



HAL
open science

Symmetric and heterogeneous population models : approximations for simulation and statistical inference

Antonin Della Noce

► **To cite this version:**

Antonin Della Noce. Symmetric and heterogeneous population models: approximations for simulation and statistical inference. Dynamical Systems [math.DS]. Université Paris-Saclay, 2021. English. NNT : 2021UPAST026 . tel-03280879

HAL Id: tel-03280879

<https://theses.hal.science/tel-03280879>

Submitted on 7 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Symmetric and heterogeneous population models: approximations for simulation and statistical inference

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n°573
Interfaces: matériaux, systèmes, usages
Spécialité de doctorat : mathématiques appliquées
Unité de recherche : Université Paris-Saclay, CentraleSupélec,
Mathématiques et Informatique pour la Complexité et les Systèmes,
91190, Gif-sur-Yvette, France
Référent : CentraleSupélec

**Thèse présentée et soutenue à Gif-sur-Yvette
le 29 mars 2021**

Antonin Della Noce

Composition du jury :

Pauline Lafitte Professeure des universités, Université Paris-Saclay, CentraleSupélec (laboratoire MICS)	Présidente
Amandine Véber Directrice de recherche CNRS, Université de Paris (UMR MAP5)	Rapportrice et examinatrice
Nicolas Vauchelet Professeur des universités, Université Sorbonne Paris Nord (UMR LAGA)	Rapporteur et examinateur
Samis Trevezas Maître de conférences, National and Kapodistrian University of Athens (Department of Mathematics)	Examineur
Paul-Henry Cournède Professeur des universités, Paris-Saclay University, CentraleSupélec (laboratoire MICS)	Directeur de thèse
Amélie Mathieu Maîtresse de conférences, INRAE AgroParisTech (UMR ÉcoSys)	Co-encadrante de thèse

This [street] is one of the principal thoroughfares of the city, and had been very much crowded during the whole day. But, as the darkness came on, the throng momentarily increased; and, by the time the lamps were well lighted, two dense and continuous tides of population were rushing past the door. At this particular period of the evening I had never before been in a similar situation, and the tumultuous sea of human heads filled me, therefore, with a delicious novelty of emotion. I gave up, at length, all care of things within the hotel, and became absorbed in contemplation of the scene without. At first my observations took an abstract and generalizing turn. I looked at the passengers in masses, and thought of them in their aggregate relations. Soon, however, I descended to details, and regarded with minute interest the innumerable varieties of figure, dress, air, gait, visage, and expression of countenance.

– Edgar Allan Poe, THE MAN OF THE CROWD, (1840)

Remerciements

*Ah c'est vrai, vous êtes pistonné, vous. Vous êtes un bourgeois. Tandis que moi,
je me suis fait tout seul, je suis un no man's land !*
—réplique de Macel Pérès, écrite par Raymond Queneau, dans le film LA CITÉ DE
L'INDICIBLE PEUR, réalisé par Jean-Pierre Mocky en 1964

Je ne suis pas un no man's land. Les lignes qui suivent se veulent un témoignage de gratitude envers les personnes qui ont fait de ces quatre années de thèse une expérience avant tout humaine, et qui m'ont permis de mener ce projet à bon port.

Mes remerciements sont en premier lieu destinés à mon directeur de thèse et ami Paul-Henry Cournède. En 2013, Paul-Henry m'introduisit dans son équipe en tant qu'apprenti chercheur, au sein du parcours ECP + R, initié par Bruno Palpant, et cette ouverture sur le monde de la recherche en mathématiques appliquées à la biologie donna à l'ensemble de ma scolarité à CentraleSupélec une vitalité bien au-delà de mes espérances (et elles étaient grandes). Pendant toute cette période qui précéda la thèse, il me proposa des projets et des stages qui, tout en me guidant sur des pistes qui avaient une grande cohérence avec la modélisation des plantes, me confrontèrent à des problématiques transposables dans bien d'autres domaines. Il fit cela par souci de me donner la plus ample liberté de choix dans mes orientations, et aussi par son naturel à transmettre la curiosité et l'éclectisme scientifique qui le caractérise. Quant au sujet de doctorat qu'il me cousit sur mesure, alliant écologie, équations aux dérivées partielles, inférence bayésienne et physique des systèmes complexes, ce dernier n'échappait pas à cette règle. Je tiens également à saluer le travail exceptionnel, qui s'apparentait souvent à un combat, réalisé en tant que directeur du laboratoire MICS, pendant la période de transformation physique et conceptuelle que marquèrent le déménagement à Saclay et l'élargissement des horizons de notre équipe, Biomathematics, aux problématiques biomédicales.

Ma co-encadrante de thèse, Amélie Mathieu, fut une très précieuse interlocutrice, et m'aida grandement à concevoir la structure de cette thèse, à trouver des exemples et des mises en situations qui me servirent tout au long de son élaboration. La feuille de route qu'elle m'a tracée continue à m'orienter dans mes recherches.

Je tiens à remercier les membres de mon jury pour l'excellent déroulement de ma soutenance. Les rapports d'Amandine Véber et de Nicolas Vauchelet furent très encourageants. Amandine me permit de comprendre que la simplification des systèmes sans limite de champ de moyen était plus complexe que je ne l'envisageais. L'échange avec Nicolas lors de la mini-soutenance de fin de deuxième année fut d'un grand enrichissement, et m'orienta dans la direction des schémas numériques lagrangiens pour la simulation de la limite de champ moyen. Ces approfondissements relatifs aux simulations numériques n'auraient pas été ce qu'ils sont sans les conseils et les recommandations bibliographiques de Pauline Lafitte, présidente de mon jury, qui, par son aide et son efficacité remarquable, me donna le courage d'approfondir des concepts essentiels. Enfin, merci à Samis Trevezas pour l'excellence de son enseignement en statistiques, pour m'avoir transmis cette jubilation à convoquer des concepts mathématiques de tous bords, à en inventer de nouveaux, pour représenter au mieux les objets sur lequel porte un problème d'inférence.

L'aboutissement de cette thèse doit aussi beaucoup aux responsables de l'école doctorale Interfaces, Pascale Le Gall, Vincent Mousseau, et Suzanne Thuron, en particulier à la confiance et à la bienveillance qu'ils me témoignèrent dans les moments où commençait à pointer le semblant d'une difficulté (le temps de la rédaction). Leurs soutiens et leurs conseils me permirent d'aborder cette période avec sérénité. Je ne peux donc que recommander cette école doctorale aux futurs chercheuses et chercheurs du Plateau, également pour les nombreuses opportunités de rencontres et d'ouvertures interdisciplinaires qu'elle offre.

Le laboratoire MICS est un lieu de synergie, et j'ai grandement bénéficié du temps que certaines et certains me consacèrent sans compter. En ce qui concerne la programmation, et plus généralement mon rapport à l'informatique, Benoît Bayol, Gautier Viaud et Jean-Christophe Attard seraient bien surpris, voire scandalisés, en apprenant que je me revendique comme leur fils spirituel, surtout au vu de ce que je suis capable de coder, et ma gestion, catastrophique, des dépôts Git qui ont la malchance de m'avoir accordé les droits Developer¹. Mais si j'utilise mon ordinateur pour des tâches autres que celles que l'on peut accomplir avec une ardoise magique, c'est grâce à ces longues heures de pianotage à quatre mains, d'apprentissage sur les travaux que Benoît, Gautier et Jean-Christophe mirent en commun et adaptèrent aux besoins de l'équipe, notamment Pygmalion et Ad-justin. Du côté des mathématiques et des statistiques, Charlotte Baey et Véronique Letort-Le Chevalier m'accompagnèrent dans la conception de mes premiers modèles dynamiques de croissance de plantes. Je reste aujourd'hui attaché aux bonnes pratiques que Charlotte et Véronique m'ont alors transmises, à ce mélange de créativité et de rigueur qui permet d'expliquer une tendance par un mécanisme

¹`git commit -m "Repartons de zéro"`

biologique, de déceler la cause d'une simulation aberrante. J'ai aussi beaucoup appris auprès de Julie Hémont, de sa grande culture mathématique et de sa capacité à se poser les bonnes questions. Le chapitre 2 de cette thèse trouve sa source dans les questionnements qui émergèrent de nos discussions et du travail impressionnant que Julie produisit en ses quelques mois de stage de M2. Il y eut aussi l'aide décisive de Brice Hannebicque et de Ludovic Goudenège pendant toutes ces années, où tant de fois j'ai demandé leurs lumières pour comprendre les preuves de mathématiciennes et mathématiciens un peu cachottiers, pour formaliser des problèmes que j'avais du mal à exprimer même en langage courant. Je tiens à témoigner de mon admiration pour l'efficacité dont Brice et Ludovic font preuve face à des problèmes qui ne sont pas les leurs. J'ai particulièrement apprécié d'assister aux groupes de travail sur la régularité des fonctions aléatoires et sur la mécanique des fluides, initiés par Alexandre Richard et Erick Herbin, qui me permirent de mieux comprendre le concept de propagation du chaos. En dehors du labo, je veux exprimer ma gratitude à Marion Carrier pour notre passionnante collaboration sur le contrôle optimal de serre, et à Divya Madhavan Brochier, initiatrice de l'Academic Writing Center, pour m'avoir tant aidé à rendre l'article associé *refusal-proof*.

Il existe des réactions de précipitation lente, comme celle du sulfate de radium dans l'acide sulfurique, étudiée par Marie Curie et Léon Kolowrat vers 1910, mais leurs temps caractéristiques restent bien inférieurs à quatre ans. Ma thèse a mis plusieurs années à précipiter dans le substrat MICS, et pour ce faire elle eut besoin de catalyseurs et de catalyseuses. Ces catalyseuses et catalyseurs m'apportèrent beaucoup d'inspiration en discutant de leurs travaux, en me conseillant de lire des ouvrages qui m'ont bouleversé. Ils me firent gagner un temps précieux en m'aidant à préparer un cours, ou en me dépêtrant d'un marasme administratif gravissime dont j'ai seul le secret. Ils firent ou continue à faire du labo un lieu de convivialité dans lequel travailler est un plaisir. Ils me firent éviter de véritables catastrophes, personnelles et collectives, dont l'éventualité hante encore mes nuits. Ils me mirent en excellentes dispositions pour faire de grandes découvertes scientifiques en partageant avec l'équipe des pâtisseries maison qui allaient si bien avec le café. Je remercie ainsi Mathilde Sautreuil, Sylvie Dervin, Fabienne Brosse, Mahmoud Bentrion, Gurvan Hermange, Chloé Adam, Kevin Arfi, Léo Chartier, Adrien Dekkers, Sarah Lemler, Zhu Jun, Elvire Roblin, Sylvain Lannuzel, Yoann Pradat, Myriam Tami, Rémi Hellequin, Walid Hammache, Guillaume Joslin, Victor Bouver, Chen Xiangtuo, Enzo Batistella, Alexandre Goy, Andreas Markoulidakis, Kevin Primicerio, Romain Pascual, Marcus Cordi, Dany Kouoh Etame, Xavier Pujos, François Poirier, Walter Jacke et, enfin et surtout, Quentin Touzé. Je tiens à remercier Anna Rozanova-Pierrat et Céline Hudelot pour leur organisation de séminaires, qui m'ont énormément apporté.

Les personnes mentionnées ci-après habitent les bons souvenirs que je garde de

l'année 2015, qui précéda celle de mon début de thèse, et elles ont beaucoup joué dans ma décision de faire un doctorat : merci à Ornwipa Thamsuwan, Charles Warner, Shawn Behling, Maureen Kennedy, Harikrishnan Murali et David Ford.

Je veux maintenant remercier ma famille et amis de longue date. Merci à ma mère, à mon père, à mon frère, à ma belle-mère et à mon beau-père pour l'affection et l'aide qu'ils nous apportent tous les jours. Merci à mon oncle, à mes cousines et à mes cousins. Merci à mes grands-parents. Merci à Pierre. Merci à Gabriel. Merci à Thierry. Enfin, à Lan et à Jules, merci de remplir ma vie.

Voilà les raisons qui font que je ne peux décemment me qualifier de no man's land.

Synthèse en français

Les mouvements collectifs décrivent des populations dans lesquelles les interactions entre individus sont le moteur de leurs déplacements dans l'espace et de leurs transformations dans le temps. La compréhension et le contrôle des mouvements collectifs constituent des enjeux majeurs dans de nombreux domaines, notamment pour l'étude des écosystèmes (dynamique des essaims d'animaux), la sécurité dans les grands rassemblements et les bâtiments (mouvement de foule), ou encore l'agriculture (étude de la croissance des plantes). Les modèles de population que nous considérons sont des systèmes d'équations différentielles ayant la propriété d'être hétérogènes, c'est-à-dire d'être constituées d'individus avec des caractéristiques différentes, et ces caractéristiques ont une influence sur la dynamique. Formellement, un système hétérogène représentant une population de taille N prend la forme suivante

$$(x_i^0, \theta_i)_{1 \leq i \leq N} \sim \mu_0^{\otimes N}$$
$$\forall i \in \llbracket 1; N \rrbracket, \quad \begin{cases} x_i(0) = x_i^0 \\ \frac{dx_i}{dt}(t) = g_N(x_i(t), \theta_i, x_{\neq i}(t), \theta_{\neq i}) \end{cases} .$$

Les conditions initiales x_i^0 et les paramètres individuels θ_i , ou caractéristiques individuelles, de ce système sont aléatoires, de distribution μ_0 , et les interactions entre les différents individus de la population se font via une fonction d'interaction g_N . Cette hypothèse d'hétérogénéité est motivée par l'application agricole, où il est question d'étudier les interactions entre plantes de différentes variétés, voire de différentes espèces. Ces systèmes sont également supposés symétriques, c'est-à-dire ayant une dynamique invariante par permutation des individus, ce qui est une caractéristique largement répandue au sein des modèles de mouvement collectif, et qui permet de nombreuses simplifications. Un certain nombre de défis restent toutefois à relever pour que ces modèles soient utilisés dans des cas d'application concrets. Nous nous concentrons en particulier sur les problèmes liés à l'inférence statistique, c'est-à-dire à la confrontation du modèle à des données expérimentales.

Un premier niveau de difficulté est d'ordre computationnel : la simulation de grandes populations en interaction peut s'avérer trop coûteuse en temps de calcul,

et elle constitue ainsi un premier obstacle à l'étude de la population à une échelle macroscopique. Un second niveau de difficulté a trait à la qualité des données : du fait de la complexité du système modélisé, les observations expérimentales ne peuvent permettre de caractériser exactement la dynamique du système (en particulier car elles ne portent généralement que sur une sous-partie de la population), et il est nécessaire de quantifier les incertitudes liées aux imperfections dans l'acquisition de ces données.

Dans cette thèse, nous caractérisons l'ensemble des sources d'incertitude liées aux observations partielles des systèmes symétriques et hétérogènes dans un cadre bayésien. Ces incertitudes se regroupent en deux catégories : les premières portent sur l'état de la population, les secondes portent sur la dynamique du système, sur lequel on ne dispose pas de modèle exact ou que l'on ne peut qu'imparfaitement simuler. En particulier, si le protocole expérimental ne permet pas de connaître exactement la taille de la population, on peut se heurter à des problèmes d'inférence particulièrement difficiles à résoudre. Nous proposons dans cette thèse, dans le cas particulier où la fonction de transition g_N peut être associée à une fonction de transition asymptotique g_∞ quand $N \rightarrow \infty$, de négliger ces sources d'incertitudes pour les grandes populations en ayant recours à une représentation macroscopique de la dynamique. Ces approximations statistiques du mouvement global de la population sont basées sur des simulations numériques des distributions limites de champ moyen associées au mouvement collectif, distribution s'exprimant comme la solution d'une équation de transport non local. Le schéma numérique proposé est du type lagrangien, c'est-à-dire que l'on cherche à approcher les trajectoires du flot caractéristique associé à l'équation de transport.

Summary

Collective motions describe populations in which individuals' interactions are the driving force behind their displacements and their transformation over time. Understanding and controlling collective motions are significant issues in many fields, especially for the study of ecosystems (swarm dynamics), safety in large gatherings and buildings (crowd movement), or agriculture (the study of plant growth). The population models we consider are systems of differential equations with the property of being heterogeneous, i.e., made up of individuals with different characteristics influencing the dynamics. This assumption is motivated by the agricultural application, to study interactions between plants of different varieties or even different species. These systems are also assumed to be symmetric, i.e., having dynamics invariant by permutation of individuals' labels, which is a widespread property within collective motion models, enabling numerous simplifications. However, several challenges remain to be addressed before these models can be used in real-life applications. We focus on the problems related to statistical inference, i.e., matching the model with experimental data and observations made on the system under study.

The first level of difficulty is computational: the simulation of a sizeable interacting population can be too costly in terms of computing time, and, therefore, it is a first impediment to the study of the population at a macroscopic scale. The second level of difficulty relates to the quality of the data: because of the complexity of the modeled system, experimental observations cannot characterize the system's dynamics exactly, in particular because they generally only concern a subset of the population. It is necessary to quantify the uncertainties related to the imperfections in the acquisition of such data.

In this thesis, we characterize all the uncertainty sources related to partial observations of symmetric and heterogeneous systems in a Bayesian framework. Some sources of uncertainty, notably the ones arising from inaccurate knowledge of the population size, result in particularly complex inference problems, which we propose to approach using a macroscopic representations of the population. This statistical approximation of the global population motion is based on numerical simulations of the mean-field limit distribution, i.e., a probability distribution

expressed as the solution of a non-local transport equation associated with the symmetric system.

Contents

1	Simulations of symmetric and heterogeneous population systems	1
1.1	Introduction	1
1.2	Symmetric and heterogeneous population systems	3
1.2.1	Heterogeneous population systems	4
1.2.2	Symmetric population systems	10
1.3	Simulation of individual-based population systems	14
1.4	The Schneider system	18
1.4.1	Context of the system	18
1.4.2	Description of the system	19
1.4.3	Properties of the system	22
1.4.4	Definition of the initial configuration distribution	25
1.4.5	Simulation of the Schneider system	26
1.5	Conclusion	45
2	Uncertainties in symmetric population systems	51
2.1	Introduction	51
2.2	Bayesian inference	54
2.3	Application to a symmetric system: Spring Cloud model with identical particles	56
2.4	Undifferentiated population	60
2.5	Uncertainty on the initial condition	62
2.6	Uncertainty on the population size	66
2.7	Uncertainty due to the inexact simulation of the system	69
2.7.1	Deterministic integrator	72
2.7.2	Stochastic integrator	75
2.8	Conclusion	83
2.9	Appendix of chapter 2	85
2.9.1	Uncertainty on the initial condition	85
2.9.2	Uncertainty on the population size	87
2.9.3	Uncertainty due to the inexact simulation of the system	90

3	Mean-field approximated inference	101
3.1	Introduction	101
3.2	Empirical measure and population distribution dynamics	104
3.3	Existence and uniqueness of the mean-field measure	111
3.4	Propagation of chaos and mean-field approximation	126
3.4.1	Propagation of chaos in the case of the Schneider system	127
3.4.2	Mean-field approximated inference in the case of the Schneider system	134
3.4.3	Second order approximation in the case of the homogeneous Spring Cloud system	140
3.5	Lagrangian simulations of the mean-field characteristic flow: two case studies	143
3.5.1	The Spring Cloud system	143
3.5.2	The Schneider system	149
3.6	Conclusion	155
3.7	Appendix of chapter 3	157

Introduction

*De vos forêts et de vos prés,
O très paisibles photographes !
La Flore est diverse à peu près
Comme des bouchons de carafes !*

–Arthur Rimbaud, CE QU’ON DIT AU POÈTE À PROPOS DE FLEURS (1871)

The increase in agricultural productivity in the second half of the twentieth century has improved food security for the entire world’s population and has managed to support its exponential growth, as one can read in the synthesis by Reid et al. (2005),³⁴ page 51. Nevertheless, many concerns emerged during the 1990s, within the agronomic community (Griffon, 1996²²) and outside the scientific community (Shiva, 1991³⁶), about the fragility of this productivist model in the long term. The article by Foley et al. (2011)¹⁸ is one of the many studies that confirms these concerns, arguing that current yields are not worth the environmental and health sacrifices they entail. The authors advocate structural changes in agricultural production management, describing how high-yield crops’ practices will be the principal causes of tomorrow’s food insecurity, particularly by contributing to climate change. The same synthesis of Reid et al.³⁴ gives on page 57 the estimate that the cost of the environmental damage caused by UK agriculture in 1996 alone would amount to 9% of agricultural income for the whole decade of the 1990s. Driven by the awareness of this long-term fragility, agricultural practices have begun to adapt, but the extent of this mutation is still too limited to qualify it as the *Double Green Revolution* Griffon²² advocates. Duru et al. (2015)¹⁶ classify these initiatives in two categories: the first, called *efficiency/substitution-based agriculture*, consists in reducing the ecological footprint of existing crops, in particular by a more parsimonious and responsible use of farm inputs; the second, called *biodiversity-based agriculture* suggests a paradigm radically different from the productivist model, by the optimization of the richness of the ecosystem constituting the rural environment, on the natural plan and the sociocultural plan. It should be stressed that the second scenario is not incompatible with yield optimization, since diversified ecosystems are characterized by a higher biomass production than homogeneous ecosystems, and that this production is more stable

and resilient to environmental disturbances, as Barot et al. (2017)^[2] point out. This argument leads to the practice of mixing varieties of different phenologies and morphologies within the same field or even mixing several species (potentially with different sowing and harvesting dates). Mixed cropping (Malézieux et al., 2009²⁸), used for vegetable gardens, has been generalized to larger-scale crops and offers promising results. If the mixed crop is *well designed*, a synergy between the different species or varieties takes place, and the resulting yield can be higher than its sole crop equivalent, with a considerably reduced use of farm inputs, such as fertilizers or pesticides. However, mixed cropping remains a marginal practice: Leff et al. (2004)²⁵ derives from satellite data that only 20 species out of the 2500 domesticated plants cover 44% of the world's arable land. This lack of enthusiasm is mainly due to the inadequacy of the technologies and infrastructures for this type of agriculture (initially designed for small crops), and the fact that the design of a successful mixed crop is still too empirical and not reproducible enough. As monocultures' modelling is still a challenge and an active topic of research (Wallach et al., 2018³⁹), it is clear that this challenge is complexified by the heterogeneous and multi-scale aspects of mixed cropping. Despite the identification of many promising studies, Gaudio et al. (2019)²⁰ conclude that the modelling of these complex systems is still in its infancy.

