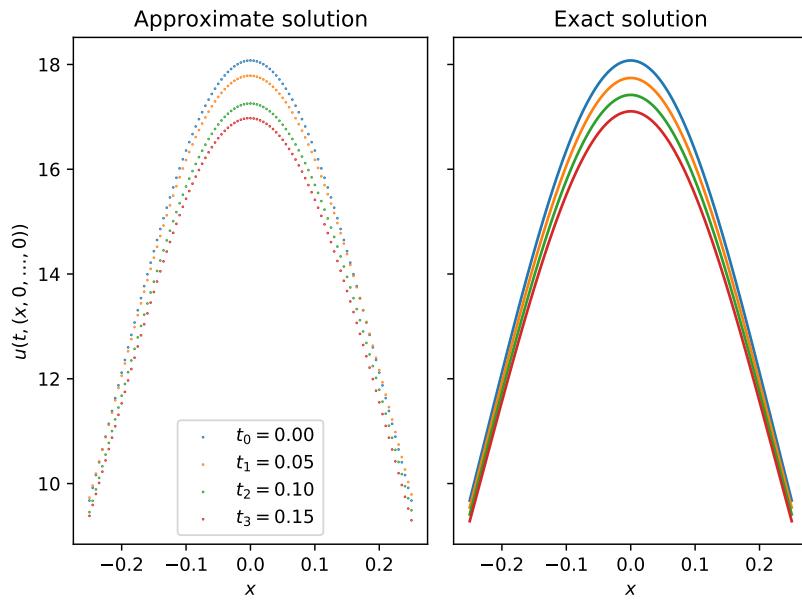
In Fig. 5.2 we use the machine learning-based method in Lemma 5.5.1 to approximate the solution  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  of the PDE in Eq. (5.52) above with  $d = 5$ ,  $T = 1/2$ , and  $\mathcal{D} = \mathbb{R}^d$ . The right-hand side of Fig. 5.2 shows a plot of  $[-1/4, 1/4] \ni x \mapsto u(t, (x, 0, \dots, 0)) \in \mathbb{R}$  for  $t \in \{0, 0.05, 0.1, 0.15\}$  where  $u$  is the exact solution of the PDE in Eq. (5.52) with  $d = 5$ ,  $T = 1/2$ , and  $\mathcal{D} = \mathbb{R}^d$  computed via Eq. (5.54) in Lemma 5.5.2 below. The left-hand side of Fig. 5.2 shows a plot of  $[-1/4, 1/4] \ni x \mapsto \mathbb{V}_n^{1,0}(\Theta_{M_n}^n(\omega), (x, 0, \dots, 0)) \in \mathbb{R}$  for  $n \in \{0, 1, 2, 3\}$  and one realization  $\omega \in \Omega$  where the functions  $\mathbb{R}^d \ni x \mapsto \mathbb{V}_n^{1,0}(\Theta_{M_n}^n(\omega), x) \in \mathbb{R}$  for  $n \in \{0, 1, 2, 3\}$ ,  $\omega \in \Omega$  were computed via Lemma 5.5.1 as an approximation of the solution of the PDE in Eq. (5.52) with  $d = 5$ ,  $T = 1/2$ , and  $\mathcal{D} = [-1/2, 1/2]^d$ . For the approximation, we take  $M_1 = M_2 = \dots = M_N = 2000$ ,  $\gamma_1 = \gamma_2 = \dots = \gamma_{2000} = 1/200$ , and  $\delta = \mathbb{1}_{\mathbb{R}^d}$  and we take  $\xi^{n,m,j}: \Omega \rightarrow \mathbb{R}^d$ ,  $n, m, j \in \mathbb{N}$ , to be independent  $\mathcal{U}_{[-1/2, 1/2]^d}$ -distributed random variables. Note that the solution of the PDE in Eq. (5.52) in the case  $\mathcal{D} = [-R, R]^d$  with  $R \in (0, \infty)$  sufficiently large is a good approximation of the solution  $u: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$  of the PDE in Eq. (5.52) in the case  $\mathcal{D} = \mathbb{R}^d$  since we have that for all  $t \in [0, T]$  the value  $u(t, x)$  of the solution  $u$  of the PDE in Eq. (5.52) in the case  $\mathcal{D} = \mathbb{R}^d$  quickly tends to 0 as  $\|x\|$  tends to  $\infty$ .

**Lemma 5.5.2.** Let  $d \in \mathbb{N}$ ,  $u_1, u_2, \dots, u_d \in \mathbb{R}$ ,  $\mathfrak{m}_1, \mathfrak{m}_2, \dots, \mathfrak{m}_d, \mathfrak{s}_1, \mathfrak{s}_2, \dots, \mathfrak{s}_d \in (0, \infty)$ , let  $a: \mathbb{R}^d \rightarrow \mathbb{R}$  satisfy for every  $x \in \mathbb{R}^d$  that  $a(x) = -\frac{1}{2}\|x\|^2$ , for every  $i \in \{1, 2, \dots, d\}$  let  $\mathfrak{S}_i: [0, \infty) \rightarrow (0, \infty)$  and  $\mathfrak{U}_i: [0, \infty) \rightarrow \mathbb{R}$  satisfy for every  $t \in [0, \infty)$  that

$$\mathfrak{S}_i(t) = \mathfrak{m}_i \left[ \frac{\mathfrak{m}_i \sinh(\mathfrak{m}_i t) + \mathfrak{s}_i \cosh(\mathfrak{m}_i t)}{\mathfrak{m}_i \cosh(\mathfrak{m}_i t) + \mathfrak{s}_i \sinh(\mathfrak{m}_i t)} \right] \quad \text{and} \quad \mathfrak{U}_i(t) = \frac{\mathfrak{m}_i u_i}{\mathfrak{m}_i \cosh(\mathfrak{m}_i t) + \mathfrak{s}_i \sinh(\mathfrak{m}_i t)}, \quad (5.53)$$

and let  $u: [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}$  satisfy for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  that

$$u(t, x) = (2\pi)^{-d/2} \left[ \prod_{i=1}^d |\mathfrak{S}_i(t)|^{-1/2} \right] \exp \left( - \sum_{i=1}^d \frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right). \quad (5.54)$$



**Fig. 5.2:** Plot of a machine learning-based approximation of the solution of the replicator-mutator PDE in Eq. (5.52) in the case  $d = 5$ ,  $T = 1/2$ , and  $\mathcal{D} = \mathbb{R}^d$ . The left-hand side shows a plot of  $[-1/4, 1/4] \ni x \mapsto \mathbb{V}_n^{1,0}(\Theta_{M_n}^n(\omega), (x, 0, \dots, 0)) \in \mathbb{R}$  for  $n \in \{0, 1, 2, 3\}$  and one realization  $\omega \in \Omega$  where the functions  $\mathbb{R}^d \ni x \mapsto \mathbb{V}_n^{1,0}(\Theta_{M_n}^n(\omega), x) \in \mathbb{R}$  for  $n \in \{0, 1, 2, 3\}$ ,  $\omega \in \Omega$  were computed via Lemma 5.5.1 as an approximation of the solution of the PDE in Eq. (5.52) with  $d = 5$ ,  $T = 1/2$ , and  $\mathcal{D} = [-1/2, 1/2]^d$  where we take  $M_1 = M_2 = \dots = M_N = 2000$ ,  $\gamma_1 = \gamma_2 = \dots = \gamma_{2000} = 1/200$ , and  $\delta = \mathbb{1}_{\mathbb{R}^d}$  and where we take  $\xi^{n,m,j} : \Omega \rightarrow \mathbb{R}^d$ ,  $n, m, j \in \mathbb{N}$ , to be independent  $\mathcal{U}_{[-1/2, 1/2]^d}$ -distributed random variables. The right-hand side of Fig. 5.2 shows a plot of  $[-1/4, 1/4] \ni x \mapsto u(t, (x, 0, \dots, 0)) \in \mathbb{R}$  for  $t \in \{0, 0.05, 0.1, 0.15\}$  where  $u$  is the exact solution of the PDE in Eq. (5.52) with  $d = 5$ ,  $T = 1/2$ , and  $\mathcal{D} = \mathbb{R}^d$ .

$d$	$T$	$N$	Mean of the approx. method	Standard deviation of the approx. method	Reference value	Relative $L^1$ -approx. error	Standard deviation of the error	Average runtime in seconds
1	$1/10$	10	1.7650547	0.0048907	1.7709574	0.0033330	0.0027616	43.949
2	$1/10$	10	3.1210874	0.0015513	3.1362901	0.0048474	0.0004946	45.002
5	$1/10$	10	17.1948978	0.0160821	17.4196954	0.0129048	0.0009232	45.934
10	$1/10$	10	295.8776489	0.0572639	303.4457874	0.0249407	0.0001887	47.750
1	$1/5$	10	1.7499938	0.0005580	1.7582066	0.0046711	0.0003174	43.129
2	$1/5$	10	3.0621917	0.0027811	3.0912904	0.0094131	0.0008996	44.443
5	$1/5$	10	16.3846066	0.0139748	16.8015567	0.0248162	0.0008318	45.019
10	$1/5$	10	268.2944397	0.0623432	282.2923073	0.0495864	0.0002208	45.612
1	$1/2$	10	1.7018557	0.0060157	1.7222757	0.0118564	0.0034929	42.092
2	$1/2$	10	2.8911286	0.0027431	2.9662336	0.0253200	0.0009248	42.657
5	$1/2$	10	14.2520916	0.1356645	15.1535149	0.0594861	0.0089527	43.338
10	$1/2$	10	201.6446228	0.3009756	229.6290127	0.1218678	0.0013107	44.190

**Tab. 5.4:** Numerical simulations for the approximation method in Lemma 5.3.1 in the case of the replicator-mutator PDEs in (5.52) in Section 5.5.4 where we assume for every  $n, m, j \in \mathbb{N}$  that  $\mathcal{D} = \mathbb{R}^d$ ,  $\xi^{n,m,j} = 0$ ,  $\gamma_m = 1/100$ , and  $M_n = 1000$  and where we assume for every  $x \in \mathbb{R}^d$  that  $\delta(x) = (2\pi)^{-d/2} t^{-d} \exp(-\frac{\|x\|^2}{2t^2})$ .

Then

(i) it holds that  $u \in C^{1,2}([0, \infty) \times \mathbb{R}^d, \mathbb{R})$ ,

(ii) it holds for every  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  that

$$u(0, x) = (2\pi)^{-d/2} \left[ \prod_{i=1}^d |\mathfrak{s}_i|^{-1/2} \right] \exp \left( - \sum_{i=1}^d \frac{(x_i - \mathfrak{u}_i)^2}{2\mathfrak{s}_i} \right), \quad (5.55)$$

and

(iii) it holds for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  that

$$\left( \frac{\partial}{\partial t} u \right)(t, x) = u(t, x) \left( a(x) - \int_{\mathbb{R}^d} u(t, \mathbf{x}) a(\mathbf{x}) d\mathbf{x} \right) + \sum_{i=1}^d \frac{1}{2} |\mathfrak{m}_i|^2 \left( \frac{\partial^2}{\partial x_i^2} u \right)(t, x). \quad (5.56)$$

*Proof of Lemma 5.5.2.* First, note that the fact that for every  $i \in \{1, 2, \dots, d\}$  it holds that  $\mathfrak{S}_i \in C^\infty([0, \infty), (0, \infty))$ , the fact that for every  $i \in \{1, 2, \dots, d\}$  it holds that  $\mathfrak{U}_i \in C^\infty([0, \infty), \mathbb{R})$ , and (5.54) establish Item (i). Moreover, observe that the fact that for every  $i \in \{1, 2, \dots, d\}$  it holds that  $\mathfrak{S}_i(0) = \mathfrak{s}_i$ , the fact that for every  $i \in \{1, 2, \dots, d\}$  it holds that  $\mathfrak{U}_i(0) = \mathfrak{u}_i$ , and (5.53) prove Item (ii). Next note that (5.54) ensures that for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$u(t, x) = \prod_{i=1}^d \left[ (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( - \frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right]. \quad (5.57)$$

The product rule hence implies that for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$\begin{aligned}
& (\frac{\partial}{\partial t} u)(t, x) \\
&= \frac{\partial}{\partial t} \left( \prod_{i=1}^d \left[ (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right] \right) \\
&= \sum_{i=1}^d \left[ \left[ \prod_{j \in \{1, \dots, d\} \setminus \{i\}} \left( (2\pi \mathfrak{S}_j(t))^{-1/2} \exp \left( -\frac{(x_j - \mathfrak{U}_j(t))^2}{2\mathfrak{S}_j(t)} \right) \right) \right] \right. \\
&\quad \cdot \left. \left[ \frac{\partial}{\partial t} \left( (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right) \right] \right]. \tag{5.58}
\end{aligned}$$

The chain rule, the product rule, and (5.57) therefore show that for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$\begin{aligned}
& (\frac{\partial}{\partial t} u)(t, x) \\
&= \sum_{i=1}^d \left[ \left[ \prod_{j \in \{1, \dots, d\} \setminus \{i\}} \left( (2\pi \mathfrak{S}_j(t))^{-1/2} \exp \left( -\frac{(x_j - \mathfrak{U}_j(t))^2}{2\mathfrak{S}_j(t)} \right) \right) \right] \right. \\
&\quad \cdot \left[ \left( \frac{\partial}{\partial t} \left( (2\pi \mathfrak{S}_i(t))^{-1/2} \right) \right) \exp \left( -\frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right. \\
&\quad \left. \left. + (2\pi \mathfrak{S}_i(t))^{-1/2} \left( \frac{\partial}{\partial t} \left( -\frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right) \exp \left( -\frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right] \right] \\
&= \sum_{i=1}^d \left[ \left[ \prod_{j \in \{1, \dots, d\} \setminus \{i\}} \left( (2\pi \mathfrak{S}_j(t))^{-1/2} \exp \left( -\frac{(x_j - \mathfrak{U}_j(t))^2}{2\mathfrak{S}_j(t)} \right) \right) \right] \right. \\
&\quad \cdot \left[ -(2\pi \mathfrak{S}_i(t))^{-1/2} \left[ \frac{(\frac{\partial}{\partial t} \mathfrak{S}_i)(t)}{2\mathfrak{S}_i(t)} \right] \exp \left( -\frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right. \\
&\quad \left. + (2\pi \mathfrak{S}_i(t))^{-1/2} \left( \frac{2(\frac{\partial}{\partial t} \mathfrak{U}_i)(t)(x_i - \mathfrak{U}_i(t))}{2\mathfrak{S}_i(t)} \right. \right. \\
&\quad \left. \left. + \frac{(x_i - \mathfrak{U}_i(t))^2(\frac{\partial}{\partial t} \mathfrak{S}_i)(t)}{2|\mathfrak{S}_i(t)|^2} \right) \exp \left( -\frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right] \right] \\
&= u(t, x) \left[ \sum_{i=1}^d \left( \frac{-(\frac{\partial}{\partial t} \mathfrak{S}_i)(t)}{2\mathfrak{S}_i(t)} + \frac{2\mathfrak{S}_i(t)(\frac{\partial}{\partial t} \mathfrak{U}_i)(t)(x_i - \mathfrak{U}_i(t)) + (x_i - \mathfrak{U}_i(t))^2(\frac{\partial}{\partial t} \mathfrak{S}_i)(t)}{2|\mathfrak{S}_i(t)|^2} \right) \right]. \tag{5.59}
\end{aligned}$$

Moreover, observe that (5.53), the chain rule, and the product rule ensure that for every  $i \in \{1, \dots, d\}$ ,  $t \in [0, \infty)$  it holds that

$$\begin{aligned}
(\frac{\partial}{\partial t} \mathfrak{U}_i)(t) &= \frac{\partial}{\partial t} \left( \frac{\mathfrak{m}_i \mathfrak{u}_i}{\mathfrak{m}_i \cosh(\mathfrak{m}_i t) + \mathfrak{s}_i \sinh(\mathfrak{m}_i t)} \right) \\
&= -|\mathfrak{m}_i|^2 \mathfrak{u}_i \left[ \frac{\mathfrak{m}_i \sinh(\mathfrak{m}_i t) + \mathfrak{s}_i \cosh(\mathfrak{m}_i t)}{[\mathfrak{m}_i \cosh(\mathfrak{m}_i t) + \mathfrak{s}_i \sinh(\mathfrak{m}_i t)]^2} \right] \\
&= -\mathfrak{S}_i(t) \mathfrak{U}_i(t)
\end{aligned} \tag{5.60}$$

and

$$\begin{aligned}
& \left( \frac{\partial}{\partial t} \mathfrak{S}_i \right) (t) \\
&= \frac{\partial}{\partial t} \left( \mathfrak{m}_i \left[ \frac{\mathfrak{m}_i \sinh(\mathfrak{m}_i t) + \mathfrak{s}_i \cosh(\mathfrak{m}_i t)}{\mathfrak{m}_i \cosh(\mathfrak{m}_i t) + \mathfrak{s}_i \sinh(\mathfrak{m}_i t)} \right] \right) \\
&= |\mathfrak{m}_i|^2 \left[ \frac{\mathfrak{m}_i \cosh(\mathfrak{m}_i t) + \mathfrak{s}_i \sinh(\mathfrak{m}_i t)}{\mathfrak{m}_i \cosh(\mathfrak{m}_i t) + \mathfrak{s}_i \sinh(\mathfrak{m}_i t)} \right] - |\mathfrak{m}_i|^2 \left[ \frac{\mathfrak{m}_i \sinh(\mathfrak{m}_i t) + \mathfrak{s}_i \cosh(\mathfrak{m}_i t)}{\mathfrak{m}_i \cosh(\mathfrak{m}_i t) + \mathfrak{s}_i \sinh(\mathfrak{m}_i t)} \right]^2 \\
&= |\mathfrak{m}_i|^2 - |\mathfrak{S}_i(t)|^2.
\end{aligned} \tag{5.61}$$

Combining this with (5.59) implies that for every  $i \in \{1, 2, \dots, d\}$ ,  $t \in [0, \infty)$  it holds that

$$\begin{aligned}
\left( \frac{\partial}{\partial t} u \right) (t, x) &= \frac{u(t, x)}{2} \sum_{i=1}^d \left[ \frac{-[|\mathfrak{m}_i|^2 - |\mathfrak{S}_i(t)|^2]}{\mathfrak{S}_i(t)} \right. \\
&\quad \left. + \frac{2|\mathfrak{S}_i(t)|^2 \mathfrak{U}_i(t)(\mathfrak{U}_i(t) - x_i) + (x_i - \mathfrak{U}_i(t))^2 (|\mathfrak{m}_i|^2 - |\mathfrak{S}_i(t)|^2)}{|\mathfrak{S}_i(t)|^2} \right] \\
&= \frac{u(t, x)}{2} \sum_{i=1}^d \left[ |\mathfrak{m}_i|^2 \left( \left( \frac{x_i - \mathfrak{U}_i(t)}{\mathfrak{S}_i(t)} \right)^2 - \frac{1}{\mathfrak{S}_i(t)} \right) \right. \\
&\quad \left. + \mathfrak{S}_i(t) + 2(|\mathfrak{U}_i(t)|^2 - \mathfrak{U}_i(t)x_i) - (|x_i|^2 - 2\mathfrak{U}_i(t)x_i + |\mathfrak{U}_i(t)|^2) \right] \\
&= \frac{u(t, x)}{2} \sum_{i=1}^d \left[ |\mathfrak{m}_i|^2 \left( \left( \frac{x_i - \mathfrak{U}_i(t)}{\mathfrak{S}_i(t)} \right)^2 - \frac{1}{\mathfrak{S}_i(t)} \right) + \mathfrak{S}_i(t) + |\mathfrak{U}_i(t)|^2 - |x_i|^2 \right].
\end{aligned} \tag{5.62}$$

Furthermore, note that (5.57) and the product rule show that for every  $i \in \{1, 2, \dots, d\}$ ,  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$\begin{aligned}
\left( \frac{\partial}{\partial x_i} u \right) (t, x) &= \frac{\partial}{\partial x_i} \left[ \prod_{j=1}^d \left[ (2\pi \mathfrak{S}_j(t))^{-1/2} \exp \left( -\frac{(x_j - \mathfrak{U}_j(t))^2}{2\mathfrak{S}_j(t)} \right) \right] \right] \\
&= \left[ \frac{\partial}{\partial x_i} \left[ (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{(x_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right] \right] \\
&\quad \cdot \prod_{j \in \{1, 2, \dots, d\} \setminus \{i\}} \left[ (2\pi \mathfrak{S}_j(t))^{-1/2} \exp \left( -\frac{(x_j - \mathfrak{U}_j(t))^2}{2\mathfrak{S}_j(t)} \right) \right] \\
&= -u(t, x) \left( \frac{x_i - \mathfrak{U}_i(t)}{\mathfrak{S}_i(t)} \right) = u(t, x) \left( \frac{\mathfrak{U}_i(t) - x_i}{\mathfrak{S}_i(t)} \right).
\end{aligned} \tag{5.63}$$

The product rule therefore assures that for every  $i \in \{1, 2, \dots, d\}$ ,  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$\begin{aligned}
\left( \frac{\partial^2}{\partial x_i^2} u \right) (t, x) &= \frac{\partial}{\partial x_i} \left( u(t, x) \left( \frac{\mathfrak{U}_i(t) - x_i}{\mathfrak{S}_i(t)} \right) \right) \\
&= \left( \frac{\partial}{\partial x_i} u \right) (t, x) \left( \frac{\mathfrak{U}_i(t) - x_i}{\mathfrak{S}_i(t)} \right) - \frac{u(t, x)}{\mathfrak{S}_i(t)} = u(t, x) \left[ \left( \frac{x_i - \mathfrak{U}_i(t)}{\mathfrak{S}_i(t)} \right)^2 - \frac{1}{\mathfrak{S}_i(t)} \right].
\end{aligned} \tag{5.64}$$

Hence, we obtain that for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$\sum_{i=1}^d \left[ \frac{1}{2} |\mathfrak{m}_i|^2 \left( \frac{\partial^2}{\partial x_i^2} u \right)(t, x) \right] = \frac{u(t, x)}{2} \sum_{i=1}^d \left[ |\mathfrak{m}_i|^2 \left( \left( \frac{x_i - \mathfrak{U}_i(t)}{\mathfrak{S}_i(t)} \right)^2 - \frac{1}{\mathfrak{S}_i(t)} \right) \right]. \quad (5.65)$$

Next observe that (5.57) and Fubini's theorem ensure that for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$\begin{aligned} & u(t, x) \left( a(x) - \int_{\mathbb{R}^d} u(t, \mathbf{x}) a(\mathbf{x}) d\mathbf{x} \right) \\ &= u(t, x) \left( -\frac{1}{2} \left[ \sum_{i=1}^d |x_i|^2 \right] - \int_{\mathbb{R}^d} -\frac{1}{2} \left[ \sum_{i=1}^d |\mathbf{x}_i|^2 \right] u(t, \mathbf{x}) d\mathbf{x} \right) \\ &= \frac{u(t, x)}{2} \left( - \left[ \sum_{i=1}^d |x_i|^2 \right] \right. \\ &\quad \left. + \sum_{i=1}^d \left[ \int_{\mathbb{R}} |\mathbf{x}_i|^2 (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{(\mathbf{x}_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) d\mathbf{x}_i \right. \right. \\ &\quad \left. \cdot \left( \prod_{j \in \{1, 2, \dots, d\} \setminus \{i\}} \int_{\mathbb{R}} (2\pi \mathfrak{S}_j(t))^{-1/2} \exp \left( -\frac{(\mathbf{x}_j - \mathfrak{U}_j(t))^2}{2\mathfrak{S}_j(t)} \right) d\mathbf{x}_j \right) \right] \right). \end{aligned} \quad (5.66)$$

This and the fact that for every  $i \in \{1, 2, \dots, d\}$ ,  $t \in [0, \infty)$  it holds that

$$\int_{\mathbb{R}} (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{(x - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) dx = 1 \quad (5.67)$$

imply that for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$\begin{aligned} & u(t, x) \left( a(x) - \int_{\mathbb{R}^d} u(t, \mathbf{x}) a(\mathbf{x}) d\mathbf{x} \right) \\ &= \frac{u(t, x)}{2} \sum_{i=1}^d \left[ -|x_i|^2 + \int_{\mathbb{R}} |\mathbf{x}_i|^2 (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{(\mathbf{x}_i - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) d\mathbf{x}_i \right]. \end{aligned} \quad (5.68)$$

Next observe that the integral transformation theorem demonstrates that for every  $i \in \{1, 2, \dots, d\}$ ,  $t \in [0, \infty)$  it holds that

$$\begin{aligned} & \int_{\mathbb{R}} x^2 \left[ (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{(x - \mathfrak{U}_i(t))^2}{2\mathfrak{S}_i(t)} \right) \right] dx \\ &= \int_{\mathbb{R}} (x + \mathfrak{U}_i(t))^2 \left[ (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{x^2}{2\mathfrak{S}_i(t)} \right) \right] dx \\ &= \int_{\mathbb{R}} x^2 \left[ (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{x^2}{2\mathfrak{S}_i(t)} \right) \right] dx \\ &\quad + \int_{\mathbb{R}} |\mathfrak{U}_i(t)|^2 \left[ (2\pi \mathfrak{S}_i(t))^{-1/2} \exp \left( -\frac{x^2}{2\mathfrak{S}_i(t)} \right) \right] dx \\ &= \mathfrak{S}_i(t) + |\mathfrak{U}_i(t)|^2. \end{aligned} \quad (5.69)$$

Combining this with (5.68) ensures that for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$u(t, x) \left( a(x) - \int_{\mathbb{R}^d} u(t, \mathbf{x}) a(\mathbf{x}) d\mathbf{x} \right) = \frac{u(t, x)}{2} \sum_{i=1}^d (\mathfrak{S}_i(t) + |\mathfrak{U}_i(t)|^2 - |x_i|^2). \quad (5.70)$$

This and (5.65) demonstrate that for every  $t \in [0, \infty)$ ,  $x = (x_1, \dots, x_d) \in \mathbb{R}^d$  it holds that

$$\begin{aligned} & u(t, x) \left( a(x) - \int_D u(t, \mathbf{x}) a(\mathbf{x}) d\mathbf{x} \right) + \sum_{i=1}^d \frac{1}{2} |\mathfrak{m}_i|^2 \left( \frac{\partial^2}{\partial x_i^2} u \right)(t, x) \\ &= \frac{u(t, x)}{2} \sum_{i=1}^d \left[ |\mathfrak{m}_i|^2 \left( \left( \frac{x_i - \mathfrak{U}_i(t)}{\mathfrak{S}_i(t)} \right)^2 - \frac{1}{\mathfrak{S}_i(t)} \right) + \mathfrak{S}_i(t) + |\mathfrak{U}_i(t)|^2 - |x_i|^2 \right]. \end{aligned} \quad (5.71)$$

Combining this with (5.62) proves Item (iii). The proof of Lemma 5.5.2 is thus complete.  $\square$

### 5.5.5 Allen–Cahn PDEs with conservation of mass

In this subsection we use the machine learning-based approximation method in Lemma 5.5.1 to approximately calculate the solutions of certain Allen–Cahn PDEs with cubic nonlinearity, conservation of mass and no-flux boundary conditions (cf., e.g., Rubinstein & Sternberg (Rubinstein and Sternberg, 1992)).

