# Thesis intro

## Preface

## Complex adaptive systems

**The dynamics of ecosystems and economic systems is the result of the interaction of heterogeneous entities structured at different levels of organizations. Those entities interact in nonlinear ways and experience evolutionary processes. Interactions and evolutionary processes may act upon individual entities at the microscopic level, but its effects are propagated at the macroscopic level, and feed back to influence the microscopic level. The interdependencies across scales make those systems difficult to understand and predict.**

- In biology is of particular interest the effects of reproduction, selection and mutation of traits, that act at the level of individual organisms but influence the populations level.
    - This is the topic of population genetics [?] and quantitative genetics [?].
- Similar is the effect of replication, selection and variations of population characteristics, acting at the population level, onto ecosystems properties.
    - trophic structure,
    - diversity-productivity relationships, and
    - patterns of nutrient flux
- Elucidating these interactions across scales is fundamental to resolving the issue of biodiversity and ecosystem functioning, and requires a blending of insights from different fields investigating different scales, such as population biology and ecosystems science [?].
- In economic systems is of recent interest the effect of replication, selection and variations of habits, customs, and organizational routines [?], acting at the level of firms, influencing the evolution of economic activities.
    - XXX
- Similar is the effect replication, selection and variations of economic activity characteristics onto economic systems characteristics, at the scale of a country
- Despite the great complexity of ecosystems, regular patterns are empirically observed.
    - For example, famous patterns that have fostered fundamental questions in biology, are
        * average densities and body sizes of species in an area decrease as energy availability increases [?]
        * Latitudinal gradients in species richness: [?]
        * Species–energy relationships
        * Relationships between local and regional richness
        * patterns in interaction networks [?] and [?]
        * in particular, diversity hostspots in mountain regions [?]
        * species abundance relationship, where most species are rare [?]
        * structured networks: [?]
    - For example, famous patterns that have fostered fundamental questions in economic systems, are
        * Economic activities such as innovation and industry concentrate disproportionately in large cities. The ten most innovative cities in the United States account for 23% of the national population, but for 48% of its patents and 33% of its gross domestic product [?].
        * diversification of economies: [?]
            · may be similar to differential regional growth pattern
        * fat tails in financial economics
        * Zipf distribution characterises firm sizes, a result holding for data from multiple years and for various definitions of firm size [?]
        * Zipf laws for cities [?]
        * economic networks that are structured [?, ?, ?, ?, ?]
        * A review of Zipf laws [?]
- The modelling of coupled phenomena is identified as a major frontier in 21st century science [?].
- *Those patterns are unique opportunities to learn the fundamental mechanisms that have generated it.*

*The close interplay between ecological and evolutionary processes has recently been majorly proposed to explain those processes*

## Eco-evolutionary processes

**For many years, biologists have separated the time scale of ecology and evolution. Recently, empirical and theoretical evidences were found that ecological and evolutionary processes may feedback on each other, with major consequences on our understanding of the processes driving their dynamics \cite{{Pelletier}}.**

- Since Darwin, biology has been divived into two distinct branches
  - Ecology: study of interactions between idividuals and biotic and abiotic environment
  - Evolution: modification of living species through reproduction, mutation and natural selection
- Eco evolution: interaction between ecology and evolution [**?**].
  - Recognized that evolution can be rapid and occur on similar time scales as ecology [**?**]
  - the effect of ecological change on evolution has long been a focus of scientific research. The reverse effect gas only recently captured our attention [**?**] –> this is eco-evolutionary dynamics
  - frequency dependence
- Empirical evidences of eco-evolutionary dynamics (in p14 if [**?**]):
  - spatial variation in biotic interactions (geographic mosaic of coevolution; Thompson, 2005),
  - impact population regulation and community dynamics (e.g., Abrams & Matsuda, 1997; Patel, Cortez, & Schreiber, 2018), in
  - species coexistence via stabilizing mechanisms
  - the equal importance of ecology and evolution on population dynamics: [**?**]
  - [**?**]:
    * In one recent example an annual plant, Brassica rapa exhibited very rapid microevolutionary change in response to a climate fluctuation that led to a multiyear drought [70]. There are likely to be important limits to the rates of climate warming that natural systems can adapt to and these should be deducible from both limits to evolutionary change and limits to the rates that species can move across the landscape tracking climate change.
- Talk about evolutionary rescue and link it with your work on spatial graphs.
- *The undersanding of eco-evolutionary feedbacks and their effect macroscopic properties remain to be understood, and is more important than ever in a rapidly changing world [?]*

## An urgent need for better models of complex adaptive systems.

**The effect of anthropogenic pressure on ecosystems, together with climate change, is stressing their functioning. Ecosystems may approach criticla shift state, which in turn may greatly affect economic systems. It is more than urgent to better understand the fundamental processes in order mitigate the consequences.**

- Increasing rates of species extinction [**?**].