In the review by Gaudio et al.,²⁰ the authors focus on mechanistic models of plants, as opposed to purely statistical models that can also be used in the field (Chen and Cournède, 2018,⁹ Galinier, 2018¹⁹). Mechanistic plant models consist of dynamical systems describing the temporal evolution of a set of coupled variables, such as biomass production, water consumption, or plant architecture, and whose transition function takes as arguments a more or less large number of parameters, which must be calibrated on experimental data so that the dynamics can reproduce the observations as well as possible. Gaudio et al.²⁰ make the distinction between individual-based models, in which each plant constituting the crop is represented, and crop models, which are designed on a more macroscopic scale, with dynamics not obtained by aggregating individual evolutions. Crop models can be practical decision support tools in the case of the sole crop, but they do not offer the possibility to represent the effects of local interactions between individuals of different varieties or species with the same accuracy as individual-based models, although Corre-Hellou et al. (2007)¹⁰ succeed in modeling legume-cereal nitrogen transfers using STICS crop model (Brisson et al., 2003⁷). Individual-based population models are more difficult to simulate than crop models, due to their finer representation of the system, especially in the case of functional-structural plant models, representing the coupling between the morphology and physiology of each plant, along with its interactions with neighboring plants (Cournède et al., 2008,¹⁴ Sievänen et al., 2008,³⁷ Hemmerling et al., 2008²³). Nevertheless, algorithmic tricks can be

considered to alleviate their computational cost (Kang et al., 2003,²⁴ Cournède et al., 2006¹³). According to Evers et al. (2019),¹⁷ functional-structural models are the proper formal framework to test hypotheses on the interactions between plants on the one hand, and, on the other hand, to discover new effects emerging from the combination of several types of plants, such emergence is impossible to predict when considering the system at a macroscopic scale only. The individual level is also the proper scale for the adequate representation of the heterogeneities within a population: within a group of individuals of the same species or the same variety, one can model inter-individual variations in the dynamics, variations that can be of genetic or environmental origin. Thus, the formalism of mixed-effect models is used by Baey et al. (2018)¹ for GreenLab functional-structural model, representing individual plants as samples from some probability distribution accounting for the diversity within the population according to specific traits. However, local interactions between individuals are not directly considered in this study.

It is necessary to find an optimal trade-off between modeling at a macroscopic scale, where all the variables of interest (field yield, water consumption) are located, and modeling at a more granular scale, to understand the phenomena determining the success or the failure of a mixed crop. The problem is how to construct, starting from an individual-based model, a model at the crop scale, with minimal information loss. Therefore, it would be a matter of approaching the whole population's dynamics while avoiding aggregating all the individual models that constitute it, these models being potentially expensive to simulate. This type of approximation would be somewhat similar to the approach implemented in Boltzmann's kinetic theory of gases (1902),⁵ where, from a microscopic description of a volume of gas as a system of colliding particles, a macroscopic representation of the fluid is constructed, by deriving the dynamics of its pressure and temperature fields. Kinetic equations theory has recently been applied to many biological systems, including animal swarming models (Carrillo et al., 2010⁸), natural neural networks (Perthame et al., 2017³²), and opinion formation phenomena (Boudin and Salvarini, 2010⁶). There are, therefore, many reasons to believe that similar reasoning can be used for plant growth models.

Several difficulties remain to be overcome before statistically approaching a population of plants represented by functional-structural models, as the ones discussed in Evers et al.¹⁷ First of all, plant simulators, which are initially computer programs, present challenges for formal analysis. Therefore, it seems complicated to obtain a simplified representation of the population if the individual scale is problematic. For instance, functional-structural models representing plant architecture, such as GreenLab (Yan et al., 2004⁴¹), require an unbounded and time-dependent number of state variables: plant compartments can appear, as well as they can disappear; each time, they are represented by a set of features, such as

their masses, their dimensions, their orientations in space,..., and these different quantities can evolve to reproduce the morphogenesis of the plant. Formally, the GreenLab model appears as a stochastic L-system, whose architectural development alone is the subject of the article by Loi and Cournède (2008).²⁶ In this article, the interaction of the architecture and the plant environment is not considered, and the authors adapt methods used for multitype branching processes to analyze this type of system. When taking into account the interaction with the environment, in particular the competitive relationship between neighboring plants (Mathieu et al., 2009²⁹), the analysis of such a complex interaction model, necessary for the construction of consistent approximations at population scale, seems very challenging to carry out. However, the model's dependence on its various input variables can be studied numerically by conducting global sensitivity analysis (Wu et al., 2012⁴⁰), and this can provide directions for simplifying the model and lightening the computational cost of the parametric identification step (Mathieu et al., 2018³⁰).

Secondly, functional-structural models are themselves of outstanding diversity, and sometimes the instances of the same model applied to different species may diverge. The existence of these variants is an indicator of the great flexibility of these models and of the community's determination to match them as closely as possible to real systems rather than to build a speculative monolith. It is nevertheless a hindrance to a unified framework for conducting methodological research with minimal impact. The first step towards a standardized formalism for functional-structural plant models was made in the design of generic platforms, such as PYGMALION (Cournède et al., 2013,¹¹ Bayol, 2016,³ chapter 3) and ADJUSTIN (Viaud, 2018,³⁸ chapter 5) which focus on parametric identification and mathematical analysis, and OpenAlea (Pradal et al., 2008³³) which focuses on model simulation and visualization.

A final level of difficulty concerns the statistical inference problems associated with this type of model, which needs to be calibrated and validated on experimental data. As a consequence of the complexity of the model, but also of the access to the data that are costly to acquire experimentally (Cournède et al., 2011¹²), the parametric identification of functional-structural models with varying architecture is still an open research question, with nevertheless substantial progress achieved in Beyer et al. (2017),⁴ where Lidar data is used to calibrate a model giving the spatial distribution of a tree crown.

This thesis restricts its scope to simpler individual-based models than functional-structural models, namely systems of smooth differential equations. Although more straightforward, this formalism is also used as population models by the community (Schneider et al., 2006,³⁵ Lv et al., 2008,²⁷ Nakagawa et al., 2015³¹), and can be more easily reconciled with the methodologies exposed in Carrillo et

al.⁸ Besides, we consider only models of populations with the property of being invariant by permutation of its individuals. This invariance has to be understood in the probabilistic sense (invariance in distribution). As we will see in chapter 3, such an assumption is essential for being able to construct a macroscopic approximation of a population from an individual-based model. Concerning the terminology, this is equivalent to assuming that the system is composed of indistinguishable particles (de Munyck, 1975¹⁵), and this assumption is common in quantum physics. In our context, the qualification of *indistinguishable* can be a source of confusion because we need to consider a system of particles having different properties, for instance, to model plants with different genotypes. Therefore, we have chosen the term *symmetric* to qualify the population model, referring to the invariance property of the probability measure describing the state of the whole population (Golse, 2013,²¹ definition 1.5.1).

This thesis is organized into three chapters. The first chapter formally defines symmetric and heterogeneous population models and presents a methodology to simulate this type of model by distinguishing the macroscopic level from the individual level. The second chapter discusses the problems of statistical inference encountered with this type of differential system and suggests a model of the different uncertainty sources that can disturb the parametric identification. The third and last chapter examines how the notion of mean-field limit, inherited from the theory of kinetic equations, can help construct consistent approximations of population dynamics and neglect some of the uncertainty sources mentioned in chapter 2.

Bibliography

- [1] Baey, C., Mathieu, A., Jullien, A., Trevezas, S., and Cournède, P.-H. “Mixed-effects estimation in dynamic models of plant growth for the assessment of inter-individual variability”. In: *Journal of agricultural, biological and environmental statistics* vol. 23, no. 2 (2018), pp. 208–232.
- [2] Barot, S., Allard, V., Cantarel, A., Enjalbert, J., Gauffreteau, A., Goldringer, I., Lata, J.-C., Le Roux, X., Niboyet, A., and Porcher, E. “Designing mixtures of varieties for multifunctional agriculture with the help of ecology. A review”. In: *Agronomy for sustainable development* vol. 37, no. 2 (2017), p. 13.
- [3] Bayol, B. “Système informatique d’aide à la modélisation mathématique basé sur un langage de programmation dédié pour les systèmes dynamiques discrets stochastiques. Application aux modèles de croissance de plantes.” PhD thesis. Université Paris-Saclay, 2016.
- [4] Beyer, R., Bayer, D., Letort, V., Pretzsch, H., and Cournède, P.-H. “Validation of a functional-structural tree model using terrestrial Lidar data”. In: *Ecological Modelling* vol. 357 (2017), pp. 55–57.
- [5] Boltzmann, L. *Leçons sur la théorie des gaz*. Vol. 1-2. Gauthier-Villars, 1902.
- [6] Boudin, L. and Salvarani, F. “Modelling opinion formation by means of kinetic equations”. In: *Mathematical modeling of collective behavior in socio-economic and life sciences*. Springer, 2010, pp. 245–270.
- [7] Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., et al. “An overview of the crop model STICS”. In: *European Journal of agronomy* vol. 18, no. 3-4 (2003), pp. 309–332.
- [8] Carrillo, J. a., Fornasier, M., Toscani, G., and Vecil, F. “Particle, kinetic, and hydrodynamic models of swarming”. In: *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences SE - 12* (2010), pp. 297–336.
- [9] Chen, X. and Cournède, P.-H. “Model-driven and data-driven approaches for crop yield prediction: analysis and comparison”. In: *International Journal of Mathematical and Computational Sciences* vol. 11, no. 7 (2018), pp. 334–342.

-
- [10] Corre-Hellou, G., Brisson, N., Launay, M., Fustec, J., and Crozat, Y. “Effect of root depth penetration on soil nitrogen competitive interactions and dry matter production in pea–barley intercrops given different soil nitrogen supplies”. In: *Field Crops Research* vol. 103, no. 1 (2007), pp. 76–85.
- [11] Cournède, P.-H., Chen, Y., Wu, Q., Baey, C., and Bayol, B. “Development and evaluation of plant growth models: Methodology and implementation in the pygmalion platform”. In: *Mathematical Modelling of Natural Phenomena* vol. 8, no. 4 (2013), pp. 112–130.
- [12] Cournède, P.-H., Letort, V., Mathieu, A., Kang, M. Z., Lemaire, S., Trevezas, S., Houllier, F., and De Reffye, P. “Some parameter estimation issues in functional-structural plant modelling”. In: *Mathematical Modelling of Natural Phenomena* vol. 6, no. 2 (2011), pp. 133–159.
- [13] Cournède, P.-H., Kang, M.-Z., Mathieu, A., Barczi, J.-F., Yan, H.-P., Hu, B.-G., and De Reffye, P. “Structural factorization of plants to compute their functional and architectural growth”. In: *Simulation* vol. 82, no. 7 (2006), pp. 427–438.
- [14] Cournède, P.-H., Mathieu, A., Houllier, F., Barthélémy, D., and De Reffye, P. “Computing competition for light in the GREENLAB model of plant growth: a contribution to the study of the effects of density on resource acquisition and architectural development”. In: *Annals of Botany* vol. 101, no. 8 (2008), pp. 1207–1219.
- [15] De Muynck, W. “Distinguishable-and indistinguishable-particle descriptions of systems of identical particles”. In: *International Journal of Theoretical Physics* vol. 14, no. 5 (1975), pp. 327–346.
- [16] Duru, M., Therond, O., and Fares, M. “Designing agroecological transitions; A review”. In: *Agronomy for Sustainable Development* vol. 35, no. 4 (2015), pp. 1237–1257.
- [17] Evers, J. B., Van Der Werf, W., Stomph, T. J., Bastiaans, L., and Anten, N. P. “Understanding and optimizing species mixtures using functional–structural plant modelling”. In: *Journal of experimental botany* vol. 70, no. 9 (2019), pp. 2381–2388.
- [18] Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O’Connell, C., Ray, D. K., West, P. C., et al. “Solutions for a cultivated planet”. In: *Nature* vol. 478, no. 7369 (2011), pp. 337–342.

-
- [19] Galinier, T. “Analyse multifactorielle de la performance des cultures-Méthodes et automatisation pour l’intégration de données agronomiques, environnementales, sociales et économiques-Exemple du maïs grain non-irrigué en Amérique du Nord”. PhD thesis. Université Paris-Saclay (ComUE), 2018.
- [20] Gaudio, N., Escobar-Gutiérrez, A. J., Casadebaig, P., Evers, J. B., Gerard, F., Louarn, G., Colbach, N., Munz, S., Launay, M., Marrou, H., et al. “Current knowledge and future research opportunities for modeling annual crop mixtures. A review”. In: *Agronomy for Sustainable Development* vol. 39, no. 2 (2019), pp. 1–20.
- [21] Golse, F. “On the Dynamics of Large Particle Systems in the Mean Field Limit”. In: *arXiv preprint arXiv:1301.5494* (2013), pp. 1–144.
- [22] Griffon, M. et al. *Vers une révolution doublement verte*. Editions Quae, 1996.
- [23] Hemmerling, R., Kniemeyer, O., Lanwert, D., Kurth, W., and Buck-Sorlin, G. “The rule-based language XL and the modelling environment GroIMP illustrated with simulated tree competition”. In: *Functional plant biology* vol. 35, no. 10 (2008), pp. 739–750.
- [24] Kang, M. Z., De Reffye, P., Barczi, J.-F., Hu, B.-G., and Houllier, F. “Stochastic 3D tree simulation using substructure instancing”. In: *Plant Growth Modeling and Applications*. Tsinghua University Press and Springer. 2003, pp. 154–168.
- [25] Leff, B., Ramankutty, N., and Foley, J. A. “Geographic distribution of major crops across the world”. In: *Global biogeochemical cycles* vol. 18, no. 1 (2004).
- [26] Loi, C. and Cournède, P.-H. “A Markovian framework to formalize stochastic L-systems and application to models of plant development”. In: *INRIA* (2008).
- [27] Lv, Q., Schneider, M. K., and Pitchford, J. W. “Individualism in plant populations: using stochastic differential equations to model individual neighbourhood-dependent plant growth”. In: *Theoretical population biology* vol. 74, no. 1 (2008), pp. 74–83.
- [28] Malézieux, E., Crozat, Y., Dupraz, C., Laurans, M., Makowski, D., Ozier-Lafontaine, H., Rapidel, B., De Tourdonnet, S., and Valantin-Morison, M. “Mixing plant species in cropping systems: concepts, tools and models: a review”. In: *Sustainable agriculture* (2009), pp. 329–353.
- [29] Mathieu, A., Cournède, P.-H., Letort, V., Barthélémy, D., and De Reffye, P. “A dynamic model of plant growth with interactions between development and functional mechanisms to study plant structural plasticity related to trophic competition”. In: *Annals of botany* vol. 103, no. 8 (2009), pp. 1173–1186.

-
- [30] Mathieu, A., Vidal, T., Jullien, A., Wu, Q., Chambon, C., Bayol, B., and Cournède, P.-H. “A new methodology based on sensitivity analysis to simplify the recalibration of functional–structural plant models in new conditions”. In: *Annals of Botany* vol. 122, no. 3 (June 2018), pp. 397–408.
- [31] Nakagawa, Y., Yokozawa, M., and Hara, T. “Competition among plants can lead to an increase in aggregation of smaller plants around larger ones”. In: *Ecological Modelling* vol. 301 (2015), pp. 41–53.
- [32] Perthame, B., Salort, D., and Wainrib, G. “Distributed synaptic weights in a LIF neural network and learning rules”. In: *Physica D: Nonlinear Phenomena* vol. 353 (2017), pp. 20–30.
- [33] Pradal, C., Dufour-Kowalski, S., Boudon, F., Fournier, C., and Godin, C. “OpenAlea: a visual programming and component-based software platform for plant modelling”. In: *Functional plant biology* vol. 35, no. 10 (2008), pp. 751–760.
- [34] Reid, W. V., Mooney, H. A., Cropper, A., Capistrano, D., Carpenter, S. R., Chopra, K., Dasgupta, P., Dietz, T., Duraiappah, A. K., Hassan, R., et al. *Ecosystems and human well-being-Synthesis: A report of the Millennium Ecosystem Assessment*. Island Press, 2005.
- [35] Schneider, M. K., Law, R., and Illian, J. B. “Quantification of Neighbourhood-Dependent Plant Growth by Bayesian Hierarchical Modelling”. In: *Journal of Ecology* vol. 94, no. 2 (2006), pp. 310–321.
- [36] Shiva, V. *The violence of the green revolution: third world agriculture, ecology and politics*. Zed Books, 1991.
- [37] Sievänen, R., Perttunen, J., Nikinmaa, E., and Kaitaniemi, P. “Toward extension of a single tree functional–structural model of Scots pine to stand level: effect of the canopy of randomly distributed, identical trees on development of tree structure”. In: *Functional Plant Biology* vol. 35, no. 10 (2008), pp. 964–975.
- [38] Viaud, G. “Méthodes statistiques pour la différenciation génotypique des plantes à l’aide des modèles de croissance”. PhD thesis. Université Paris-Saclay, 2018.
- [39] Wallach, D., Makowski, D., Jones, J. W., and Brun, F. *Working with dynamic crop models: methods, tools and examples for agriculture and environment*. Academic Press, 2018.
- [40] Wu, Q.-L., Cournède, P.-H., and Mathieu, A. “An efficient computational method for global sensitivity analysis and its application to tree growth modelling”. In: *Reliability Engineering & System Safety* vol. 107 (2012), pp. 35–43.

- [41] YAN, H.-P., Kang, M. Z., De Reffye, P., and Dingkuhn, M. “A dynamic, architectural plant model simulating resource-dependent growth”. In: *Annals of botany* vol. 93, no. 5 (2004), pp. 591–602.

Chapter 1

Simulations of symmetric and heterogeneous population systems

*All these people that you mention, yes, I know them, they're quite lame
I had to rearrange their faces and give them all another name*
—Bob Dylan, DESOLATION ROW, (1965)

1.1 Introduction

When modelling a population of living forms, making the assumption that it consists only of similar individuals is somewhat restrictive. Such an assumption neglects all eventual specificities at the individual level that may impact the studied phenomenon. It makes no sense to make such an assumption if one intends to model a multi-species population, but these individual specificities might also need to be taken into account even in the case of a mono-species population. These specificities might have a genetic origin, or they might come from micro-environmental variations affecting each individual in different ways.

The importance of integrating this representation of population diversity even in a monospecific population was formulated in biological/ecological articles such as Liu et al. (2004),²⁷ which underlines that the variability of emergence within a maize crop has an impact on the final yield. Emergence is related to the thermal and mechanical properties of the soil, and of course, to some genetic factors of the maize plant (see, for instance, Claverie, 2018,⁸ chapter 4, or Feng et al., 2014¹⁹). Such variability can only be accurately represented by individual-based population models, such as GreenLab (Cournède et al., 2008¹²), Lignum (Sievänen et al., 2008³⁷) or GroIMP (Hemmerling et al., 2008²³). These different individual-based models consist of dynamical systems, representing both plant inner processes and interactions with their direct environment. Estimating some parameters has to

be carried out, at the individual level, so that the model reproduces better the dynamics of the specimens under study (concerning the methodology for the parameter estimation on the individual plant, see Cournède et al., 2011¹⁰). In a first approach, a population can be derived by simulating all individual models jointly (Cournède et al., 2009¹¹). The population is then said to be heterogeneous when the parameters change from one individual to another. The observed variability of the parameters related to the functioning processes of the considered plant is generally explained by genetic factors (Letort et al., 2008²⁵).