Assume Lemma 5.5.1, let  $\epsilon \in (0, \infty)$  satisfy  $\epsilon = \frac{1}{10}$ , assume that  $d \in \{1, 2, 5, 10\}$ ,  $\mathbb{D} = [-1/2, 1/2]^d$ ,  $T \in \{1/5, 1/2, 1\}$ ,  $N = 10$ ,  $K_1 = K_2 = \dots = K_N = 5$ , and  $M_1 = M_2 = \dots = M_N = 500$ , assume that  $\xi^{n,m,j}$ ,  $n, m, j \in \mathbb{N}$ , are independent  $\mathcal{U}_D$ -distributed random variables, assume for every  $m \in \mathbb{N}$  that  $\gamma_m = 10^{-2}$ , and assume for every  $s, t \in [0, T]$ ,  $x, \mathbf{x} \in \mathbb{D}$ ,  $y, \mathbf{y} \in \mathbb{R}$ ,  $v \in \mathbb{R}^d$ ,  $A \in \mathcal{B}(D)$  that  $\nu_x(A) = \int_A d\mathbf{x}$ ,  $g(x) = \exp(-\frac{1}{4}\|x\|^2)$ ,  $\mu(x) = (0, \dots, 0)$ ,  $\sigma(x)v = \epsilon v$ ,  $f(t, x, \mathbf{x}, y, \mathbf{y}) = y - y^3 - (\mathbf{y} - \mathbf{y}^3)$ , and

$$H(t, s, x, v) = R(x, x + \mu(x)(t - s) + \sigma(x)v) = R(x, x + \epsilon v) \quad (5.72)$$

(cf. (5.6) and (5.17)). The solution  $u: [0, T] \times \mathbb{D} \rightarrow \mathbb{R}$  of the PDE in (5.41) then satisfies that for every  $t \in (0, T]$ ,  $x \in \partial_{\mathbb{D}}$  it holds that  $\langle \mathbf{n}(x), (\nabla_x u)(t, x) \rangle = 0$  and that for every  $t \in [0, T]$ ,  $x \in \mathbb{D}$  it holds that  $u(0, x) = \exp(-\frac{1}{4}\|x\|^2)$  and

$$(\frac{\partial}{\partial t} u)(t, x) = \frac{\epsilon^2}{2} (\Delta_x u)(t, x) + u(t, x) - [u(t, x)]^3 - \int_{[-1/2, 1/2]^d} u(t, \mathbf{x}) - [u(t, \mathbf{x})]^3 d\mathbf{x}. \quad (5.73)$$

In Table 5.5 we use the machine learning-based approximation method in Lemma 5.5.1 to approximately calculate the mean of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the standard deviation of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the relative  $L^1$ -approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , the uncorrected sample standard deviation of the approximation error associated to  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$ , and the average runtime in seconds needed for calculating one realization of  $\mathbb{V}_N^{1,0}(\Theta_{M_N}^N, (0, \dots, 0))$  based on 5 independent realizations (5 independent runs). The reference value, which is used as an approximation for the unknown value  $u(T, (0, \dots, 0))$  of the exact solution of (5.73), has been calculated via the MLP approximation method

$d$	$T$	$N$	Mean of the approx. method	Standard deviation of the approx. method	Reference value	Relative $L^1$ -approx. error	Standard deviation of the error	Average runtime in seconds
1	$1/5$	10	0.9947184	0.0021832	0.9932255	0.0015709	0.0021380	31.417
2	$1/5$	10	0.9908873	0.0027061	0.9868883	0.0040521	0.0027421	35.069
5	$1/5$	10	0.9942151	0.0052064	0.9710707	0.0238340	0.0053615	38.363
10	$1/5$	10	0.9792556	0.0203935	0.9514115	0.0292661	0.0214350	42.782
1	$1/2$	10	0.9870476	0.0014673	0.9880013	0.0014996	0.0007477	30.297
2	$1/2$	10	0.9763564	0.0030895	0.9750274	0.0024841	0.0021561	34.922
5	$1/2$	10	0.9518845	0.0051304	0.9431354	0.0092766	0.0054398	37.963
10	$1/2$	10	0.9249420	0.0052786	0.9063239	0.0205424	0.0058242	43.139
1	1	10	0.9823494	0.0003647	0.9780817	0.0043633	0.0003729	29.250
2	1	10	0.9659823	0.0004128	0.9658025	0.0003195	0.0003137	34.485
5	1	10	0.9209547	0.0019223	0.9158821	0.0055385	0.0020988	39.318
10	1	10	0.8693402	0.0029947	0.8683143	0.0030165	0.0015052	44.258

**Tab. 5.5:** Numerical simulations for the approximation method in Lemma 5.3.1 in the case of the Allen–Cahn PDEs with conservation of mass in (5.73) in Section 5.5.5.

for non-local nonlinear PDEs in Lemma 5.4.1 (cf. Lemma 5.4.6 and Beck et al., 2019, Remark 3.3).

## 5.6 Code availability

The machine learning-based method and the MLP method are implemented in the Julia package **HighDimPDE.jl** available at <https://github.com/SciML/HighDimPDE.jl>, and the simulation code is available at [https://github.com/vboussange/HighDimPDE\\_examples](https://github.com/vboussange/HighDimPDE_examples).

## 5.7 Acknowledgements

This project has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy EXC 2044-390685587, Mathematics Münster: Dynamics-Geometry-Structure. This project has been partially supported by the startup fund project of Shenzhen Research Institute of Big Data under grant No. T00120220001. L.P. was supported by the Swiss National Science Foundation grant (N<sup>o</sup> 310030\_188550).





# Discussion

“ To be but one with all living things, to return, by a radiant self-forgetfulness, to the All of Nature.

— Friedrich Hölderlin (1770-1843)

Understanding ecological and economic systems involves the underpinning of general organizational principles at the origin of invariant patterns (Levin, 2002, Fig. 1.3). Aiming at advancing our understanding of eco-evolutionary dynamics in ecological and economic systems, this thesis contributed to

- (i) a general understanding of the role of eco-evolutionary processes in shaping the dynamics of biological populations structured in complex landscapes (Chapter 2),
- (ii) the quantification of the effect of eco-evolutionary processes on the dynamics of economic systems at the country level (Chapter 4), and
- (iii) methodological advances in the forward and inverse modelling of eco-evolutionary dynamics (Chapters 2, 3 and 5).

In the following, I discuss the chapters of this thesis collectively, highlighting how they contribute to advance our current understanding of the dynamics of ecological and economic systems, and how they improve the current eco-evolutionary modelling paradigm. I further highlight current limitations, and propose future research directions.

## 6.1 Contributions

### 6.1.1 Linking eco-evolutionary processes to patterns of differentiation

Phenotypic differentiation arises from feedbacks between population dynamics, dispersal and mutations (Hamilton, 2021), and Chapter 2 determines how these feedbacks are modulated by landscape features. Mutations act upon individual organisms, and result in genetic drift in populations of finite size, causing stochastic variations in allelic proportions and phenotypes (Slatkin, 1987). In geographically structured populations, drift results in patterns of neutral differentiation (Slatkin, 1987). Migration fluxes reduce neutral differentiation by homogenizing differences in allelic proportions and phenotypes (Slatkin, 1987), and this effect is further increased by landscape connectivity (Wright, 1943; McRae, 2006; McRae and Beier, 2007) through the mechanism of "isolation by limited dispersal" (Lande, 1991; Orsini et al., 2013). When landscapes present heterogeneous habitats, natural selection can supplement the effect of genetic drift and increase the sole effect of stochasticity on differentiation (Fisher, 1958). Under this scenario, local environmental conditions

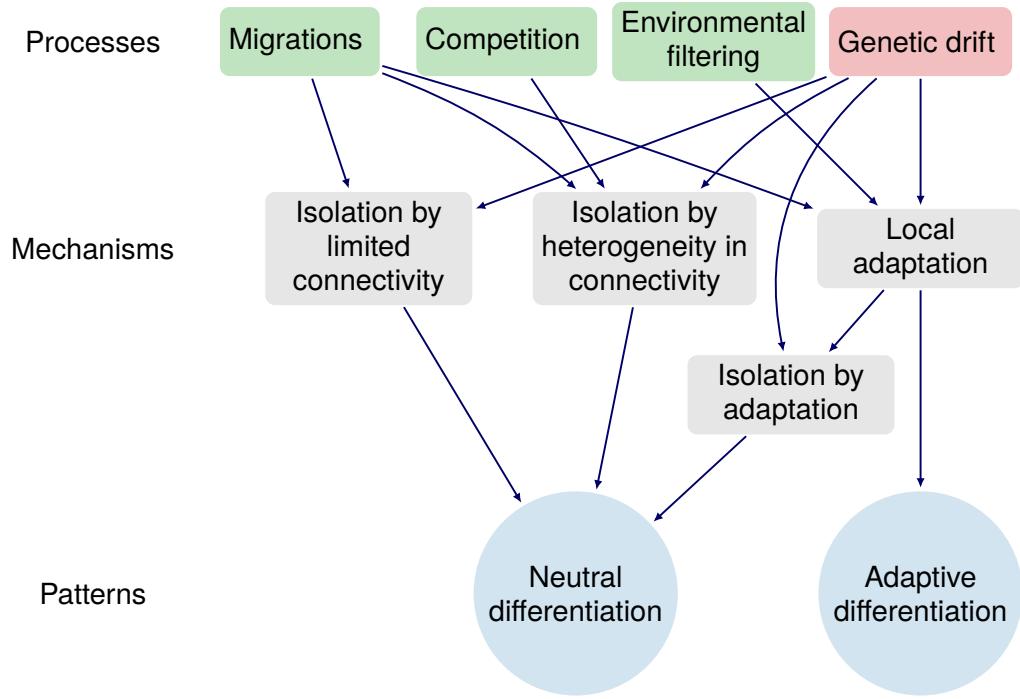
select individuals with traits that provide them higher fitness (Gaither et al., 2018). At the population level, this results in populations adapting to their local environment, a mechanism coined "local adaptation" (Kawecki and Ebert, 2004) and resulting in patterns of "adaptive differentiation". Adaptive differentiation is hindered by migration fluxes, which prevent local adaptation by destabilizing the evolution of traits towards the optimal (Meszéna et al., 1997; Débarre et al., 2013; Mirrahimi and Gandon, 2020). While adaptive differentiation concerns traits under selection, it indirectly affects the differentiation of neutral traits, that are co-evolving with traits under selection through linkages (Billiard et al., 2015; Lepers et al., 2021). This results in turn to the mechanism of "isolation by adaptation", where habitat heterogeneity, rather than landscape connectivity, increases neutral differentiation (Nosil et al., 2008). Simple mechanisms resulting in neutral and adaptive differentiation are identified, but how they are modulated by eco-evolutionary feedbacks and the topological features of realistic landscapes, presenting irregularities in connectivity (Dale and Fortin, 2010; Lieberman et al., 2005), is unclear.

In Chapter 2, I demonstrate a novel mechanism, involving both the heterogeneity in connectivity of the landscape and the process of intra-specific competition, that considerably affects neutral differentiation. Through the creation of unbalanced migration fluxes, heterogeneity in connectivity increases the intensity of competition in highly connected populations. This, in turn, reduces gene flow and reinforces neutral differentiation. I also investigate the mechanism of local adaptation when habitat connectivity is irregular (Dale and Fortin, 2010; Lieberman et al., 2005). I show that the complexity of habitat spatial distribution can be reduced to a measure of habitat spatial auto-correlation, coined the "habitat assortativity". Landscapes characterized by a high habitat assortativity support populations that are systematically better adapted than in landscape with low assortativity. Specifically, I provide an analytical condition for local adaptation (Eq. (2.7)), that sheds light on how it relates to dispersal intensity, selection strength, habitat heterogeneity, and habitat assortativity.

Because habitat assortativity affects local adaptation, it must also affect neutral differentiation through the mechanism of isolation by adaptation (Orsini et al., 2013). Closing the loop, I demonstrate that habitat assortativity affects neutral differentiation through two antagonistic effects. By favoring local adaptation, it promotes isolation by adaptation, therefore increasing neutral differentiation. In parallel, it favors gene flow within clusters of similar environmental conditions, decreasing isolation by limited dispersal. This results in habitat assortativity decreasing neutral differentiation for low dispersal intensity, and increasing neutral differentiation for high dispersal intensity (Fig. 2.5). This complex feedback is essential to understand population differentiation in complex landscapes. Figure 6.1 summarizes the feedback mechanisms underpinned in Chapter 2. Overall, Chapter 2 establishes a comprehensive map of causal pathways involved in the phenotypic differentiation of populations structured in complex landscapes.

### 6.1.2 Linking economic patterns to eco-evolutionary processes

Combining fine-grained economic data and process-based models, Chapter 4 bridges evolutionary economics, complexity economics and biology, advancing our general understanding



**Fig. 6.1:** Graphical representation of the eco-evolutionary mechanisms underpinned in Chapter 2 and involved in neutral and adaptive differentiation. Ecological processes are displayed in green boxes, evolutionary processes are displayed in red boxes.

of the endogenous forces driving the dynamics of economic systems. Neoclassical economics and evolutionary economics seek to explain economic change with process-based models. While Neoclassical economics explains economic change by exogenous changes (Boschma and Frenken, 2005), evolutionary economics rather focuses on endogenous forces, considering interactions between firms and economic activities, and evolutionary processes acting upon them as fundamental drivers of economic change (Metcalfe et al., 2006). Yet, evolutionary economics has up to now restricted its attention to subnational industrial dynamics (Hodgson, 2019). In contrast, complexity economics (Hidalgo, 2021) predicts economic change at the national level with endogenous predictors, but is agnostic to the underlying processes (Hidalgo, 2021). Rather, it relies on statistical techniques to process fine-grained economic data (Mealy et al., 2019), and a current concern is to unfold the causal processes (Hidalgo, 2021).

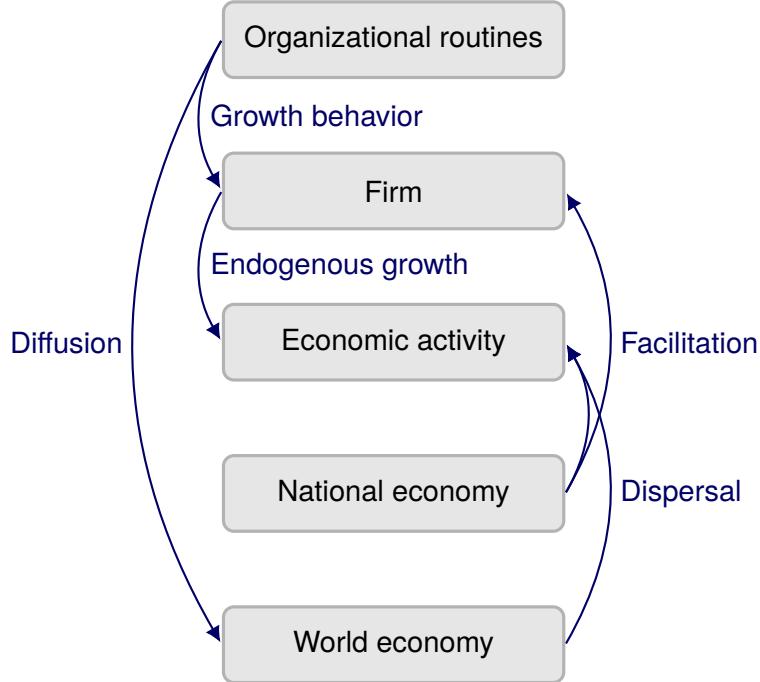
Relying on inverse modelling, Chapter 4 combines process-based models and fine-grained data to underpin the processes responsible for economic change. The approach undertaken relies on deep connections between processes acting upon economic activities and biological populations. Analogously to biological populations that are characterized by genes, economic activities are characterized by organizational routines (Nelson, 1985), which experience evolutionary processes and define how they engage in interactions with other economic activities (Nelson, 1985). As a result, economic activities can be considered as autonomous entities, whose dynamics is determined by its characteristics and the processes acting upon

them (Boschma and Frenken, 2005). The processes at stake leave characteristic signatures on the temporal dynamics of economic activities, consisting in distinctive temporal variations and couplings. These signatures can be identified with inverse modelling methods, estimating the support of population dynamic models embedding alternative hypothetical processes.

Chapter 4 tests whether the dynamics of economic activities at the national scale can be explained by positive and negative interactions between them, spatial dispersal processes, and/or economic transformations. Using population dynamic models capturing the different interdependencies, Chapter 4 provides quantitative evidence that economic activities engage in positive interactions and disperse across countries. Positive interactions may arise from a variety of processes operating on firms, such as beneficial business interactions through supply chains (Ozman, 2009; Saavedra et al., 2009) and agglomeration externalities (Van Der Panne, 2004). Its support implies that the dynamics of economic activities are highly inter-dependent, and suggests that diversity promotes economic development (Hidalgo, 2018). Spatial diffusion of routines (Hodgson and Knudsen, 2004) and knowledge spillovers (Caragliu and Nijkamp, 2016) may be at the origin of the spatial dispersal of economic activities (Zahra et al., 2000; Rogers, 2003; Boschma et al., 2008). We found discrepancies across countries in the strength-of-evidence for spatial dispersal, suggesting that economic activities are more akin to disperse in some countries than others. As an explanation, we propose that transfers of knowledge and routines may be blocked by barriers due to differences in cognitive, organizational, social, institutional or geographic proximity between countries (Boschma, 2005; Caragliu and Nijkamp, 2016). Figure 6.2 summarizes the ecological mechanisms evidenced in Chapter 4. Overall, Chapter 4 provides quantitative evidences that, akin to biological systems, processes of interaction and dispersal shape the dynamics of economic systems at the country scale.

### 6.1.3 Advances in the modelling of realistic spatial and phenotypic structures

Chapters 2 and 5 deliver new methods to incorporate important features of empirical systems within eco-evolutionary models. Evolutionary dynamics have traditionally been studied in the context of regular population structures (Lieberman et al., 2005). For instance, to investigate differentiation in biological populations, a multitude of studies have considered regular spatial structures, missing the effect of spatial complexity on the underlying mechanisms (Slatkin, 1973; Slatkin, 1978; Kirkpatrick and Barton, 1997; Polechová and Barton, 2015; Polechová, 2018; AndradeRestrepo et al., 2019; Doebeli and Dieckmann, 2003; Meszéna et al., 1997; Yeaman and Otto, 2011; Débarre et al., 2013; Mirrahimi and Gandon, 2020). Biological habitats differ in their connectivity (Dale and Fortin, 2010), and economic entities are structured through complex networks (Schweitzer et al., 2009). Lieberman et al., 2005 and subsequent studies of "evolutionary dynamics on graphs" (e.g., Tkadlec et al., 2019) show that this complexity considerably affects the interplay between selection and drift. However, evolutionary dynamics on graphs does not consider eco-evolutionary feedbacks (Govaert et al., 2019). Thus far, models that capture eco-evolutionary feedbacks together with realistic, complex population structures were missing.



**Fig. 6.2:** Graphical representation of the eco-evolutionary processes evidenced in Chapter 5, and how they affect the different organizational levels in economic systems. Chapter 5 provides quantitative evidences for the processes of mutualistic interactions (facilitation) and spatial dispersal, respectively coupling the world economy organizational level and the national economy organizational level, to the economic activity level. Directly depending on firms, the economic activity level is also coupled to the firm organizational level. Furthermore, in Chapter 5, we hypothesize that dispersal processes originate from the diffusion of organizational routines, coupling the organizational routines level to the world economy level.

Another important feature of realistic biological populations that may affect their dynamics is the variety of traits that characterize them (Doebeli and Ispolatov, 2010b). While a vast majority of studies on eco-evolutionary feedbacks has focused on the evolution of scalar phenotypes (Doebeli and Ispolatov, 2010b), in most organisms, many phenotypic properties combine in complicated ways to determine ecological processes (Doebeli and Ispolatov, 2014). For instance, Doebeli and Ispolatov, 2010b shows that the consideration of multiple traits is likely to generate more phenotypic diversity than expected with one dimensional models. Trade-offs in traits are also essential features shaping the evolutionary dynamics of biological populations, with consequences on the dynamics of e.g. cancer cell evolution (Fiandaca et al., 2021) and plankton dynamics (Le Gland et al., 2020). Yet the simulation of eco-evolutionary models of biological populations structured over high dimensional phenotypic spaces is tremendously difficult, since the numerical cost of traditional methods grows exponentially in the number of dimensions considered (Bellman, 2010).