- [**?**]: prediction based on habitat models on the basis of mid-range climate-warming scenarios for 2050 that 15– 37% of species would be 'committed to extinction

- meta analysis of [**?**]: 20–30% of plant and animal species would be at increasingly high risk of extinction as global mean temperatures exceed a warming of 2–3 8C above preindustrial levels.

- [**?**]: shifts of species : higher elevations at a median rate of 11.0 meters per decade, higher latitudes at a median rate of 16.9 kilometers per decade. [**?**]

- The plausibility of a planetary-scale 'tipping point' highlights the need to improve biological forecasting by detecting early warning signs of critical transitions on global as well as local scales, and by detecting feedbacks that promote such transitions. It is also necessary to address root causes of how humans are forcing biological changes.

- Developing sustainable approaches to system use implies understanding what maintains resilience and how human intervention might affect it [**?**].

- Current modelling approach are based on spatial distribution models, which learn the ecological niches of species are reproject it given environmental predictors. This entails important assumptions that may be not justified: [**?**]
    - three main critics
        * biotic interactions: novel biotic interactions represent a neglected but major driver of ecosystem modifications under climate change
        * evolutioanry change: e.g. two species of bush cricket were show to have increased fractions of longer-winged (more disper- sive) individuals in recently founded populations, whilst two butterfly species ( Hesperia comma and Aricia agestis ) have increased the variety of habitat types that they can colonize. [**?**]
        * species dispersal: Bioclimate envelope models do not account for species dispersal, but instead aim to predict the potential range of organisms under changed climate

- *The understanding of the key mechanisms that affect ecosystem properties is of utmost importance to anticipate ecosystem shifts.*

- 

## Forward modelling of complex systems

**Scientists have traditionally generated theories about laws underlying phenomena in nature using intuition, which predictions are validated against empirical data [?] [?] [?] −> Deduction process. The key difficulty has been to derive predictions from assumptions, leading to simplifications that has prevented the investigation of the effect of eco-evolutionary processes**.

- model idea: experience of thought,to understand the cause and consequences of miscroscopic rules and laws
    - Simple mathematical models to investigate some long standing questions in ecology
        * [**?**]: theoretical has focused on the spatio temporal dynamics generated by simple ecological models - simple rules are able to produce complex spatio teporal patterns. This is improving oru understanding of a bunch of problems.
        * [**?**]: simple mathematical models with very complicated dynamics
- Fisher, Wright and Haldane, the founders of evolutionary dynamics, have participated in generating the new synthesis, assuming strong assumptions to derive predictions from rule based modelling.
    - A major limitation of current quantitative genetics theory is the reliance on simplified ecological scenarios that are not representative ofthe complexityofeco-evolutionary feedbacks innature [**?**]
- in parallel, work of Volterra:
    - mechanisms underlying the observed fluctuations in the Adriatic fishers. Volterra demonstrated that his equations would have a natural tendency to oscillate.
- adaptive dynamics: fitness landscape, and analysis of invasion, introduce in [**?**]. Provides a mathematical framework to study the interplay between ecological and evolutioanry processes
    - major contribution: diversification due to evolutionary branching sheds a new light on the funda- mental quesetion of the origin of species. Before, allopatric speciation.
    - controverse relative to the biological hypotheses
    - while adaptive dynamics has been successful in taking into account environmental feedbacks and the emergence of polymorphism under frequency- dependent disruptive selection, it does so by assuming strong constraints on the mutation process and standing variation in the population [**?**]
- Computational approaches have allowed to investigate more realistic scenarios
    - Cellular automata, developed by John von Neumann in the 1940
    - network, with network science (Barabasi: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life, Barabasi: Network Science)
    - ABM [**?, ?, ?**]
        * mainly numerical studies

∗ analysis of spatial population dynamics:
　　　　　　· moment equations with closures [**?**, **?**] –> permit to reduce the dimensionality
　　　　　　· scaling of demographic parameters, temporal or spatial
　　　− It remains difficult to understand mechanisms from the sole observations of numerical simulations.
- Renormalization group analysis allowed to derive macroscopic laws from microscopic phenomena in physics, and is an appealing paradigm for biology.
　　− iterative map is iterated over spatial scales to understand how property of a portion of a system scales.
- Work of Meleard and colleagues: renormalization group analysis to understand the emergence of macroscopic biological variables (population size) from simple, well established rules (reproduction, competition, variation).
　　− [**?**] argues that it is the only way to find out the appropriate equations
　　− This work provides powerful mathematical objects and techniques to simulate complex adaptive systems, from a rule-based modelling perspective
- This paradigm could be used to understand how population charactersitics emerge from the interplay between complex spatial structure and eco-evolutionary processes.
- *Our understanding of the evolution of quantitative traits in nature is still limited by the challenge of including realistic trait distributions in the context of frequency-dependent selection and ecological feedbacks [**?**]*
- *Implications on the prediction and understanding of ecosystems has remained limited, because prediction vs mechanistic modelling is different. But prediction is the final step of the scientific method!*