Baey et al. (2013)² underline the problem that such a methodology is limited by the requirement of calibrating individual models composing the population, which appears to be too costly to implement when the number of individuals is significant. The authors advocate that any heterogeneous population model should incorporate some assumptions on the distribution of parameters from one individual to another so that observations collected at the individual scale might be connected and bring information to the population scale. The authors rely on the formalism of mixed-effect models, first introduced by Fisher (1919).²⁰ Within this formalism, a population is not only a collection of individuals, but it is also a representation of how the individuals are generated or instantiated. It consists of defining a probability distribution, or a set of possible probability distributions, explaining the variability of parameter values at the individual level.

Mixed-effect models are rarely used for populations where individuals interact with each other, with the notable exception of the article by Schneider et al. (2006),³⁶ studying the growth dynamics of competing plants. This relative lack of coverage of this subject may come from the fact that the modelling of interactions leads to non-analytical dynamical systems, whose numerical resolution can be costly in computational time. This difficulty can be a hindrance to parameter identification, and it can lead the authors to neglect or to give up taking into account heterogeneity. For instance, Lv et al. (2008)²⁸ and Nakagawa et al. (2015)³¹ choose to simplify the model introduced by Schneider et al. (2006)³⁶ by considering that all individuals are identical, whereas in the original model all plants have intrinsic growth parameters.

In this chapter, we restrict ourselves to individual-based models taking the form of differential equations systems with random initial conditions (Strand, 1970³⁸). This formalism can lead to technical issues if the distribution of heterogeneities is not compatible with the transition function of the system. If the model is non-linear, an unwise choice of the distribution of parameters and initial conditions can lead to biologically (or physically) aberrant behaviour of the system: one may observe finite time blow-ups or abnormal values of state variables. Therefore, the relationship between the initial distribution and the system dynamics function must be investigated beforehand to ensure that individual trajectories remain

within an acceptable domain of the phase space.

These dynamical systems use labels to distinguish members of the population; most often, these labels are numbers from 1 to N , where N is the population's size. Except in some rare cases, the assignment of these labels is arbitrary, and such indices should have no impact on the population dynamics, for the sake of reproducibility of the results. A population model is said to be symmetric when any permutation of individual labels does not change the overall evolution. As our objective is to perform a statistical inference, it is essential to make sure that the system presents this property of invariance by relabelling; without this property, the study of a population from a subgroup of individuals requires first to address the question of the selection of this subgroup, and this selection is entirely model-dependent.

Section 1.2 introduces the notion of distribution providing global solutions to a differential equation and the notion of symmetric population model in a probabilistic setting. A specific class of symmetric population model is considered, characterized by dynamics depending on the population's empirical measure, i.e., the uniform distribution over the individuals in the population. Section 1.3 suggests possible approaches to reduce the computational time of the simulations of symmetric ODE-based population models. If the state of the population, represented by the empirical measure, can be approximated by a piecewise polynomial process, then individual trajectories can be computed separately, in parallel, at least over a short period of time. Section 1.4 applies this methodology for the simulation of the plant population model introduced by Schneider et al. (2006).³⁶

1.2 Symmetric and heterogeneous population systems

The population models studied in this chapter consist of differential systems with random initial conditions. Let us start by introducing some notations concerning the phase space, where the dynamics take place, and the probability space, over which the random variables of the systems are defined.

notation for the state space: \mathcal{X} is an Euclidean vector space, endowed with a dot product $(x, y) \in \mathcal{X}^2 \mapsto x \cdot y \in \mathbb{R}$ and the associated Euclidean norm $x \in \mathcal{X} \mapsto |x| = \sqrt{x \cdot x} \in \mathbb{R}_+$. Typically $\mathcal{X} = \mathbb{R}^{d_{\mathcal{X}}}$ for some dimension $d_{\mathcal{X}}$ and the dot product is the canonical dot product.

notation for the space of heterogeneities: Θ is a metric space subset of a finite-dimensional vector space. Θ is endowed with the metric $m_{\Theta} : \Theta^2 \rightarrow \mathbb{R}_+$.

Typically $\Theta \subset \mathbb{R}^{d_\Theta}$ for some dimension d_Θ and the chosen metric is the one defined from the 1-norm or the 2-norm on \mathbb{R}^{d_Θ} .

notations for the probability space: $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. We assume that the sample space (Ω, \mathcal{F}) is rich enough so that all random variables defined in what follows do not have existence problem. The set of random variables taking values in a Borel space $(\mathcal{Z}, \mathcal{E})$, i.e. the set of measurable functions from (Ω, \mathcal{F}) to $(\mathcal{Z}, \mathcal{E})$, is $\mathcal{R}(\Omega, \mathcal{F}, \mathcal{Z}, \mathcal{E})$. The set of probability measures over $(\mathcal{Z}, \mathcal{E})$ is $\mathcal{P}(\mathcal{Z})$.

Let Z be a random variable in $\mathcal{R}(\Omega, \mathcal{F}, \mathcal{Z}, \mathcal{E})$ and $\mathbb{P}_Z \in \mathcal{P}(\mathcal{Z})$. $Z \sim \mathbb{P}_Z$ denotes the fact that Z is of distribution \mathbb{P}_Z , i.e. for all $A \in \mathcal{E}$, $\mathbb{P}(Z \in A) = \mathbb{P}_Z(A)$. If Y is another random variable taking values in the same space, we note $Z \sim Y$ the fact that Z and Y have the same probability distribution.

Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be two measurable spaces. Then $\mu_1 \otimes \mu_2$ denotes the product measure over the product of σ -algebras $\mathcal{A}_1 \times \mathcal{A}_2$, defined by

$$\forall (A_1, A_2) \in \mathcal{A}_1 \times \mathcal{A}_2, \quad (\mu_1 \otimes \mu_2)(A_1, A_2) = \mu_1(A_1)\mu_2(A_2) \quad (1.1)$$

If $\mu_1 = \mu_2 = \mu$ then the product measure is $\mu^{\otimes 2}$.

1.2.1 Heterogeneous population systems

In this section, we consider a population system having the following expression

$$\begin{cases} (x_i^0, \theta_i)_{1 \leq i \leq N} \sim \mu_0^N \\ \forall t \geq 0, \quad \frac{dx_i^t}{dt} = g_N^i(t, x_i^t, \theta_i, (x_j^t, \theta_j)_{j \neq i}) \end{cases} \quad (1.2)$$

N is the size of the population. Each individual i is represented at time t by a state variable x_i^t and a set of constant parameters θ_i . The initial configuration of the population, assigning to each individual an initial state x_i^0 and a constant parameter θ_i , is sampled from a probability measure $\mu_0^N \in \mathcal{P}((\mathcal{X} \times \Theta)^N)$. The time evolution of the state variables is then completely determined by the transition functions g_N^i , which take into account the eventual interactions between individuals. The following definitions aim at providing some conditions ensuring that the system 1.2 is well-defined. Some elementary properties of the resulting trajectories are also recalled.

Let us start by a reminder on the terminologies of maximal solutions and global solutions, whose existence and uniqueness is given by the Cauchy-Lipschitz theorem in the deterministic case.

Theorem 1.1. Cauchy-Lipschitz

Let $F : \mathbb{R} \times \mathcal{X} \rightarrow \mathcal{X}$ a locally Lipschitz continuous function with respect to its second argument, i.e. such that

$$\forall t \in \mathbb{R}, \forall x \in \mathcal{X}, \exists K_{x,t} > 0, \exists R_{x,t} > 0, \forall x_1, x_2 \in \mathcal{X}, \\ |x_1 - x_2| \leq R_{x,t} \Rightarrow |F(t, x_1) - F(t, x_2)| \leq K_{x,t} |x_1 - x_2| \quad (1.3)$$

Then for all initial condition $(t_0, x_0) \in \mathbb{R} \times \mathcal{X}$, there exists a unique interval \mathcal{J} of \mathbb{R} containing t_0 and a function $x : \mathcal{J} \rightarrow \mathcal{X}$ continuously differentiable such that:

1.
$$\begin{cases} x(t_0) = x_0 \\ \forall t \in \mathcal{J}, \frac{dx}{dt}(t) = F(t, x(t)) \end{cases} \quad (1.4)$$

x is solution of the Cauchy problem associated to (t_0, x_0) over interval \mathcal{J} .

2. for all interval \mathcal{J}' containing t_0 , and $y : \mathcal{J}' \rightarrow \mathcal{X}$ satisfying the Cauchy problem (t_0, x_0) over \mathcal{J}' , we have

$$\mathcal{J}' \subset \mathcal{J} \text{ and } x|_{\mathcal{J}'} = y \quad (1.5)$$

We say that (J, x) is the maximal solution of the Cauchy problem associated to (t_0, x_0) . Besides if there exists a function $K : \mathbb{R} \rightarrow \mathbb{R}_+$ continuous such that $\forall x \in \mathcal{X}, |F(t, x)| \leq K(t)(1 + |x|)$, then the maximal solution is defined over \mathbb{R} . We then say that such solution is a global solution.

Because of the biological context of our study, we only consider systems generating trajectories defined for all time $t \geq 0$ for some specific initial conditions. The following definition formalizes this constraint on the initial configuration distribution.

Definition 1.1. distribution providing global solutions to a differential equation

Let $f : \mathbb{R}_+ \times \mathcal{X} \times \Theta \rightarrow \mathcal{X}$ the function associated to the parametric differential equation $y' = f(t, y, \theta)$ and let $\mu_0 \in \mathcal{P}(\mathcal{X} \times \Theta)$ a probability distribution. Then we say that μ_0 is compatible with the differential equation of function f if it provides global solution to the equation, i.e. if for all random variable $z_0 = (x_0, \theta)$ of distribution μ_0 and for \mathbb{P} -almost every $\omega \in \Omega$, the differential equation

$$x(0) = x_0(\omega) \\ \forall t \in \mathbb{R}_+, \frac{dx}{dt}(t) = f(t, x(t), \theta(\omega)) \quad (1.6)$$

has a unique global solution defined over \mathbb{R}_+ .

Definition 1.2. sample path integrable process, from Strand (1970)³⁸

A stochastic process $(x_t)_{t \geq 0}$ taking values in \mathcal{X} is said to be sample path integrable if for all $t \geq 0$ and for \mathbb{P} -almost every $\omega \in \Omega$, the integral $\int_0^t x_s(\omega) ds$ exists.

Ordinary differential equations with random initial conditions are studied in Strand (1970).³⁸ In the cases that interest us in here, having a probability distribution providing global solutions to a differential equation implies that there exists a stochastic process $(x_t)_{t \geq 0}$ taking values in \mathcal{X} , adapted to the constant filtration $(\mathcal{F}_t = \mathcal{F})_{t \geq 0}$, and a random variable θ , such that the process $(f(t, x_t, \theta))_{t \geq 0}$ is sample path integrable and such that the following equality holds for all $t \in \mathbb{R}_+$ and for \mathbb{P} -almost every $\omega \in \Omega$

$$x_t(\omega) = x_0(\omega) + \int_0^t f(s, x_s(\omega), \theta(\omega)) ds \quad (1.7)$$

As the choice of the initial random variable $z_0 = (x_0, \theta)$ fully determines the system (in terms of sample path uniqueness), there exists a function $(t, z) \in \mathbb{R}_+ \times \mathcal{Z} \mapsto x^{z_0}(t, z)$ such that for \mathbb{P} -almost every $\omega \in \Omega$, we have:

1. $t \in \mathbb{R}_+ \mapsto x^{z_0}(t, z_0(\omega))$ is differentiable
2. $\forall t \in \mathbb{R}_+, \quad x^{z_0}(t, z_0(\omega)) = x_0(\omega) + \int_0^t f(s, x^{z_0}(s, z_0(\omega)), \theta(\omega)) ds$

When such a relationship between μ_0 and f holds, we characterize the dynamics of the system by the following notation:

$$\begin{cases} (x_0, \theta) \sim \mu_0 \\ \forall t \in \mathbb{R}_+, \quad \frac{dx_t}{dt} = f(t, x_t, \theta) \end{cases} \quad (1.8)$$

As the initial condition is only characterized by its distribution, and as the choice of a random variable of a given distribution is not unique, a solution process satisfying (1.8) is not pathwise unique. However, the law of the solution processes, i.e., processes satisfying equation (1.7), is uniquely defined in terms of finite-dimensional distributions. In particular, if the processes $(x_t)_{t \geq 0}$ and $(y_t)_{t \geq 0}$ are two solution processes, then for all $t \geq 0$, $x_t \sim y_t$. However, pathwise uniqueness holds conditionally to the initial configuration of the population.

Example 1.1. distribution providing global solutions to a differential equation

The probability measure μ_0 of the uniform distribution $\mathcal{U}([-1; 1])$ is compatible

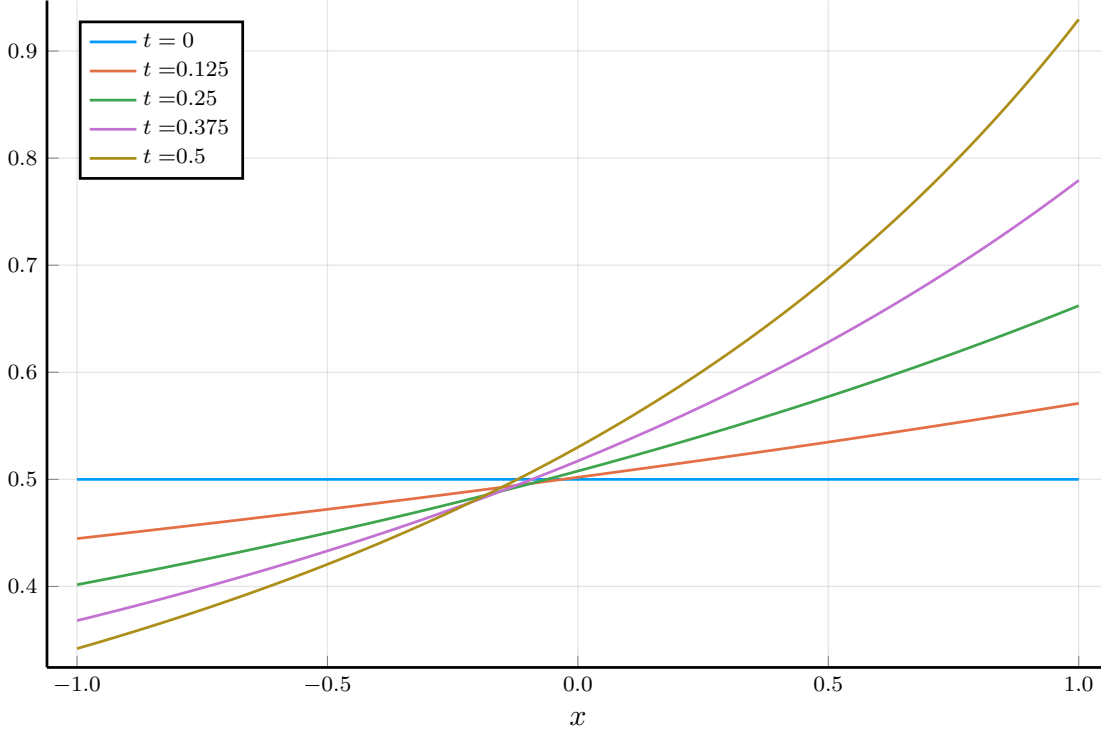


Figure 1.1: Density f_t , defined in equation (1.11), for different values of t

with the differential equation $y' = 1 - y^2$. For any $y_0 \in [-1; 1]$, the solution of the initial value problem $(0, y_0)$ is given by

$$\forall t \in \mathbb{R}_+, \quad y(t) = \frac{(1 + y_0)e^t - 1 + y_0}{(1 + y_0)e^t + 1 - y_0} \quad (1.9)$$

A way to visualize the law of the solution processes of the system

$$\begin{cases} y_0 \sim \mathcal{U}([-1; 1]) \\ \frac{dy_t}{dt} = 1 - y_t^2 \end{cases} \quad (1.10)$$

is to represent the evolution of the density of any solution process $(y_t)_{t \geq 0}$ over time. For any time $t \geq 0$, the expression of the density is

$$f_t : x \in \mathbb{R} \mapsto \mathbb{I}\{-1 \leq x \leq 1\} \cdot \frac{(e^t + 1)^2 + (1 - e^t)^2}{2(e^t + 1 - x(e^t - 1))^2} \quad (1.11)$$

Figure 1.1 displays the graph of the density f_t for different t . As the solutions of the differential equation are time-increasing, the mass of the density is shifted to the right.

Individual trajectories of system (1.2) are therefore well-defined when μ_0^N is compatible with the vector field $(g_N^i)_{1 \leq i \leq N}$. Under this guarantee of existence and uniqueness (in process distribution), the system (1.2) is referred to as an individual-based population with individual characteristics. This statement is recapped in the following definition.

Definition 1.3. individual-based population system with individual parameters

Let $N > 1$ an integer, let $\mathcal{Z} = \mathcal{X} \times \Theta$, $(g_N^i : \mathbb{R}_+ \times \mathcal{Z} \times \mathcal{Z}^{N-1} \rightarrow \mathcal{X})_{1 \leq i \leq N}$ a collection of functions. Let $\mu_0^N \in \mathcal{P}(\mathcal{Z}^N)$ providing global solutions to the differential system (1.2). Such differential system with random initial condition is called an individual-based population system with individual characteristics. A set of individual trajectories is a stochastic process $(x_{1:N}^t)_{t \geq 0} = ((x_i^t)_{t \geq 0})_{1 \leq i \leq N}$ such that there exists a random variable $\theta_{1:N} = (\theta_i)_{1 \leq i \leq N}$ satisfying the following properties

$$\begin{aligned} (x_i^0, \theta_i)_{1 \leq i \leq N} &\sim \mu_0^N \\ \text{and for } \mathbb{P}\text{-almost every } \omega \in \Omega, \forall i \in \{1, \dots, N\}, \\ x_i^t(\omega) &= x_i^0(\omega) + \int_0^t g_N^i(s, x_i^s(\omega), \theta_i(\omega), (x_j^s(\omega), \theta_j(\omega))_{j \neq i}) ds \end{aligned} \tag{1.12}$$

θ remains constant throughout the evolution of the system. It represents individual characteristics or constant state variables. In keeping with the terminology used in mixed-effect models literature (Davidian and Giltinan, 1993,¹⁴ Baey, 2014¹), we refer to θ as the **individual parameter**, as opposed to x which is the individual state. Depending on the context, some components of θ represent some intrinsic features of individuals. The variability of θ from one individual to another gives a heterogeneous population.

Definition 1.4. Marginal distribution

Let $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ a probability measure on the space on the Cartesian product of two measurable spaces. Then the x -marginal distribution μ^x is the distribution defined by its action on measurable and bounded functions:

$$\forall h \in \mathcal{M}_b(\mathcal{X} \rightarrow \mathbb{R}), \quad \mathbb{E} \{h(x), x \sim \mu^x\} = \int_{\mathcal{X} \times \mathcal{Y}} h(x) \mu(dx, dy) \tag{1.13}$$

Definition 1.5. heterogeneous population system

An individual-based population system with individual parameters is said to be heterogeneous if the marginal distribution $\mu_0^{N,\theta}$ is such that

$$\mathbb{P} \left\{ \theta_{1:N} \sim \mu_0^{N,\theta} \mid \exists (i, j) \in \llbracket 1; N \rrbracket^2, \theta_i \neq \theta_j \right\} > 0 \tag{1.14}$$

Example 1.2. heterogeneous population system

Let $x_0 \in \mathcal{X}$, $\theta, \theta' \in \Theta$ two distinct points. Let us consider the initial distribution $\mu_0^N = \left(\delta_{x_0} \otimes \frac{\delta_\theta + \delta_{\theta'}}{2} \right)^{\otimes N}$. Then the population system is heterogeneous. Nevertheless, we have that

$$\mathbb{P} \left\{ \theta_{1:N} \sim \mu_0^{N,\theta} \mid \forall (i, j) \in \llbracket 1; N \rrbracket^2, \theta_i = \theta_j \right\} = \frac{1}{2^{N-1}} \quad (1.15)$$

Definition 1.3 constitutes a restriction of the concept of *individual-based model*.¹⁵ In the literature, the concept of *individual-based model* encompasses a wide variety of formal objects, such as cellular automata or multi-agent systems in a stochastic environment. In this thesis, the scope is restricted to time-continuous dynamical systems in a deterministic environment, and the only source of randomness is the distribution of parameters and initial conditions, i.e., what monitors the diversity of the population. A concern for simplicity mainly drives the choice of this restriction. The assumption of a known environment is not too restrictive in our case, as we study statistical inference on data collected in the past. Indeed, we can assume that if certain features in the environment are considered relevant to explain the evolution of the system, they have been previously measured during the experiment. Therefore, the environment is incorporated within the transition function g_N , more specifically in the time dependency of this function. If the population does not evolve in a controlled environment, or if this environment is only partially observed, a stochastic process $(e_t)_{t \geq 0}$, modelling the variations of the environmental features, needs to be added to the arguments of the transition function, constituting another source of uncertainty for the statistical inference problem. An example of such system for plant growth can be found in Della Noce et al. (2019).¹⁷ Apart from the results related to statistical inference (chapter 2 and chapter 4), we postulate that most of the assertions concerning the dynamics of the system can easily be generalized to the case of a stochastic environment.