From first principles, Chapter 2 derives a stochastic individual-based model capturing eco-evolutionary feedbacks in populations structured in complex landscapes and high dimensional phenotypic space. Chapter 5 further provides tools to efficiently simulate a

deterministic approximation of the individual-based model in high dimensional phenotypic spaces. The model presented in Chapter 2 involves the combination of graphs and high dimensional phenotypic spaces to model realistic population structures. Naturally capturing eco-evolutionary feedbacks, the model can readily be generalized to include other eco-evolutionary processes (see Section 2.A.3 for an extended variant with trait-based competition). The Julia library **Evoid.jl**, that I have written for the purpose of the numerical experiments in Chapter 2 and that is publicly available (Boussange, 2021), already implements a generic version of the model. As such, the model presented in Chapter 2 and the Julia library **Evoid.jl** may be used to investigate other questions involving complex population structures (Lieberman et al., 2005) and the co-evolution of characteristics (Doebeli and Ispolatov, 2010b). While they can realistically reproduce the discrete and stochastic nature of ecological and evolutionary processes (DeAngelis and Mooij, 2005), numerical simulations of individual-based model may not provide a general understanding of system investigated (Lion, 2016; Hodgson, 2019), and cannot be scaled to simulate large systems involving a large number of individuals (DeAngelis and Mooij, 2005). Yet, the individual-based model proposed in Chapter 2 is mathematically tractable, and can be approximated with a deterministic PDE approximation under simplifying assumptions. The mathematical tractability allows obtaining analytical insights on how the interplay between the processes acting upon the individuals transform into mechanisms affecting the population level (Chapter 2). The PDE approximation, combined with the numerical methods presented in Chapter 5, further allow efficient simulations. The numerical methods proposed in Chapter 5 are now implemented in a new registered Julia package that I released under the name of **HighDimPDE.jl** (Boussange, 2022a), now belonging to the SciML organisation (Rackauckas et al., 2020a). The package aims at hosting many more solver algorithms that can efficiently simulate high dimensional PDEs, and has, as of September 2022, already received contributions from 5 independent developers (see *Contributions to the main branch of HighDimPDE.jl* 2022). These contributions may greatly enhance **HighDimPDE.jl** over the years, promising efficient simulations of eco-evolutionary models. Together, Chapters 2 and 5 deliver novel tools to advance our understanding of the effect of the complexity of spatial and phenotypic structures on eco-evolutionary dynamics.

### 6.1.4 Advances in inverse modelling for identifying eco-evolutionary processes in empirical systems

Our understanding and prediction of eco-evolutionary dynamics in ecological and economic systems critically depends on the confrontation of process-based models with empirical data (Pelletier et al., 2009; Hidalgo, 2021). The most celebrated inference methods for inverse modelling in biology are Bayesian inference methods with Markov Chain Monte Carlo (Lignell et al., 2013; Higgins et al., 2010; Xu et al., 2006; Fiechter et al., 2013; Rosenbaum et al., 2019) and variational methods (Schartau et al., 2017). Bayesian inference methods require numerous forward model integrations (Schneider et al., 2017), and are highly affected by the number of model parameters (Csilléry et al., 2010). Variational methods require the model sensitivity to its parameters (Schartau et al., 2017) and are prone to converge to local minima, especially with complex models (Gábor and Banga,

2015). Those central issues likely explain the very limited use of inverse modelling to further our knowledge on eco-evolutionary processes in ecological systems (but see Sukumaran et al., 2016; Skeels and Cardillo, 2019; Skeels et al., 2022 that use approximate Bayesian computation methods).

Chapter 3 presents a novel inverse modelling framework that allows to estimate the parameter values and the support of complex eco-evolutionary models from temporal data. The framework is based on a variational method, but resolves its main shortcomings by heavily relying on automatic differentiation (Rackauckas et al., 2020a), state-of-the-art optimizers (Kingma and Ba, 2014), and a learning strategy based on a mini-batch method. The use of automatic differentiation simply eliminates the effort required to obtain the model sensitivity to its parameters, and the state-of-the-art optimizers, together with the mini-batch method, ensure the efficiency and robustness of the method in handling highly nonlinear models. Chapter 3 takes part in an ongoing effort to blend machine learning (ML) and traditional models to gain scientific understanding and extrapolability (Karpatne et al., 2017; Rackauckas et al., 2020a; Schneider et al., 2017; Rolnick et al., 2022; Kashinath et al., 2021; Yazdani et al., 2020; Raissi et al., 2019a). In physical systems such as oceanic and atmospheric systems, general organizational principles are known and formulated in general circulation models, where ML is mostly used to accelerate simulations (Kashinath et al., 2021) and improve model forecast skill (Schneider et al., 2017). In contrast, general models of ecological and economic systems are yet to be formulated, and methods such as the ML framework presented in Chapter 3 can greatly contribute to identifying the general organizational principles required to reach this goal (Karpatne et al., 2017). By contrasting competing hypotheses embedded in alternative models, Chapters 3 and 5 provide concrete examples, both with synthetic and empirical data, that the inverse modelling framework in Chapter 3 can successfully elucidate eco-evolutionary mechanisms. Integrating the practical constraints of current ecological datasets (Dornelas et al., 2018), the inverse modelling framework may also be relevant for improving the forecast ability of existing eco-evolutionary model (Norberg et al., 2012). Built thanks to the composability of the celebrated differential equation solver **DifferentialEquations.jl** and the deep learning library **Flux.jl**, the inverse modelling framework is implemented in a new, multipurpose Julia package readily available to the scientific community, that I released under the name of **MiniBatchInference.jl** (Boussange, 2022b). Together, the inverse modelling framework proposed in Chapter 3 successfully blends ML methods with mechanistic ecosystem models to gain scientific knowledge from observation data.

## 6.2 Limitations

### 6.2.1 Computational complexity for modelling the evolution of full phenotypic densities

Alternative methods to those presented in Chapters 2 and 5 may be more appropriate for the forward modelling of eco-evolutionary dynamics. While individual-based models are interesting tools to investigate stochastic drift in finite size populations, the Gillespie

algorithm (Gillespie, 1976) employed to simulate the individual-based model in Chapter 2 is computationally intensive, and requires to compute the fitness of all individuals at each birth or death event, which depends on the characteristics of all the other individuals. The resulting computational complexity scales poorly with the number of individuals involved, preventing its use to model large populations. PDE approximations are computationally more efficient in large populations for low dimensional phenotypic spaces ( $\lesssim$  3-dimensional). The methods presented in Chapter 5 can efficiently simulate PDE models in higher dimensions (demonstrated up to 10 traits), but still suffer from a number of issues that may prevent their practical use. First, the MLP method can only provide the population number for one single trait value in one run. Consequently, the MLP method cannot characterize the total population density with a reasonable computational complexity. In contrast, the full population density can be obtained with the ML-based approximation method, but the method involves the training of many neural networks (one at each time step). This is worrying, since long time horizons are required by practitioners, and that the training of a neural network is numerically costly. Another problem with the numerical methods proposed in Chapter 5 is that they involve the tuning of meta parameters, including the choice of a measure for the integration of the non-local term ( $\nu_x$  in Lemmas 5.3.1 and 5.4.1). Though this choice is critical to the success of the numerical simulations, it is unclear how it can be determined. While greatly decreasing the computational cost of traditional methods, the novel methods proposed in Chapter 5 are still computationally demanding and require a delicate tuning of the meta-parameters.

Because PDE models track the evolution of the full phenotypic density of populations, PDE models inevitably require a considerable computational effort – irrespective of the numerical method used. Nevertheless, only the first three moments of the population density are usually of interest, namely population size, trait mean and trait variance (Nordbotten et al., 2020). Instead of seeking to numerically approximate the full phenotypic density, moment closure approximation methods (Wickman et al., 2021; Lion et al., 2022; Nordbotten et al., 2020) may instead be considered. Those approaches consist in approximating the population density with a multivariate Gaussian distribution. This, in turn, allows to transform the PDE problem into a system of coupled differential equations involving the time evolution of the population size (1 variable for a single species population), the mean trait values ( $d$  scalar variables), and the variance-covariance matrix of the multidimensional trait density ( $d^2$  variables). As such, the computational cost of this method only scales polynomially with the number of dimension ( $\mathcal{O}(d^2)$ ), while providing the sufficient information required to investigate eco-evolutionary dynamics in high dimensional phenotypic spaces. It is worth noting that instead of using neural networks, multivariate Gaussian functions could also be used within the ML-based method for simulating eco-evolutionary models. Equivalent to the simplifying assumption taken with moment closure methods, we expect this approach to greatly improve the computational efficiency of the ML-based method, as they involve fewer free parameters ( $d(d + 1) + 1$ , i.e. 111 parameters for  $d = 10$ ) than neural networks ( $(d + 50)(2d + 50) + 3(d + 50) + 1$  parameters used for the examples in Sections 5.5.1 to 5.5.5, i.e. 4381 parameters for  $d = 10$ ). In summary, using the ML-based approximation presented in Chapter 5 with multivariate Gaussian functions may considerably lower the number of iterations required in the training process, while reducing the computational cost.

The proposed methods in Chapters 2 and 5 suffer from considerable computational cost, because they seek to simulate PDE models which track the evolution of the full phenotypic distribution of populations. Because only the population number, the mean and the variance-covariance matrix of the phenotypic distributions are of interest, closure approximation methods could be considered. Those methods are compatible with the ML-based method proposed in Chapter 5.

### 6.2.2 Uncertainty estimation and model selection bias

The inverse modelling framework proposed in Chapter 3 and used in Chapter 4 also presents pitfalls. First, the mini-batching learning strategy requires the choice of a mini-batch size to ensure the convergence to the maximum likelihood estimate. This choice should be motivated by the roughness of the model likelihood landscape (see Section 3.2.2), but may affect the model selection process: a small batch size implies that the model goodness-of-fit is evaluated on the fast dynamics of the data, while the resulting support could differ, were the model fitted with a higher mini-batch size. Theoretical developments are required to provide statistically justified guidance for the choice of the mini-batch size. Second, the maximum likelihood estimate of models with complex likelihood landscape may be underestimated, because not correctly identified. As a result, the model selection process may be biased towards models which associated likelihood landscape is easier to navigate. Third, the information criterion-based model selection procedure used in Chapters 3 and 4 is uniquely based on a trade-off between the goodness-of-fit and the number of parameters of the model, which may not be satisfactory to characterize process-based models (Clermont and Zenker, 2015). Other criterion, for instance based on the complexity of the dynamical behavior of the model (such as, e.g., its Lyapunov exponent), could be developed. Fourth, the inverse modelling framework developed in Chapter 3 requires a differentiable model, a strong prerequisite that may not be met by, e.g. stochastic models. Fifth, the inverse modelling framework only provides the maximum of the posterior distribution, while the posterior distribution may be multimodal, with alternative modes carrying valuable information (Wilson, 2020).

Alternatively, Sukumaran et al., 2016; Skeels and Cardillo, 2019; Skeels et al., 2022 employ variants of approximate Bayesian computation methods (Csilléry et al., 2010) for eco-evolutionary model selection. Their approach consists in aggregating model simulation outputs into summary statistics, used to train classifier algorithms (such as, e.g., random forests or neural networks) in recognizing the signatures of the competing models. Once trained, the classifier algorithms are used on the summary statistics obtained from the empirical data, discriminating between the alternative hypotheses. This approach does not require model differentiability, and is consequently more flexible than the method proposed in Chapter 3. Also, the use of summary statistics can elucidate which particular feature of the empirical data is better explained by a given model. Nevertheless, this strength is a pitfall: summary statistics necessarily reduce the information contained in empirical data, which can prevent to correctly discriminate between models (Csilléry et al., 2010). Together, the mini-batch inverse modelling framework in Chapter 3 is sensitive to the mini-batch size, requires models to be differentiable, and does not provide uncertainty estimation. While approximate

Bayesian computation may be a valuable alternative, it also presents restrictive shortcomings. Still, the mini-batch learning strategy extends beyond the framework proposed, and could be used in combination with novel approaches in Bayesian computation to combine the best of both worlds.

## 6.3 Perspectives

### 6.3.1 Development opportunities in inverse modelling

The mini-batch method presented in Chapter 3 and the ML-based approximation method developed in Chapter 5 offer unique development opportunities to leverage inverse modelling. The mini-batch method is relevant beyond the ML framework presented in Chapter 3, and could be used within a fully Bayesian framework, where the full posterior distribution of the model is estimated. Compared to considering a standard formulation of the likelihood function such as in Eq. (3.4), using the mini-batch loglikelihood formulation proposed in Eq. (3.8) results in a smoother posterior distribution, which should reduce the large number of forward model integration required by Bayesian inference. While this number could still be prohibitively expensive for Bayesian inference with MCMC chains, automatic differentiation variational inference (ADVI, Morningstar et al., 2020; Gosh et al., 2021) offers an appealing alternative. In ADVI, the posterior distribution is approximated by a Gaussian distribution (Morningstar et al., 2020), significantly reducing the number of model integration (Morningstar et al., 2020). Improving the ML framework presented in Chapter 3, ADVI could capture multimodality in the model posterior distribution (by approximating the multimodal distribution with a gaussian distribution with large variance). This, in turn, could improve the robustness of model selection (Section 6.2), and provide uncertainties estimations on the parameters. Bayesian Learning via Stochastic Gradient Langevin Dynamics (Welling and Teh, 2011) is an alternate solution to ADVI, that could readily be used with the ML framework proposed in Chapter 3 instead of the optimization algorithms suggested. This algorithm builds upon recent advances in Bayesian deep learning (Wilson, 2020), and interprets the iterative gradient-based optimization procedure as a Markov chain with an equilibrium distribution over the posterior distribution of the model parameters. It therefore comes with the scalability of variational methods and the interpretability of Bayesian methods, providing uncertainty estimates while ensuring computational efficiency.

For inverse modelling with PDE models, an extension of the ML-based approximation method presented in Chapter 5 could yield interesting performance. In Lemma 5.3.1, the parameters of the PDE model are assumed fixed, but could be set as free parameters, analogously to the parameters of the neural networks used for approximating the PDE solution. The loss function in Eq. (5.18) would then take the PDE model parameters as additional arguments, and include an additional term, penalizing the distance between the PDE model solution and the data. This term, analogous to Eq. (3.8) in Chapter 3, would constrain the PDE parameters, similarly to the training of physics informed neural networks (Raissi et al., 2019a; Yazdani et al., 2020). Yet in contrast to Raissi et al., 2019a; Yazdani et al., 2020, a major advantage of this approach is to efficiently perform inverse modelling

with high-dimensional dynamical models. Because Julia is a programming language with pervasive automatic differentiation (see Section 1.2.4 in Chapter 1), this development would require little effort with the Julia library **HighDimPDE.jl**. Together, the ML methods developed in Chapters 3 and 5 offer unique opportunities to grant more robustness and efficiency to inverse modelling methods, providing uncertainty estimation and the possibility to handle high dimensional models.

### 6.3.2 Confronting eco-evolutionary model on spatial graphs and empirical data

The confrontation of the predictions obtained in Chapter 2 with empirical data, and the use of empirical data to constrain the eco-evolutionary model on spatial graphs presented in Chapter 2, could advance our understanding of eco-evolutionary dynamics in empirical systems. In Chapter 2, we predicted three topology metrics to correlate with standard population differentiation metrics ( $Q_{ST}$  metrics). By projecting real landscapes on spatial graphs (see Dale and Fortin, 2010 and Fig. S2.13), the topology metrics, together with empirical data on population differentiation, could be used to verify our predictions. Discrepancies may indicate that additional processes are involved in population differentiation in empirical systems. On the other hand, a validation of the predictions from Chapter 2 could help to predict population differentiation at a global scale. These predictions could, in turn, be linked to patterns of species richness, in order to underpin how population differentiation may lead to speciation over time (Templeton, 1981).

Hindcasts with the eco-evolutionary model on spatial graphs over geological time, using paleo-climatic data (Straume et al., 2020) as environmental condition forcings, could also help to address fundamental questions on the processes involved in spatial biodiversity patterns. In contrast to deep-time spatially explicit eco-evolutionary models (see Hagen, 2022 for a review), the model developed in Chapter 2 does not assume a timescale separation between ecological and evolutionary processes. This would allow disentangling, for instance, the role of feedbacks between trait-based competition (Dieckmann and Doebeli, 1999) – succinctly considered in Chapter 2 (see Section 2.A.3) –, and environmental filtering (Doebeli and Dieckmann, 2003), in shaping patterns of species distribution (Rahbek et al., 2019a).

Along shorter time scales, the eco-evolutionary model on spatial graph could be calibrated on empirical data of species distribution (see, e.g., Abegg et al., 2020; GBIF: The Global Biodiversity Information Facility, 2022b) with the inverse modelling method proposed in Chapter 3, to better predict how biological populations will adapt to climate change (Norberg et al., 2012; Urban et al., 2016). Using climate projections (Kriticos et al., 2012) as environmental forcings, the model could capture the processes of spatial dispersal and adaptation, that are fundamental yet neglected in models currently used for biodiversity forecasts (Pearson and Dawson, 2003). Last but not least, multiple species or functional groups could also be accounted for in the model proposed in Chapter 2. Tracking suitable environmental conditions, species ranges are expected to shift towards higher elevations and higher latitudes (Chen et al., 2011), but how this will modify trophic interactions and ecosystem dynamics is unclear (Descombes et al., 2020). The model would provide a

framework to explore how global warming may generate feedbacks between the processes of dispersal, species' evolutionary response, and trophic interactions, collectively affecting ecosystem dynamics. In summary, confronting the predictions of Chapter 2 with empirical data, extensions of the model developed in Chapter 2, together with the use of the ML framework presented in Chapter 3, could bring further insights on the actual mechanisms involved in empirical systems, and help to predict their responses to climate change.

### 6.3.3 Econobiology, a new venue to understanding economic systems, and designing more appropriate governance

The success of the biologically inspired model presented in Chapter 4 in characterizing the dynamics of economic activities calls for a deeper investigation of commonalities and differences in organizational principles in ecological and economic systems. As a first step, I suggest investigating whether the results of Chapter 4 hold for finer levels of economic activity aggregation, as more detailed data on economic activities is available (Hausmann and Hidalgo, 2011). Second, the eco-evolutionary models should be complexified, to investigate whether it can capture more detailed aspects of the dynamics. In parallel, a promising research direction is to better understand how capital fluxes (King and Levine, 1993) compare with fluxes of matter and energy (Veldhuis et al., 2018) in ecological systems. Veldhuis et al., 2018 provides an excellent synthesis of our understanding of how ecosystem organization emerges through self-reinforcing eco-evolutionary mechanisms. This synthesis is an interesting roadmap to formulate hypotheses and investigate analogous self-reinforcing mechanisms determining fluxes of capital in economic systems. There may be deep connections between nutrient cycles and capital cycles, where ecological compartments and financial intermediaries develop to improve the efficiency of nutrient flows and capital allocations (Odum, 1969). Veldhuis et al., 2018; King and Levine, 1993 show that, in both ecological and economic systems, this efficiency sustains productivity and growth, but that the developmental process may fail. In the latter case, ecosystems may stay in a state of lower maturity (Margalef, 1963) and economies may be stuck in a technological lock-in (Hidalgo et al., 2007). I expect that this research venue will pinpoint key organizational differences between ecological and economic systems. Since they have undergone evolution for a much longer time than economic systems, ecological systems may be more effective and resilient than economic systems, providing guidance for economic policies. Together, future directions should seek to understand similarities and differences in organizational principles in ecological and economic systems, which may eventually help improving economic regulations.

## 6.4 Concluding remarks

Bridging biology, mathematical modelling, machine learning and economics, this thesis advances our general understanding of eco-evolutionary processes and feedbacks in ecological and economic systems. It also provides novel modelling tools, that were publicly released in Julia packages to join efforts and collectively advance this understanding. While we now

better understand eco-evolutionary feedbacks in spatially structured populations, we are still far from a general understanding of their role at the ecosystem level. The challenge is enormous, because anthropogenic pressure and climate change are rapidly affecting ecosystems. Methods similar to those presented in this thesis, combining machine learning and process-based models, can massively accelerate the underpinning of general principles at the ecosystem level, and provide reliable forecasts. This, in turn, should help designing adequate management of ecosystem services to mitigate the consequences of human development. Machine learning, together with evolutionary biology, can also help us to underpin general principles in economic systems. Paving the way, this thesis demonstrates that analogous processes to those shaping ecosystems have major effects on economic change. Future research will tell whether economic and ecological systems have deeper organizational principles in common. Realizing overlaps and differences can help us to design better economic policies. Our young economic systems can surely learn from ecological systems, that have survived major crises for more than 3.5 billion years. Philosophically speaking, this realization invites us to demonstrate more humility towards ecological systems, and "*[...] to return, by a radiant self-forgetfulness, to the All of Nature*".



# CV

## Personal Information

<b>Residence</b>	Zürich, Switzerland
<b>E-mail</b>	bvictor@ethz.ch
<b>Website</b>	vbuossange.github.io
<b>Github</b>	github.com/vbuossange
<b>Age</b>	Born 1995 (age 27)
<b>Citizenship</b>	France citizen

## Personal skills

<b>Languages</b>	English (fluent) French (native) Spanish (B2) German (B1)
<b>Programming languages</b>	Julia Python C++ Java Matlab R Bash VBA
<b>Sports</b>	Ski mountaineering Alpinism Rock climbing Enduro mountainbiking Surfing
<b>Alpine CV</b>	[vbuossange.github.io/pages/alpine_cv/]

## Education

- 10.2022 **Ph.D in Environmental Sciences**, Swiss Federal Institute for Forest, Snow and Landscape (WSL | Swiss Federal Institute of Technology Zurich, ETH), Switzerland  
 09.2018

*Forward and inverse modelling of eco-evolutionary processes in ecological and economic systems.* Under the guidance of Prof. Dr. Loïc Pellissier.

06.2017 **Full year academic exchange**, University of New South Wales (UNSW Sydney),  
09.2016 Australia

06.2017 **Master thesis in theoretical geomechanics**, UNSW Sydney | CSIRO, Australia  
02.2017

*Numerical continuation and bifurcation analysis for unconventional geomechanics.*  
Under the guidance of Dr. Thomas Poulet.

08.2018 **M.S. in Energy and Environmental Engineering**, Institut National Des Sciences

09.2013 Appliquées de Lyon (INSA Lyon), France

Three-year undergraduate engineering course in Energy and Environmental Systems, focused on Advanced Energy Systems and Efficiency.

08.2018 **B.S. in Mathematics and Physics**, Institut National Des Sciences Appliquées de

09.2013 Lyon (INSA Lyon), France

Ranking : 21/650 students.

## Professional appointments

08.2018 **R&D intern**, Compagnie National du Rhône (CNR), France

03.2018 Development of an Energy Management System based on various optimisation techniques for optimal production of renewable resources. Applications to EU sponsored projects: **Jupiter1000** (power-to-gas), **Move in pure** (vehicle-to-grid), **Marie-Galante island** (micro-grid)

## Publications

### Peer-reviewed

1. **Boussange, V.** & Pellissier, L., *Eco-evolutionary model on spatial graphs reveals how habitat structure affects phenotypic differentiation.* **Commun Biol** 5, 668 (2022). [[bioRxiv](#)]

### Preprints

1. **Boussange, V.**, Vilimelis-Aceituno, P., Pellissier, L., *Mini-batching ecological data to improve ecosystem models with machine learning* [[bioRxiv](#)] (2022), 46 pages. In review.

- 
2. **Boussange, V.**, Becker, S., Jentzen, A., Kuckuck, B., Pellissier, L., *Deep learning approximations for non-local nonlinear PDEs with Neumann boundary conditions*. [[arXiv](#)] (2022), 59 pages. Revision requested from Partial Differential Equations and Applications.

## Proceedings

1. Poulet, T., Alevizos, S., Veveakis, M., **Boussange, V.**, Regenauer-Lieb, K., *Episodic mineralising fluid injection through chemical shear zones*, ASEG Extended Abstracts (2018), 5 pages.

## In preparation

1. **Boussange, V.**, Sornette, D., Lischke, H., Pellissier, L., *Analogous forces to ecological interactions, dispersal and mutations shape the dynamics of economic activities*.

## Talks

- 07.2022 **Speaker**, HIGHDIMPDE.JL: A Julia package for solving high-dimensional PDEs, JuliaCon2022, online. [youtube.com/watch?v=4sXqGhhknT4](https://youtube.com/watch?v=4sXqGhhknT4)
- 06.2022 **Speaker**, Interpretable machine learning for forecasting dynamical processes in ecosystems, World Biodiversity Forum, Davos, Switzerland.
- 06.2022 **Invited speaker**, Investigating empirical patterns of biodiversity with mechanistic eco-evolutionary models, Seminar at the Theoretical Ecology and Evolution group, Universität Bern.
- 11.2021 **Invited speaker**, Numerical approximations of solutions of highly dimensional, non-local nonlinear PDEs, StAMBio seminar, St Andrews, UK.
- 10.2021 **Speaker**, Graph topology and habitat assortativity drive phenotypic differentiation in an eco-evolutionary model, Conference on Complex Systems, Lyon, France.
- 10.2021 **Speaker**, Using graph-based metrics to assess the effect of landscape topography on diversification, ECBC, Amsterdam, Netherlands.
- 09.2021 **Speaker**, Solving non-local nonlinear Partial Differential Equations in high dimensions with HighDimPDE.jl, International Conference on Computational Methods in Systems Biology, Bordeaux, France.
- 04.2021 **Speaker**, Responses of neutral and adaptive diversity to complex geographic population structure, Mathematical Population Dynamics, Ecology and Evolution, CIRM Marseille, France.