## Machine learning, and scientific machine learning

**Machine learning has made enormous progresses in the recent years, computers learn abstraction of mechanisms from data to make predictions about the world –> induction process.**

- Short review on ML
　　− LeCun: advances in artificial intelligence [**?**]
　　　　∗ computers can learn to predcit the future using vast datasets, too large for humans to grasp [**?**]
　　　　∗ image recognition, and natural language processing.
　　　　∗ natural language processing
- Problem:
　　− the structure learnt is inscrutable for humans, and accuracy, not insight, is often the goal.
　　− ML is limited by interpretability and extrapolability
　　− Scientific knowledge is best expressed in formal logical languages (3) (from [**?**]).
- Short review on SciML:
　　− Scientific Machine Learning is an emerging discipline aiming at combining the best of ML and scientific models, providing unique opportunities to better understand ecosystems and economic systems.
　　− How machine learning is helping scientific computing
　　　　∗ surrogate modelling: simulations can be accelerated by having machine learning models approximate the input-output relation.
　　　　　　· Arnulf and co.
　　　　∗ Recent advances in interpretable machine learning are enabling the generation of theories to be automated: machines have identified mathematical laws in physical and biological systems that took many years for scientists to solve manually
　　　　∗ for many parameterized scientific simulations, one can ask "what parameters would make my model best fit the data?" This inverse problem is pervasive yet difficult because it requires the ability to efficiently compute the gradient of a large existing simulation.
　　　　∗ Method for automatically generate symbolic equations for nonlinear coupled dymamical systems directly from time series data [**?**] –> Symbolic regression. Cons: this method is "model-free", and is rather more relevant in engineering. It generates equation from a bank of polynomials, but is not suited for testing a bunch of different models. It also handle model with simple dynamics, while real-world systems often show chaotic dynamics. Identifying only the useful

analytical relations that are related to the system dynamics. still faced with the challenge of justifying and giving words to their meaning. One difficulty is that we cannot know with cer- tainty the units of bulk constants in the law ex- pressions (for example, combinations of masses, lengths, etc. embodied in the system). Second, the equation may model something that is in- herently difficult to observe directly, such as total energy. Requiring equations to maintain consist- ent physical units still leaves room for ambiguity. [**?**] [**?**]

- As of now, the percolation of mathematical models and machine learning in empirical evolutionary biology and complexity economics is weak.
  - first principles are blurry
  - we cannot repeat indefinitely the experience, and data is limited and limiting. Therefore, need to develop other techniques to tune different candidate of eco-evolutionary model.
    * basic scientific knowledge must be incorporated in SciML to extract meaningful information from observations.
- application of ML to ecosystem
  - analysis oflarge-scale simulations can illuminate the evolution ofecosystems across potential climate futures
- *Scientific Machine Learning is yet to be developed in the field of complex adaptive systems, and help tackle climate change [**?**].*

## Programming languages

**Computational tools critically affects the productivity and advances of science. Scientific Machine Learning requires highly performant computational environment, which is interoperable, i.e. where the machine learning and scientific computing tools can communicate. The pressing challenges to better understand the fundamental processes shaping the dynamics of complex adaptive systems also requires the cooperation of other non-computer scientists coming from different fields.**

- 2 language problem:
  - productivity languages (high level languages, dynamic) (Python, Matlab, R) for easy development
    * Features of productivity languages, such as dynamic typing or garbage collection, make exploratory and iterative development simple –> scientific applications often begin their lives in a productivity language.
  - performance languages (low level, compiled languages)(C, C++, Fortran) for speed and a predictable mapping to hardware [**?**].
- productivity languages are the most common scientific programming languages. They often rely on bindings with performance language, with negative externalities
  - no handling of automatic differentiation, at the heart of the training of neural networks, which is likely to advance theoretical and statistical studies of biological problems [**?**]
  - no understanding and participation of the development of the libraries
- converting an existing application (or some subset thereof) to a different language requires significant programmer involvement; features previously handled by the language (e.g. memory management) now have to be emulated by hand. As a result, porting software from a high level to a low level language is often a daunting task.
  - this is not only required for performance gain, but also compatibility between different libraries.
- Julia, high performance programming language [**?**, **?**]
  - decrease the gap between productivity and performance languages. On one hand, it provides productivity features like dynamic typing, garbage collection, and multiple dispatch. On the other, it has a type-specializing just-in-time compiler and lets programmers control the layout of data structure in memory.
  - start prototypes in Julia, but also solve their large-scale models within the same language, instead of resorting to two language solutions when performance is needed.
    * Julia's unique features make it easy to make packages which are type-generic and parallel (https://www.stochasticlifestyle.com/like-julia-scales-productive-insights-julia-developer/)
  - essential to science: can combine analytical tools with machine learning.