Another source of uncertainty, often considered in the literature, is the modelling noise, i.e., when dynamics are random. The population model is then a system of stochastic differential equations (Lv et al., 2008,²⁸ Bolley et al., 2011,⁵ Degond et al., 2014¹⁶).

$$\left\{ \begin{array}{l} (x_i^0, \theta_i)_{1 \leq i \leq N} \sim \mu_0^N \\ \forall i \in \{1, \dots, N\}, \quad x_0^i = x_i^0 \\ \forall t \geq 0, \quad dx_t^i = g_N^i(t, x_t^i, \theta_i, (x_t^j, \theta_j)_{j \neq i}) dt + \sigma(t, x_t^i, (x_t^j, \theta_j)_{j \neq i}) dB_t^i + \\ \sigma_e(t, x_t^i, (x_t^j, \theta_j)_{j \neq i}) dB_t \end{array} \right. \quad (1.16)$$

where $((B_t^i)_{t \geq 0})_{1 \leq i \leq N}, (B_t)_{t \geq 0}$ are $d_{\mathcal{X}}$ -dimensional independent Wiener processes. The design of the diffusion coefficients σ_i and σ_e has to be made wisely so that

the individual trajectories remain within an acceptable domain of the phase space. Modelling noises can constitute an alternative hypothesis to individual heterogeneity to account for the variability in the data collected in the population (Lv et al., 2008,²⁸ Trevezas and Cournède, 2013³⁹). In this thesis, we mainly focus on models, whose randomness is due to the initial configuration of the population.

1.2.2 Symmetric population systems

A population model is said to be symmetric if the dynamics are invariant by permutation of individual indices. This invariance is to be understood in the probabilistic sense: for all $\sigma : \{1, \dots, N\} \rightarrow \{1, \dots, N\}$ a permutation, $(x_1^t, \theta_1, \dots, x_N^t, \theta_N)_{t \geq 0}$, a set of individual trajectories, has the same distribution as $(x_{\sigma(1)}^t, \theta_{\sigma(1)}, \dots, x_{\sigma(N)}^t, \theta_{\sigma(N)})$. We can consider a less restrictive condition for the definition of a symmetric system, consisting of a punctual equality in time of the processes' distributions.

Definition 1.6. *symmetric population system*

An individual-based population system of size N is said to be symmetric if for all bijection $\sigma \in \mathfrak{S}_N$ and for all time $t \geq 0$

$$(x_{\sigma(1:N)}^t, \theta_{\sigma(1:N)}) \sim (x_{1:N}^t, \theta_{1:N}) \quad (1.17)$$

This definition is consistent with the definition of symmetric N -particle distribution introduced in Golse (2013).²¹

notation: \mathfrak{S}_N the set of bijective maps from $\{1, \dots, N\}$ to $\{1, \dots, N\}$. For $\sigma \in \mathfrak{S}_N$, we can consider the linear map consisting in permuting the order of components of a list of element of \mathcal{X}^N

$$P_\sigma^{\mathcal{X}^N} : x_{1:N} \in \mathcal{X}^N \mapsto x_{\sigma(1:N)} = (x_{\sigma(i)})_{1 \leq i \leq N} \in \mathcal{X}^N \quad (1.18)$$

Example 1.3. *symmetric population systems*

1. Let $\mu_0 \in \mathcal{P}(\mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}_+^*)$ and the system introduced in Cucker and Smale (2007).¹³

$$\forall i \in \{1, \dots, N\}, \quad \begin{cases} (x_i^0, v_i^0, m_i) \sim \mu_0 \\ \frac{dx_i}{dt}(t) = v_i(t) \\ \frac{dv_i}{dt}(t) = \frac{H}{m_i} \sum_{j=1}^N \frac{v_j(t) - v_i(t)}{(1 + |x_i(t) - x_j(t)|/\sigma_x)^\beta} \end{cases} \quad (1.19)$$

This system models the motion of animals of various masses, tending to align their speeds with those of their congeners in the surrounding area. We can check easily that this system is symmetric.

2. Let $\mu_0 \in \mathcal{P}(\mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+)$ and the system referred to as Kuramoto-Vicsek system in Degond et al. (2014).¹⁶

$$\forall i \in \{1, \dots, N\}, \quad \begin{cases} (x_i^0, \alpha_i^0, \nu_i) \sim \mu_0 \\ \frac{dx_i}{dt}(t) = c(\cos(\alpha_i(t)), \sin(\alpha_i(t))) \\ \frac{d\alpha_i}{dt}(t) = \frac{\nu_i \sum_{j=1}^N \sin(\alpha_j(t) - \alpha_i(t))}{\sqrt{\left(\sum_{j=1}^N \cos(\alpha_j(t))\right)^2 + \left(\sum_{j=1}^N \sin(\alpha_j(t))\right)^2}} \end{cases} \quad (1.20)$$

The individuals move at a constant speed, and their directions tend to align with the mean direction of the population. It is also easy to check that this system is symmetric.

Example 1.4. a non-symmetric population system

Let $\lambda \neq \mu$ two real numbers.

$$\begin{cases} (x_1^0, x_2^0) \sim \mathcal{U}([0; 1])^{\otimes 2} \\ x_1(t) = x_1^0 e^{\lambda t} \\ x_2(t) = x_2^0 e^{\mu t} \end{cases} \quad (1.21)$$

It is clear that $(x_1(t), x_2(t))$ does not have the same distribution as $(x_2(t), x_1(t))$. As a consequence, there is no way to estimate parameter λ if one only observes the trajectory $x_2(t)$. This illustrates the fact that the information given by the observation of a subgroup of a non-symmetric population depends on the choice of the subgroup.

The symmetry of systems (1.19) and (1.20) is ensured by the initial distribution $\mu_0^N = \mu_0^{\otimes N}$, which is a factorized distribution, and by the invariance of the transition function by indices' permutation. The last invariance is ensured by the relation linking the interaction and the empirical measure of the population, given by the following definition.

Definition 1.7. empirical measure of a collection of points (from Mischler, 2011,²⁹ definition 1.2.4)

Let \mathcal{Z} be a metric space and $z_{1:N} = (z_i)_{1 \leq i \leq N}$ a collection of points in the space. Then the probability distribution

$$\mu[z_{1:N}] = \frac{1}{N} \sum_{i=1}^N \delta_{z_i} \in \mathcal{P}(\mathcal{Z}) \quad (1.22)$$

is the empirical measure of the collection of points. The set of empirical measures of N masses is denoted

$$\mathcal{P}_N(\mathcal{Z}) = \left\{ \frac{1}{N} \sum_{i=1}^N \delta_{z_i}, (z_1, \dots, z_N) \in \mathcal{Z}^N \right\} \quad (1.23)$$

The empirical measure is the probability measure corresponding to the uniform distribution over the population at time t . Systems (1.19) and (1.20) can be written more generally as

$$\forall i \in \{1, \dots, N\}, \quad \begin{cases} (x_i^0, \theta_i) \sim \mu_0 \in \mathcal{P}(\mathcal{X} \times \Theta) \\ \frac{dx_i}{dt}(t) = h_N \left(t, x_i(t), \theta_i, \frac{1}{N} \sum_{j=1}^N \delta_{(x_j(t), \theta_j)} \right) \end{cases} \quad (1.24)$$

Indeed, system (1.19) depends on the empirical measure linearly.

$$\frac{H}{m_i} \sum_{j=1}^N \frac{v_j(t) - v_i(t)}{(1 + |x_i(t) - x_j(t)|)^\beta} = \frac{NH}{m_i} \int_{\mathbb{R}^6} \frac{v - v_i(t)}{(1 + |x_i(t) - x|)^\beta} \hat{\mu}[t](dx, dv) \quad (1.25)$$

$$\text{where } \hat{\mu}[t] = \frac{1}{N} \sum_{j=1}^N \delta_{(x_j(t), v_j(t))}$$

System (1.20), however, is nonlinearly dependent on the empirical measure.

$$\begin{aligned} & \frac{\nu_i \sum_{j=1}^N \sin(\alpha_j(t) - \alpha_i(t))}{\sqrt{\left(\sum_{j=1}^N \cos(\alpha_j(t)) \right)^2 + \left(\sum_{j=1}^N \sin(\alpha_j(t)) \right)^2}} \\ &= \frac{\nu_i \int_{\mathbb{R}} \sin(\alpha - \alpha_i(t)) \hat{\mu}[t](d\alpha)}{\sqrt{\left(\int_{\mathbb{R}} \cos(\alpha) \hat{\mu}[t](d\alpha) \right)^2 + \left(\int_{\mathbb{R}} \sin(\alpha) \hat{\mu}[t](d\alpha) \right)^2}} \quad (1.26) \\ & \text{where } \hat{\mu}[t] = \frac{1}{N} \sum_{j=1}^N \delta_{\alpha_j(t)} \end{aligned}$$

The generic expression (1.24) leads us to consider a specific class of population systems whose dynamics are functions of the empirical measure. The following propositions give the proof that systems of the form (1.24) are symmetric.

Proposition 1.1. relation between empirical measure and permutation-invariant class

Let $N \geq 1$, \mathcal{Z} a metric space. We consider the quotient set $\mathcal{Z}^N / \mathfrak{S}_N$ associated with the equivalence relation

$$\forall z_{1:N}, z'_{1:N} \in \mathcal{Z}^N, z_{1:N} \mathcal{R} z'_{1:N} \Leftrightarrow \exists \sigma \in \mathfrak{S}_N, z_{1:N} = z'_{\sigma(1:N)} = (z'_{\sigma(i)})_{1 \leq i \leq N} \quad (1.27)$$

Then the map $\bar{z}_{1:N} \in \mathcal{Z}^N / \mathfrak{S}_N \mapsto \mu[z_{1:N}] \in \mathcal{P}_N(\mathcal{Z})$ is a bijection.

Proof. The map is clearly surjective. Let $z_{1:N}, y_{1:N} \in \mathcal{Z}^N$ such that $\mu[z_{1:N}] = \mu[y_{1:N}]$. Let us consider the equivalence relation over the set $\{1, \dots, N\}$

$$\forall j_1, j_2 \in \{1, \dots, N\}, \quad j_1 \mathcal{R} j_2 \Leftrightarrow z_{j_1} = z_{j_2} \quad (1.28)$$

The associated partition is denoted $\{1, \dots, N\} = C_1^z \cup \dots \cup C_r^z$ for some $r \geq 1$. Let $k \in \{1, \dots, r\}$ and $\ell \in C_k^z$, we can define the set $C_k^y = \{j \in \{1, \dots, N\} | y_j = z_\ell\}$ independently of the choice of ℓ . Then for all $k \in \{1, \dots, r\}$ and $\ell \in C_k^z$, we have

$$\text{Card}(C_k^z) = N\mu[z_{1:N}](\{z_\ell\}) = N\mu[y_{1:N}](\{z_\ell\}) = \text{Card}(C_k^y) \quad (1.29)$$

It follows that C_k^z and C_k^y are in bijection. Let us choose for all $k \in \{1, \dots, r\}$ $\gamma_k : C_k^z \rightarrow C_k^y$ a bijection. We then define the map σ by

$$\sigma : i \in \{1, \dots, N\} \mapsto \gamma_k(i) \text{ with } k \text{ such that } i \in C_k^z \quad (1.30)$$

Let $i_1 \in C_{k_1}^z$ and $i_2 \in C_{k_2}^z$ such that $\sigma(i_1) = \sigma(i_2)$. Then $k_1 = k_2$, since $C_{k_1}^y$ and $C_{k_2}^y$ are either disjoint or equal, and $i_1 = i_2$ by injectivity of γ_{k_1} . So σ is injective and therefore bijective, and we have $z_{1:N} = y_{\sigma(1:N)}$ by construction. \square

That means that we can represent graphically an empirical distribution as a cloud of points, where the points are undifferentiated, of the same colour for instance.

Proposition 1.2. specific class of symmetric population system

Let us consider the individual-based population system of initial configuration distribution $\mu_0^N = \mu_0^{\otimes N}$ where $\mu_0 \in \mathcal{P}(\mathcal{X} \times \Theta)$ and of interaction function

$$\begin{aligned} g_N : (t, x_1, \theta_1, (x_j, \theta_j)_{2 \leq j \leq N}) \in \mathbb{R}_+ \times \mathcal{X} \times \Theta \times (\mathcal{X} \times \Theta)^{N-1} \\ \mapsto h_N \left(t, x_1, \theta_1, \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, \theta_i)} \right) \end{aligned} \quad (1.31)$$

for some function $h_N : \mathbb{R}_+ \times \mathcal{X} \times \Theta \times \mathcal{P}_N(\mathcal{X} \times \Theta) \rightarrow \mathcal{X}$. Then the individual-based population system is symmetric.

Proof. Let $(x_{1:N}^t)_{t \geq 0}$ be a set of individual trajectories of the system. We note $H_N : (t, x_{1:N}, \theta_{1:N}) \in \mathbb{R}_+ \times \mathcal{X}^N \times \Theta^N \mapsto \left(h_N(t, x_i, \theta_i, \frac{1}{N} \sum_{j=1}^N \delta_{(x_j, \theta_j)}) \right)_{1 \leq i \leq N} \in \mathcal{X}^N$. By definition, we have for all $t \geq 0$ and for \mathbb{P} -almost every $\omega \in \Omega$

$$x_{1:N}^t = x_{1:N}^0 + \int_0^t H_N(s, x_{1:N}^s, \theta_{1:N}) ds \quad (1.32)$$

Let $\sigma \in \mathfrak{S}_N$. By composing the previous relation by $P_\sigma^{\mathcal{X}^N}$ we obtain

$$\begin{aligned} P_\sigma^{\mathcal{X}^N}(x_{1:N}^t) &= P_\sigma^{\mathcal{X}^N}(x_{1:N}^0) + \int_0^t P_\sigma^{\mathcal{X}^N}(H_N(s, x_{1:N}^s, \theta_{1:N})) ds \\ P_\sigma^{\mathcal{X}^N}(x_{1:N}^t) &= P_\sigma^{\mathcal{X}^N}(x_{1:N}^0) + \int_0^t H_N(s, P_\sigma^{\mathcal{X}^N}(x_{1:N}^s), P_\sigma^{\Theta^N}(\theta_{1:N})) ds \end{aligned} \quad (1.33)$$

It follows that the $(P_\sigma^{\mathcal{X}^N}(x_{1:N}^t))_{t \geq 0}$ is a set of individual trajectories for the system of initial configuration distribution $P_\sigma^{(\mathcal{X} \times \Theta)^N}(\mu_0^{\otimes N}) = \mu_0^{\otimes N}$ and having the same interaction function. Therefore $P_\sigma^{\mathcal{X}^N}(x_{1:N}^t) \sim x_{1:N}^t$ for all $t \geq 0$ by uniqueness of the distribution of solution processes. \square

The proposition also holds in the case of stochastic dynamics, i.e. for systems of the form (1.16), if the diffusion coefficients (denoted σ and σ_e in equation (1.16)) can be expressed as functions depending only on $(t, x, \theta, \hat{\mu}_N)$ where $\hat{\mu}_N$ is the empirical measure of the population.

1.3 Simulation of individual-based population systems

The simulation of symmetric population systems taking the form

$$\forall i \in \{1, \dots, N\}, \quad \begin{cases} (x_i^0, \theta_i) \sim \mu_0 \in \mathcal{P}(\mathcal{X} \times \Theta) \\ \frac{dx_i}{dt}(t) = h_N \left(t, x_i(t), \theta_i, \frac{1}{N} \sum_{j=1}^N \delta_{(x_j(t), \theta_j)} \right) \end{cases} \quad (1.34)$$

involves a numerical approximation of trajectories' distribution $(x_{1:N}(t))_{t \geq 0}$. The marginal distribution of a single individual state in the population, denoted by $\mu_x^{N:1}[t]$, is particularly useful as it enables a visualization of the system evolution.

Definition 1.8. marginal distribution of a symmetric population system

Let us consider a symmetric population system of the form (1.24) and let $\mu^N[t]$ be the distribution of $(x_i^t, \theta_i)_{1 \leq i \leq N}$ where $(x_{1:N}^s)_{s \geq 0}$ is any set of individual trajectories solution of the system. Let $k \leq N$. The marginal distribution $\mu_x^{N:k}[t]$ is the one defined by:

$$\forall \varphi \in \mathcal{C}_b^0(\mathcal{X}^k \rightarrow \mathbb{R}), \quad \int_{\mathcal{X}^k} \varphi(x_{1:k}) \mu_x^{N:k}[t](dx_{1:k}) = \int_{(\mathcal{X} \times \Theta)^N} \varphi(x_{1:k}) \mu^N[t](dx_1, d\theta_1, \dots, dx_N, d\theta_N) \quad (1.35)$$

where $\mathcal{C}_b^0(\mathcal{X}^k \rightarrow \mathbb{R})$ is the space of continuous and bounded functions defined over \mathcal{X}^k and taking values in \mathbb{R} .

Because the considered systems are symmetric, marginal distributions $\mu_x^{N:k}[t]$ represent the state variables of any group of size k in a population of size N . In order to compute the distribution $\mu_x^{N:1}[t]$ numerically, a classical method consists in approximating it by the empirical measure associated to independent and identically distributed (i.i.d.) samples $(x_1^{t,1}, \dots, x_1^{t,n})$ from $\mu_x^{N:1}[t]$. Obtaining a single simulation from $\mu_x^{N:1}[t]$ requires going through the following steps:

1. simulate from $\mu_0^{\otimes N}$ an initial configuration of a population of size N ;
2. solve *exactly* the differential system (1.24) over the interval $[0; t]$;
3. take a single individual in the population, e.g. $x_1(t)$.

The convergence of the empirical measure to the original distribution is a result that can be found for instance in Legland and Oudjane (2004),²⁴ expressed with a weak convergence metric.

$$\sup_{\varphi \in C_b^0(\mathcal{X}): \|\varphi\|_\infty=1} \mathbb{E} \left\{ \left| \frac{1}{n} \sum_{k=1}^n \varphi(x_1^{t,k}) - \int_{\mathcal{X}} \varphi(x) \mu_x^{N:1}[t](dx) \right|, x_1^{t,1:n} \sim (\mu_x^{N:1}[t])^{\otimes n} \right\} \leq \frac{1}{\sqrt{n}} \quad (1.36)$$

It is important to note that this convergence does not necessarily hold if the n -sample is composed of individuals coming from the same simulations, because of the correlation between the members of a given population. The bias due to individuals' interdependence can be seen when we consider the empirical measure

$\hat{\mu}_x^{N,n}[t] = \frac{1}{Nn} \sum_{k=1}^n \sum_{i=1}^N \delta_{x_i^{t,k}}$ where $x_i^{t,k}$ is the state of individual i obtained at the k^{th} simulation. Then for any continuous and bounded function φ , we have

$$\begin{aligned} & \mathbb{E} \left\{ \left| \frac{1}{Nn} \sum_{k=1}^n \sum_{i=1}^N \varphi(x_i^{t,k}) - \int_{\mathcal{X}} \varphi(x) \mu_x^{N:1}[t](dx) \right|^2, x_{1:N}^{t,1:n} \sim (\mu_x^N[t])^{\otimes n} \right\} \\ &= \frac{1}{N^2 n^2} \sum_{k_1=1}^n \sum_{k_2=1}^n \sum_{i_1=1}^N \sum_{i_2=1}^N \mathbb{E} \left\{ \left(\varphi(x_{i_1}^{t,k_1}) - \int_{\mathcal{X}} \varphi(x) \mu_x^{N:1}[t](dx) \right) \times \right. \\ & \quad \left. \left(\varphi(x_{i_2}^{t,k_2}) - \int_{\mathcal{X}} \varphi(x) \mu_x^{N:1}[t](dx) \right), x_{1:N}^{t,1:n} \sim (\mu_x^N[t])^{\otimes n} \right\} \\ &= \frac{1}{Nn} \text{Var} \left\{ \varphi(x), x \sim \mu_x^{N:1}[t] \right\} + \frac{N-1}{Nn} \text{Cov} \left\{ \varphi(x_1), \varphi(x_2); (x_1, x_2) \sim \mu_x^{N:2}[t] \right\} \end{aligned} \quad (1.37)$$

If it is not established that the covariance terms in equation (1.37) is zero or can be neglected, then we need to simulate an entire population of size N to obtain a single sample from $\mu_x^{N:1}[t]$.