## Softwares

- |      |   |       |
|------|---|-------|
| 2022 | <b>MiniBatchInference.jl</b><br><a href="https://github.com/vboussange/MiniBatchInference.jl">github.com/vboussange/MiniBatchInference.jl</a> | Julia |
|------|---|-------|

A Julia package for maximum likelihood estimation and model selection of strongly nonlinear dynamical models.

2021	<b>HighDimPDE.jl</b> <a href="https://github.com/vboussange/HighDimPDE.jl">github.com/vboussange/HighDimPDE.jl</a>	Julia
	A Julia package that breaks down the curse of dimensionality in solving non local, non linear PDEs.	
2021	<b>EvoId.jl</b>	Julia
2019	<a href="https://github.com/vboussange/EvoId.jl">github.com/vboussange/EvoId.jl</a>	
	Evolutionary individual based modelling, mathematically grounded.	
2018	<b>OptiVPP</b> <a href="#">confidential</a>	Python, GAMS
	Energy Management System for Virtual Power Plants.	

## Open source software contributions

SciML  
DiffEqFlux.jl  
CUDA.jl  
Flux.jl  
LightGraphs.jl

## Teaching and supervision

12.2020	<b>701-3001-00L Environmental Systems Data Science</b> , ETH Zürich, D-USYS,
09.2020	Switzerland
06.2020	<b>262-0100-00L Lab rotation</b> , ETH Zürich, D-BSSE, Switzerland
04.2020	
12.2020	<b>Taste of research internship</b> , Polytech Nice-Sophia, France
09.2020	

## Reviews

2022	<b>Journal of Open Source Software</b>
2019	<b>Journal of Theoretical Biology</b>





# Bibliography

- Abegg, M., U.-B. Brändli, F. Cioldi, et al. (2020). *Swiss national forest inventory NFI* (cit. on p. 165).
- Abergel, Frederic and Remi Tachet (2010). “A nonlinear partial integro-differential equation from mathematical finance”. In: *Discrete Contin. Dyn. Syst.* 27.3, pp. 907–917 (cit. on p. 125).
- Acemoglu, Daron, Simon Johnson, and James A. Robinson (2005). “Chapter 6 Institutions as a Fundamental Cause of Long-Run Growth”. In: *Handbook of Economic Growth*. Vol. 1. SUPPL. PART A, pp. 385–472 (cit. on p. 101).
- Acemoglu, Daron, Simon Johnson, and James A Robinson (2001). “The colonial origins of comparative development: An empirical investigation”. In: *American economic review* 91.5, pp. 1369–1401 (cit. on p. 1).
- Aguilée, Robin, David Claessen, and Amaury Lambert (2012). “Adaptive radiation driven by the interplay of eco-evolutionary and landscape dynamics”. In: *Evolution* 67.5, no–no (cit. on p. 14).
- Aide, T. Mitchell, Carlos Corrada-Bravo, Marconi Campos-Cerqueira, et al. (2013). “Real-time bioacoustics monitoring and automated species identification”. In: *PeerJ* 1.1, e103 (cit. on p. 72).
- Al-Aradi, Ali, Adolfo Correia, Danilo de Frietas Naiff, Gabriel Jardim, and Yuri Saporito (2019). “Extensions of the Deep Galerkin Method”. In: *arXiv:1912.01455*, 27 pp. (Cit. on p. 127).
- Alber, Mark, Adrian Buganza Tepole, William R. Cannon, et al. (2019). “Integrating machine learning and multiscale modeling perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences”. In: *npj Digital Medicine* 2.1, p. 115. arXiv: 1910.01258 (cit. on p. 56).
- Aldrich, Howard E., Geoffrey M. Hodgson, David L. Hull, et al. (2008). “In defence of generalized Darwinism”. In: *Journal of Evolutionary Economics* 18.5, pp. 577–596 (cit. on p. 86).
- Alfaro, Matthieu and Rémi Carles (2017). “Replicator-mutator equations with quadratic fitness”. In: *Proc. Amer. Math. Soc.* 145.12, pp. 5315–5327 (cit. on p. 125).
- Alfaro, Matthieu and Mario Veruete (2019). “Evolutionary branching via replicator-mutator equations”. In: *J. Dynam. Differential Equations* 31.4, pp. 2029–2052 (cit. on p. 125).
- Altermatt, Florian and Emanuel A. Fronhofer (2018). “Dispersal in dendritic networks: Ecological consequences on the spatial distribution of population densities”. In: *Freshwater Biology* 63.1, pp. 22–32 (cit. on p. 29).

- Amadori, Anna Lisa (2003). "Nonlinear integro-differential evolution problems arising in option pricing: a viscosity solutions approach". In: *Differential Integral Equations* 16.7, pp. 787–811 (cit. on p. 125).
- Anceschi, Niccolo, Jorge Hidalgo, Carlos A. Plata, et al. (2019). "Neutral and niche forces as drivers of species selection". In: *Journal of Theoretical Biology* 483, p. 109969 (cit. on pp. 14, 30).
- Andersen, Otto (1993). "On the Internationalization Process of Firms: A Critical Analysis". In: *Journal of International Business Studies* 24.2, pp. 209–231 (cit. on pp. 86, 100).
- AndradeRestrepo, Martín, Nicolas Champagnat, and Régis Ferrière (2019). "Local adaptation, dispersal evolution, and the spatial ecoevolutionary dynamics of invasion". In: *Ecology Letters* 22.5. Ed. by Vincent Calcagno, pp. 767–777 (cit. on pp. 29, 158).
- Applegate, J. M. and Adam Lampert (2021). "Firm size populations modeled through competition-colonization dynamics". In: *Journal of Evolutionary Economics* 31.1, pp. 91–116 (cit. on pp. 4, 86).
- Bahar, Dany, Ricardo Hausmann, and Cesar A. Hidalgo (2014). "Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion?" In: *Journal of International Economics* 92.1, pp. 111–123 (cit. on pp. 86, 100, 101).
- Banerjee, Malay, Sergei V. Petrovskii, and Vitaly Volpert (2021). "Nonlocal reaction-diffusion models of heterogeneous wealth distribution". In: *Mathematics* 9.4, Article No. 351, 18 pp. (Cit. on p. 125).
- Banks, H.T., J.E. Banks, R. Bommarco, et al. (2017). "PARAMETER ESTIMATION FOR AN ALLOMETRIC FOOD WEB MODEL". In: *International Journal of Pure and Applied Mathematics* 114.1, pp. 143–160 (cit. on p. 71).
- Bansaye, Vincent and Sylvie Méléard (2015). "Some stochastic models for structured populations : scaling limits and long time behavior". In: *Stochastic Models for Structured Populations: Scaling Limits and Long Time Behavior*, pp. 1–107. arXiv: 1506.04165 (cit. on pp. 14, 16, 34).
- Barnosky, Anthony D., Elizabeth A. Hadly, Jordi Bascompte, et al. (2012). "Approaching a state shift in Earth's biosphere". In: *Nature* 486.7401, pp. 52–58. arXiv: 9605103 [cs] (cit. on pp. 7, 55).
- Barone, A., F. Esposito, C. J. Magee, and A. C. Scott (1971). "Theory and applications of the sine-Gordon equation". In: *La Rivista del Nuovo Cimento* 1.2, pp. 227–267 (cit. on pp. 125, 137, 143).
- Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen (2003). "The nested assembly of plant-animal mutualistic networks". In: *Proceedings of the National Academy of Sciences* 100.16, pp. 9383–9387. arXiv: arXiv:1207.6416 (cit. on p. 101).
- Bauer, W. F. (1958). "The Monte Carlo Method". In: *J. Soc. Ind. Appl. Math.* 6.4, pp. 438–451 (cit. on p. 126).
- Beaumont, Mark A. (2010). "Approximate Bayesian Computation in Evolution and Ecology". In: *Annual Review of Ecology, Evolution, and Systematics* 41.1, pp. 379–406. arXiv: 1212.1417 (cit. on p. 71).

- Beck, Christian, Sebastian Becker, Patrick Cheridito, Arnulf Jentzen, and Ariel Neufeld (2021a). “Deep splitting method for parabolic PDEs”. In: *SIAM J. Sci. Comput.* 43.5, A3135–A3154 (cit. on pp. 126–128).
- Beck, Christian, Sebastian Becker, Philipp Grohs, Nor Jaafari, and Arnulf Jentzen (2021b). “Solving the Kolmogorov PDE by means of deep learning”. In: *J. Sci. Comput.* 88, Article No. 73, 28 pp. (Cit. on pp. 126, 128).
- Beck, Christian, Weinan E, and Arnulf Jentzen (2019). “Machine learning approximation algorithms for high-dimensional fully nonlinear partial differential equations and second-order backward stochastic differential equations”. In: *J. Nonlinear Sci.* 29.4, pp. 1563–1619 (cit. on pp. 141, 142, 144, 152).
- Beck, Christian, Martin Hutzenthaler, Arnulf Jentzen, and Benno Kuckuck (2020). “An overview on deep learning-based approximation methods for partial differential equations”. In: *Revision requested from Discrete Contin. Dyn. Syst., arXiv:2012.12348*, 22 pp. (Cit. on pp. 126, 128, 135).
- Becker, Sebastian, Ramon Braunwarth, Martin Hutzenthaler, Arnulf Jentzen, and Philippe von Wurstemberger (2020). “Numerical Simulations for Full History Recursive Multilevel Picard Approximations for Systems of High-Dimensional Partial Differential Equations”. In: *Commun. Comput. Phys.* 28.5, pp. 2109–2138 (cit. on p. 135).
- Becks, Lutz, Stephen P. Ellner, Laura E. Jones, and Nelson G. Hairston (2012). “The functional genomics of an eco-evolutionary feedback loop: linking gene expression, trait evolution, and community dynamics”. In: *Ecology Letters* 15.5, pp. 492–501 (cit. on p. 4).
- Bell, Graham (2017). “Evolutionary Rescue”. In: *Annual Review of Ecology, Evolution, and Systematics* 48.1, pp. 605–627 (cit. on p. 86).
- Bellman, Richard (2010). *Dynamic Programming*. Princeton Landmarks in Mathematics. Reprint of the 1957 edition. Princeton University Press, Princeton, NJ, pp. xxx+340 (cit. on pp. 6, 126, 159).
- Benincà, Elisa, Jef Huisman, Reinhard Heerkloss, et al. (2008). “Chaos in a long-term experiment with a plankton community”. In: *Nature* 451.7180, pp. 822–825 (cit. on pp. 7, 55, 59, 69).
- Benth, Fred Espen, Kenneth Hvistendahl Karlsen, and Kristin Reikvam (2001). “Optimal portfolio selection with consumption and nonlinear integro-differential equations with gradient constraint: a viscosity solution approach”. In: *Finance Stoch.* 5.3, pp. 275–303 (cit. on p. 125).
- Berestycki, Henri, Tianling Jin, and Luis Silvestre (2016). “Propagation in a non local reaction diffusion equation with spatial and genetic trait structure”. In: *Nonlinearity* 29.4, pp. 1434–1466 (cit. on p. 125).
- Berestycki, Henri, Grégoire Nadin, Benoit Perthame, and Lenya Ryzhik (2009). “The non-local Fisher-KPP equation: travelling waves and steady states”. In: *Nonlinearity* 22.12, pp. 2813–2844 (cit. on pp. 125, 136, 141).
- Bertino, Laurent, Geir Evensen, and Hans Wackernagel (2003). “Sequential Data Assimilation Techniques in Oceanography”. In: *International Statistical Review* 71.2, pp. 223–241 (cit. on p. 56).

- Bezanson, Jeff, Jiahao Chen, Benjamin Chung, et al. (2018). "Julia: dynamism and performance reconciled by design". In: *Proceedings of the ACM on Programming Languages* 2.OOPSLA, pp. 1–23 (cit. on p. 8).
- Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B. Shah (2017). "Julia: A fresh approach to numerical computing". In: *SIAM Review* 59.1, pp. 65–98. arXiv: 1411.1607 (cit. on p. 8).
- Bian, Shen, Li Chen, and Evangelos A. Latos (2017). "Global existence and asymptotic behavior of solutions to a nonlocal Fisher-KPP type problem". In: *Nonlinear Anal.* 149, pp. 165–176 (cit. on pp. 125, 136, 141).
- Billiard, Sylvain, Régis Ferrière, Sylvie Méléard, and Viet Chi Tran (2015). "Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks". In: *Journal of Mathematical Biology* 71.5, pp. 1211–1242. arXiv: 1310.6274 (cit. on pp. 14, 30, 156).
- Bjørnstad, Ottar N. and Bryan T. Grenfell (2001). "Noisy Clockwork: Time Series Analysis of Population Fluctuations in Animals". In: *Science* 293.5530, pp. 638–643 (cit. on pp. 59, 63, 69).
- Bocquet, Marc, Julien Brajard, Alberto Carrassi, and Laurent Bertino (2019). "Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models". In: *Nonlinear Processes in Geophysics* 26.3, pp. 143–162 (cit. on p. 58).
- Bonan, Gordon B. (2008). "Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests". In: *Science* 320.5882, pp. 1444–1449 (cit. on p. 55).
- Boschma, Ron (2005). "Proximity and Innovation: A Critical Assessment". In: *Regional Studies* 39.1, pp. 61–74 (cit. on pp. 101, 158).
- Boschma, Ron, Rikard Eriksson, and Urban Lindgren (2008). "How does labour mobility affect the performance of plants? The importance of relatedness and geographical proximity". In: *Journal of Economic Geography* 9.2, pp. 169–190 (cit. on pp. 86, 100, 158).
- Boschma, Ron A and Koen Frenken (2005). "Why is economic geography not an evolutionary science? Towards an evolutionary economic geography". In: (cit. on pp. 157, 158).
- Bottou, Léon (2012). "Stochastic gradient descent tricks". In: *Neural networks: Tricks of the trade*. Springer, pp. 421–436 (cit. on p. 71).
- Bounova, Gergana and Olivier de Weck (2012). "Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles". In: *Physical Review E* 85.1, p. 016117 (cit. on p. 18).
- Boussange, Victor, Pau Vilimelis, and Loïc Pellissier (2022). "Mini-batching ecological data to improve ecosystem models with machine learning". In: (cit. on pp. 85, 87, 91, 92).
- Boyd, Ian L. (2012). "The Art of Ecological Modeling". In: *Science* 337.6092, pp. 306–307 (cit. on pp. 7, 55).
- Bromberger, Seth and other Contributors (2017). "JuliaGraphs/LightGraphs.jl". In: (cit. on p. 33).

- Brummitt, Charles D., Andrés Gómez-Liévano, Ricardo Hausmann, and Matthew H. Bonds (2020). "Machine-learned patterns suggest that diversification drives economic development". In: *Journal of The Royal Society Interface* 17.162, p. 20190283. arXiv: 1812.03534 (cit. on p. 5).
- Bunin, Guy (2017). "Ecological communities with Lotka-Volterra dynamics". In: *Physical Review E* 95.4, p. 042414 (cit. on pp. 87, 100).
- Bürger, Reinhart (2000). *The mathematical theory of selection, recombination, and mutation*. eng. Wiley series in mathematical and computational biology. Chichester [etc]: J. Wiley (cit. on pp. 14, 18, 28).
- Burger, Reinhart and Josef Hofbauer (1994). "Mutation load and mutation-selection-balance in quantitative genetic traits". In: *J. Math. Biol.* 32.3, pp. 193–218 (cit. on p. 125).
- Burnham, Kenneth P. and David R. Anderson (2002). *Model Selection and Multimodel Inference*. Ed. by Kenneth P. Burnham and David R. Anderson. New York, NY: Springer New York (cit. on pp. 68, 71, 91, 93).
- Burrows, Michael T., Amanda E. Bates, Mark J. Costello, et al. (2019). "Ocean community warming responses explained by thermal affinities and temperature gradients". In: *Nature Climate Change* 9.12, pp. 959–963 (cit. on p. 70).
- Bustos, Sebastián, Charles Gomez, Ricardo Hausmann, and César A. Hidalgo (2012). "The Dynamics of Nestedness Predicts the Evolution of Industrial Ecosystems". In: *PLoS ONE* 7.11. Ed. by Luís A. Nunes Amaral, e49393. arXiv: 1203.3796 (cit. on p. 101).
- Cabral, Juliano Sarmento, Luis Valente, and Florian Hartig (2017). "Mechanistic simulation models in macroecology and biogeography: state-of-art and prospects". In: *Ecography* 40.2, pp. 267–280 (cit. on p. 13).
- Caglioti, E., P.-L. Lions, C. Marchioro, and M. Pulvirenti (1995). "A special class of stationary flows for two-dimensional Euler equations: a statistical mechanics description. II". In: *Comm. Math. Phys.* 174.2, pp. 229–260 (cit. on p. 125).
- Callaway, Ragan M., R. W. Brooker, Philippe Choler, et al. (2002). "Positive interactions among alpine plants increase with stress". In: *Nature* 417.6891, pp. 844–848 (cit. on p. 85).
- Cao, Jiguo, Gregor F. Fussmann, and James O. Ramsay (2008). "Estimating a Predator-Prey Dynamical Model with the Parameter Cascades Method". In: *Biometrics* 64.3, pp. 959–967 (cit. on p. 70).
- Caragliu, Andrea and Peter Nijkamp (2016). "Space and knowledge spillovers in European regions: the impact of different forms of proximity on spatial knowledge diffusion". In: *Journal of Economic Geography* 16.3, pp. 749–774 (cit. on pp. 85, 101, 158).
- Carrara, Francesco, Florian Altermatt, Ignacio Rodriguez-Iturbe, and Andrea Rinaldo (2012). "Dendritic connectivity controls biodiversity patterns in experimental metacommunities". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.15, pp. 5761–5766 (cit. on pp. 13, 18, 28, 29).
- Case, T. J. (1990). "Invasion resistance arises in strongly interacting species-rich model competition communities." In: *Proceedings of the National Academy of Sciences* 87.24, pp. 9610–9614 (cit. on pp. 87, 100).

- Castro, Javier (2021). "Deep Learning Schemes For Parabolic Nonlocal Integro-Differential Equations". In: *arXiv:2103.15008*, 28 pp. (Cit. on p. 127).
- Cauwels, Peter and Didier Sornette (2012). "Quis Pedit Ipsa Pretia: Facebook Valuation and Diagnostic of a Bubble Based on Nonlinear Demographic Dynamics". In: *The Journal of Portfolio Management* 38.2, pp. 56–66 (cit. on pp. 4, 86).
- Champagnat, Nicolas, Régis Ferrière, and Sylvie Méléard (2006). "Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models". In: *Theoretical Population Biology* 69.3, pp. 297–321 (cit. on pp. 5, 14, 15, 28, 35).
- Chan, Terence (1999). "Pricing contingent claims on stocks driven by Lévy processes". In: *Ann. Appl. Probab.* 9.2, pp. 504–528 (cit. on p. 125).
- Chapin, Francis Stuart, Pamela A Matson, Harold A Mooney, and Peter Morrison Vitousek (2002). "Principles of terrestrial ecosystem ecology". In: (cit. on p. 1).
- Chen, I-Ching, Jane K. Hill, Ralf Ohlemüller, David B. Roy, and Chris D. Thomas (2011). "Rapid Range Shifts of Species Associated with High Levels of Climate Warming". In: *Science* 333.6045, pp. 1024–1026 (cit. on p. 165).
- Chen, Jingrun, Rui Du, and Keke Wu (2020a). "A Comparison Study of Deep Galerkin Method and Deep Ritz Method for Elliptic Problems with Different Boundary Conditions". In: *Comm. Math. Res.* 36.3, pp. 354–376 (cit. on p. 127).
- Chen, Li, Kevin Painter, Christina Surulescu, and Anna Zhigun (2020b). "Mathematical models for cell migration: a non-local perspective". In: *Philos. Trans. Roy. Soc. B* 375.1807, Article No. 20190379, 9 pp. arXiv: 1911.05200 (cit. on p. 125).
- Clark, Adam Thomas, Hao Ye, Forest Isbell, et al. (2015). "Spatial convergent cross mapping to detect causal relationships from short time series". In: *Ecology* 96.5, pp. 1174–1181 (cit. on p. 70).
- Clermont, Gilles and Sven Zenker (2015). "The inverse problem in mathematical biology". In: *Mathematical Biosciences* 260, pp. 11–15 (cit. on pp. 6, 163).
- Cohendet, Patrick, David Grandadam, Chahira Mehouachi, and Laurent Simon (2018). "The local, the global and the industry common: the case of the video game industry". In: *Journal of Economic Geography* 18.5, pp. 1045–1068 (cit. on p. 85).
- Coleman, Sidney (1994). "Quantum sine-Gordon equation as the massive Thirring model". In: *Bosonization*. Vol. 1. World Scientific, pp. 128–137 (cit. on pp. 125, 137, 143).
- Colizza, Vittoria, Romualdo Pastor-Satorras, and Alessandro Vespignani (2007). "Reaction-diffusion processes and metapopulation models in heterogeneous networks". In: *Nature Physics* 3.4, pp. 276–282. arXiv: 0703129 [cond-mat] (cit. on pp. 17, 31).
- Cont, Rama and Peter Tankov (2004). *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, pp. xvi+535 (cit. on p. 125).
- Cordes, Christian (2006). "Darwinism in economics: from analogy to continuity". In: *Journal of Evolutionary Economics* 16.5, pp. 529–541 (cit. on p. 86).

- Cox, Sonja and Jan van Neerven (2013). "Pathwise Hölder convergence of the implicit-linear Euler scheme for semi-linear SPDEs with multiplicative noise". In: *Numer. Math.* 125.2, pp. 259–345 (cit. on p. 126).
- Cristelli, Matthieu, Andrea Tacchella, and Luciano Pietronero (2015). "The Heterogeneous Dynamics of Economic Complexity". In: *PLOS ONE* 10.2. Ed. by Cathy W.S. Chen, e0117174 (cit. on p. 101).
- Cruz, José M. T. S. and Daniel Ševčovič (2020). "On solutions of a partial integro-differential equation in Bessel potential spaces with applications in option pricing models". In: *Jpn. J. Ind. Appl. Math.* 37.3, pp. 697–721 (cit. on p. 125).
- Csilléry, Katalin, Michael G.B. Blum, Oscar E. Gaggiotti, and Olivier François (2010). "Approximate Bayesian Computation (ABC) in practice". In: *Trends in Ecology & Evolution* 25.7, pp. 410–418 (cit. on pp. 7, 160, 163).
- Curtsdotter, Alva, H. Thomas Banks, John E. Banks, et al. (2019). "Ecosystem function in predator-prey food websconfronting dynamic models with empirical data". In: *Journal of Animal Ecology* 88.2. Ed. by Daniel Stouffer, pp. 196–210 (cit. on pp. 6, 56, 71, 72).
- Dale, M. R.T. and M.J. Fortin (2010). "From graphs to spatial graphs". In: *Annual Review of Ecology, Evolution, and Systematics* 41, pp. 21–38 (cit. on pp. 13, 29, 156, 158, 165).
- Darwin, Charles (1859). *On the origin of species*. Routledge (cit. on p. 2).
- Davies, Richard G., C. David L. Orme, David Storch, et al. (2007). "Topography, energy and the global distribution of bird species richness". In: *Proceedings of the Royal Society B: Biological Sciences* 274.1614, pp. 1189–1197 (cit. on p. 28).
- DeAngelis, Donald L and Wolf M Mooij (2005). "Individual-based modeling of ecological and evolutionary processes". In: *Annual Review of Ecology, Evolution, and Systematics*, pp. 147–168 (cit. on pp. 5, 160).
- DeAngelis, Donald L. and Simeon Yurek (2015). "Equation-free modeling unravels the behavior of complex ecological systems". In: *Proceedings of the National Academy of Sciences* 112.13, pp. 3856–3857 (cit. on p. 55).
- Débarre, F., O. Ronce, and S. Gandon (2013). "Quantifying the effects of migration and mutation on adaptation and demography in spatially heterogeneous environments". In: *Journal of Evolutionary Biology* 26.6, pp. 1185–1202 (cit. on pp. 5, 14, 22, 29, 37, 156, 158).
- D'Elia, Marta, Qiang Du, Christian Glusa, et al. (2020). "Numerical methods for nonlocal and fractional models". In: *Acta Numerica* 29, pp. 1–124 (cit. on pp. 125, 127).
- DeLong, John P., Torrance C. Hanley, and David A. Vasseur (2014). "Predator-prey dynamics and the plasticity of predator body size". In: *Functional Ecology* 28.2. Ed. by Michael Pfrender, pp. 487–493 (cit. on p. 56).
- Deneu, Benjamin, Maximilien Servajean, Pierre Bonnet, et al. (2021). "Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment". In: *PLOS Computational Biology* 17.4. Ed. by Adam C. Martiny, e1008856 (cit. on p. 55).