- community of modellers: collective intelligence.
  * compare to python, general collection of repositories,
  * Nb on Julia lang Discourse
- Julia characterisitics
  * built in multithreading and parallellism
  * abstraction for generalized array format –> user is required to tranfrom equation into a vector of equation
- Automatic differentiation: take gradients of programs making Automatic Differentiation a first class language feature
- How julia works: extensive use of
  * multiple dispatch : allows a function to compile to different outputs dependent on the types of the inputs, effectively allowing choices of input types to trigger forms of code generation. –> This e.g. allows automatic differentiation with Dual numbers (see e.g. ForwardDiff.jl)
  * metaprogramming
  * This makes it easy for user to use and easy for compiler to understand.
- Interoperability: combination of Flux.jl and Turing.jl (probabilistic programming) to construct probabilistic variants of traditional machine learning models.
- Julia achievements
  - start prototypes in Julia, but also solve their large-scale models within the same language, instead of resorting to two language solutions when performance is needed.
  - Experimental design [**?**]
  - DifferentialEquations.jl [**?**]
    * through extensive use of multiple dispatch, metaprogramming, and call overloading, DifferentialEquations.jl offers a unified user interface to solve and analyze various form of DE while not sacrificing features and performance.
  - Optim.jl [**?**]
  - StructuralIdentifiability.jl
  - Zygote.jl [**?**]
  - SciML: a library of softwares for machine learning [**?**]
  - the Climate Modeling Alliance is building an entirely new climate model that learns continuously from data and from high-resolution simulations, in julia, in contrast to C++/Fortran : At the cost of a daunting translation workload, they aim to build a model that is more accessible to new developers and more compatible with ML libraries. [**?**]
- *Julia's unique features make it the right platform for developing computational tools, that can readily be used by the scientific community, and that are composable, making the sum more than its parts. This has direct benefits to the community, accelerating research.*

## Thesis outline

**In this thesis, I develop novel forward and inverse modelling approaches that account for the interplay between eco-evolutionary processes, and utilise them to shed light on fundamental mechanisms that influence the dynamics of biological and economic systems.**

- We adress the following questions:
  - Can we gain a quantitative understanding of the implications and effects of microscopic processes acting on non random structures (Chap1)
  - Can we develop tools to automatically generate an quantiative paradigm to characterise a system ? (Chap2)
  - Can we find general quantitative organizing principles in biological and economic systems? (Chap3)
    * Answers questions about the relationship between the organization of economic activities and the functioning of those economic systems.
  - Can we develop better computational tools to model the necessary complexity of things.
  - to what extent, for example, do the characteristics of a developing economy simply represent properties that emerge from the selfish behaviors of individual agents, and to what extent are they imposed at higher levels of organization? [**?**]

- This work consists of 4 chapters, that can be grouped into two methodological and two investigative chapters. In the method chapters (Chap 2 and Chap 4) I present novel approaches for respecively inverse and forward modelling. In the two investigative chapters, I seek to undersand how eco-evolutioanry processes drive fundamental mechanisms affecting the dynamics of biological and economic systems. ### Chapter 1 In chapter 1, I investigate how eco-evolutionary processes, in combination with complex spatial structures, influence the trait distribution of a biological population. I build a stochastic eco-evolutionary individual based model, where individuals are structured over a spatial graph. I proceed using a forward modelling approach, where the fundamental processes of reproduction, competition, mutation and migration operate at the individual level. I seek to undersand how those microscopic forces result in patterns at the population level. I derive a PDE approximation of the IBM, providing analytical insights on how the graph affects the population macroscopic properties. Together with numerical simulations, I show that three main graph properties shape the trait distribution of the biological population, establishing mechanistic links between landscape features and population eco-evolutionary dynamics.

## Chapter 2

In chapter 2, I develop an inverse modelling method, where observation data is used to recover the most probable dynamical model describing the system under consideration. I propose a machine learning framework relying on a singular learning strategy, where the maximum likelihood of a model is obtained by training the model against mini-batches of data with short time horizon. I show that this strategy, when combined with the proposed numerical implementation, results in a ML framework that can provide statistical support for the true generating model among several candidates. The ML framework is implemented in the Julia library MiniBatchInference.jl, and can to elucidate mechanistic pathways to improve our undersanding and predictions of ecosystem dynamics.

## Chapter 3

In chapter 3, I investigate whether the biological process of ecological interactions spatial transfers and mutations. I use MiniBatchInference.jl to test