The convergence (1.36) also implies an exact resolution of system (1.24), which is only possible in the event of an analytical differential system. In the general case, we have to resort to numerical solvers. The error induced by the numerical resolution has to be small enough so that the resulting n -sample can be serenely considered as an n -sample from the original distribution. Moreover, if it is established that the distribution $\mu_x^{N:1}[t]$ is absolutely continuous for the Lebesgue measure, then the approximation of this distribution can be refined by computing a kernel density estimate of its density.

The different simulations of the population can be carried out in parallel. Nevertheless, the simulation of a single population of individuals in interaction can be expensive in terms of computational time, especially when N is large. The computation time varies according to the specific form of the interaction linking each individual to the whole population.

Example 1.5. Cucker-Smale

In system (1.19), the computation time used to compute the velocity of all the individuals in the population is proportional to N^2 .

Example 1.6. Spring Cloud

Let us consider a system composed of N punctual masses, mutually connected by springs of stiffnesses $\kappa_i\kappa_j$, with κ_i and κ_j being individual parameters of particle i and particle j respectively. The evolution of the system is given by the following linear system of equations:

$$\forall i \in \{1, \dots, N\}, \quad \begin{cases} \frac{dx_i}{dt}(t) = v_i(t) \\ \frac{dv_i}{dt}(t) = \frac{1}{m_i} \sum_{j=1}^N \kappa_i\kappa_j(x_j(t) - x_i(t)) \end{cases} \quad (1.38)$$

It can be noticed that the location of the barycenter determines all the interactions in the population:

$$\frac{d^2x_i}{dt^2}(t) = -\frac{\kappa_i}{m_i} \left(\sum_{j=1}^N \kappa_j \right) x_i(t) + \frac{\kappa_i}{m_i} \sum_{j=1}^N \kappa_j x_j(t) \quad (1.39)$$

The computation time of the acceleration for all particles in the system is proportional to N .

Parallelization of the computation is conceivable if the interactions between individuals are short-range. This idea is commonly used for simulations of molecular dynamics (Plimpton, 1993³⁴). A space discretization can be conducted, and the cells' sizes should not exceed the typical length of interaction (Harvey et al., 2015,²² Cornell et al., 2019⁹).

Another type of methodology to earn some efficiency in the computation consists in freezing interactions over a short period, leaving the individuals to evolve separately within a constant population. This approach is used in Li et al. (2019)²⁶ for the simulation of a microbial population: higher scales are first simulated independently of the lower scales, then lower scales are updated while keeping constant the higher scales. In the case of system (1.24), two scales can be distinguished: the population level, represented by the empirical measure, and the individual level. Fixing the state of the population over some interval $[t_0; t_1]$ boils down to considering the following dynamical system

$$\forall t \in [t_0; t_1], \forall i \in \{1, \dots, N\}, \quad \begin{cases} \frac{dx_i}{dt}(t) = h_N(t, x_i(t), \theta_i, \hat{\mu}_N[t_0]) \\ \mu_N[t_0] = \frac{1}{N} \sum_{j=1}^N \delta_{(x_j(t_0), \theta_j)} \end{cases} \quad (1.40)$$

If the transition function h_N is smooth enough and if the length of the time interval is small enough, then there are reasons to believe that system (1.40) constitutes a good approximation of system (1.24). We can also consider a more accurate approximation of the trajectory $t \mapsto \hat{\mu}_N[t]$. Instead of considering a piecewise constant approximation of the trajectory, we can consider a piecewise polynomial approximation $t \mapsto \tilde{\mu}_N[t] \in \mathcal{P}(\mathcal{Z})$. This approximation is particularly useful from the computational point of view when fixing the empirical measure leads to analytical differential systems.

Example 1.7. Spring-Cloud

The system obtained by fixing the empirical measure in population model (1.38) is

$$\forall t \in [t_0; t_1], \forall i \in \{1, \dots, N\}, \quad \frac{d^2 x_i}{dt^2}(t) = -\frac{\kappa_i}{m_i} \left(\sum_{j=1}^N \kappa_j \right) x_i(t) + \frac{\kappa_i}{m_i} \sum_{j=1}^N \kappa_j x_j(t_0)$$

$$x_i(t) = \frac{\sum_{j=1}^N \kappa_j x_j(t_0)}{\sum_{j=1}^N \kappa_j} (1 - \cos(\omega_i(t - t_0))) + x_i(t_0) \cos(\omega_i(t - t_0)) + \frac{v_i(t_0)}{\omega_i} \sin(\omega_i(t - t_0))$$

where $\omega_i = \sqrt{\frac{\kappa_i \sum_{j=1}^N \kappa_j}{m_i}}$

(1.41)

The motion of each particle is along an ellipse. Notice that this approximation provides the exact dynamics of the system when the individual parameters κ_i, m_i are constant over the population (homogeneous population).

Example 1.8. Cucker-Smale

Fixing the empirical measure in system (1.19) leads to

$$\forall t \in [t_0; t_1], \forall i \in \{1, \dots, N\}, \quad \begin{cases} \frac{dx_i}{dt}(t) = v_i(t) \\ \frac{dv_i}{dt}(t) = \frac{H}{m_i} \sum_{j=1}^N \frac{v_j(t_0) - v_i(t)}{(1 + |x_i(t) - x_j(t_0)|)^\beta} \end{cases} \quad (1.42)$$

This nonlinear ODE can be solved in parallel for all individuals in the population. The linearized version of the system can also be considered

$$\forall t \in [t_0; t_1], \forall i \in \{1, \dots, N\}, \quad \begin{cases} \frac{dx_i}{dt}(t) = v_i(t) \\ \frac{dv_i}{dt}(t) = \frac{H}{m_i} \sum_{j=1}^N \frac{v_j(t_0) - v_i(t)}{(1 + |x_i(t_0) - x_j(t_0)|)^\beta} \end{cases} \quad (1.43)$$

which leads to analytical solutions over $[t_0; t_1]$.

$$\forall t \in [t_0; t_1], \quad v_i(t) = \frac{1}{\bar{w}_i^0} \sum_{j=1}^N w_{ij}^0 v_j(t_0) \left(1 - \exp\left(-\frac{t-t_0}{\tau_i^0}\right)\right) + v_i(t_0) \exp\left(-\frac{t-t_0}{\tau_i^0}\right)$$

$$\text{where } w_{ij}^0 = \frac{1}{(1 + |x_i(t_0) - x_j(t_0)|)^\beta}, \quad \bar{w}_i^0 = \sum_{j=1}^N w_{ij}^0, \quad \text{and } \tau_i^0 = \frac{H \bar{w}_i^0}{m_i}$$

$$x_i(t) = x_i(t_0) + \int_{t_0}^t v_i(s) ds \quad (1.44)$$

The choice of the period length $t_1 - t_0$ must take into account the dynamics of the original model. The time step must be small if the system is in a transitional regime, whereas it can be chosen large if the population is close to an eventual stationary state or equilibrium. The section 1.4 applies this methodology for the model of plant competition introduced by Schneider et al. (2006).³⁶

1.4 The Schneider system

1.4.1 Context of the system

According to the simulation study mimicking natural selection in Eloy et al. (2017),¹⁸ competition for light shapes the aerial part of woody plants. In agriculture, competition models are of prime importance to find the optimal density within a crop and ensure a good yield. This type of interaction is better represented in the plant modelling literature than *collaboration*-like interactions. Mutualistic

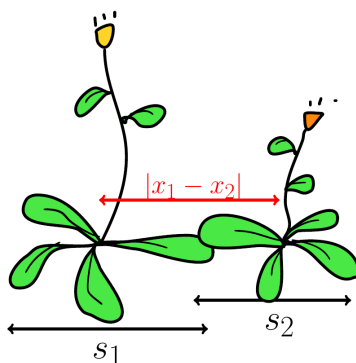


Figure 1.2: Representation of the size variable s in *Arabidopsis thaliana* used in Schneider et al. (2006).³⁶

relationships, like the transfer of nitrogen from legume to cereals studied in Patra et al. (1986),³³ explain nevertheless the advantages of multispecific crops. We have chosen to illustrate our methodology with a model of competition for light introduced by Schneider et al. (2006),³⁶ hoping that our approach could be generalised to systems with positive interactions between plants.

Competition models for light can provide a more or less accurate description of plant morphology, according to the objectives of modellers (Berger et al., 2008³). The estimation of the influence of plant organs and compartments over the local light environment is necessary to account for the variability in the architecture of the aerial parts (Clark and Bullock, 2007,⁷ Cournède et al., 2008,¹² Beyer et al., 2014⁴). In the model considered here, designed initially to study competition in a monospecific population of annual plants, arabidopsis (*Arabidopsis thaliana*), the plant morphology is only described by a characteristic dimension, and this considerably alleviates the experimental protocol to monitor the evolution of the population. The inter-plants competition is expressed via an empirical potential, which depends only on the observed individual features, referred to as a competition index (see Weigelt and Jolliffe, 2003,⁴¹ for a review of competition indices used in forestry).

1.4.2 Description of the system

In this system, the plant state is uniquely described by its rosette's diameter, denoted by s , to which we also refer as the *sizes* of the plants (see figure 1.2). If the competition exerted on an individual plant can be neglected, the evolution of the size of the plant of label i through time is given by the following differential

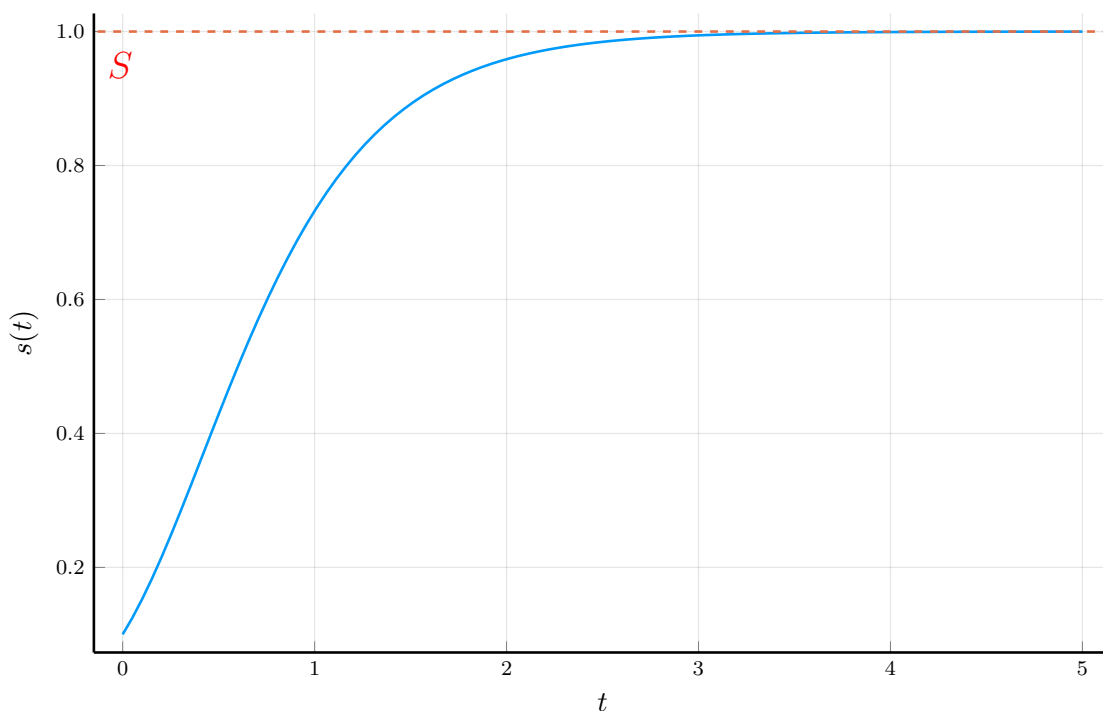


Figure 1.3: Evolution of the size of a plant without competition, according to equation (1.46), with the configuration $S = 1$, $\gamma = 2$, $s_0 = 0.1$.

equation, corresponding to a Gompertz growth function (Paine et al., 2012³²).

$$\frac{ds_i}{dt}(t) = \gamma_i s_i(t) (\log(S_i/s_m) - \log(s_i(t)/s_m)) \quad (1.45)$$

In the above equation, S_i, γ_i are intrinsic parameters of the individual plant, representing its asymptotic size and growth rate. The variation of these parameters from one plant to another can be due to genetic variability or micro-variations of the environment not covered by the experimental protocol. s_m is the minimal size of the plants, principally used as a normalisation constant. Notice that the differential equation (1.45) can be solved analytically for any given initial condition s_0^i , which has to be chosen, such as $s_m \leq s_0^i \leq S_i$.

$$\forall t \geq 0, \quad s_i(t) = S_i \left(\frac{s_0^i}{S_i} \right)^{e^{-\gamma_i t}} \quad (1.46)$$

The typical shape of the growth curve in the absence of competition is given in figure 1.3. In the presence of neighbouring plants exerting competition on plant

i , the growth equation (1.45) is modified by adding a time-dependent competition factor to the asymptotic size S_i , leading to the equation

$$\frac{ds_i}{dt}(t) = \gamma_i s_i(t) \left(\log(S_i/s_m)(1 - C_N^i(t)) - \log(s_i(t)/s_m) \right) \quad (1.47)$$

The competition index of plant i is a function of the surrounding plant states. In Schneider et al. (2006)³⁶ and in Lv et al. (2008),²⁸ the following expression is proposed for the competition index:

$$C_N^i(t) = \frac{1}{N-1} \sum_{j \neq i} C(s_i(t), s_j(t), |x_i - x_j|)$$

with $C(s_i(t), s_j(t), |x_i - x_j|) = \frac{\log(s_j(t)/s_m)}{2R_M \left(1 + \frac{|x_i - x_j|^2}{\sigma_x^2}\right)} \left(1 + \tanh\left(\frac{1}{\sigma_r} \log(s_j(t)/s_i(t))\right)\right)$

(1.48)

The competition potential $C(s_i, s_j, d_{ij})$ expresses the contribution of plant j to the competition exerted on plant i at a distance $d_{ij} = |x_i - x_j|$. In this model, competition is, therefore, just to be understood as a negative perturbation of the development a plant would theoretically have in optimal conditions. The competition index is composed of three factors that can be interpreted separately:

1. $\log(s_j(t)/s_m)/R_M$: the larger plant j , the stronger the competition it exerts on plant i ;
2. $1/2 \left(1 + \tanh\left(\frac{1}{\sigma_r} \log(s_j(t)/s_i(t))\right)\right)$: the larger plant i in comparison with plant j , the weaker the competition exerted on plant i by j , the parameter σ_r monitors the effect of the relative size;
3. $\frac{1}{1 + \frac{|x_i - x_j|^2}{\sigma_x^2}}$: the further apart plants i and j are, the less competition there is.

The normalisation constant R_M is chosen such that all competition indices in the population remain in the interval $[0; 1]$. For instance, if we have the prior knowledge that plant sizes cannot exceed some maximal size S_M , we can set $R_M = \log(S_M/s_m)$.

In the system taking into account the competition, each individual i is described by its size s_i , which is the only state variable, its position $x_i = (x_i^1, x_i^2) \in \mathbb{R}^2$, and its intrinsic parameters S_i, γ_i . Using the notations in the previous section, the individual parameter θ consists of four components (x^1, x^2, S, γ) and the state variable s is univariate.

1.4.3 Properties of the system

In this section, we shall establish some results to prove that the system defined above leads to global solutions over \mathbb{R}_+ . It is also essential to check whether the differential equation's behaviour is consistent with the biology of the system, i.e., sizes cannot be negative, and competition cannot have a positive effect on growth. Let us first start by proving the existence of global solutions for the Schneider system.

The existence of global solutions requires the following intuitive lemma, which is a variation of Grönwall lemma.

Lemma 1.1. Variation of Grönwall lemma

Let $y : [0; T] \rightarrow \mathbb{R}$ a continuously differentiable function such that there exists $\gamma \in \mathbb{R}$ and $\bar{y} \in \mathbb{R}$ such that $\forall t \in [0; T]$, $-\gamma y(t) \leq y'(t) \leq \gamma(\bar{y} - y(t))$. Then $\forall t \in [0; T]$, $y(0)e^{-\gamma t} \leq y(t) \leq \bar{y} - (\bar{y} - y(0))e^{-\gamma t}$.

Proof. Let us consider the function $\phi_+ : t \in [0; T] \mapsto (\bar{y} - y(t))e^{\gamma t}$. The derivative of the function is $\forall t \in [0; T]$, $\frac{d\phi_+}{dt}(t) = (\gamma(\bar{y} - y(t)) - y'(t))e^{\gamma t} \geq 0$. So ϕ_+ is increasing over $[0; T]$. In particular, we have for all $t \in [0; T]$:

$$\begin{aligned} \phi_+(t) &\geq \phi_+(0) \\ (\bar{y} - y(t))e^{\gamma t} &\geq \bar{y} - y(0) \\ y(t) &\leq \bar{y} - (\bar{y} - y(0))e^{-\gamma t} \end{aligned} \tag{1.49}$$

Similarly, we can prove that the function $\phi_- : t \in [0; T] \mapsto y(t)e^{\gamma t}$ is increasing over $[0; T]$ since $\frac{d\phi_-}{dt}(t) = (y'(t) + \gamma y(t))e^{\gamma t} \geq 0$. Therefore $\forall t \in [0; T]$, $y(t) \geq y(0)e^{-\gamma t}$. \square

Proposition 1.3. sufficient condition for the existence of a global solution to the Schneider system

Let us consider the system of N differential equations, with $N > 1$, defined by the dynamics

$$\forall i \in \{1, \dots, N\}, \left\{ \begin{array}{l} s_i(0) = s_i^0 \\ \forall t \geq 0, \frac{ds_i}{dt}(t) = \gamma_i s_i(t) \\ \times \left(\log(S_i/s_m) \left(1 - \frac{1}{N-1} \sum_{j \neq i} C(s_i(t), s_j(t), |x_i - x_j|) \right) - \log(s_i(t)/s_m) \right) \end{array} \right. \tag{1.50}$$

and by the constants $(s_i^0, x_i, S_i, \gamma_i)_{1 \leq i \leq N}$. The function $C : \mathbb{R}_+^3 \rightarrow \mathbb{R}$ is defined in equation (1.48) and $s_m, R_M > 0$. If $\forall i \in \{1, \dots, N\}$, $s_m < s_i^0 < S_i$, $s_m < S_i < s_m e^{R_M}$ and $\gamma_i > 0$, then the system has a global solution, defined over \mathbb{R}_+ .

Proof. Let us introduce the notation for the domain containing the initial configuration of the system

$$\mathcal{D} = \{(s, x, S, \gamma) \in \mathcal{Z} \mid s_m \leq s \leq S, s_m \leq S \leq s_m e^{R_M}, \gamma \geq 0\} \quad (1.51)$$

with $\mathcal{Z} = \mathbb{R}_+^* \times \mathbb{R}^2 \times \mathbb{R}_+ \times \mathbb{R}_+$. The set of individual parameters is $\Theta = \mathbb{R}^2 \times \mathbb{R}_+ \times \mathbb{R}_+$. Let $z_{1:N}^0 = (s_i^0, x_i, S_i, \gamma_i)_{1 \leq i \leq N} = (s_i^0, \theta_i)_{1 \leq i \leq N}$ be an initial configuration of the system chosen within the interior of the domain \mathcal{D} , i.e. $z_{1:N}^0 \in (\mathring{\mathcal{D}})^N$.