- Descombes, Patrice, Camille Pitteloud, Gaëtan Glauser, et al. (2020). "Novel trophic interactions under climate change promote alpine plant coexistence". In: *Science* 370.6523, pp. 1469–1473 (cit. on p. 165).
- Deyle, Ethan R., Robert M. May, Stephan B. Munch, and George Sugihara (2016). "Tracking and forecasting ecosystem interactions in real time". In: *Proceedings of the Royal Society B: Biological Sciences* 283.1822, p. 20152258 (cit. on pp. 55, 63).
- Dias, Murilo S., Thierry Oberdorff, Bernard Hugueny, et al. (2014). "Global imprint of historical connectivity on freshwater fish biodiversity". In: *Ecology Letters* 17.9. Ed. by Marti Anderson, pp. 1130–1140 (cit. on pp. 13, 28).
- Dieckmann, Ulf and Michael Doebeli (1999). "On the origin of species by sympatric speciation". In: *Nature* 400.6742, pp. 354–357 (cit. on pp. 3, 5, 13, 14, 165).
- Ding, Wen-Na, Richard H. Ree, Robert A. Spicer, and Yao-Wu Xing (2020). "Ancient orogenic and monsoon-driven assembly of the world's richest temperate alpine flora". In: *Science* 369.6503, pp. 578–581 (cit. on pp. 13, 46, 208).
- Doebeli, Michael (2011). *Adaptive diversification*. Monographs in population biology. Princeton, N.J: Princeton University Press (cit. on p. 38).
- Doebeli, Michael and Ulf Dieckmann (2003). "Speciation along environmental gradients". In: *Nature* 421.6920, pp. 259–264 (cit. on pp. 3, 5, 14, 29, 30, 158, 165).
- Doebeli, Michael and Iaroslav Ispolatov (2014). "Chaos and unpredictability in evolution". In: *Evolution* 68.5, pp. 1365–1373. arXiv: 1309.6261 (cit. on p. 159).
- (2010a). "Complexity and diversity". In: *Science* 328.5977, pp. 494–497 (cit. on pp. 5, 125, 126, 136, 141).
  - (2010b). "Complexity and Diversity". In: *Science* 328.5977, pp. 494–497 (cit. on pp. 9, 159, 160).
- Doney, Scott C. (2010). "The Growing Human Footprint on Coastal and Open-Ocean Biogeochemistry". In: *Science* 328.5985, pp. 1512–1516 (cit. on p. 55).
- Dong, Ruiwen, Christian Goodbrake, Heather A Harrington, and Gleb Pogudin (2021). "Differential elimination for dynamical models via projections with applications to structural identifiability". In: pp. 1–37. arXiv: 2111.00991 (cit. on p. 63).
- Dopfer, Kurt and Jason Potts (2007). *The General Theory of Economic Evolution*. Vol. 4. Routledge, pp. 2–4 (cit. on pp. 1, 4, 85).
- Dornelas, Maria, Laura H. Antão, Faye Moyes, et al. (2018). "BioTIME: A database of biodiversity time series for the Anthropocene". In: *Global Ecology and Biogeography* 27.7, pp. 760–786 (cit. on pp. 69, 70, 161).
- Doron, Maéva, Pierre Brasseur, Jean Michel Brankart, Svetlana N. Losa, and Angélique Melet (2013). "Stochastic estimation of biogeochemical parameters from Globcolour ocean colour satellite data in a North Atlantic 3D ocean coupled physical-biogeochemical model". In: *Journal of Marine Systems* 117-118, pp. 81–95 (cit. on p. 56).
- Duchi, John, Elad Hazan, and Yoram Singer (2011). "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization". In: *J. Mach. Learn. Res.* 12.61, pp. 2121–2159 (cit. on p. 134).

- E, Weinan, Jiequn Han, and Arnulf Jentzen (2021a). “Algorithms for Solving High Dimensional PDEs: From Nonlinear Monte Carlo to Machine Learning”. In: *Nonlinearity* 35.1, pp. 278–310 (cit. on p. 135).
- (2017). “Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations”. In: *Commun. Math. Stat.* 5.4, pp. 349–380 (cit. on p. 126).
- E, Weinan, Martin Hutzenthaler, Arnulf Jentzen, and Thomas Kruse (2021b). “Multilevel Picard iterations for solving smooth semilinear parabolic heat equations”. In: *Partial Differ. Equ. Appl.* 2.6, Article No. 80, 31 pp. (Cit. on pp. 126, 135).
- (2019a). “On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations”. In: *J. Sci. Comput.* 79.3, pp. 1534–1571 (cit. on p. 126).
  - (2019b). “On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations”. In: *J. Sci. Comput.* 79.3, pp. 1534–1571 (cit. on p. 135).
- E, Weinan and Bing Yu (2018). “The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems”. In: *Commun. Math. Stat.* 6.1, pp. 1–12 (cit. on p. 127).
- Econo, Evan P. and Timothy H. Keitt (2010). “Network isolation and local diversity in neutral metacommunities”. In: *Oikos* 119.8, pp. 1355–1363 (cit. on pp. 13, 28).
- (2007). “Species diversity in neutral metacommunities: a network approach”. In: *Ecology Letters* 11.1, 071117033013001–???. (Cit. on pp. 13, 18, 28, 29).
- Eigen, Manfred, John McCaskill, and Peter Schuster (1988). “Molecular quasi-species”. In: *The Journal of Physical Chemistry* 92.24, pp. 6881–6891 (cit. on pp. 88, 100).
- Ellis, Erle C. (2011). “Anthropogenic transformation of the terrestrial biosphere”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369.1938, pp. 1010–1035 (cit. on pp. 3, 55).
- Ezard, Thomas H.G., Steeve D. Côté, and Fanie Pelletier (2009). “Eco-evolutionary dynamics: disentangling phenotypic, environmental and population fluctuations”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1523, pp. 1491–1498 (cit. on p. 3).
- Farmer, J. D. and A. W. Lo (1999). “Frontiers of finance: Evolution and efficient markets”. In: *Proceedings of the National Academy of Sciences* 96.18, pp. 9991–9992 (cit. on pp. 4, 86).
- Fennel, Katja, Martin Losch, Jens Schröter, and Manfred Wenzel (2001). “Testing a marine ecosystem model: sensitivity analysis and parameter optimization”. In: *Journal of Marine Systems* 28.1-2, pp. 45–63 (cit. on p. 56).
- Fiandaca, Giada, Marcello Delitala, and Tommaso Lorenzi (2021). “A Mathematical Study of the Influence of Hypoxia and Acidity on the Evolutionary Dynamics of Cancer”. In: *Bulletin of Mathematical Biology* 83.7, pp. 1–29 (cit. on p. 159).

- Fiechter, J., R. Herbei, W. Leeds, et al. (2013). "A Bayesian parameter estimation method applied to a marine ecosystem model for the coastal Gulf of Alaska". In: *Ecological Modelling* 258, pp. 122–133 (cit. on pp. 56, 69, 160).
- FISHER, R. A. (1937). "THE WAVE OF ADVANCE OF ADVANTAGEOUS GENES". In: *Annals of Eugenics* 7.4, pp. 355–369 (cit. on p. 125).
- Fisher, Ronald Aylmer (1958). *The genetical theory of natural selection*. (cit. on p. 155).
- Fisher, Rosie A., Charles D. Koven, William R.L. Anderegg, et al. (2018). "Vegetation demographics in Earth System Models: A review of progress and priorities". In: *Global Change Biology* 24.1, pp. 35–54 (cit. on pp. 7, 55).
- Fletcher, Roger (2013). *Practical methods of optimization*. John Wiley & Sons (cit. on p. 92).
- Foster, John and J. Stan Metcalfe (2012). "Economic emergence: An evolutionary economic perspective". In: *Journal of Economic Behavior & Organization* 82.2-3, pp. 420–432 (cit. on p. 1).
- Frank, Steven A. (2022). "Automatic differentiation and the optimization of differential equation models in biology". In: 10. arXiv: 2207.04487 (cit. on p. 8).
- Freeman, Christopher (2002). *As time goes by : from the industrial revolutions to the information revolution*. Oxford: Oxford University Press (cit. on p. 86).
- Frey, Rüdiger and Verena Köck (2021). "Deep Neural Network Algorithms for Parabolic PIDEs and Applications in Insurance Mathematics". In: *arXiv:2109.11403*, 24 pp. (Cit. on p. 127).
- (2022). "Deep Neural Network Algorithms for Parabolic PIDEs and Applications in Insurance Mathematics". In: *Mathematical and Statistical Methods for Actuarial Sciences and Finance*. Ed. by Marco Corazza, Cira Perna, Claudio Pizzi, and Marilena Sibillo. Cham: Springer International Publishing, pp. 272–277 (cit. on p. 127).
- Gábor, Attila and Julio R. Banga (2015). "Robust and efficient parameter estimation in dynamic models of biological systems". In: *BMC Systems Biology* 9.1, p. 74 (cit. on pp. 7, 56, 160).
- Gaither, Michelle R., Georgios A. Gkafas, Menno de Jong, et al. (2018). "Genomics of habitat choice and adaptive evolution in a deep-sea fish". In: *Nature Ecology & Evolution* 2.4, pp. 680–687. arXiv: NIHMS150003 (cit. on p. 156).
- Gajewski, Herbert and Klaus Zacharias (2003). "On a nonlocal phase separation model". In: *J. Math. Anal. Appl.* 286.1, pp. 11–31 (cit. on p. 125).
- Gan, Xiaoting, Ying Yang, and Kun Zhang (2020). "A robust numerical method for pricing American options under Kou's jump-diffusion models based on penalty method". In: *J. Appl. Math. Comput.* 62.1–2, pp. 1–21 (cit. on p. 125).
- Gandhi, Amar, Simon Levin, and Steven Orszag (2000). "Moment expansions in spatial ecological models and moment closure through Gaussian approximation". In: *Bulletin of Mathematical Biology* 62.4, pp. 595–632 (cit. on p. 5).
- Garant, Dany, Samantha E. Forde, and Andrew P. Hendry (2007). "The multifarious effects of dispersal and gene flow on contemporary adaptation". In: *Functional Ecology* 21.3, pp. 434–443 (cit. on pp. 13, 30).

- Gatabazi, P., J.C. Mba, E. Pindza, and C. Labuschagne (2019). "Grey LotkaVolterra models with application to cryptocurrencies adoption". In: *Chaos, Solitons & Fractals* 122, pp. 47–57 (cit. on pp. 4, 86).
- GBIF: The Global Biodiversity Information Facility (2022a). "What is GBIF?" In: (cit. on p. 72).
- (2022b). "What is GBIF?" In: (cit. on p. 165).
- Geary, William L., Michael Bode, Tim S. Doherty, et al. (2020). "A guide to ecosystem models and their environmental applications". In: *Nature Ecology & Evolution* 4.11, pp. 1459–1471 (cit. on p. 55).
- Gehlen, M., R. Barciela, L. Bertino, et al. (2015). "Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments". In: *Journal of Operational Oceanography* 8.sup1, s168–s187 (cit. on p. 56).
- Génieys, S., V. Volpert, and P. Auger (2006). "Pattern and waves for a model in population dynamics with nonlocal consumption of resources". In: *Math. Model. Nat. Phenom.* 1.1, pp. 65–82 (cit. on pp. 125, 136, 141).
- Gentleman, Wendy, Andrew Leising, Bruce Frost, Suzanne Strom, and James Murray (2003). "Functional responses for zooplankton feeding on multiple resources: a review of assumptions and biological dynamics". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 50.22-26, pp. 2847–2875 (cit. on p. 56).
- Gharamti, M.E., J. Tjiputra, I. Bethke, et al. (2017). "Ensemble data assimilation for ocean biogeochemical state and parameter estimation at different sites". In: *Ocean Modelling* 112, pp. 65–89 (cit. on p. 56).
- Gholaminejad, Amir, Kurt Keutzer, and George Biros (2019). "ANODE: Unconditionally Accurate Memory-Efficient Gradients for Neural ODEs". In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. Vol. 2019-Augus. California: International Joint Conferences on Artificial Intelligence Organization, pp. 730–736. arXiv: 1902.10298 (cit. on p. 62).
- Gilarranz, Luis J. and Jordi Bascompte (2012). "Spatial network structure and metapopulation persistence". In: *Journal of Theoretical Biology* 297, pp. 11–16 (cit. on pp. 13, 28).
- Gillespie, Daniel T. (1976). "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of Computational Physics* 22.4, pp. 403–434 (cit. on pp. 15, 162).
- Giuliani, Elisa (2007). "The selective nature of knowledge networks in clusters: evidence from the wine industry". In: *Journal of Economic Geography* 7.2, pp. 139–168 (cit. on p. 101).
- Glorot, Xavier and Yoshua Bengio (2010). "Understanding the difficulty of training deep feedforward neural networks". In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Yee Whye Teh and Mike Titterington. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, pp. 249–256 (cit. on p. 139).

- Gonon, Lukas and Christoph Schwab (2021). “Deep ReLU neural networks overcome the curse of dimensionality for partial integrodifferential equations”. In: *arXiv:2102.11707*, 35 pp. arXiv: 2102.11707 (cit. on p. 127).
- Gosh, Sanmitra, Paul Birrell, and Daniela De Angelis (2021). “Variational inference for non-linear ordinary differential equations”. In: *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics* 130.29, pp. 2719–2727 (cit. on pp. 56, 164).
- Gould, Stephen Jay and Niles Eldredge (1972). “Punctuated equilibria: an alternative to phyletic gradualism”. In: *Models in paleobiology* 1972, pp. 82–115 (cit. on p. 85).
- Govaert, Lynn, Emanuel A. Fronhofer, Sébastien Lion, et al. (2019). “Eco-evolutionary feedbacksTheoretical models and perspectives”. In: *Functional Ecology* 33.1, pp. 13–30. arXiv: 1806.07633 (cit. on pp. 3, 5, 158).
- Gravel, Dominique, François Massol, and Mathew A. Leibold (2016). “Stability and complexity in model meta-ecosystems”. In: *Nature Communications* 7.1, p. 12457 (cit. on pp. 13, 28).
- Grime, John Philip (1973). “Competitive Exclusion in Herbaceous Vegetation”. In: *Nature* 242.5396, pp. 344–347 (cit. on p. 85).
- Grohs, Philipp and Felix Voigtlaender (2021). “Proof of the Theory-to-Practice Gap in Deep Learning via Sampling Complexity bounds for Neural Network Approximation Spaces”. In: *arXiv:2104.02746*, 42 pp. arXiv: 2104.02746 (cit. on p. 127).
- Guégan, Jean-François, Sovan Lek, and Thierry Oberdorff (1998). “Energy availability and habitat heterogeneity predict global riverine fish diversity”. In: *Nature* 391.6665, pp. 382–384 (cit. on pp. 13, 28).
- Guo, Ling, Hao Wu, Xiaochen Yu, and Tao Zhou (2022). “Monte Carlo PINNs: deep learning approach for forward and inverse problems involving high dimensional fractional partial differential equations”. In: *arXiv:2203.08501*, 18 pp. (Cit. on p. 127).
- Gutenkunst, Ryan N., Joshua J. Waterfall, Fergal P. Casey, et al. (2007). “Universally Sloppy Parameter Sensitivities in Systems Biology Models”. In: *PLoS Computational Biology* 3.10. Ed. by Adam P Arkin, e189 (cit. on p. 55).
- Gyöngy, István and Nicolai Krylov (2003). “On the splitting-up method and stochastic partial differential equations”. In: *Ann. Probab.* 31.2, pp. 564–591 (cit. on p. 126).
- Hagen, Oskar (2022). “Coupling ecoevolutionary mechanisms with deeptime environmental dynamics to understand biodiversity patterns”. In: *Ecography*, pp. 1–16 (cit. on pp. 2–4, 165).
- Hairer, Martin and Hao Shen (2016). “The dynamical sine-Gordon model”. In: *Comm. Math. Phys.* 341.3, pp. 933–989 (cit. on pp. 125, 137, 143).
- Hairston, Nelson G., Stephen P. Ellner, Monica A. Geber, Takehito Yoshida, and Jennifer A. Fox (2005). “Rapid evolution and the convergence of ecological and evolutionary time”. In: *Ecology Letters* 8.10, pp. 1114–1127 (cit. on p. 3).
- Hall, Brian Keith (2013). *Strickberger’s evolution*. eng. Ed. by Brian Keith Hall. 5th editio. Sudbury: Jones & Bartlett Learning (cit. on p. 2).

- Hamel, F., F. Lavigne, G. Martin, and L. Roques (2020). “Dynamics of adaptation in an anisotropic phenotype-fitness landscape”. In: *Nonlinear Anal. Real World Appl.* 54, Article No. 103107, 33 pp. (Cit. on pp. 125, 137, 144).
- Hamel, François and Nikolaï Nadirashvili (2001). “Travelling fronts and entire solutions of the Fisher-KPP equation in  $\mathbb{R}^N$ ”. In: *Arch. Ration. Mech. Anal.* 157.2, pp. 91–163 (cit. on p. 125).
- Hamilton, Matthew B (2021). *Population genetics*. John Wiley & Sons (cit. on p. 155).
- Han, Jiequn, Arnulf Jentzen, and Weinan E (2018). “Solving high-dimensional partial differential equations using deep learning”. In: *Proc. Natl. Acad. Sci. USA* 115.34, pp. 8505–8510 (cit. on pp. 7, 126).
- Hastings, Alan, Carole L. Hom, Stephen Ellner, Peter Turchin, and H. Charles J. Godfray (1993). “Chaos in Ecology: Is Mother Nature a Strange Attractor?” In: *Annual Review of Ecology and Systematics* 24.1, pp. 1–33 (cit. on pp. 7, 55, 59, 69).
- Hastings, Alan and Thomas Powell (1991). “Chaos in a Three-Species Food Chain”. In: *Ecology* 72.3, pp. 896–903 (cit. on pp. 63, 64, 78, 80, 81, 209).
- Hausmann, Ricardo ; and César A. Hidalgo (2011). *The Atlas of Economic Complexity* (cit. on p. 166).
- Heinrich, Stefan (1998). “Monte Carlo complexity of global solution of integral equations”. In: *J. Complex.* 14.2, pp. 151–175 (cit. on p. 127).
- Heinrich, Stefan and Eugène Sindambiwe (1999). “Monte Carlo complexity of parametric integration”. In: *J. Complex.* 15.3, pp. 317–341 (cit. on p. 127).
- Hendry, Andrew P (2016). *Eco-evolutionary Dynamics*. Princeton: Princeton University Press (cit. on p. 2).
- Henry-Labordère, Pierre (2012). “Counterparty Risk Valuation: A Marked Branching Diffusion Approach”. In: *arXiv:1203.2369*, 17 pp. arXiv: 1203.2369 (cit. on p. 125).
- Hidalgo, C. A., B. Klinger, A.-L. Barabasi, and R. Hausmann (2007). “The Product Space Conditions the Development of Nations”. In: *Science* 317.5837, pp. 482–487 (cit. on pp. 101, 166).
- Hidalgo, César A. (2021). “Economic complexity theory and applications”. In: *Nature Reviews Physics* 3.2, pp. 92–113 (cit. on pp. 86, 91, 157, 160).
- (2018). “From useless to keystone”. In: *Nature Physics* 14.1, pp. 9–10 (cit. on p. 158).
- Higgins, Steven I., Simon Scheiter, and Mahesh Sankaran (2010). “The stability of African savannas: insights from the indirect estimation of the parameters of a dynamic model”. In: *Ecology* 91.6, pp. 1682–1692 (cit. on pp. 6, 56, 69, 160).
- Hochbruck, Marlis and Alexander Ostermann (2005). “Explicit exponential Runge–Kutta methods for semilinear parabolic problems”. In: *SIAM J. Numer. Anal.* 43.3, pp. 1069–1090 (cit. on p. 126).
- Hodgson, Geoffrey M. (2002). “Darwinism in economics: from analogy to ontology”. In: *Journal of Evolutionary Economics* 12.3, pp. 259–281 (cit. on p. 85).

- Hodgson, Geoffrey M. (2019). *Evolutionary Economics*. Vol. 66. Cambridge University Press, pp. 37–39 (cit. on pp. 4, 9, 85, 157, 160).
- Hodgson, Geoffrey M. and Thorbjorn Knudsen (2004). “The firm as an interactor: firms as vehicles for habits and routines”. In: *Journal of Evolutionary Economics* 14.3, pp. 281–307 (cit. on pp. 86, 100, 158).
- Holderegger, Rolf, Urs Kamm, and Felix Gugerli (2006). “Adaptive vs. neutral genetic diversity: implications for landscape genetics”. In: *Landscape Ecology* 21.6, pp. 797–807 (cit. on pp. 13, 30, 33).
- Holland, Matthew D. and Alan Hastings (2008). “Strong effect of dispersal network structure on ecological dynamics”. In: *Nature* 456.7223, pp. 792–794 (cit. on pp. 13, 28).
- Houchmandzadeh, B. and M. Vallade (2017). “Fisher waves: An individual-based stochastic model”. In: *Physical Review E* 96.1, pp. 1–13. arXiv: 1703.02835 (cit. on p. 125).
- Hsieh, Chihhao, Christian Anderson, and George Sugihara (2008). “Extending Nonlinear Analysis to Short Ecological Time Series”. In: *The American Naturalist* 171.1, pp. 71–80 (cit. on p. 70).
- Huang, Jian, Zhongdi Cen, and Anbo Le (2013). “A finite difference scheme for pricing American put options under Kou’s jump-diffusion model”. In: *J. Funct. Spaces Appl.*, Article No. 651573, 11 pp. (Cit. on p. 125).
- Hubbell, Stephen P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in Population Biology 32. Princeton [etc: Princeton University Press (cit. on p. 13).
- Huisman, Jef and Franz J. Weissing (1999). “Biodiversity of plankton by species oscillations and chaos”. In: *Nature* 402.6760, pp. 407–410 (cit. on pp. 7, 55, 59, 69).
- Huré, Côme, Huyê Pham, and Xavier Warin (2020). “Deep backward schemes for high-dimensional nonlinear PDEs”. In: *Math. Comp.* 89, pp. 1547–1579 (cit. on p. 127).
- Hutzenthaler, Martin, Arnulf Jentzen, Thomas Kruse, Tuan Anh Nguyen, and Philippe von Wurstemberger (2020). “Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations”. In: *Proc. A.* 476.2244, Article No. 20190630, 25 pp. (Cit. on pp. 126, 135).
- Huxley, Julian (1942). “Evolution: The Modern Synthesis”. In: (cit. on p. 5).
- Innes, Mike, Alan Edelman, Keno Fischer, et al. (2019). “A Differentiable Programming System to Bridge Machine Learning and Scientific Computing”. In: arXiv: 1907.07587 (cit. on p. 8).
- Ioffe, Sergey and Christian Szegedy (2015). “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, pp. 448–456 (cit. on p. 134).
- Jetz, Walter, Melodie A. McGeoch, Robert Guralnick, et al. (2019). “Essential biodiversity variables for mapping and monitoring species populations”. In: *Nature Ecology and Evolution* 3.4, pp. 539–551 (cit. on p. 72).

- Johnson, Jerald B. and Kristian S. Omland (2004). "Model selection in ecology and evolution". In: *Trends in Ecology & Evolution* 19.2, pp. 101–108 (cit. on pp. 6, 7).
- Jost, Jürgen (2005). *Dynamical systems: examples of complex behaviour*. Springer Science & Business Media (cit. on p. 73).
- Jung, Martin, Prabhat Raj Dahal, Stuart H. M. Butchart, et al. (2020). "A global map of terrestrial habitat types". In: *Scientific Data* 7.1, p. 256 (cit. on pp. 46, 208).
- K Mogensen, Patrick and Asbjørn N Røiseth (2018). "Optim: A mathematical optimization package for Julia". In: *Journal of Open Source Software* 3.24, p. 615. arXiv: arXiv: 1710.07708 (cit. on p. 62).
- Kaplan, Norman L., R.R. Hudson, and C. H. Langley (1989). "The hitchhiking effect revisited." In: *Genetics* 123.4, pp. 887–899 (cit. on p. 13).
- Karger, Dirk Nikolaus, Olaf Conrad, Jürgen Böhner, et al. (2017). "Climatologies at high resolution for the earth's land surface areas". In: *Scientific Data* 4.1, p. 170122. arXiv: 1607.00217 (cit. on pp. 46, 208).
- Karpatne, Anuj, Gowtham Atluri, James H. Faghmous, et al. (2017). "Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data". In: *IEEE Transactions on Knowledge and Data Engineering* 29.10, pp. 2318–2331. arXiv: 1612.08544 (cit. on pp. 7, 161).
- Kashinath, K., M. Mustafa, A. Albert, et al. (2021). "Physics-informed machine learning: Case studies for weather and climate modelling". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2194 (cit. on pp. 7, 56, 70, 161).
- Kavallaris, Nikos I., Johannes Lankeit, and Michael Winkler (2017). "On a degenerate nonlocal parabolic problem describing infinite dimensional replicator dynamics". In: *SIAM J. Math. Anal.* 49.2, pp. 954–983 (cit. on p. 125).
- Kavallaris, Nikos I. and Takashi Suzuki (2018). *Non-local partial differential equations for engineering and biology*. Vol. 31. Mathematics for Industry (Tokyo). Springer, Cham, pp. xix+300 (cit. on p. 125).
- Kawecki, Tadeusz J. and Dieter Ebert (2004). "Conceptual issues in local adaptation". In: *Ecology Letters* 7.12, pp. 1225–1241 (cit. on p. 156).
- Keller, Wolfgang (2004). "International Technology Diffusion". In: *Journal of Economic Literature* 42.3, pp. 752–782 (cit. on pp. 86, 100).
- Kimura, M. (1965). "A stochastic model concerning the maintenance of genetic variability in quantitative characters." In: *Proceedings of the National Academy of Sciences of the United States of America* 54.3, pp. 731–736 (cit. on p. 37).
- Kimura, M and G H Weiss (1964). "The stepping stone model of population structure and the decrease of genetic correlation with distance." In: *Genetics* 49.4, pp. 561–76 (cit. on pp. 16, 28).
- King, R. G. and R. Levine (1993). "Finance and Growth: Schumpeter Might Be Right". In: *The Quarterly Journal of Economics* 108.3, pp. 717–737 (cit. on p. 166).

- Kingma, Diederik P. and Jimmy Ba (2014). "Adam: A Method for Stochastic Optimization". In: *arXiv:1412.6980*, 15 pp. arXiv: 1412.6980 (cit. on pp. 62, 69, 92, 134, 138, 161).
- Kirkpatrick, Mark and N. H. Barton (1997). "Evolution of a Species' Range". In: *The American Naturalist* 150.1, pp. 1–23 (cit. on pp. 29, 30, 158).
- Klebanoff, Aaron and Alan Hastings (1994). "Chaos in three species food chains". In: *Journal of Mathematical Biology* 32.5, pp. 427–451 (cit. on p. 63).
- Kou, S. G. (2002). "A Jump-Diffusion Model for Option Pricing". In: *Manag. Sci.* 48.8, pp. 1086–1101. eprint: <https://doi.org/10.1287/mnsc.48.8.1086.166> (cit. on p. 125).
- Kreft, Holger and Walter Jetz (2007). "Global patterns and determinants of vascular plant diversity". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.14, pp. 5925–5930 (cit. on p. 28).
- Kremen, Claire (2005). "Managing ecosystem services: what do we need to know about their ecology?" In: *Ecology Letters* 8.5, pp. 468–479 (cit. on p. 55).
- Kriticos, Darren J, Bruce L Webber, Agathe Leriche, et al. (2012). "CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling". In: *Methods in Ecology and Evolution* 3.1, pp. 53–64 (cit. on p. 165).
- Lacey, A. A. (1995). "Thermal runaway in a non-local problem modelling Ohmic heating. I. Model derivation and some special cases". In: *European J. Appl. Math.* 6.2, pp. 127–144 (cit. on p. 125).
- Lagaris, I.E., A.C. Likas, and D.G. Papageorgiou (2000). "Neural-network methods for boundary value problems with irregular boundaries". In: *IEEE Trans. Neural Netw.* 11.5, pp. 1041–1049 (cit. on p. 127).
- Lagaris, Isaac E, Aristidis Likas, and Dimitrios I Fotiadis (1998). "Artificial neural networks for solving ordinary and partial differential equations". In: *IEEE Trans. Neural Netw.* 9.5, pp. 987–1000 (cit. on p. 127).
- Lande, R. (1991). "Isolation by distance in a quantitative trait". In: *Genetics* 128.2, pp. 443–452 (cit. on pp. 14, 28, 155).
- Lande, Russell (1992). "NEUTRAL THEORY OF QUANTITATIVE GENETIC VARIANCE IN AN ISLAND MODEL WITH LOCAL EXTINCTION AND COLONIZATION". In: *Evolution* 46.2, pp. 381–389 (cit. on p. 15).
- Laubmeier, A. N., Kate Wootton, J. E. Banks, et al. (2018). "From theory to experimental designQuantifying a trait-based theory of predator-prey dynamics". In: *PLOS ONE* 13.4. Ed. by Mathew S. Crowther, e0195919 (cit. on p. 71).
- Law, Richard and Ulf Dieckmann (1999). "Moment approximations of individual-based models". In: (cit. on p. 5).
- Lawson, Linda M., Yvette H. Spitz, Eileen E. Hofmann, and Robert Bryan Long (1995). "A data assimilation technique applied to a predator-prey model". In: *Bulletin of Mathematical Biology* 57.4, pp. 593–617 (cit. on p. 56).
- Lawson, Tony (2013). "What is this school' called neoclassical economics?" In: *Cambridge Journal of Economics* 37.5, pp. 947–983 (cit. on pp. 1, 4, 86).

- Le Gland, Guillaume, Sergio Vallina, S. Lan Smith, and Pedro Cermeño (2020). “SPEAD 1.0 A model for Simulating Plankton Evolution with Adaptive Dynamics in a two-trait continuous fitness landscape applied to the Sargasso Sea”. In: *Geoscientific Model Development Discussions*, pp. 1–54 (cit. on p. 159).
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). “Deep learning”. In: *Nature* 521.7553, pp. 436–444 (cit. on pp. 7, 8, 126).
- Leibold, M. A., M. Holyoak, N. Mouquet, et al. (2004). “The metacommunity concept: a framework for multi-scale community ecology”. In: *Ecology Letters* 7.7, pp. 601–613 (cit. on p. 85).
- Lepers, Clotilde, Sylvain Billiard, Matthieu Porte, Sylvie Méléard, and Viet Chi Tran (2021). “Inference with selection, varying population size, and evolving population structure: application of ABC to a forwardbackward coalescent process with interactions”. In: *Heredity* 126.2, pp. 335–350. arXiv: 1910.10201 (cit. on pp. 28, 156).
- Levin, Simon A. (2002). “Complex adaptive systems: Exploring the known, the unknown and the unknowable”. In: *Bulletin of the American Mathematical Society* 40.01, pp. 3–20 (cit. on pp. 1, 2, 13, 155).
- Levin, Simon A (1998). “Ecosystems and the Biosphere as Complex Adaptive Systems Simon”. In: *Ecosystems* 1.5, pp. 431–436 (cit. on p. 1).
- Liao, Yulei and Pingbing Ming (2021). “Deep Nitsche Method: Deep Ritz Method with Essential Boundary Conditions”. In: *Comm. Comput. Phys.* 29.5, pp. 1365–1384 (cit. on p. 127).
- Lieberman, Erez, Christoph Hauert, and Martin A. Nowak (2005). “Evolutionary dynamics on graphs”. In: *Nature* 433.7023, pp. 312–316 (cit. on pp. 9, 156, 158, 160).
- Lignell, Risto, Heikki Haario, Marko Laine, and T. Frede Thingstad (2013). “Getting the right parameter values for models of the pelagic microbial food web”. In: *Limnology and Oceanography* 58.1, pp. 301–313 (cit. on pp. 56, 69, 160).
- Linke, Heiner (2015). “Applications of Brownian Motion”. In: 5114, pp. 199–213 (cit. on p. 34).
- Lion, Sébastien (2016). “Moment equations in spatial evolutionary ecology”. In: *Journal of Theoretical Biology* 405, pp. 46–57 (cit. on pp. 5, 9, 13, 160).
- Lion, Sébastien, Mike Boots, and Akira Sasaki (2022). “Multimorph Eco-Evolutionary Dynamics in Structured Populations”. In: *The American Naturalist* 200.3, pp. 345–372 (cit. on pp. 3, 5, 162).
- Liu, Cong, Kenneth L. Dudley, Zheng-Hui Xu, and Evan P. Economo (2018). “Mountain metacommunities: climate and spatial connectivity shape ant diversity in a complex landscape”. In: *Ecography* 41.1, pp. 101–112 (cit. on p. 28).
- Liu, Dong C. and Jorge Nocedal (1989). “On the limited memory BFGS method for large scale optimization”. In: *Mathematical Programming* 45.1-3, pp. 503–528 (cit. on p. 62).

- Lorz, Alexander, Tommaso Lorenzi, Michael E. Hochberg, Jean Clairambault, and Benoît Perthame (2013). “Populational adaptive evolution, chemotherapeutic resistance and multiple anti-cancer therapies”. In: *ESAIM Math. Model. Numer. Anal.* 47.2, pp. 377–399 (cit. on p. 125).
- Lu, Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis (2021). “DeepXDE: A Deep Learning Library for Solving Differential Equations”. In: *SIAM Review* 63.1, pp. 208–228. eprint: <https://doi.org/10.1137/19M1274067> (cit. on p. 127).
- Luo, Yadong and Rosalie L. Tung (2007). “International expansion of emerging market enterprises: A springboard perspective”. In: *Journal of International Business Studies* 38.4, pp. 481–498 (cit. on pp. 86, 100).
- Ma, Yingbo, Vaibhav Dixit, Michael J. Innes, Xingjian Guo, and Chris Rackauckas (2021). “A Comparison of Automatic Differentiation and Continuous Sensitivity Analysis for Derivatives of Differential Equation Solutions”. In: *2021 IEEE High Performance Extreme Computing Conference (HPEC)*. 2. IEEE, pp. 1–9. arXiv: 1812.01892 (cit. on p. 62).
- Manel, Stéphanie, Michael K. Schwartz, Gordon Luikart, and Pierre Taberlet (2003). “Landscape genetics: combining landscape ecology and population genetics”. In: *Trends in Ecology & Evolution* 18.4, pp. 189–197 (cit. on pp. 14, 28).
- Mangan, N. M., J. N. Kutz, S. L. Brunton, and J. L. Proctor (2017). “Model selection for dynamical systems via sparse regression and information criteria”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473.2204, p. 20170009. arXiv: 1701.01773 (cit. on pp. 68, 92).
- Marasco, A., A. Picucci, and A. Romano (2016). “Market share dynamics using LotkaVolterra models”. In: *Technological Forecasting and Social Change* 105, pp. 49–62 (cit. on pp. 4, 86).
- Margalef, R (1963). “On Certain Unifying Principles in Ecology”. In: *The American Naturalist* 97.897, pp. 357–374 (cit. on p. 166).
- Mari, Lorenzo, Renato Casagrandi, Enrico Bertuzzo, Andrea Rinaldo, and Marino Gatto (2014). “Metapopulation persistence and species spread in river networks”. In: *Ecology Letters* 17.4. Ed. by Ferenc Jordán, pp. 426–434 (cit. on pp. 13, 28).
- Matear, Richard J. (1995). “Parameter optimization and analysis of ecosystem models using simulated annealing: A case study at Station P”. In: *Journal of Marine Research* 53.4, pp. 571–607 (cit. on p. 56).
- Maufrais, Raymond (1997). *Aventures en Guyane. Chefs-d'œuvre de l'aventure humaine*. Ramsay (cit. on p. ix).
- May, Robert M. (2004). “Uses and Abuses of Mathematics in Biology”. In: *Science* 303.5659, pp. 790–793 (cit. on p. 5).
- McCann, Kevin and Alan Hastings (1997). “Reevaluating the omnivorystability relationship in food webs”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 264.1385, pp. 1249–1254 (cit. on pp. 63, 64, 78, 209).
- McCann, Kevin and Peter Yodzis (1994a). “Biological Conditions for Chaos in a Three-Species Food Chain”. In: *Ecology* 75.2, pp. 561–564 (cit. on pp. 63, 78).

- (1994b). “Nonlinear Dynamics and Population Disappearances”. In: *The American Naturalist* 144.5, pp. 873–879 (cit. on p. 63).
- McFall, Kevin Stanley and James Robert Mahan (2009). “Artificial Neural Network Method for Solution of Boundary Value Problems With Exact Satisfaction of Arbitrary Boundary Conditions”. In: *IEEE Trans. Neural Netw.* 20.8, pp. 1221–1233 (cit. on p. 127).
- McRae, Brad H. (2006). “ISOLATION BY RESISTANCE”. In: *Evolution* 60.8, p. 1551 (cit. on pp. 29, 155).
- McRae, Brad H. and Paul Beier (2007). “Circuit theory predicts gene flow in plant and animal populations”. In: *Proceedings of the National Academy of Sciences* 104.50, pp. 19885–19890 (cit. on pp. 28, 29, 155).
- Mealy, Penny, J. Doyne Farmer, and Alexander Teytelboym (2019). “Interpreting economic complexity”. In: *Science Advances* 5.1, eaau1705 (cit. on pp. 86, 157).
- Merton, Robert C. (1976). “Option pricing when underlying stock returns are discontinuous”. In: *J. Financ. Econ.* 3.1–2, pp. 125–144 (cit. on p. 125).
- Meszéna, Géza, István Czibula, and Stefan Geritz (1997). “Adaptive Dynamics in a 2-Patch Environment: A Toy Model for Allopatric and Parapatric Speciation”. In: *Journal of Biological Systems* 05.02, pp. 265–284 (cit. on pp. 3, 5, 14, 21, 22, 29, 31, 32, 156, 158).
- Metcalfe, J. Stan, John Foster, and Ronnie Ramlogan (2006). “Adaptive economic growth”. In: *Cambridge Journal of Economics* 30.1, pp. 7–32 (cit. on p. 157).
- Metropolis, Nicholas and S. Ulam (1949). “The Monte Carlo Method”. In: *J. Amer. Statist. Assoc.* 44.247, pp. 335–341. eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1949.10483310> (cit. on p. 126).
- Metz, JA, SAH Geritz, and G Meszéna (1995). “Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction”. In: *International Institute for Applied Systems Analysis*, p. 42 (cit. on p. 5).
- Michalakelis, Christos, Thomas Sphicopoulos, and Dimitris Varoutas (2011). “Modeling competition in the telecommunications market based on concepts of population biology”. In: *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 41.2, pp. 200–210 (cit. on pp. 4, 86).
- Midgley, Guy and Lee Hannah (2019). “Extinction Risk from Climate Change”. In: *Biodiversity and Climate Change*. Yale University Press, pp. 294–296 (cit. on pp. 3, 55).
- Mirrahimi, Sepideh and Sylvain Gandon (2020). “Evolution of specialization in heterogeneous environments: equilibrium between selection, mutation and migration”. In: *Genetics* 214.2, pp. 479–491 (cit. on pp. 5, 14, 16, 21, 22, 29, 37, 156, 158).
- Mitchell, Melanie (2009). *Complexity: A guided tour*. Oxford university press (cit. on p. 1).
- Modis, Theodore (1997). “Genetic re-engineering of corporations”. In: *Technological Forecasting and Social Change* 56.2, pp. 107–118 (cit. on pp. 4, 86).
- Morningstar, Warren R., Sharad M. Vikram, Cusuh Ham, Andrew Gallagher, and Joshua V. Dillon (2020). “Automatic Differentiation Variational Inference with Mixtures”. In: 18, pp. 1–45. arXiv: 2003.01687 (cit. on p. 164).

- Muneepeerakul, Rachata, Enrico Bertuzzo, Heather J. Lynch, et al. (2008). "Neutral meta-community models predict fish diversity patterns in MississippiMissouri basin". In: *Nature* 453.7192, pp. 220–222 (cit. on p. 14).
- Nagylaki, T. (1994). "Geographical variation in a quantitative character." In: *Genetics* 136.1, pp. 361–81 (cit. on pp. 14, 28).
- Nelson, Richard R (1985). *An evolutionary theory of economic change*. harvard university press (cit. on pp. 1, 4, 85, 157).
- Nelson, Richard R, Sidney G Winter, and T H E Belknap Press (2014). "Towards an Evolutionary Theory of Economic Change". In: *Long-run Economics : An Evolutionary Approach to Economic Growth*. Bloomsbury Academic (cit. on p. 4).
- Newman, M. E. J. (2003). "Mixing patterns in networks". In: *Physical Review E* 67.2, p. 026126. arXiv: 0209450 [cond-mat] (cit. on pp. 23, 31, 38).
- Norberg, Jon, Mark C. Urban, Mark Vellend, Christopher A. Klausmeier, and Nicolas Loeuille (2012). "Eco-evolutionary responses of biodiversity to climate change". In: *Nature Climate Change* 2.10, pp. 747–751 (cit. on pp. 3, 55, 161, 165).
- Nordbotten, Jan M., Simon A. Levin, Eörs Szathmáry, and Nils C. Stenseth (2018). "Ecological and evolutionary dynamics of interconnectedness and modularity". In: *Proc. Natl. Acad. Sci. USA* 115.4, pp. 750–755 (cit. on p. 125).
- Nordbotten, Jan M. and Nils C. Stenseth (2016). "Asymmetric ecological conditions favor Red-Queen type of continued evolution over stasis". In: *Proc. Natl. Acad. Sci. USA* 113.7, pp. 1847–1852 (cit. on p. 125).
- Nordbotten, Jan Martin, Folmer Bokma, Jo Skeie Hermansen, and Nils Chr Stenseth (2020). "The dynamics of trait variance in multi-species communities". In: *R. Soc. Open Sci.* 7.8, Article No. 200321, 20 pp. (Cit. on pp. 5, 125, 162).
- Nosil, Patrik, Scott P Egan, and Daniel J Funk (2008). "HETEROGENEOUS GENOMIC DIFFERENTIATION BETWEEN WALKING-STICK ECOTYPES: "ISOLATION BY ADAPTATION" AND MULTIPLE ROLES FOR DIVERGENT SELECTION". In: *Evolution* 62.2, pp. 316–336 (cit. on p. 156).
- Nowak, Martin A and Natalia L Komarova (2001). "Towards an evolutionary theory of language". In: *Trends in Cognitive Sciences* 5.7, pp. 288–295 (cit. on p. 5).
- Odum, Eugene P. (1969). "The strategy of ecosystem development". In: *Science* 164.3877, pp. 262–270 (cit. on p. 166).
- Oechssler, Jörg and Frank Riedel (2001). "Evolutionary dynamics on infinite strategy spaces". In: *Econom. Theory* 17.1, pp. 141–162 (cit. on p. 125).
- Olff, Han, David Alonso, Matty P. Berg, et al. (2009). "Parallel ecological networks in ecosystems". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1524, pp. 1755–1779 (cit. on pp. 1, 2).
- Orsini, Luisa, Joost Vanoverbeke, Ine Swillen, Joachim Mergeay, and Luc De Meester (2013). "Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization". In: *Molecular Ecology* 22.24, pp. 5983–5999 (cit. on pp. 13, 29, 30, 155, 156).

- Ozman, M. (2009). "Inter-firm networks and innovation: a survey of literature". In: *Economics of Innovation and New Technology* 18.1, pp. 39–67 (cit. on pp. 85, 100, 158).
- Pahlow, Markus, Alain F. Vézina, Benoit Casault, et al. (2008). "Adaptive model of plankton dynamics for the North Atlantic". In: *Progress in Oceanography* 76.2, pp. 151–191 (cit. on p. 71).
- Pájaro, Manuel, Antonio A. Alonso, Irene Otero-Muras, and Carlos Vázquez (2017). "Stochastic modeling and numerical simulation of gene regulatory networks with protein bursting". In: *J. Theoret. Biol.* 421, pp. 51–70 (cit. on pp. 125, 126).
- Pang, Guofei, Lu Lu, and George Em Karniadakis (2019). "fPINNs: Fractional Physics-Informed Neural Networks". In: *SIAM Journal on Scientific Computing* 41.4, A2603–A2626. eprint: <https://doi.org/10.1137/18M1229845> (cit. on p. 127).
- Pearson, Richard G. and Terence P. Dawson (2003). "Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?" In: *Global Ecology and Biogeography* 12.5, pp. 361–371 (cit. on p. 165).
- Pelc, Joanna S., Ehouarn Simon, Laurent Bertino, Ghada El Serafy, and Arnold W. Heemink (2012). "Application of model reduced 4D-Var to a 1D ecosystem model". In: *Ocean Modelling* 57-58, pp. 43–58 (cit. on p. 56).
- Pelletier, F., D. Garant, and A.P. Hendry (2009). "Eco-evolutionary dynamics". eng. In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364.1523, pp. 1483–1489 (cit. on pp. 3, 9, 13, 160).
- Peng, Grace C. Y., Mark Alber, Adrian Buganza Tepole, et al. (2021). "Multiscale Modeling Meets Machine Learning: What Can We Learn?" In: *Archives of Computational Methods in Engineering* 28.3, pp. 1017–1037. arXiv: 1911.11958 (cit. on p. 56).
- Perkel, Jeffrey M. (2019). "Julia: come for the syntax, stay for the speed". In: *Nature* 572.7767, pp. 141–142 (cit. on p. 8).
- Perretti, Charles T., George Sugihara, and Stephan B. Munch (2013). "Nonparametric forecasting outperforms parametric methods for a simulated multispecies system". In: *Ecology* 94.4, pp. 794–800 (cit. on p. 63).
- Perthame, Benoît and Stephane Génieys (2007). "Concentration in the nonlocal Fisher equation: the Hamilton-Jacobi limit". In: *Math. Model. Nat. Phenom.* 2.4, pp. 135–151 (cit. on pp. 125, 136, 141).
- Petkova, Desislava, John Novembre, and Matthew Stephens (2015). "Visualizing spatial population structure with estimated effective migration surfaces". In: *Nature Genetics* 48.1, pp. 94–100 (cit. on p. 28).
- Pham, Huyêñ (2009). *Continuous-time stochastic control and optimization with financial applications*. Vol. 61. Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, pp. xviii+232 (cit. on p. 125).
- Pinsky, M. L., B. Worm, M. J. Fogarty, J. L. Sarmiento, and S. A. Levin (2013). "Marine Taxa Track Local Climate Velocities". In: *Science* 341.6151, pp. 1239–1242 (cit. on p. 70).