For a fixed initial configuration, the transition function of the system is

$$G_N : (s_{1:N}, \theta_{1:N}) \in (\mathbb{R}_+^*)^N \times \mathcal{Z}^N \mapsto \frac{1}{N-1} \left(\sum_{j \neq i} g(s_i, \theta_i, s_j, \theta_j) \right)_{1 \leq i \leq N} \in \mathbb{R}^N$$

with $g : (s_1, (x_1, S_1, \gamma_1), s_2, (x_2, S_2, \gamma_2)) \in \mathbb{R}_+^* \times \Theta \times \mathbb{R}_+^* \times \Theta$
 $\mapsto \gamma_1 s_1 (\log(S_1/s_m)(1 - C(s_1, s_2, |x_1 - x_2|)) - \log(s_1/s_m)) \in \mathbb{R}$

(1.52)

By Cauchy-Lipschitz theorem (1.1), as $G_N(\cdot, \theta_{1:N})$ is continuously differentiable with respect to $s_{1:N}$, the system (1.50) has a unique maximal solution defined over an interval $[0; t_m)$ with $t_m \in \overline{\mathbb{R}_+^*}$. For all time $t \in [0; t_m)$, we note $s_{1:N}(t)$ the maximal solution associated to the initial configuration $z_{1:N}^0$, and $z_{1:N}(t) = (s_i(t), x_i, S_i, \gamma_i)_{1 \leq i \leq N}$. We consider the interval $I_{\mathcal{D}} = \{t \in [0; t_m) \mid \forall \tau \in [0; t], z_{1:N}(\tau) \in (\mathring{\mathcal{D}})^N\}$. This interval is not reduced to the singleton $\{0\}$, since $t \in [0; t_m) \mapsto z_{1:N}(t)$ is continuous with $z_{1:N}(0) = z_{1:N}^0 \in \mathring{\mathcal{D}}^N$ and $\mathring{\mathcal{D}}^N$ is an open set. Then $t^* = \sup I_{\mathcal{D}}$ is strictly positive. Let $t \in [0; t^*)$ and $i, j \in \{1, \dots, N\}$, we have the following inequality on the competition index:

$$0 \leq C(s_i(t), s_j(t), |x_i - x_j|) \leq 1$$

so $-\gamma_i \log(s_i(t)/s_m) \leq \frac{d}{dt} \log(s_i(t)/s_m) \leq \gamma_i (\log(S_i/s_m) - \log(s_i(t)/s_m))$

(1.53)

By applying lemma 1.1, we obtain that for all $t \in [0; t^*)$ and for all $i \in \{1, \dots, N\}$

$$s_m (s_i^0/s_m)^{e^{-\gamma_i t}} \leq s_i(t) \leq S_i (s_i^0/S_i)^{e^{-\gamma_i t}} \quad (1.54)$$

If t^* is finite, we have therefore for all $t \in [0; t^*)$ and for all $i, j \in \{1, \dots, N\}$

$$\int_0^t |g(s_i(\tau), \theta_i, s_j(\tau), \theta_j)| d\tau \leq \gamma_i S_i \int_0^t \log \left(\frac{S_i}{s_m} \cdot \left(\frac{s_m}{s_i^0} \right)^{e^{-\gamma_i \tau}} \right) d\tau$$

$$\leq \gamma_i S_i t^* \log(S_i/s_m) + S_i \log(s_m/s_i^0) (1 - e^{-\gamma_i t^*})$$

(1.55)

so the integral $\int_0^t g(s_i(\tau), \theta_i, s_j(\tau), \theta_j) d\tau$ is absolutely convergent at point t^* . It

follows that for all $i \in \{1, \dots, N\}$ $\lim_{t \rightarrow (t^*)^-} s_i(t) = s_i^0 + \int_0^{t^*} g(s_i(\tau), \theta_i, s_j(\tau), \theta_j) d\tau$

exists. As $t \in [0; t_m) \mapsto s_{1:N}(t)$ is the unique maximal solution, then $t^* < t_m$ and we can evaluate the $s_{1:N}$ at time t^* . If we do so, we can notice that $s_m < s_i(t^*) < S_i$, which is in contradiction with the definition of $t^* = \sup I_{\mathcal{D}}$. As a consequence $t^* = t_m = +\infty$. \square

This result shows that the evolution of the population model takes place without finite-time blow-up, but it also shows that the dynamics are consistent with the initial assumption on the competition indices, which must remain bounded between 0 and 1. This sufficient condition on the existence of global solutions directs us toward a choice for the initial configuration distribution support.

Definition 1.9. support of a probability distribution

Let \mathcal{Z} be a metric space and $\mu \in \mathcal{P}(\mathcal{Z})$ a probability measure. We consider the set $\mathcal{S}_\mu = \{K \in \mathcal{B}(\mathcal{Z}) \mid K \text{ is closed and } \mu(K) = 1\}$. Then the support of the distribution μ is $\text{supp}(\mu) = \bigcap_{K \in \mathcal{S}_\mu} K$.

Corollary 1.1. The Schneider system is a symmetric population model

Let $\mu_0 \in \mathcal{P}(\mathbb{R}^5)$ such that $\text{supp}(\mu_0) \subset \mathring{\mathcal{D}}$ and $N > 1$. Let us consider the differential system with random initial configurations

$$\begin{aligned} (s_i^0, x_i, S_i, \gamma_i)_{1 \leq i \leq N} &\sim \mu_0^{\otimes N} \\ \forall i \in \{1, \dots, N\}, \quad \frac{ds_i}{dt}(t) &= \frac{1}{N-1} \sum_{j \neq i} g(s_i(t), \theta_i, s_j(t), \theta_j) \end{aligned} \quad (1.56)$$

where g is the function defined in equation (1.52). Then the Schneider system is a symmetric individual-based population models.

Proof. First, as a direct consequence of the proposition 1.3, $\mu_0^{\otimes N}$ is compatible with the differential system (1.50): for any random variable $z_{1:N}^0 \sim \mu_0^{\otimes N}$, there exists $\Omega' \in \mathcal{F}$ such that $\mathbb{P}(\Omega') = 1$ and $\forall \omega \in \Omega', \quad z_{1:N}^0(\omega) \in \mathcal{D}^N$, and we can apply the previous result on any $z_{1:N}^0(\omega)$. Besides, we have for all $t \in \mathbb{R}_+$ and for all $i \in \{1, \dots, N\}$, along any set of individual trajectories $s_{1:N}(t)$,

$$\begin{aligned} \frac{1}{N-1} \sum_{j \neq i} g(s_i(t), \theta_i, s_j(t), \theta_j) &= \frac{N}{N-1} \int_{\mathcal{Z}} g(s_i(t), \theta_i, s', \theta') \frac{1}{N} \sum_{j=1}^N \delta_{(s_j(t), \theta_j)}(ds', d\theta') \\ &\quad - \frac{g(s_i(t), \theta_i, s_i(t), \theta_i)}{N-1} \end{aligned} \quad (1.57)$$

Therefore, according to proposition 1.2, the Schneider system is a symmetric population model. \square

1.4.4 Definition of the initial configuration distribution

In this section, we choose a parametric expression for the initial configuration distribution μ_0 . First of all, its support must guarantee that the distribution is compatible with the Schneider system. This property can be verified if the support meets the sufficient condition of proposition 1.3. The distribution must also be parameterised in a way that it can generate a wide variety of initial configurations. In our case, we are interested in a population having a spatial distribution of individual parameters S and γ , meaning that these parameters are chosen with a high correlation with the position variable x .

In keeping with Schneider et al.³⁶ and Lv et al.,²⁸ the initial sizes of the plants are fixed to a constant $s_0 > s_m$ over the population, and the positions of the plants are uniformly distributed over a square domain $[0; L]^2$ for some distance L . Lv et al.²⁸ also consider a Poisson point-process distribution of the plants over the plane, which still leads to a symmetric population model, but it does not correspond to a distribution of the form $\mu_0^{\otimes N}$.

The distribution of the individual parameters S and γ is determined by two parametric surfaces $x \in [0; L]^2 \mapsto \bar{S}(x) \in \mathbb{R}_+$ and $x \in [0; L]^2 \mapsto \bar{\gamma}(x) \in \mathbb{R}_+$ defined by

$$\begin{aligned}\bar{S}(x) &= S_0 + (S_M - S_0) \exp\left(-\frac{1}{2}(x - x_1^S)^\top H_1^S (x - x_1^S)\right) - (S_0 - S_m) \exp\left(-\frac{1}{2}(x - x_2^S)^\top H_2^S (x - x_2^S)\right) \\ \bar{\gamma}(x) &= \gamma_0 + (\gamma_M - \gamma_0) \exp\left(-\frac{1}{2}(x - x_1^\gamma)^\top H_1^\gamma (x - x_1^\gamma)\right) - (\gamma_0 - \gamma_m) \exp\left(-\frac{1}{2}(x - x_2^\gamma)^\top H_2^\gamma (x - x_2^\gamma)\right)\end{aligned}\tag{1.58}$$

In the above equations, the surfaces are parameterised by their offsets S_0 or γ_0 , the location of high values of the parameter x_1^S or x_1^γ , the location of low values of the parameter x_2^S or x_2^γ , typical high values S_M or γ_M , and typical low values S_m or γ_m . The four matrices $H_1^S, H_2^S, H_1^\gamma, H_2^\gamma$ are symmetric positive and they monitor the shape of the surface in the neighbourhoods of $x_1^S, x_2^S, x_1^\gamma, x_2^\gamma$ respectively.

Individual parameters S and γ are chosen independent conditionally to the position x , with the conditional distributions

$$\begin{aligned}S \mid x &\sim \mathcal{U}([\bar{S}(x) - \sigma_S; \bar{S}(x) + \sigma_S]) \\ \gamma \mid x &\sim \mathcal{U}([\bar{\gamma}(x) - \sigma_\gamma; \bar{\gamma}(x) + \sigma_\gamma])\end{aligned}\tag{1.59}$$

In particular, the conditional expectations $\mathbb{E}[S \mid x]$ and $\mathbb{E}[\gamma \mid x]$ are equal to $\bar{S}(x)$ and $\bar{\gamma}(x)$ respectively.

Figure 1.4 represents a typical shape of the surface \bar{S} representing the mean value of parameter S over the domain $[0; L]^2$ for the following configuration of parameters.

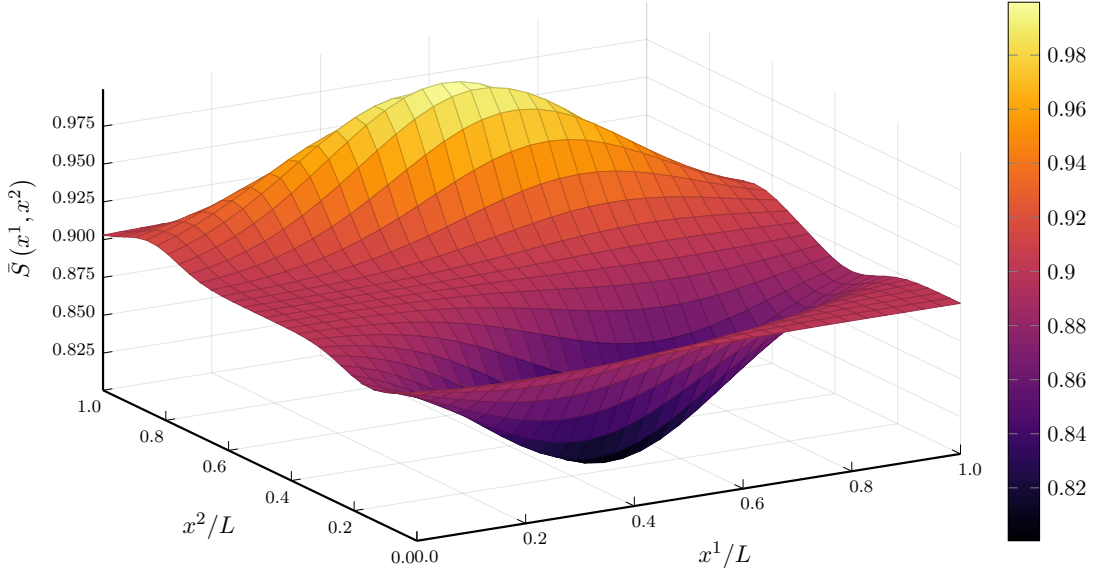


Figure 1.4: Surface $x \in [0; L]^2 \mapsto \bar{S}(x) \in \mathbb{R}$, defined in equation (1.58), with parameters in table 1.1.

In summary, if a random variable (s_0, x, S, γ) is of distribution μ_0 , then its components satisfy the following relation:

$$s_0 \sim \delta_{s_0}, \quad x \sim \mathcal{U}([0; L]^2), \quad S | x \sim \mathcal{U}([\bar{S}(x) - \sigma_S; \bar{S}(x) + \sigma_S]), \quad \gamma | x \sim \mathcal{U}([\bar{\gamma}(x) - \sigma_\gamma; \bar{\gamma}(x) + \sigma_\gamma]) \quad (1.60)$$

Equivalently, the distribution is defined by its action on continuous and bounded functions. Let $h : \mathcal{Z} \rightarrow \mathbb{R}$ be continuous and bounded, we have

$$\mathbb{E} \{h(s, x, S, \gamma), (s, x, S, \gamma) \sim \mu_0\} = \int_{[0; L]^2} \int_{\bar{S}(x) - \sigma_S}^{\bar{S}(x) + \sigma_S} \int_{\bar{\gamma}(x) - \sigma_\gamma}^{\bar{\gamma}(x) + \sigma_\gamma} \frac{h(s_0, x, S, \gamma)}{4L^2 \sigma_S \sigma_\gamma} d\gamma dS dx \quad (1.61)$$

1.4.5 Simulation of the Schneider system

Resolution using high-order numerical schemes

To approximate the distribution, we first need to solve numerically the differential system (1.50) for fixed initial conditions. We have used a Runge-Kutta method of 5th order with 4th order free interpolation (Tsitouras, 2011⁴⁰) implemented in the package `DifferentialEquations.jl` of Julia language (Rackauckas and Nie, 2017³⁵).

Parameters of μ_0	Values
$(L, \sigma_S, \sigma_\gamma)$	(1,0.1,0.1)
$(S_0, S_m, S_M), (\gamma_0, \gamma_m, \gamma_M)$	(0.9, 0.8, 1.0)
x_1^S, x_2^S	(0.5,0.75), (0.5,0.25)
x_1^γ, x_2^γ	(0.75,0.5), (0.25,0.5)
H_1^S, H_2^S	$\begin{pmatrix} 21.3 & 26.7 \\ 26.7 & 133.3 \end{pmatrix}, \begin{pmatrix} 21.3 & 26.7 \\ -26.7 & 133.3 \end{pmatrix}$
H_1^γ, H_2^γ	$\begin{pmatrix} 133.3 & 26.7 \\ 26.7 & 21.3 \end{pmatrix}, \begin{pmatrix} 133.3 & -26.7 \\ -26.7 & 21.3 \end{pmatrix}$
s_0	0.1

Table 1.1: Parameters of the initial distribution μ_0

Competition Parameters	Values
s_m	5×10^{-2}
R_M	3.09
(σ_r, σ_x)	(0.69, 1.0)

Table 1.2: Competition Parameters and their chosen values for the simulations

We consider solving the differential system of N equations

$$\forall i \in \{1, \dots, N\}, \quad \begin{cases} s_i(0) = s_i^0(\omega) \\ \forall t \geq 0, \quad \frac{ds_i}{dt}(t) = \frac{1}{N-1} \sum_{j \neq i} g(s_i(t), \theta_i(\omega), s_j(t), \theta_j(\omega)) \end{cases} \quad (1.62)$$

where $(s_i^0, \theta_i) \sim \mu_0^{\otimes N}$ is a random variable and $\omega \in \Omega' = \{\omega' \in \Omega \mid (s_i^0(\omega'), \theta_i(\omega')) \in (\mathcal{D})^N\}$. μ_0 is the distribution defined in the previous section. Tables 1.1 and 1.2 give the values of the parameters used for the simulation. The accuracy of the approximated solution is monitored by the relative tolerance (argument `reltol` in the solver), which is used to adapt the time step throughout the simulation. The smaller the tolerance, the more accurate the approximation, but of course the longer the simulation. We did not investigate the relation between the tolerance ϵ of the solver and eventual estimation of classical convergence metrics, such as

$$\frac{1}{N} \sum_{i=1}^N \sup_{0 \leq t \leq T} |s_i(t) - \hat{s}_i^\epsilon(t)| = \frac{1}{N} \sum_{i=1}^N \|s_i - \hat{s}_i^\epsilon\|_\infty \quad (1.63)$$

with $\hat{s}_{1:N}^\epsilon$ being the solution returned by the solver and $s_{1:N}$ being the true solution. However, by considering a decreasing sequence of tolerances $(\epsilon_k = 10^{-k})_{1 \leq k \leq K}$, we

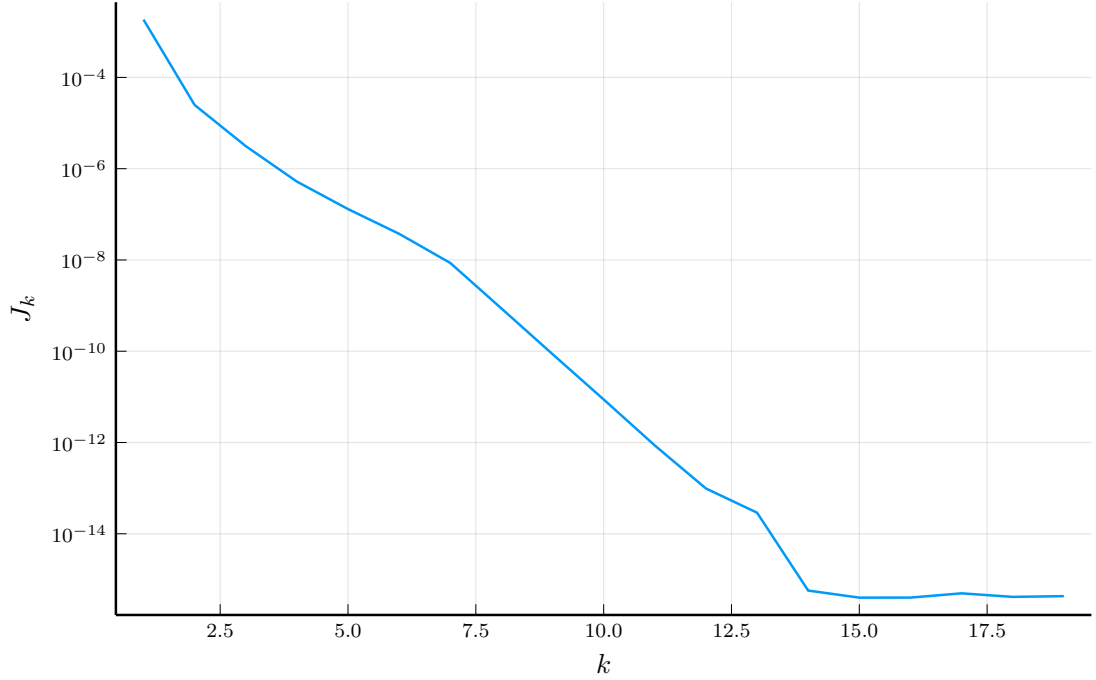


Figure 1.5: Evolution of the metric J_k , defined in equation (1.64), for decreasing tolerance $\epsilon_k = 10^{-k}$.

can select the tolerance using the metric

$$J_k = \frac{1}{N} \sum_{i=1}^N \|s_i^{\epsilon_k} - s_i^{\epsilon_K}\|_{\infty} \quad (1.64)$$

with K chosen large enough. We simulate the system over the interval $[0; T]$ with T chosen larger than the time it takes for the slowest plant (of growth rate γ_m) to reach 99% of the largest size S_M in the absence of competition, i.e. $\frac{1}{\gamma_m - \sigma_{\gamma}} \log\left(\frac{\log(s_0/(S_M + \sigma_S))}{\log(0.99)}\right) = 7.82$ in our case. $T = 8$ is our choice and the population size is set to $N = 50$ individuals. The initial configuration is fixed for the different simulations and the reference solution is the one obtained for a tolerance $\epsilon_{20} = 10^{-20}$. The sup in equation (1.63) is computed using a Brent univariate optimisation method (Brent, 1973⁶) implemented in the `Julia` package `Optim.jl` (Mogensen and Riseth, 2018³⁰). The results are given in figure 1.5. We can observe in figure 1.5 that the numerical solution obtained for the relative tolerance 10^{-2} is close to the reference solution by 2.5×10^{-5} . We therefore choose the value 10^{-2} for the relative tolerance for the next simulations, which are carried out for different

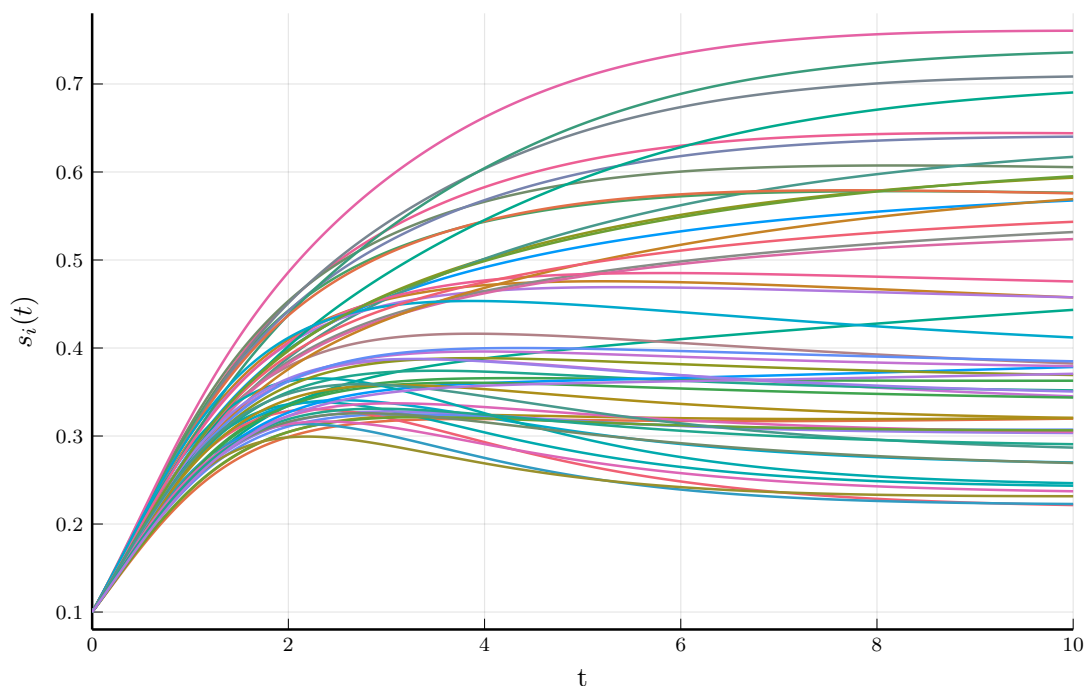


Figure 1.6: Set of individual trajectories $t \in [0; 10] \mapsto s_{1:N}(t)$ obtained with Julia package `DifferentialEquations.jl`³⁵ for a population of size $N = 50$.

population sizes N , different initial configurations and different time horizons T .