- Pisarenko, V. F. and D. Sornette (2004). "Statistical methods of parameter estimation for deterministically chaotic time series". In: *Physical Review E* 69.3, p. 036122. arXiv: 0308059 [physics] (cit. on pp. 56, 69).
- Pistorius, C.W.I. and J.M. Utterback (1997). "Multi-mode interaction among technologies". In: *Research Policy* 26.1, pp. 67–84 (cit. on pp. 85, 101).
- Polechová, Jitka (2018). "Is the sky the limit? On the expansion threshold of a species' range". In: *PLoS Biology* 16.6, pp. 1–18 (cit. on pp. 14, 29, 30, 158).
- Polechová, Jitka and Nicholas H. Barton (2015). "Limits to adaptation along environmental gradients". In: *Proceedings of the National Academy of Sciences of the United States of America* 112.20, pp. 6401–6406 (cit. on pp. 29, 30, 158).
- Pontarp, Mikael, Åke Brännström, and Owen L. Petchey (2019). "Inferring community assembly processes from macroscopic patterns using dynamic ecoevolutionary models and Approximate Bayesian Computation (ABC)". In: *Methods in Ecology and Evolution* 10.4. Ed. by Timothée Poisot, pp. 450–460 (cit. on pp. 4, 6, 85).
- Post, David M., M. Elizabeth Conners, and Debra S. Goldberg (2000). "Prey preference by a top predator and the stability of linked food chains". In: *Ecology* 81.1, pp. 8–14 (cit. on p. 63).
- Purves, Drew, Jörn P. W. Scharlemann, Mike Harfoot, et al. (2013). "Time to model all life on Earth". In: *Nature* 493.7432, pp. 295–297 (cit. on p. 56).
- Rackauckas, Christopher, Yingbo Ma, Julius Martensen, et al. (2020a). "Universal Differential Equations for Scientific Machine Learning". In: arXiv: 2001.04385 (cit. on pp. 7, 8, 57, 62, 69, 160, 161).
- Rackauckas, Christopher, Yingbo Ma, Julius Martensen, et al. (2020b). "Universal Differential Equations for Scientific Machine Learning". In: arXiv:2001.04385v3, 18 pages (cit. on p. 8).
- Rackauckas, Christopher and Qing Nie (2017). "DifferentialEquations.jl a performant and feature-rich ecosystem for solving differential equations in Julia". In: *Journal of Open Research Software* 5 (cit. on pp. 33, 62).
- Rahbek, Carsten, Michael K. Borregaard, Alexandre Antonelli, et al. (2019a). "Building mountain biodiversity: Geological and evolutionary processes". In: *Science* 365.6458, pp. 1114–1119 (cit. on pp. 2, 3, 13, 165).
- Rahbek, Carsten, Michael K. Borregaard, Robert K. Colwell, et al. (2019b). "Humboldt's enigma: What causes global patterns of mountain biodiversity?" In: *Science* 365.6458, pp. 1108–1113 (cit. on p. 1).
- Rahbek, Carsten and Gary R. Graves (2001). "Multiscale assessment of patterns of avian species richness". In: *Proceedings of the National Academy of Sciences of the United States of America* 98.8, pp. 4534–4539 (cit. on p. 28).
- Raissi, M., P. Perdikaris, and G.E. Karniadakis (2019a). "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In: *Journal of Computational Physics* 378, pp. 686–707 (cit. on pp. 70, 161, 164).

- Raissi, Maziar, Paris Perdikaris, and George E Karniadakis (2019b). “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations”. In: *J. Comput. Phys.* 378, pp. 686–707 (cit. on p. 127).
- Ramsay, J. O., G. Hooker, D. Campbell, and J. Cao (2007). “Parameter estimation for differential equations: a generalized smoothing approach”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.5, pp. 741–796 (cit. on p. 70).
- Rangel, Thiago F., Neil R. Edwards, Philip B. Holden, et al. (2018). “Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves”. In: *Science* 361.6399 (cit. on pp. 2, 3).
- Raue, Andreas, C. Kreutz, T. Maiwald, et al. (2009). “Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood”. In: *Bioinformatics* 25.15, pp. 1923–1929 (cit. on p. 58).
- Richardson, Jonathan L., Mark C. Urban, Daniel I. Bolnick, and David K. Skelly (2014). “Microgeographic adaptation and the spatial scale of evolution”. In: *Trends in Ecology & Evolution* 29.3, pp. 165–176 (cit. on pp. 29, 30).
- Rodriguez-Fernandez, Maria, Pedro Mendes, and Julio R. Banga (2006). “A hybrid approach for efficient and robust parameter estimation in biochemical pathways”. In: *Biosystems* 83.2-3, pp. 248–265 (cit. on p. 69).
- Rogers, Everett M (2003). *Diffusion of Innovations*. 5th ed. Riverside: Free Press (cit. on pp. 86, 100, 158).
- Rolnick, David, Priya L. Donti, Lynn H. Kaack, et al. (2022). “Tackling Climate Change with Machine Learning”. In: *ACM Computing Surveys* 55.2, pp. 1–96. arXiv: 1906.05433 (cit. on pp. 7, 161).
- Romer, Paul M (1986). “Increasing Returns and Long-Run Growth”. In: *Journal of Political Economy* 94.5, pp. 1002–1037 (cit. on p. 4).
- Roques, Lionel and Olivier Bonnefon (2016). “Modelling Population Dynamics in Realistic Landscapes with Linear Elements: A Mechanistic-Statistical Reaction-Diffusion Approach”. In: *PLoS ONE* 11.3. Ed. by Zhen Jin, Article No. e0151217, 20 pp. (Cit. on p. 125).
- Rosenbaum, Benjamin, Michael Raatz, Guntram Weithoff, Gregor F. Fussmann, and Ursula Gaedke (2019). “Estimating Parameters From Multiple Time Series of Population Dynamics Using Bayesian Inference”. In: *Frontiers in Ecology and Evolution* 6.JAN (cit. on pp. 56, 69, 160).
- Rubinstein, Jacob and Peter Sternberg (1992). “Nonlocal reaction-diffusion equations and nucleation”. In: *IMA J. Appl. Math.* 48.3, pp. 249–264 (cit. on pp. 125, 138, 151).
- Ruder, Sebastian (2016). “An overview of gradient descent optimization algorithms”. In: pp. 1–14. arXiv: 1609.04747 (cit. on p. 62).
- Ruppert, Krista M., Richard J. Kline, and Md Saydur Rahman (2019). “Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA”. In: *Global Ecology and Conservation* 17, e00547 (cit. on p. 72).

- Ruth, Matthias (1996). "Evolutionary economics at the crossroads of biology and physics". In: *Journal of Social and Evolutionary Systems* 19.2, pp. 125–144 (cit. on p. 85).
- Saavedra, S., R. P. Rohr, L. J. Gilarranz, and J. Bascompte (2014). "How structurally stable are global socioeconomic systems?" In: *Journal of The Royal Society Interface* 11.100, pp. 20140693–20140693 (cit. on pp. 4, 86).
- Saavedra, Serguei, Felix Reed-tsochas, and Brian Uzzi (2009). "A simple model of bipartite cooperation for ecological and organizational networks". In: *Nature* 457.7228, pp. 436–466 (cit. on pp. 4, 85, 86, 100, 101, 158).
- Saviotti, Pier-Paolo, Andreas Pykad', and Bogang Jun (2020). "Diversification, structural change, and economic development". In: *Journal of Evolutionary Economics* 30.5, pp. 1301–1335 (cit. on p. 100).
- Savolainen, Outi, Tanja Pyhäjärvi, and Timo Knürr (2007). "Gene Flow and Local Adaptation in Trees". In: *Annual Review of Ecology, Evolution, and Systematics* 38.1, pp. 595–619 (cit. on p. 3).
- Sayama, Hiroki (2015). *Introduction to the Modeling and Analysis of Complex Systems*. Tech. rep. (cit. on p. 5).
- Schartau, Markus, Philip Wallhead, John Hemmings, et al. (2017). "Reviews and syntheses: parameter identification in marine planktonic ecosystem modelling". In: *Biogeosciences* 14.6, pp. 1647–1701 (cit. on pp. 6, 55, 56, 58, 70, 71, 91, 160).
- Scheffer, Marten, Steve Carpenter, Jonathan A Foley, Carl Folke, and Brian Walker (2001). "Catastrophic shifts in ecosystems". In: *Nature* 413.6856, pp. 591–596 (cit. on p. 55).
- Scheffer, Marten and Egbert H van Nes (2006). "Self-organized similarity, the evolutionary emergence of groups of similar species." In: *Proceedings of the National Academy of Sciences of the United States of America* 103.16, pp. 6230–5 (cit. on pp. 87, 100).
- Scheiter, Simon, Liam Langan, and Steven I. Higgins (2013). "Nextgeneration dynamic global vegetation models: learning from community ecology". In: *New Phytologist* 198.3, pp. 957–969 (cit. on p. 55).
- Schneider, Tapio, Shiwei Lan, Andrew Stuart, and João Teixeira (2017). "Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted HighResolution Simulations". In: *Geophysical Research Letters* 44.24, pp. 12,396–12,417. arXiv: 1709.00037 (cit. on pp. 7, 58, 70, 160, 161).
- Schoener, T. W. (2011). "The Newest Synthesis: Understanding the Interplay of Evolutionary and Ecological Dynamics". In: *Science* 331.6016, pp. 426–429 (cit. on p. 3).
- Scholl, Maarten Peter, Anisoara Calinescu, and J. Doyne Farmer (2021). "How market ecology explains market malfunction". In: *Proceedings of the National Academy of Sciences* 118.26, e2015574118. arXiv: 2009.09454 (cit. on pp. 4, 86).
- Schumpeter, Joseph A (1934). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle* (cit. on pp. 1, 85).
- Schweitzer, Frank, Giorgio Fagiolo, Didier Sornette, and Douglas R. White (2009). "Economic Networks : The New Challenges". In: *Science* July, pp. 422–426 (cit. on pp. 101, 158).

- Silverberg, Gerald and Bart Verspagen (2005). “Evolutionary theorizing on economic growth”. In: *The Evolutionary Foundations of Economics* August, pp. 506–539 (cit. on p. 101).
- Sirignano, Justin and Konstantinos Spiliopoulos (2018). “DGM: A deep learning algorithm for solving partial differential equations”. In: *J. Comput. Phys.* 375, pp. 1339–1364 (cit. on pp. 7, 127).
- Skeels, A, W Bach, O Hagen, W Jetz, and L Pellissier (2022). “Temperature-Dependent Evolutionary Speed Shapes the Evolution of Biodiversity Patterns Across Tetrapod Radiations”. In: *Systematic Biology* 0.0. Ed. by Luke Harmon, pp. 1–16 (cit. on pp. 7, 85, 161, 163).
- Skeels, Alexander and Marcel Cardillo (2019). “Reconstructing the Geography of Speciation from Contemporary Biodiversity Data”. In: *The American Naturalist* 193.2, pp. 240–255 (cit. on pp. 161, 163).
- Slatkin, M (1987). “Gene flow and the geographic structure of natural populations”. In: *Science* 236.4803, pp. 787–792 (cit. on p. 155).
- Slatkin, Montgomery (1973). “GENE FLOW AND SELECTION IN A CLINE”. In: *Genetics* 75.4, pp. 733–756 (cit. on pp. 29, 158).
- (1993). “Isolation by distance in equilibrium and non-equilibrium populations”. In: *Evolution* 47.1, pp. 264–279 (cit. on p. 13).
  - (1978). “Spatial patterns in the distributions of polygenic characters”. In: *Journal of Theoretical Biology* 70.2, pp. 213–228 (cit. on pp. 14, 29, 158).
- Sornette, Didier (2014). “Physics and financial economics (1776–2014): puzzles, Ising and agent-based models”. In: *Reports on progress in physics* 77.6, p. 62001 (cit. on p. 102).
- Spitz, Y.H., J.R. Moisan, M.R. Abbott, and J.G. Richman (1998). “Data assimilation and a pelagic ecosystem model: parameterization using time series observations”. In: *Journal of Marine Systems* 16.1-2, pp. 51–68 (cit. on p. 56).
- Stein, Anke, Katharina Gerstner, and Holger Kreft (2014). “Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales”. In: *Ecology Letters* 17.7, pp. 866–880 (cit. on pp. 28, 30).
- Stoleriu, Iulian (2011). “Non-local models for solid-solid phase transitions”. In: *ROMAI J.* 7.1, pp. 157–170 (cit. on p. 125).
- Straume, E.O., C Gaina, S Medvedev, and K.H. Nisancioglu (2020). “Global Cenozoic Paleobathymetry with a focus on the Northern Hemisphere Oceanic Gateways”. In: *Gondwana Research* 86, pp. 126–143 (cit. on p. 165).
- Strogatz, Steven H (2018). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC press (cit. on p. 73).
- Sukumar, N. and Ankit Srivastava (2022). “Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks”. In: *Comput. Methods Appl. Mech. Engrg.* 389, Article No. 114333 (cit. on p. 127).

- Sukumaran, Jeet, Evan P. Economo, and L. Lacey Knowles (2016). "Machine learning biogeographic processes from biotic patterns: A new trait-dependent dispersal and diversification model with model choice by simulation-Trained discriminant analysis". In: *Systematic Biology* 65.3, pp. 525–545 (cit. on pp. 161, 163).
- Sunderasan, Srinivasan (2020). "Financial Modeling". In: *Long-Term Investments*. Routledge India, pp. 33–51 (cit. on p. 125).
- Suweis, Samir, Joel A Carr, Amos Maritan, Andrea Rinaldo, and Paolo D'Odorico (2015). "Resilience and reactivity of global food security". In: *Proceedings of the National Academy of Sciences* 112.22, pp. 6902–6907 (cit. on pp. 4, 86).
- Suzuki, Yuka and Evan P. Economo (2021). "From species sorting to mass effects: spatial network structure mediates the shift between metacommunity archetypes". In: *Ecography*, p. 05453 (cit. on pp. 13, 28).
- Tacchella, A., D. Mazzilli, and L. Pietronero (2018). "A dynamical systems approach to gross domestic product forecasting". In: *Nature Physics* 14.8, pp. 861–865 (cit. on pp. 4, 86, 91).
- Templeton, A R (1981). "Mechanisms of Speciation - A Population Genetic Approach". In: *Annual Review of Ecology and Systematics* 12.1, pp. 23–48 (cit. on p. 165).
- Thompson, Patrick L., Bronwyn Rayfield, and Andrew Gonzalez (2017). "Loss of habitat and connectivity erodes species diversity, ecosystem functioning, and stability in metacommunity networks". In: *Ecography* 40.1, pp. 98–108 (cit. on pp. 13, 18, 28).
- Tilman, David (1994). "Competition and Biodiversity in Spatially Structured Habitats". In: *Ecology* 75.1, pp. 2–16 (cit. on pp. 88, 100).
- Tkadlec, Josef, Andreas Pavlogiannis, Krishnendu Chatterjee, and Martin A. Nowak (2019). "Population structure determines the tradeoff between fixation probability and fixation time". In: *Communications Biology* 2.1, p. 138 (cit. on pp. 13, 158).
- Urban, M. C., G. Bocedi, A. P. Hendry, et al. (2016). "Improving the forecast for biodiversity under climate change". In: *Science* 353.6304 (cit. on pp. 3, 7, 9, 55, 69, 70, 72, 165).
- Van Der Panne, Gerben (2004). "Agglomeration externalities: Marshall versus Jacobs". In: *Journal of Evolutionary Economics* 14.5, pp. 593–604 (cit. on pp. 85, 100, 158).
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, et al. (2017). "Attention is All you Need". In: *Advances in Neural Information Processing Systems*. Ed. by I Guyon, U Von Luxburg, S Bengio, et al. Vol. 30. Curran Associates, Inc. (cit. on p. 56).
- Veblen, Thorstein (1898). "Why is Economics not an Evolutionary Science?" In: *The Quarterly Journal of Economics* 12.4, p. 373 (cit. on pp. 1, 85).
- Veech, Joseph A. and Thomas O. Crist (2007). "Habitat and climate heterogeneity maintain beta-diversity of birds among landscapes within ecoregions". In: *Global Ecology and Biogeography* 16.5, pp. 650–656 (cit. on p. 28).
- Veldhuis, Michiel P., Matty P. Berg, Michel Loreau, and Han Olff (2018). "Ecological autocatalysis: a central principle in ecosystem organization?" In: *Ecological Monographs* 88.3, pp. 304–319 (cit. on pp. 2, 166).

- Vellend, Mark (2010a). "Conceptual Synthesis in Community Ecology". In: *The Quarterly Review of Biology* 85.2, pp. 183–206 (cit. on p. 2).
- (2010b). "Conceptual Synthesis in Community Ecology". In: *The Quarterly Review of Biology* 85.2, pp. 183–206. arXiv: arXiv:1011.1669v3 (cit. on p. 85).
- Villa, Chiara, Mark A. J. Chaplain, and Tommaso Lorenzi (2021). "Evolutionary dynamics in vascularised tumours under chemotherapy: mathematical modelling, asymptotic analysis and numerical simulations". In: *Vietnam J. Math.* 49.1, pp. 143–167 (cit. on p. 125).
- Voulodimos, Athanasios, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis (2018). "Deep learning for computer vision: A brief review". In: *Computational intelligence and neuroscience* 2018 (cit. on p. 7).
- Wang, Fang, Ling Xue, Kun Zhao, and Xiaoming Zheng (2021). "Global stabilization and boundary control of generalized Fisher/KPP equation and application to diffusive SIS model". In: *J. Differential Equations* 275, pp. 391–417 (cit. on pp. 125, 136, 141).
- Wang, Ian J. and Gideon S. Bradburd (2014). "Isolation by environment". In: *Molecular Ecology* 23.23, pp. 5649–5662 (cit. on pp. 13, 30).
- Wang, Sifan and Paris Perdikaris (2020). "Deep learning of free boundary and Stefan problems". In: *J. Comput. Phys.*, Article No. 109914, 27 pp. (Cit. on p. 127).
- Ward, Ben A., Marjorie A.M. Friedrichs, Thomas R. Anderson, and Andreas Oschlies (2010). "Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models". In: *Journal of Marine Systems* 81.1-2, pp. 34–43 (cit. on p. 56).
- Welling, Max and Yee Whye Teh (2011). "Bayesian Learning via Stochastic Gradient Langevin Dynamics". In: *ICML* (cit. on p. 164).
- Wernerfelt, Birger (1989). "From Critical Resources to Corporate Strategy". In: *Journal of General Management* 14.3, pp. 4–12 (cit. on p. 85).
- Whitlock, Michael C. (2008). "Evolutionary inference from Q ST". In: *Molecular Ecology* 17.8, pp. 1885–1896 (cit. on p. 15).
- Wickman, Jonas, Sebastian Diehl, Bernd Blasius, et al. (2017). "Determining selection across heterogeneous landscapes: A perturbation-based method and its application to modeling evolution in space". In: *The American Naturalist* 189.4, pp. 381–395 (cit. on p. 14).
- Wickman, Jonas, Thomas Koffel, and Christopher A Klausmeier (2021). "A general theoretical framework for trait-based eco-evolutionary dynamics : population structure , intraspecific variation , and community assembly". In: *bioRxiv* (cit. on p. 162).
- Wilke, Claus O., Jia Lan Wang, Charles Ofria, Richard E. Lenski, and Christoph Adami (2001). "Evolution of digital organisms at high mutation rates leads to survival of the flattest". In: *Nature* 412.6844, pp. 331–333 (cit. on p. 69).
- Willard, Jared, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar (2020). "Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems". In: 1.1, pp. 1–35. arXiv: 2003.04919 (cit. on p. 56).
- Wilson, Andrew Gordon (2020). "The Case for Bayesian Deep Learning". In: pp. 1–7. arXiv: 2001.10995 (cit. on pp. 163, 164).

- Wright, Sewal (1943). "Isolation By Distance". In: *Genetics* (cit. on p. 155).
- Xiao, Yongjin and Marjorie A. M. Friedrichs (2014). "The assimilation of satellite-derived data into a onedimensional lower trophic level marine ecosystem model". In: *Journal of Geophysical Research: Oceans* 119.4, pp. 2691–2712 (cit. on p. 56).
- Xu, Tao, Luther White, Dafeng Hui, and Yiqi Luo (2006). "Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction". In: *Global Biogeochemical Cycles* 20.2, n/a–n/a (cit. on pp. 56, 69, 160).
- Yazdani, Alireza, Lu Lu, Maziar Raissi, and George Em Karniadakis (2020). "Systems biology informed deep learning for inferring parameters and hidden dynamics". In: *PLOS Computational Biology* 16.11. Ed. by Vassily Hatzimanikatis, e1007575 (cit. on pp. 7, 70, 161, 164).
- Ye, Hao, Richard J. Beamish, Sarah M. Glaser, et al. (2015). "Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling". In: *Proceedings of the National Academy of Sciences* 112.13, E1569–E1576 (cit. on p. 55).
- Ye, Hao and George Sugihara (2016). "Information leverage in interconnected ecosystems: Overcoming the curse of dimensionality". In: *Science* 353.6302, pp. 922–925 (cit. on pp. 55, 63, 70).
- Yeaman, Sam and Sarah P. Otto (2011). "Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift". In: *Evolution* 65.7, pp. 2123–2129 (cit. on pp. 28, 29, 158).
- Young, Tom, Devamanyu Hazarika, Soujanya Poria, and Erik Cambria (2018). "Recent trends in deep learning based natural language processing". In: *ieee Computational intelligenCe magazine* 13.3, pp. 55–75 (cit. on p. 7).
- Yuan, Lei, Yi-Qing Ni, Xiang-Yun Deng, and Shuo Hao (2022). "A-PINN: Auxiliary physics informed neural networks for forward and inverse problems of nonlinear integro-differential equations". In: *J. Comput. Phys.* Early access version available online, Article No. 111260 (cit. on p. 127).
- Zahra, Shaker A., R. Duane Ireland, and Michael A. Hitt (2000). "International Expansion by New Venture Firms: International Diversity, Mode of Market Entry, Technological Learning, and Performance". In: *Academy of Management Journal* 43.5, pp. 925–950 (cit. on pp. 86, 100, 158).
- Zang, Yaohua, Gang Bao, Xiaojing Ye, and Haomin Zhou (2020). "Weak adversarial networks for high-dimensional partial differential equations". In: *Journal of Computational Physics* 411, Article No. 109409 (cit. on p. 127).
- Zhang, Guanglu, Daniel A. McAdams, Venkatesh Shankar, and Milad Mohammadi Darani (2018). "Technology Evolution Prediction Using LotkaVolterra Equations". In: *Journal of Mechanical Design* 140.6, pp. 1–9 (cit. on pp. 4, 86).
- Zhu, Qing and Qianlai Zhuang (2015). "Ecosystem biogeochemistry model parameterization: Do more flux data result in a better model in predicting carbon flux?" In: *Ecosphere* 6.12, art283 (cit. on p. 56).

## Webpages

Boussange, Victor (2021). *EvoId.jl*. URL: <https://github.com/vboussange/EvoId.jl> (visited on Sept. 19, 2022) (cit. on p. 160).

- (2022a). *HighDimPDE.jl*. URL: <https://github.com/SciML/HighDimPDE.jl> (visited on Sept. 19, 2022) (cit. on p. 160).
- (2022b). *MiniBatchInference.jl*. URL: <https://github.com/vboussange/MiniBatchInference.jl> (visited on Sept. 19, 2022) (cit. on p. 161).

*Contributions to the main branch of HighDimPDE.jl* (2022). URL: <https://github.com/SciML/HighDimPDE.jl/graphs/contributors> (visited on Sept. 19, 2022) (cit. on p. 160).

*TensorFlow* (2015). URL: <https://github.com/tensorflow/tensorflow> (visited on Sept. 19, 2022) (cit. on p. 8).

The Linux Foundation (2016). *PyTorch*. URL: <https://github.com/pytorch/pytorch> (visited on Sept. 19, 2022) (cit. on p. 8).



# List of Figures

1.1	<b>Graphical representation of organizational levels and their couplings in ecological and economic systems.</b> An arrow indicates that the organizational level at its tail can influence the organizational level at its head. Left diagram is inspired from Hendry, 2016. . . . .	2
1.2	<b>Graphical representation of the eco-evolutionary processes determining eco-evolutionary dynamics in ecological systems.</b> By extension, I use this terminology to designate interaction and evolutionary processes in economic systems. . . . .	3
1.3	<b>Forward and inverse modelling approaches for the understanding of complex adaptive systems.</b> A forward modelling approach consists in deriving a model, embedding a set of processes inspired from prior knowledge. The objective is to understand how the interplay between the processes considered transforms in (feedback) mechanisms that are associated with a prediction. An inverse modelling approach integrates empirical observation within the modelling process. Posterior knowledge on the empirical system is obtained whether form the interpretation of the estimated parameter values, or through model selection. . . . .	6
2.1	<b>Graphical representation of the structure of individuals in the eco-evolutionary model. (a) Setting with no selection, where individuals are characterized by a set of neutral traits <math>u \in \mathcal{U}</math>.</b> The scatter plots represent a projection of the first two components of $u$ for the individuals present on the designated vertices at time $t = 1000$ , obtained from one simulation of the IBM. <b>(b) Setting with heterogeneous selection.</b> In this setting, individuals are additionally characterized by adaptive traits $s \in \mathcal{S}$ . Blue vertices favour the optimal adaptive trait value $\theta_\bullet$ , while red vertices favour $\theta_\circ$ . The scatter plots represent a projection of the first component of $u$ and $s$ for the individuals present on the designated vertices at time $t = 1000$ , obtained from one simulation. The majority of individuals are locally well-adapted and have an adaptive trait close to the optimal value, but some maladaptive individuals originating from neighbouring vertices are also present. $m = 0.05$ . . . . .	17

- 2.2 Effect of  $\langle l \rangle$  and  $h_d$  on average population size  $\bar{N}$  and neutral differentiation  $Q_{ST,u}$  in the setting with no selection. (a) Response of  $\bar{N}$  to homogeneity in degree  $h_d = \langle \sqrt{k} \rangle^2 / \langle k \rangle$  for all undirected connected graphs with  $M = 7$  vertices and  $m = 0.5$ . (b) Response of  $Q_{ST,u}$  to average path length  $\langle l \rangle$  for similar simulations obtained with  $m = 0.01$ . (c) Response of  $Q_{ST,u}$  to homogeneity in degree  $h_d$  for the same data. In (a), (b) and (c), each dot represents average results from 5 replicate simulations of the IBM, the colour scale corresponds to the proportion of the graphs with similar  $x$  and  $y$  axis values (graph density), and the blue line corresponds to a linear fit. (d) Standardized effect of  $h_d$  and  $\langle l \rangle$  on  $Q_{ST,u}$ , obtained from multivariate regression models independently fitted on similar data obtained for  $m = 0.01$  and  $m = 0.5$ . The contributions of  $\langle l \rangle$  and  $h_d$  to  $Q_{ST,u}$  are alike for low migration regimes. Error bars show 95% confidence intervals. Analogous results on graphs with  $M = 9$  vertices are presented in Fig. S2.6 and all regression details can be found in Table S2.2. . . . . 20
- 2.3 Effect of habitat assortativity  $r_\Theta$  and migration  $m$  on the local adaptive trait distribution  $\bar{n}^\bullet$  and on the adaptive differentiation level  $Q_{ST,s}$  under the mean field, deterministic approximation Eq. (2.5). (a) Effect of  $m$  and  $r_\Theta$  on  $\bar{n}^\bullet$ . Migration induces the apparition of maladaptive individuals (centred around  $\theta_\bullet = 0.5$ ), which destabilize local adaptation by displacing the mean value of the well-adapted individuals (centred around  $\theta_\bullet = -0.5$ ). Together with the decrease in local adaptation, migration causes a displacement of the mean value of the local trait distribution (represented by the vertical dashed lines), which decreases local population size and adaptive differentiation  $Q_{ST,s}$ . (b) Similar data for higher  $r_\Theta$ . Increasing  $r_\Theta$  increases population size and  $Q_{ST,s}$ . (c) Effect of  $r_\Theta$  on  $Q_{ST,s}$ . The red line indicates the critical migration threshold  $m^*$  predicted by Eq. (2.7);  $Q_{ST,s}$  vanishes when  $m > m^*$ . . . . . 23
- 2.4 Effect of habitat heterogeneity  $r_\Theta$  on  $Q_{ST,s}$  and average population size  $\bar{N}$  for general graph ensembles. (a) Effect of  $r_\Theta$  on  $Q_{ST,s}$  for all undirected connected graphs with  $M = 7$  vertices and varying  $r_\Theta$ , for  $m = 0.1$ . (b) Effect of  $r_\Theta$  on average population size  $\bar{N}$  for the same simulations. In (a) and (b), each dot represents average results from 5 replicate simulations of the IBM, the colour scale corresponds to the proportion of the graphs with similar  $x$  and  $y$  axis values (graph density), and the blue lines correspond to results obtained from the mean field approximation Eq. (2.5). Insights from Eq. (2.5) are congruent with the IBM simulations for complex habitat connectivity patterns at low  $m$ . Similar results with  $m = 0.5$  are presented in Fig. S2.11. . . . . 25

2.5	<b>Effect of <math>r_\Theta</math>, <math>\langle l \rangle</math> and <math>h_d</math> on <math>Q_{ST,s}</math> and <math>Q_{ST,u}</math> in the setting with heterogeneous selection.</b> (a) Comparison of the response of $Q_{ST,u}$ to migration with the response of $Q_{ST,u}$ in the setting with no selection for the complete graph. The dashed vertical blue line corresponds to the critical migration regime $m^*$ predicted by Eq. (2.7). Heterogeneous selection increases $Q_{ST,u}$ when $m < m^*$ , but local adaptation is lost when $m > m^*$ , and in this case $Q_{ST,u}$ reaches similar levels as $Q_{ST,u}$ in the setting with no selection. (b) Response of $Q_{ST,u}$ to $r_\Theta$ and migration for the path graph. $r_\Theta$ correlates positively with $Q_{ST,u}$ for high $m$ , but correlates negatively for low $m$ . In (a–b), each plain dot represents average results from 5 replicate simulations, the bars represent one standard deviation, and each fade dot represents a single replicate value. (c–d) Standardized effect of $h_d$ , $\langle l \rangle$ and $r_\Theta$ on $Q_{ST,s}$ and $Q_{ST,u}$ obtained from a multivariate regression model independently fitted for low and high migration regimes on average results from 5 replicate simulations of the IBM on all undirected connected graphs with $M = 7$ vertices and varying $r_\Theta$ (see Methods). The ambivalence of the effect of $r_\Theta$ on $Q_{ST,u}$ found for the path graph holds for general graph ensembles and adds up to that of $\langle l \rangle$ and $h_d$ . Error bars show 95% confidence intervals. Analogous results on graphs with $M = 9$ vertices are presented in Fig. S2.12 and all regression details can be found in Table S2.3. . . . .	27
S2.6	<b>Effect of <math>\langle l \rangle</math> and <math>h_d</math> on average population size <math>\bar{N}</math> and neutral differentiation <math>Q_{ST,u}</math> under the setting with no selection, analogous to Fig. 2.2 but for 1126 of the 261,080 undirected connected graphs with <math>M = 9</math> vertices.</b> . . . . .	40
S2.7	<b>Effect of the carrying capacity <math>K</math> on <math>Q_{ST,u}</math>, <math>Q_{ST,s}</math> and metapopulation size <math>N</math> for the line graph with <math>M = 7</math> vertices for <math>m = 0.1</math>.</b> Decreasing $K$ increases $Q_{ST,u}$ as it favours drift, but it does not influence $Q_{ST,s}$ . Each boxplot is based on 5 replicate simulations of the IBM, and fade dots represent single values for each replicate. . . . .	41
S2.8	<b>Comparison of the adaptive trait density on one vertex obtained from Eq. (S2.26) (left) and from the IBM simulations (right) in the setting with heterogeneous selection, for the star graph with <math>M = 7</math> vertices.</b> The densities obtained from Eq. (S2.26) and from the IBM are qualitatively similar. . . . .	42
S2.9	<b>Comparison of <math>Q_{ST,s}</math> and <math>N</math> obtained from the deterministic approximation Eq. (S2.26) and from IBM simulations in the setting with heterogeneous selection, on the star graph with <math>M = 7</math> vertices.</b> $Q_{ST,s}$ and population size obtained from Eq. (S2.26) closely match the IBM simulations. Each plain dot represents average results from 5 replicate simulations of the IBM, bars represent one standard deviation, and each fade dot represents a single replicate value. . . . .	43
S2.10	<b>Graphs with spatial distribution of habitat types corresponding to different habitat assortativity <math>r_\Theta</math>.</b> Graphs (ad) can be described exactly with a mean field approach, as blue and red vertices have an equivalent position on the graph. . . . .	43



S2.14 Standardized effects of $h_d$ , $\langle l \rangle$ and $r_\Theta$ on $Q_{ST,u}$ in the setting with no selection and in the setting with heterogeneous selection for the time horizons $t = 1000$ and $t = 2000$ , obtained from multivariate regression models independently fitted for low and high migration regimes to average results from 5 replicate simulations of the IBM on all undirected connected graphs with $M = 7$ vertices and varying $r_\Theta$ (see Methods for details). (a–c) illustrate that the effects of the topology metrics on $Q_{ST,u}$ and $Q_{ST,s}$ remain constant for $t > 1000$ in both the settings without selection and with heterogeneous selection. Error bars show 95% confidence intervals. . . . .	47
S2.15 Comparison of the neutral trait density on one vertex obtained from Eq. (S2.21) (left) and from the IBM simulations (right) in the setting with no selection, for the chain graph. The densities obtained from Eq. (S2.21) and from the IBM are dissimilar. . . . .	48
S2.16 Comparison of results obtained from the deterministic approximations Eqs. (S2.20) and (S2.21) and from IBM simulations in the setting with no selection, on the star graph with $M = 7$ vertices. While Eq. (S2.20) can capture population size, Eq. (S2.21) is not able to capture $Q_{ST,u}$ . Each plain dot represents average results from 5 replicate simulations, bars represent one standard deviation, and each fade dot represents a single replicate value. . . . .	49
<b>3.1 Illustration of the mini-batch method.</b> A Characteristic features of the naive loss function $L_M(\theta)$ and the mini-batch loss function $L_M^*(\theta)$ . The blue and orange lines correspond to cross sections of $L_M(\theta)$ and $L_M^*(\theta)$ , respectively, obtained from the ecosystem model presented in the section Simulated food-web model as a case study. While the cross section of $L_M(\theta)$ presents many local minima and a very large gradient in the neighbourhood of the true parameter, which renders the navigability of the loss surface difficult, the cross section of $L_M^*(\theta)$ is smooth and shows a single minimum, illustrating the regularization induced by the mini-batch method. B Convergence of gradient descent algorithms applied to $L_M$ and $L_M^*$ . The blue, orange and green lines correspond to the loss function evaluated against the epochs (number of parameter updates) using $L_M$ , $L_M$ starting from a parameter value close to the true parameters, and $L_M^*$ , respectively. The ill-behaviour of $L_M$ leads to the convergence to a local minimum, while $L_M^*$ is associated with a smooth convergence to the true parameters. C Graphical representation of the proposed mini-batch method. To improve the navigability of the posterior landscape, the algorithm splits the time series into mini-batches with short time horizons (blue and red portions of the time series). Since mini-batches are treated independently, the method naturally extends to independent time series. . . . .	61
<b>3.2 Reference food web systems considered.</b> In Section 3.3.2, the blue model $M_1$ from (Hastings and Powell, 1991) is considered, where a resource is eaten by consumers, which are themselves eaten by predators. In Section 3.3.3, the orange model $M_2$ , corresponding to the omnivory variant introduced in (McCann and Hastings, 1997), is also considered. . . . .	64

3.3	<b>Performance of the proposed ML framework for varying noise levels, under the complete and the partial observation setting.</b> A True parameter $\tilde{x}_P$ against estimated parameter $\hat{x}_P$ for $r = 0.2$ under the complete and the partial observation setting. Although the parameter estimation is more accurate when complete observations of abundance are available, parameters show a fair fit when estimated with partial observations. B Relative parameter error $ (\hat{p} - \bar{p})/\bar{p} $ for varying noise levels under the setting of complete or partial observations. B supports the above observation for varying noise levels. C Forecast skill $\rho^2$ of the trained model under the complete observation setting. D Analogous data under the partial observation setting. In A–D, the batch size is set to $m = 6$ . . . . .	66
3.4	<b>Performance of the ML framework in processing and combining the information of multiple independent data sets.</b> A Relative parameter error $ (\hat{p} - \bar{p})/\bar{p} $ for different numbers of time series and levels of noise, under the complete and the partial observation setting. B Forecast skill for different numbers of time series and time horizons of the forecasts, under the complete and the partial observation setting. $r = 0.1$ . In A–B, $ (\hat{p} - \bar{p})/\bar{p} $ decreases while $\rho^2$ increases as the number of time series processed increases, demonstrating the capacity of the ML framework to process and combine the information from independent time series. In A–B, each box plot corresponds to 100 independent simulations where $\tilde{x}_P$ varies $\tilde{x}_P \in [0.071, 0.225]$ , and the batch size is set to $m = 6$ . . . . .	67
3.5	<b>Performance of the ML framework in supporting the predator omnivory hypothesis in a food web.</b> A–C Hypothesis testing for levels of noise $r = 0.1, 0.2, 0.3$ . Blue dots correspond to $w_{\mathcal{M}_1}$ , the Akaike weights of the simple food-web model, and orange dots correspond to $w_{\mathcal{M}_2} = 1 - w_{\mathcal{M}_1}$ , the Akaike weights of the omnivory model, which can be interpreted as model probabilities. A–C indicate that the ML framework can detect omnivory, as the omnivory model $\mathcal{M}_2$ is given strong support ( $w_{\mathcal{M}_2} > 99\%$ ) for most of the $\omega$ range investigated. . . . .	69
S3.6	<b>Performance of the ML framework in supporting the predator omnivory hypothesis in a food web for the partial observation setting.</b> In A, B and C where $r = 0.1, 0.2, 0.3$ , the lack of data prevents the correct estimation of the omnivory variant model parameters, leading model $\mathcal{M}_1$ to be supported for a wider range of $\omega$ in contrast to the complete observation setting (Fig. 3.5A). . . . .	79
S3.7	<b>Divergence between the trajectory <math>x(t)</math> and a perturbed trajectory <math>x_{\delta x_0}(t)</math>, obtained from the reference food-web model from Hastings and Powell, 1991 and detailed in Section 3.B.</b> For $t \lesssim 100$ , $x(t)$ and $x_{\delta x_0}(t)$ are correlated and the divergence regime is informative, but for $t \gtrsim 100$ the trajectories become essentially uncorrelated, corresponding to the mixed divergence regime. . . . .	80
S3.8	<b>Divergence between the trajectory <math>x(t)</math> and a perturbed trajectory <math>x_{\delta p}(t)</math>, obtained from the reference food-web model from Hastings and Powell, 1991 and detailed in Section 3.B.</b> For $t \lesssim 40$ , $x(t)$ and $x_{\delta p}(t)$ are correlated and the divergence regime is informative, but for $t \gtrsim 40$ the trajectories become essentially uncorrelated, corresponding to the mixed divergence regime. . . . .	81

**4.1 General population dynamic model used to characterize the dynamics of economic activities, capturing interactions between themselves, their spatial dispersal and their transformations into other economic activities.** For each country  $c$ , each one of the 9 economic activities considered is assumed to be driven by replication and self limitation processes, characterized respectively by a growth rate  $r_i^{(c)}$  and a self-limitation coefficient  $b_i^{(c)}$ , which values are assumed to be country-dependent and determined by the institutional context and the resources available (in terms of human capital, knowledge capital and natural resources). In addition, interactions between activities within each country are determined by  $\alpha^{(c)}$  and can be mutualistic ( $\alpha^{(c)} > 0$ ) or competitive ( $\alpha^{(c)} < 0$ ). Due to spatial dispersal, the capital of an economic activity can experience positive (resp. negative) fluxes of capital when it is lower (resp. higher) than the average global capital for the economic activity at time  $t$ , arising at rate  $\delta^{(c)}$ . Due to transformations into other economic activities, the capital of an economic activity can experience positive (resp. negative) fluxes of capital when it is lower (resp. higher) than the capital of other economic sectors within the country at time  $t$ , arising at rate  $\mu^{(c)}$ . The parameters  $r_i^{(c)}$ ,  $b_i^{(c)}$ ,  $\alpha^{(c)}$ ,  $\delta^{(c)}$  and  $\mu^{(c)}$  that best fit the empirical data are estimated in each country for each sub-model in Eqs. (4.2) to (4.6), to investigate whether the proposed processes have a significant effect on the temporal dynamics of economic activities.

This diagram has been designed using resources from Flaticon.com. . . . . 90

**4.2 Validation of the model selection procedure.** In **A** and **B** we generate data with models  $\mathcal{M}_{\alpha^-}$  and  $\mathcal{M}_{\alpha^+}$  with varying  $\alpha$ , in **C** we generate data with model  $\mathcal{M}_\delta$  with varying  $\delta$  and in **D** we generate data with model  $\mathcal{M}_\mu$  with varying  $\mu$ . We then contaminate the generated data with realistic noise ( $\sigma = 0.2$ ), and seek for the maximum likelihood of all models  $\mathcal{M}_{null}$ ,  $\mathcal{M}_{\alpha^-}$ ,  $\mathcal{M}_{\alpha^+}$ ,  $\mathcal{M}_\delta$  and  $\mathcal{M}_\mu$ . The maximum likelihood of each model is then used to compute difference in BIC for each model,  $\Delta BIC$ , which comparison provides a strength-of-evidence for each model. **A-D** show that the model selection procedure is valid, since only the true models are placed in the "Strong support" category when no other model is given support (equivalently, when all other model are placed in the "No support" category). . . . . 94

**4.3 Best model fits and associated predictions for DEU, FRA and ARE, and effect of GDP on best model loglikelihoods.** **A-C** Best model fits and associated predictions for DEU, FRA and ARE. Scatter dots represent the empirical data, plain lines correspond to the best model fits, and dotted lines correspond to the best model predictions. The inset plots provide the  $\Delta BIC$  for each country. **D** Effect of GDP on model loglikelihood. The plain line corresponds to a linear fit. 96

4.4 Statistical support or the population dynamic models $\mathcal{M}_{\alpha^-}$ , $\mathcal{M}_{\alpha^+}$ , $\mathcal{M}_\delta$ and $\mathcal{M}_\mu$ , based on Bayesian information criterion f. A Strength-of-evidence comparison for $\mathcal{M}_{\alpha^-}$ , $\mathcal{M}_{\alpha^+}$ , $\mathcal{M}_\delta$ and $\mathcal{M}_\mu$ against $\mathcal{M}_{null}$ , for countries where $\mathcal{M}_{null}$ is rejected, sorted in decreasing order. B Number of countries where models $\mathcal{M}_{null}$ , $\mathcal{M}_{\alpha^+}$ , $\mathcal{M}_{\alpha^-}$ , $\mathcal{M}_\delta$ , and $\mathcal{M}_\mu$ are ranked as best models. C Number of countries where models $\mathcal{M}_{\alpha^-}$ , $\mathcal{M}_{\alpha^+}$ , $\mathcal{M}_\delta$ and $\mathcal{M}_\mu$ are accepted against $\mathcal{M}_{null}$ . D Global strength-of-evidence comparison for $\mathcal{M}_{\alpha^-}$ , $\mathcal{M}_{\alpha^+}$ , $\mathcal{M}_\delta$ , and $\mathcal{M}_\mu$ against $\mathcal{M}_{null}$ . A-D show that $\mathcal{M}_{\alpha^+}$ and $\mathcal{M}_\delta$ stand out as the most supported models, both in term of number of countries where they are given support, and in term of strength-of-evidence. . . . .	98
4.5 Differences in growth rates $r_i$ and coefficients of self-limitation $b_i$ between $\mathcal{M}_{\alpha^-}$ , $\mathcal{M}_{\alpha^+}$ , $\mathcal{M}_\delta$ , $\mathcal{M}_\mu$ , and $\mathcal{M}_{null}$ . In A-B, we aggregate the parameters $r_i$ and $b_i$ across all economic activities and all countries where models are supported. A shows that $\mathcal{M}_\delta$ overestimates $r_i$ against the other models. This is because for a dispersal rate $\delta > 0$ , most economic activities experience a negative capital flux, and consequently need to grow more to compensate for the losses. B shows that $\mathcal{M}_{\alpha^+}$ tends to overestimate $b_i$ relative to the other models, which negative effect on growth dynamics compensates the beneficial effect of positive interactions. . . . .	99
S4.6 Validation of the model selection procedure. The figure is analogous to Fig. 4.2 but with $\sigma = 0.3$ . . . . .	103
S4.7 Fits of model $\mathcal{M}_{\alpha^-}$ for ARE and HRV, where $\mathcal{M}_{\alpha^-}$ is given support against $\mathcal{M}_{null}$ . . . . .	104
S4.8 Fits of model $\mathcal{M}_{\alpha^+}$ for DEU and HUN, where $\mathcal{M}_{\alpha^+}$ is given support against $\mathcal{M}_{null}$ . . . . .	105
S4.9 Fits of model $\mathcal{M}_\delta$ for GBR and BEL, where $\mathcal{M}_\delta$ is given support against $\mathcal{M}_{null}$ . . . . .	106
S4.10 Fits of model $\mathcal{M}_\mu$ for NZL and KOR, where $\mathcal{M}_\mu$ is given support against $\mathcal{M}_{null}$ . . . . .	107
S4.11 Contributions of fluxes due to spatial dispersal on growth dynamics. A shows the growth dynamics for KOR, together with the best model fit. B-C show the positive and negative fluxes, respectively, due to spatial dispersal for each economic activity, relative to the capital size of economic activity. The contributions of spatial dispersal to growth decreases over time. . . . .	108
5.1 Graphical illustration of the neural network architecture used in the numerical simulations. In Sections 5.5.1 to 5.5.5 we employ neural networks with 4 layers (corresponding to 3 affine linear transformations in the neural networks) with $d$ neurons on the input layer (corresponding to a $d$ -dimensional input layer), with $d + 50$ neurons on the 1st hidden layer (corresponding to a $(d + 50)$ -dimensional 1st hidden layer), with $d + 50$ neurons on the 2nd hidden layer (corresponding to a $(d + 50)$ -dimensional 2nd hidden layer), and with 1 neuron on the output layer (corresponding to a 1-dimensional output layer) in the numerical simulations. . . . .	140

- 5.2 **Plot of a machine learning-based approximation of the solution of the replicator-mutator PDE in Eq. (5.52) in the case  $d = 5$ ,  $T = 1/2$ , and  $\mathcal{D} = \mathbb{R}^d$ .** The left-hand side shows a plot of  $[-1/4, 1/4] \ni x \mapsto \mathbb{V}_n^{1,0}(\Theta_{M_n}^n(\omega), (x, 0, \dots, 0)) \in \mathbb{R}$  for  $n \in \{0, 1, 2, 3\}$  and one realization  $\omega \in \Omega$  where the functions  $\mathbb{R}^d \ni x \mapsto \mathbb{V}_n^{1,0}(\Theta_{M_n}^n(\omega), x) \in \mathbb{R}$  for  $n \in \{0, 1, 2, 3\}$ ,  $\omega \in \Omega$  were computed via Lemma 5.5.1 as an approximation of the solution of the PDE in Eq. (5.52) with  $d = 5$ ,  $T = 1/2$ , and  $\mathcal{D} = [-1/2, 1/2]^d$  where we take  $M_1 = M_2 = \dots = M_N = 2000$ ,  $\gamma_1 = \gamma_2 = \dots = \gamma_{2000} = 1/200$ , and  $\delta = \mathbb{1}_{\mathbb{R}^d}$  and where we take  $\xi^{n,m,j}: \Omega \rightarrow \mathbb{R}^d$ ,  $n, m, j \in \mathbb{N}$ , to be independent  $\mathcal{U}_{[-1/2, 1/2]^d}$ -distributed random variables. The right-hand side of Fig. 5.2 shows a plot of  $[-1/4, 1/4] \ni x \mapsto u(t, (x, 0, \dots, 0)) \in \mathbb{R}$  for  $t \in \{0, 0.05, 0.1, 0.15\}$  where  $u$  is the exact solution of the PDE in Eq. (5.52) with  $d = 5$ ,  $T = 1/2$ , and  $\mathcal{D} = \mathbb{R}^d$ . . . . . 146
- 6.1 **Graphical representation of the eco-evolutionary mechanisms underpinned in Chapter 2 and involved in neutral and adaptive differentiation.** Ecological processes are displayed in green boxes, evolutionary processes are displayed in red boxes. . . . . 157
- 6.2 **Graphical representation of the eco-evolutionary processes evidenced in Chapter 5, and how they affect the different organizational levels in economic systems.** Chapter 5 provides quantitative evidences for the processes of mutualistic interactions (facilitation) and spatial dispersal, respectively coupling the world economy organizational level and the national economy organizational level, to the economic activity level. Directly depending on firms, the economic activity level is also coupled to the firm organizational level. Furthermore, in Chapter 5, we hypothesize that dispersal processes originate from the diffusion of organizational routines, coupling the organizational routines level to the world economy level. . . . . 159