Figure 1.6 represents the evolution of a population of size $N = 50$ individuals over the time interval $[0; 10]$ for a given initial configuration. We can notice by looking at the final slopes of the growth curves that the sizes of the different plants have, for a vast majority, reached stationary sizes at time $T = 10$, leading to think that the whole population may have a stationary distribution. This intuition is confirmed later by figure 1.8. The proof of the existence of this stationary distribution has not yet been investigated.

By iterating the simulations, we can obtain an approximation of the distribution $\mu_s^{N:1}[t, \cdot]$, corresponding to the marginal distribution of the individual trajectory s_1^t , or by symmetry, the marginal distribution of a single individual size in the whole population. Figure 1.7 compares 100 different realisations of the individual trajectory s_1^t . By looking at this graph, we can speculate that the distribution $\mu_s^{N:1}[t, \cdot]$ is absolutely continuous for the Lebesgue measure for all time $t > 0$. This graphical intuition can also be substantiated by the fact that the distribution of the sizes in the absence of competition has a density for all time $t > 0$, having the

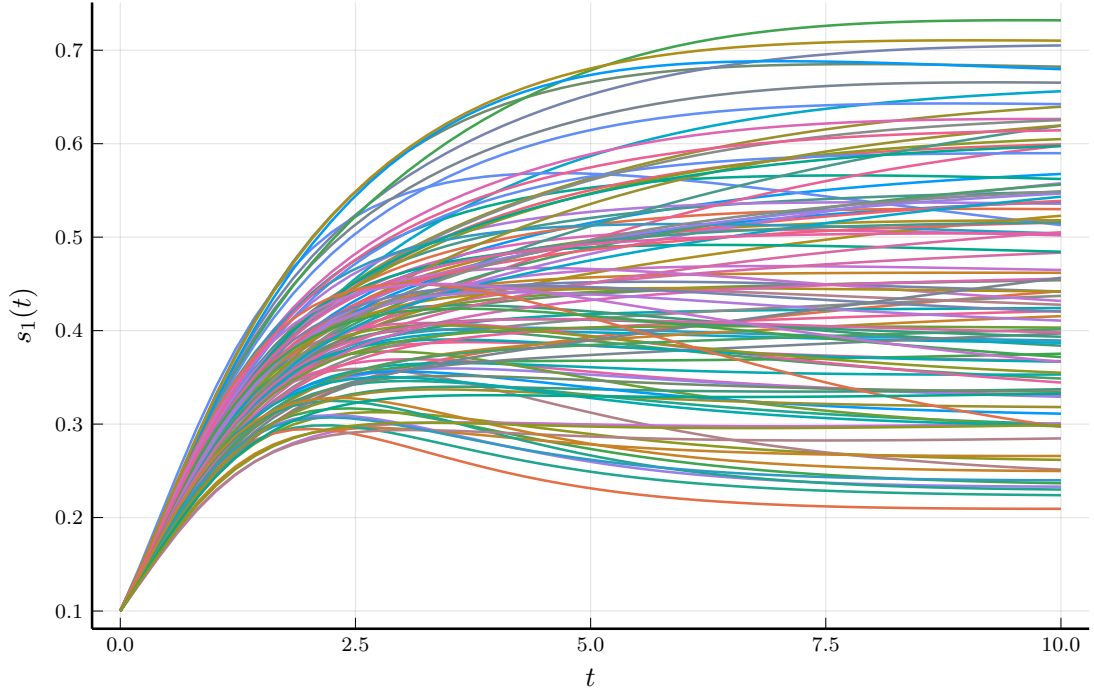


Figure 1.7: 100 independent simulations of the evolution $t \in [0; 10] \mapsto s_1(t)$ within a population of size $N = 50$ using `DifferentialEquations.jl`.³⁵

following expression

$$f_t^{nc} : s \in \mathbb{R} \mapsto \mathbb{I}\{s_0 < s\} \frac{d}{ds} \left[\int_s^{+\infty} F_\gamma^{\mu_0} \left(-\frac{1}{t} \log \left(\frac{\log(s/S)}{\log(s_0/S)} \right) \right) p_S^{\mu_0}(S) dS \right] \quad (1.65)$$

with $F_\gamma^{\mu_0}$ being the cumulative distribution function of μ_0^γ , the marginal distribution of γ , and $p_S^{\mu_0}$ being the marginal density of S . To visualise the evolution of size density, we carried out a kernel density estimation over the different realisations of s_1^t , using `KernelEstimation.jl` package. A Gaussian kernel is used for all density estimations and the support boundaries are specified, taking into account the fact that the support of the density is included within the segment $[s_m(s_0/s_m)^{\exp(-\gamma M t)}; S_M(s_0/S_M)^{\exp(-\gamma M t)}]$. As expected, figure 1.8 shows that the variance of the size density increases with time. The distribution at time $t = 10$ has a mode around $s = 0.3$, which is far below the minimal value of the asymptotic size $S_m - \sigma_S = 0.7$, meaning that inter-plant competition has a significant impact on the development of the population.

We want now to consider the influence of the population size N on the shape of the state distribution to assess the importance of this parameter graphically. This

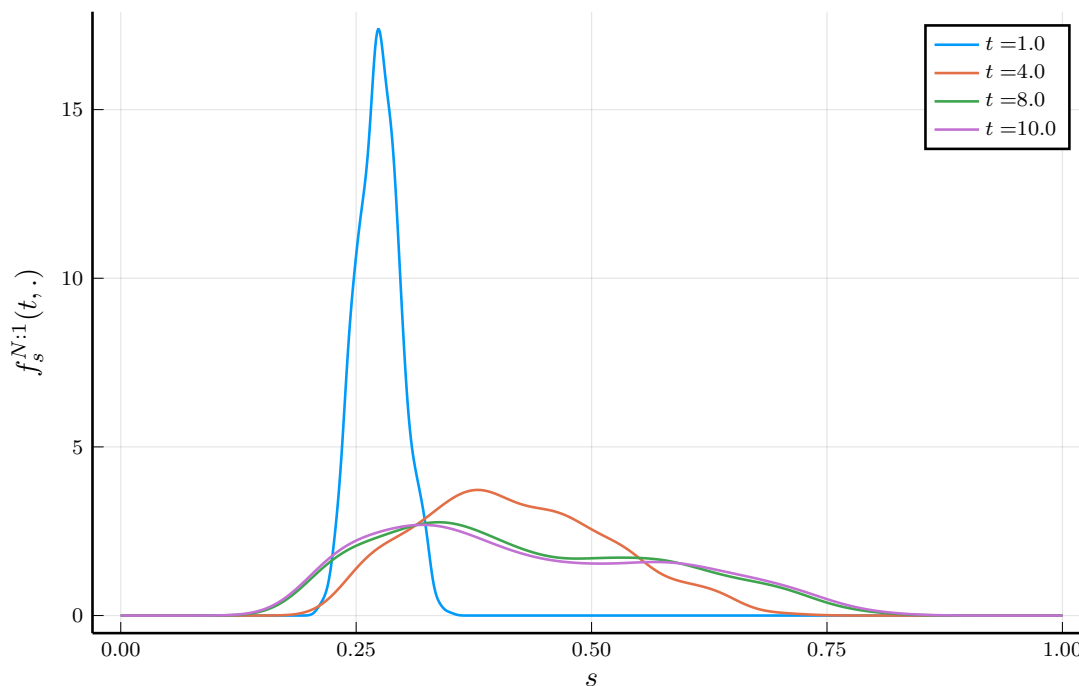


Figure 1.8: Kernel estimates of the densities $f_s^{N:1}(t, \cdot)$ at times $t = 1, 4, 10$ from a sample of 1 000 realisations of s_1^t , obtained by iterating the solver `DifferentialEquations.jl` over a population of size $N = 50$.

analysis has to face the problem of the computational cost induced by the simulation of large populations. Indeed, because of the pairwise interactions between individuals, the evolution of the computational time is asymptotically quadratic, as illustrated in figure 1.9. We can notice in figure 1.9 that the computation time required to simulate a population of 1 000 individuals over this short period $[0; 1]$ is roughly 10 s, which begins to be problematically long if we want to approximate the distribution of individual sizes. As we are especially interested in the system's behaviour for a large population, we need to find a more efficient approximation strategy for the Schneider system.

Fixed competition simulation

The stationary behaviour of the population implies that the competition indices are almost constant after some time. Lv et al. (2008)²⁸ consider an approximation of the system dynamics by fixing over sub-intervals of the simulation period, leading to piecewise constant competition indices. We give the proof of the consistency of this approximation, which does not figure in the original article, and we use the

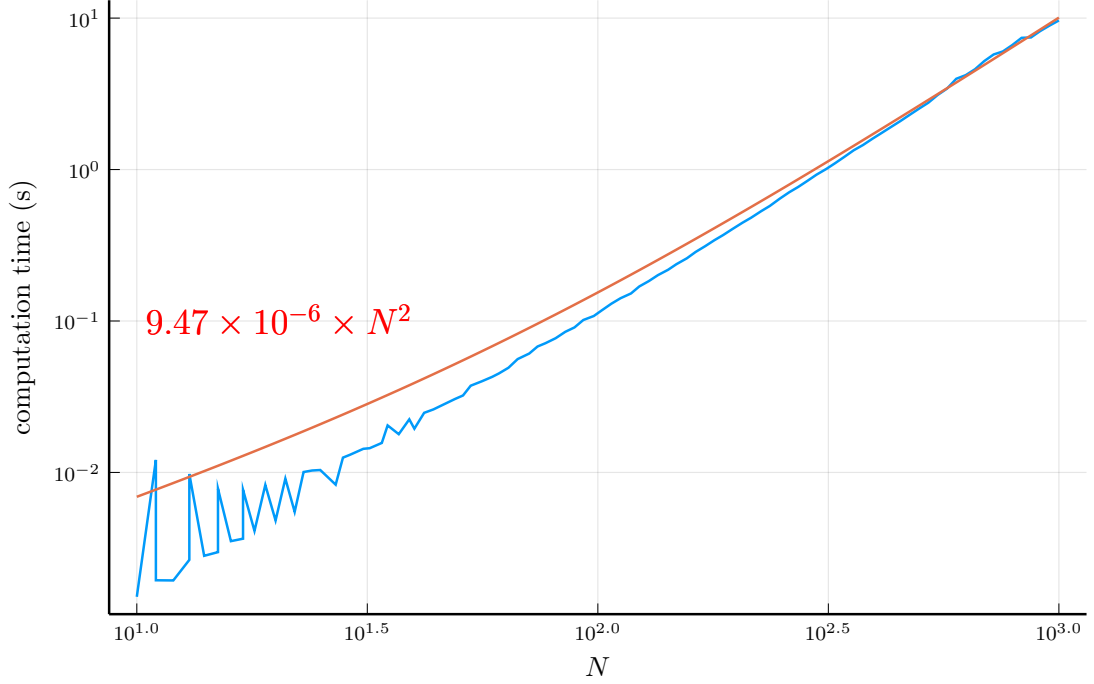


Figure 1.9: Evolution of the computation time for the simulation of the Schneider system using `DifferentialEquations.jl` over the time interval $[0; 1]$ for increasing population size N .

proof to derive an adaptive time-step method.

Proposition 1.4. consistency of fixed competition approximation

Let $T > 0$ and $(\Delta_n)_{n \in \mathbb{N}^*}$ a sequence of subdivisions of the interval $[0; T]$ with $\forall n \in \mathbb{N}^* t_0^n = 0, t_1^n, \dots, t_n^n = T$ such that $\max_{1 \leq k \leq n} (t_k^n - t_{k-1}^n) \xrightarrow[n \rightarrow \infty]{} 0$ and such that $\forall n \in \mathbb{N}^*, n \max_{1 \leq k \leq n} (t_k^n - t_{k-1}^n) \leq M$ for some $M > 0$. Let $z_{1:N}^0 = (s_i^0, \theta_i)_{1 \leq i \leq N} = (s_i^0, x_i, S_i, \gamma_i) \in (\tilde{\mathcal{D}})^N$, where \mathcal{D} is defined by equation (1.51). Let $s_{1:N} : t \in [0; T] \mapsto (s_i(t))_{1 \leq i \leq N} \in (\mathbb{R}_+^*)^N$ be the solution of the Schneider system (1.50) for the initial configuration $z_{1:N}^0$. Let us consider the function $\tilde{s}_{1:N}^n : t \in [0; T] \mapsto (\tilde{s}_i^n(t))_{1 \leq i \leq N} \in (\mathbb{R}_+^*)^N$ defined by, $\forall i \in \{1, \dots, N\}$,

$$\left\{ \begin{array}{l} \tilde{s}_i^n(0) = s_i^0 \\ \forall k \in \{1, \dots, n\}, \forall t \in [t_{k-1}^n; t_k^n), \quad \frac{d\tilde{s}_i^n}{dt}(t) = \gamma_i \tilde{s}_i^n(t) (\log(S_i/s_m) (1 - \tilde{C}_i^{k-1})) \\ - \log(\tilde{s}_i^n(t)/s_m) \text{ with } \tilde{C}_i^{k-1} = \frac{1}{N-1} \sum_{j \neq i} C(\tilde{s}_i^n(t_{k-1}^n), \tilde{s}_j^n(t_{k-1}^n), |x_i - x_j|) \end{array} \right. \quad (1.66)$$

Then: $\frac{1}{N} \sum_{i=1}^N \sup_{0 \leq t \leq T} |s_i(t) - \tilde{s}_i^n(t)| \xrightarrow{n \rightarrow \infty} 0$.

The proof of this result uses the regularity of the transition function with respect to the competition indices. An estimation of the error induced by keeping the competition constant over a short period of time is derived by evaluating $C(\tilde{s}_i^n(t), \tilde{s}_j^n(t), |x_i - x_j|)$ and by using the fact that $\tilde{s}_i^n(t)$ has analytical expressions over all subintervals of the subdivision.

Proof. The function $\tilde{s}_{1:N}^n$ is well-defined by the differential system (1.66) and has the following expression

$$\forall i \in \{1, \dots, N\}, \forall k \in \{1, \dots, n\}, \forall t \in [t_{k-1}^n; t_k^n],$$

$$\tilde{s}_i^n(t) = s_m \exp \left(\log(S_i/s_m) \left(1 - \tilde{C}_i^{k-1}\right) (1 - e^{-\gamma_i(t-t_{k-1}^n)}) + \log(\tilde{s}_i^n(t_{k-1}^n)/s_m) e^{-\gamma_i(t-t_{k-1}^n)} \right) \quad (1.67)$$

It can be proved by direct induction that $\tilde{C}_i^k \in [0; 1]$ for all $k \in \{1, \dots, n\}$ and $i \in \{1, \dots, N\}$, and that $s_m \leq \tilde{s}_i^n \leq S_i$. Let $k \in \{1, \dots, n\}$ and $t \in [t_{k-1}^n; t_k^n]$, and $i \in \{1, \dots, N\}$. We have:

$$s_i(t) = s_i(t_{k-1}^n) + \int_{t_{k-1}^n}^t g_C(s_i(\tau), \theta_i, C_N^i(\tau)) d\tau$$

where $g_C(s_i(t), \theta_i, C_N^i(t)) = \gamma_i s_i(t) (\log(S_i/s_m)(1 - C_N^i(t)) - \log(s_i(t)/s_m))$

and $\tilde{s}_i^n(t) = \tilde{s}_i^n(t_{k-1}^n) + \int_{t_{k-1}^n}^t g_C(\tilde{s}_i^n(\tau), \theta_i, \tilde{C}_i^{k-1}) d\tau$.

$$(1.68)$$

Thus: $|s_i(t) - \tilde{s}_i^n(t)| \leq |s_i(t_{k-1}^n) - \tilde{s}_i^n(t_{k-1}^n)| + \int_{t_{k-1}^n}^t (|g_C(s_i(\tau), \theta_i, C_N^i(\tau)) - g_C(\tilde{s}_i^n(\tau), \theta_i, C_N^i(\tau))| + |g_C(\tilde{s}_i^n(\tau), \theta_i, C_N^i(\tau)) - g_C(\tilde{s}_i^n(\tau), \theta_i, \tilde{C}_i^{k-1})|) d\tau$

$$(1.69)$$

We then estimate the quantities $|g_C(s_i(t), \theta_i, C_N^i(t)) - g_C(\tilde{s}_i^n(t), \theta_i, C_N^i(t))|$ and $|g_C(\tilde{s}_i^n(t), \theta_i, C_N^i(t)) - g_C(\tilde{s}_i^n(t), \theta_i, \tilde{C}_i^{k-1})|$ as functions of $|s_i(t) - \tilde{s}_i^n(t)|$ in order to resort to Grönwall lemma. Concerning the first term,

$$|g_C(s_i(t), \theta_i, C_N^i(t)) - g_C(\tilde{s}_i^n(t), \theta_i, C_N^i(t))| = \left| \int_{s_i(t)}^{\tilde{s}_i^n(t)} \partial_s g_C(s, \theta_i, C_N^i(t)) ds \right| \quad (1.70)$$

As the order relation between $s_i(t)$ and $\tilde{s}_i^n(t)$ is unknown, we consider all the values in between these two variables as elements of their convex hull, denoted by $\text{Conv}(\{s_i(t), \tilde{s}_i^n(t)\})$. So, for any $s \in \text{Conv}(\{s_i(t), \tilde{s}_i^n(t)\})$, we have:

$$\begin{aligned} \partial_s g_C(s, \theta_i, C_N^i(t)) &= \gamma_i [\log(S_i/s_m)(1 - C_N^i(t)) - \log(s/s_m) - 1] \\ |\partial_s g_C(s, \theta_i, C_N^i(t))| &\leq \gamma_i (1 + 2 \log(S_i/s_m)) \end{aligned} \quad (1.71)$$

$$\text{thus: } |g_C(s_i(t), \theta_i, C_N^i(t)) - g_C(\tilde{s}_i^n(t), \theta_i, C_N^i(t))| \leq \gamma_i(1 + 2 \log(S_i/s_m)) |s_i(t) - \tilde{s}_i^n(t)| \quad (1.72)$$

For the second term, we have:

$$\begin{aligned} & |g_C(\tilde{s}_i^n(\tau), \theta_i, C_N^i(\tau)) - g_C(\tilde{s}_i^n(\tau), \theta_i, \tilde{C}_i^{k-1})| = \gamma_i \log(S_i/s_m) \tilde{s}_i^n(t) \left| C_N^i(t) - \tilde{C}_i^{k-1} \right| \\ & \left| C_N^i(t) - \tilde{C}_i^{k-1} \right| \leq \frac{1}{N-1} \sum_{j \neq i} |C(s_i(t), s_j(t), |x_i - x_j|) - C(\tilde{s}_i^n(t_{k-1}^n), \tilde{s}_j^n(t_{k-1}^n), |x_i - x_j|)| \end{aligned} \quad (1.73)$$

We estimate the variation of the competition function by an upper-bound expressed as the sum of three terms. For any $(i, j) \in \{1, \dots, N\}^2$, we have

$$\begin{aligned} & |C(s_i(t), s_j(t), |x_i - x_j|) - C(\tilde{s}_i^n(t_{k-1}^n), \tilde{s}_j^n(t_{k-1}^n), |x_i - x_j|)| \\ & \leq |C(s_i(t), s_j(t), |x_i - x_j|) - C(\tilde{s}_i^n(t), s_j(t), |x_i - x_j|)| + \\ & |C(\tilde{s}_i^n(t), s_j(t), |x_i - x_j|) - C(\tilde{s}_i^n(t), \tilde{s}_j^n(t), |x_i - x_j|)| + \\ & |C(\tilde{s}_i^n(t), \tilde{s}_j^n(t), |x_i - x_j|) - C(\tilde{s}_i^n(t_{k-1}^n), \tilde{s}_j^n(t_n^k), |x_i - x_j|)| \end{aligned} \quad (1.74)$$

For the first term in the majorization (1.74),

$$\begin{aligned} & |C(s_i(t), s_j(t), |x_i - x_j|) - C(\tilde{s}_i^n(t), s_j(t), |x_i - x_j|)| \\ & = \left| \int_{s_i(t)}^{\tilde{s}_i^n(t)} \partial_{s_1} C(s_1, s_j(t), |x_i - x_j|) ds_1 \right| \\ & \forall s_1 \in \text{Conv}(\{s_i(t), \tilde{s}_i^n(t)\}), \\ & \partial_{s_1} C(s_1, s_j(t), |x_i - x_j|) \\ & = -\frac{\log(s_j(t)/s_m)}{2R_M s_1 \sigma_r (1 + |x_i - x_j|^2/\sigma_x^2)} \left(1 - \tanh\left(\frac{\log(s_j(t)/s_1)}{\sigma_r}\right) \right)^2 \end{aligned} \quad (1.75)$$

$$\begin{aligned} & |\partial_{s_1} C(s_1, s_j(t), |x_i - x_j|)| \leq \frac{\log(S_j/s_m)}{2R_M \sigma_r s_m (1 + |x_i - x_j|^2/\sigma_x^2)} \\ & |C(s_i(t), s_j(t), |x_i - x_j|) - C(\tilde{s}_i^n(t), s_j(t), |x_i - x_j|)| \\ & \leq \frac{\log(S_j/s_m)}{2R_M \sigma_r s_m (1 + |x_i - x_j|^2/\sigma_x^2)} |s_i(t) - \tilde{s}_i^n(t)| \end{aligned}$$

For the second term in the majorization (1.74),

$$\begin{aligned} & |C(\tilde{s}_i^n(t), s_j(t), |x_i - x_j|) - C(\tilde{s}_i^n(t), \tilde{s}_j^n(t), |x_i - x_j|)| \\ & = \left| \int_{s_j(t)}^{\tilde{s}_j^n(t)} \partial_{s_2} C(\tilde{s}_i^n(t), s_2, |x_i - x_j|) ds_2 \right| \end{aligned} \quad (1.76)$$

and we have:

$$\forall s_2 \in \text{Conv}(\{s_j(t), \tilde{s}_j^n(t)\}),$$

$$\partial_{s_2} C(s_i^n(t), s_2, |x_i - x_j|) = \frac{\sigma_r + \log(s_2/s_m)(1 - \tanh(\frac{1}{\sigma_r} \log(s_2/\tilde{s}_i^n(t))^2)) + \sigma_r \tanh(\frac{1}{\sigma_r} \log(s_2/\tilde{s}_i^n(t)))}{2R_M \sigma_r s_2 (1 + |x_i - x_j|^2/\sigma_x^2)}$$

$$|\partial_{s_2} C(s_i^n(t), s_2, |x_i - x_j|)| \leq \frac{2\sigma_r + \log(S_j/s_m)}{2R_M \sigma_r s_m (1 + |x_i - x_j|^2/\sigma_x^2)}$$

$$|C(\tilde{s}_i^n(t), s_j(t), |x_i - x_j|) - C(\tilde{s}_i^n(t), \tilde{s}_j^n(t), |x_i - x_j|)| \leq \frac{2\sigma_r + \log(S_j/s_m)}{2R_M \sigma_r s_m (1 + |x_i - x_j|^2/\sigma_x^2)} |s_j(t) - \tilde{s}_j^n(t)| \quad (1.77)$$

For the third term in the majorization (1.74),

$$\begin{aligned} & |C(\tilde{s}_i^n(t), \tilde{s}_j^n(t), |x_i - x_j|) - C(\tilde{s}_i^n(t_{k-1}^n), \tilde{s}_j^n(t_{k-1}^n), |x_i - x_j|)| = \left| \int_{t_{k-1}^n}^t \frac{d}{d\tau} [C(\tilde{s}_i^n(\tau), \tilde{s}_j^n(\tau), |x_i - x_j|)] d\tau \right| \\ & = \left| \int_{t_{k-1}^n}^t \left[\partial_{s_1} C(\tilde{s}_i^n(\tau), \tilde{s}_j^n(\tau), |x_i - x_j|) g_C(\tilde{s}_i^n(\tau), \theta_i, \tilde{C}_i^{k-1}) \right. \right. \\ & \quad \left. \left. + \partial_{s_2} C(\tilde{s}_i^n(\tau), \tilde{s}_j^n(\tau), |x_i - x_j|) g_C(\tilde{s}_j^n(\tau), \theta_j, \tilde{C}_j^{k-1}) \right] d\tau \right| \end{aligned}$$

We also that:

$$|g_C(\tilde{s}_i^n(t), \theta_i, \tilde{C}_i^{k-1})| \leq \gamma_i S_i \log(S_i/s_m)$$

and thus:

$$\begin{aligned} & |C(\tilde{s}_i^n(t), \tilde{s}_j^n(t), |x_i - x_j|) - C(\tilde{s}_i^n(t_{k-1}^n), \tilde{s}_j^n(t_{k-1}^n), |x_i - x_j|)| \leq \\ & \frac{(\gamma_i S_i \log(S_i/s_m) \log(S_j/s_m) + \gamma_j S_j \log(S_j/s_m) (2\sigma_r + \log(S_j/s_m)))(t - t_{k-1}^n)}{2R_M \sigma_r s_m (1 + |x_i - x_j|^2/\sigma_x^2)} \end{aligned} \quad (1.78)$$

We use all the inequalities from (1.71) to (1.78) to derive an upper-bound of

$$\frac{1}{N} \sum_{i=1}^N |s_i(t) - \tilde{s}_i^n(t)|. \text{ For this purpose, we introduce the following coefficients,}$$

indexed by the individuals $i, j \in \{1, \dots, N\}$

$$a_i = \gamma_i (1 + 2 \log(S_i/s_m))$$

$$b_{ij} = \frac{\gamma_i S_i \log(S_i/s_m) \log(S_j/s_m)}{2R_M \sigma_r s_m (1 + |x_i - x_j|^2/\sigma_x^2)}$$

$$c_{ij} = \frac{\gamma_i S_i \log(S_i/s_m) (2\sigma_r + \log(S_j/s_m))}{2R_M \sigma_r s_m (1 + |x_i - x_j|^2/\sigma_x^2)}$$

$$d_{ij} = \frac{\gamma_i S_i \log(S_i/s_m) (\gamma_i S_i \log(S_i/s_m) \log(S_j/s_m) + \gamma_j S_j \log(S_j/s_m) (2\sigma_r + \log(S_j/s_m)))}{2R_M \sigma_r s_m (1 + |x_i - x_j|^2/\sigma_x^2)} \quad (1.79)$$

These coefficients enables an estimation of the error $\frac{1}{N} \sum_{i=1}^N |s_i(t) - \tilde{s}_i^n(t)|$ using the error at time $t = t_{k-1}^n$ and the error integrated over the interval $[t_{k-1}^n; t]$.

$$\begin{aligned}
 \frac{1}{N} \sum_{i=1}^N |s_i(t) - \tilde{s}_i^n(t)| &\leq \frac{1}{N} \sum_{i=1}^n |s_i(t_{k-1}^n) - \tilde{s}_i^n(t_{k-1}^n)| + \frac{1}{N} \sum_{i=1}^N a_i \int_{t_{k-1}^n}^t |s_i(\tau) - \tilde{s}_i^n(\tau)| d\tau \\
 &+ \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i} b_{ij} \int_{t_{k-1}^n}^t |s_i(\tau) - \tilde{s}_i^n(\tau)| d\tau + \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j \neq i} c_{ij} \int_{t_{k-1}^n}^t |s_j(\tau) - \tilde{s}_j^n(\tau)| d\tau \\
 &+ \frac{1}{N(N-1)} \sum_{i=1} \sum_{j \neq i} d_{ij} \frac{(t - t_{k-1}^n)^2}{2}
 \end{aligned} \tag{1.80}$$

Let us introduce the characteristic time τ_1 and the constant $K_{\gamma^2_s}$, depending only on the individuals parameters in the population

$$\begin{aligned}
 \tau_1 &= \left(\max_{1 \leq i \leq N} a_i + \frac{1}{N-1} \left(\max_{1 \leq i \leq N} \sum_{j \neq i} b_{ij} + \max_{1 \leq j \leq N} \sum_{i \neq j} c_{ij} \right) \right)^{-1} \\
 K_{\gamma^2_s} &= \frac{1}{N(N-1)} \sum_{i=1} \sum_{j \neq i} d_{ij}
 \end{aligned}$$

We can write:

$$\begin{aligned}
 \frac{1}{N} \sum_{i=1}^N |s_i(t) - \tilde{s}_i^n(t)| &\leq \frac{1}{N} \sum_{i=1}^N |s_i(t_{k-1}^n) - \tilde{s}_i^n(t_{k-1}^n)| + \frac{1}{\tau_1} \int_{t_{k-1}^n}^t \frac{1}{N} \sum_{i=1}^N |s_i(\tau) - \tilde{s}_i^n(\tau)| d\tau + \\
 &K_{\gamma^2_s} \int_{t_{k-1}^n}^t (\tau - t_{k-1}^n) d\tau
 \end{aligned} \tag{1.81}$$

We introduce the local errors on the subintervals $[t_{k-1}^n; t_k^n]$.

$$\begin{aligned}
 E_k^n &= \frac{1}{N} \sum_{i=1}^N \sup_{t_{k-1}^n \leq t \leq t_k^n} |s_i(t) - \tilde{s}_i^n(t)| \\
 \Delta t_n &= \max_{1 \leq k \leq n} (t_k^n - t_{k-1}^n) \\
 E_k^n &\leq E_{k-1}^n + \frac{\Delta t_n}{\tau_1} E_k^n + K_{\gamma^2_s} \Delta t_n^2
 \end{aligned} \tag{1.82}$$

As $\Delta t_n \xrightarrow{n \rightarrow \infty} 0$ there exists $n_0 \in \mathbb{N}^*$ such that $\forall n \geq n_0$, $\Delta t_n < \tau_1$. So for $n \geq n_0$,

and for all $k \in \{1, \dots, n\}$,

$$E_k^n \leq \frac{E_{k-1}^n + K_{\gamma^2 s} \Delta t_n^2}{1 - \Delta t_n / \tau_1}$$

with $E_0^n = 0$ (1.83)

therefore $E_k^n \leq K_{\gamma^2 s} \Delta t_n \tau_1 \frac{1 - (1 - \Delta t_n / \tau_1)^k}{(1 - \Delta t_n / \tau_1)^k}$ by induction

We can use the estimation on the local errors E_k^n to derive an upper-bound for the global error over the whole interval $[0; T]$:

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \sup_{0 \leq t \leq T} |s_i(t) - \tilde{s}_i^n(t)| &\leq \max_{0 \leq k \leq n} E_k^n \\ &\leq K_{\gamma^2 s} \Delta t_n \tau_1 \frac{1 - (1 - \Delta t_n / \tau_1)^n}{(1 - \Delta t_n / \tau_1)^n} \\ &\leq K_{\gamma^2 s} \Delta t_n \tau_1 \left(\exp\left(\frac{M}{\tau_1 - M/n}\right) - 1 \right) \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

(1.84)

□

We can notice that the error estimate in inequality (1.78) can be improved by replacing the fixed competition term by a Taylor expansion up to order p of the competition potential $\tilde{C}_{ij}^m = C(\tilde{s}_i^n(t), \tilde{s}_j^n(t), |x_i - x_j|)$.

$$\tilde{C}_{ij}^{n,k}(t) = \tilde{C}_{ij}^n(t_{k-1}^n) + \frac{d\tilde{C}_{ij}^n}{dt}(t_{k-1}^n)(t - t_{k-1}^n) + \dots + \frac{d^p \tilde{C}_{ij}^n}{dt^p}(t_{k-1}^n) \frac{(t - t_{k-1}^n)^p}{p!} \quad (1.85)$$

If the competition term is a polynomial function of t , the trajectories of the system (1.66) are still analytical. We have therefore a methodology to obtain arbitrarily accurate approximation of the true solution of the system (1.50).

The choice of a subdivision of the interval $[0; T]$ before the system simulation is problematic, as one does not know in advance the quality of approximation consisting of fixing the competition for a while. That is why it is preferable to build the subdivision dynamically by estimating whether we are in a phase where the competition changes quickly or slowly. The proof of the proposition 1.4 involves a quantity that gives the part of the total error that can be attributed to the action of fixing competition. This is the quantity estimated in inequality 1.78. We can also consider the following quantity:

$$\begin{aligned} C_{fc}(t_0, t, \tilde{s}_{1:N}^0, \theta_{1:N}) = \\ \frac{1}{N-1} \sum_{i=1}^N \frac{\gamma_i S_i \log(S_i / s_m)}{\sum_{i'=1}^N \gamma_{i'} S_{i'} \log(S_{i'} / s_m)} \sum_{j \neq i} \sup_{t_0 \leq \tau \leq t} |C(\tilde{s}_i^n(\tau), \tilde{s}_j^n(\tau), |x_i - x_j|) - C(\tilde{s}_i^0, \tilde{s}_j^0, |x_i - x_j|)| \end{aligned}$$

(1.86)

Intuitively, if the approximated trajectories $\tilde{s}_{1:N}^n$ of the population evolve so that there are substantial variations of the competition potentials for each individual over the interval $[t_0; t]$, then the quantity (1.86) tends to be large, and this indicates that the approximation of a constant competition over $[t_0; t]$ is poor. Each variation of the competition potential is weighted by $\gamma \log(S/s_m)$ in order to pay particular attention to plants having faster growth rates and larger asymptotic sizes. Similarly as in inequality (1.78), it can be shown that quantity (1.86) is upper-bounded by a quadratic function of t , but this approximation is rather rough, as the size remains bounded for all time.

Algorithm 1 describes the adaptive time step method we propose. It requires the minimisation of a continuous and non convex function over a segment.

notation: Let $f : K \rightarrow \mathbb{R}$ a continuous function over a compact set. We use the notation

$$\text{Argmin} \{f(x), x \in K\} = \left\{ x \in K \mid f(x) = \min_{y \in K} f(y) \right\} \quad (1.87)$$

Algorithm 1 Adaptive time step simulation using piecewise constant competition

Inputs: a tolerance $\varepsilon > 0$, the size of subpopulation $N_{sp} \leq N$, a sample $(s_i^0, \theta_i)_{1 \leq i \leq N}$ from the initial distribution μ_0 , time horizon T

Initialisation:

1. $t = 0$
2. compute competition indices for all individuals in the population $(C_i^0)_{1 \leq i \leq N}$
3. initialisation of the individual sizes $\forall i \in \{1, \dots, N\}$, $\tilde{s}_i = s_i^0$

while $t < T$ **do**

1. select a subpopulation of size N_{sp} by sampling a discrete uniform distribution $\mathcal{U}(\{1, \dots, N\})$, the subpopulation obtained is denoted by $(\tilde{s}'_i, \theta'_i)_{1 \leq i \leq N_{sp}}$.
2. compute the time update t' given by the following formula

$$t^* = \max \text{Argmin} \left\{ (C_{fc}(t, t', \tilde{s}'_{1:N_{sp}}, \theta_{1:N_{sp}}) - \varepsilon)^2, t \leq t' \leq T \right\} \quad (1.88)$$

3. update all individual sizes over the interval $[t; t^*]$ with

$$\forall \tau \in [t; t^*], \tilde{s}_i(\tau) = s_m \exp \left(\log(S_i/s_m) (1 - C_i^t) (1 - e^{-\gamma_i(\tau-t)}) + \log(\tilde{s}'_i/s_m) e^{-\gamma_i(\tau-t)} \right) \quad (1.89)$$

4. set $t := t^*$.
5. update the values of all competition indices $(C_i^t)_{1 \leq i \leq N}$.

end while

Outputs: set of individual trajectories $(t \in [0; T] \mapsto \tilde{s}_i(t))_{1 \leq i \leq N}$.

The consistency of the simulation method provided by algorithm 1 is a corollary of the proposition 1.4.

Corollary 1.2. consistency of the adaptive method with piecewise constant competition

Let $T > 0$. Let $\varepsilon > 0$ and $(\Delta t_n)_{n \in \mathbb{N}^*}$ a sequence of strictly positive real numbers such that $\Delta t_n \leq M/n$ for some $M > 0$. Let $z_{1:N}^0 = (s_i^0, \theta_i)_{1 \leq i \leq N} = (s_i^0, x_i, S_i, \gamma_i) \in (\mathring{\mathcal{D}})^N$, where \mathcal{D} is defined by equation (1.51). Let $s_{1:N} : t \in [0; T] \mapsto (s_i(t))_{1 \leq i \leq N} \in (\mathbb{R}_+^*)^N$ be the solution of the Schneider system (1.50) for the initial configuration $z_{1:N}^0$. We consider for all $n \in \mathbb{N}^*$ the sequences $(t_k^n)_{k \in \mathbb{N}}$, $(\tilde{s}_{1:N}^{n,k})_{k \in \mathbb{N}}$ and $(\tilde{C}_{1:N}^{n,k})_{k \in \mathbb{N}}$ defined by

$$t_0^n = 0, \quad \tilde{s}_{1:N}^{n,0} = s_{1:N}^0, \quad \tilde{C}_{1:N}^{n,0} = \left(\frac{1}{N-1} \sum_{j \neq i} C(s_i^0, s_j^0, |x_i - x_j|) \right)_{1 \leq i \leq N}$$

and $\forall k \in \mathbb{N}, \forall i \in \{1, \dots, N\}$,

$$t_{k+1}^n = \max \text{Argmin} \left\{ (C_{fc}(t_k^n, t, \tilde{s}_{1:N}^{n,k}, \theta_{1:N}) - \varepsilon_n)^2, t_k^n \leq t \leq \min(t_k^n + \Delta t_n, T) \right\}$$

$$\tilde{s}_i^{n,k+1} = s_m \exp \left(\log(S_i/s_m) \left(1 - \tilde{C}_i^{n,k} \right) (1 - e^{-\gamma_i(t_{k+1}^n - t_k^n)}) + \log(\tilde{s}_i^{n,k}/s_m) e^{-\gamma_i(t_{k+1}^n - t_k^n)} \right)$$

$$\tilde{C}_i^{n,k+1} = \frac{1}{N-1} \sum_{j \neq i} C(\tilde{s}_i^{n,k+1}, \tilde{s}_j^{n,k+1}, |x_i - x_j|)$$
(1.90)

Then for all $n \in \mathbb{N}^*$, there exists $K_n \in \mathbb{N}^*$ such that $t_{K_n}^n = T$ (and therefore $\forall k \geq K_n, t_k^n = T$). Moreover, the sequence of functions $(\tilde{s}_{1:N}^n)_{n \in \mathbb{N}^*}$ defined by equation

(1.66) over the subdivision $(t_0^n, t_1^n, \dots, t_{K_n}^n)$ is such that $\lim_{n \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \sup_{0 \leq t \leq T} |s_i(t) - \tilde{s}_i^n(t)| = 0$.

Proof. To simplify the notations, we use the following abbreviation: $C_{fc}(t_k^n, t) := C_{fc}(t_k^n, t, \tilde{s}_{1:N}^{n,k}, \theta_{1:N})$. By construction of $(t_k^n)_{k \in \mathbb{N}}$, the sequence is non-decreasing and $\forall k \in \mathbb{N}, t_k^n \leq T$. We want to prove that this sequence is stationary and that it defines a subdivision of the interval $[0; T]$. Because of the regularity of the competition function, it is possible to prove that the time increment $t_{k+1}^n - t_k^n$ are above a minimal value $\alpha > 0$, until $t_k^n + \alpha > T$. Using the upper-bound of the function $t \in [t_k^n; t_{k+1}^n] \mapsto C_{fc}(t_k^n, t, \tilde{s}_{1:N}^{n,k}, \theta_{1:N})$ defined in inequalities (1.78)

$$\forall t \in [t_k^n; t_{k+1}^n], C_{fc}(t_k^n, t, \tilde{s}_{1:N}^{n,k}, \theta_{1:N}) \leq \frac{K_{\gamma^2 s}(t - t_k^n)}{\frac{1}{N} \sum_{i=1}^N \gamma_i S_i \log(S_i/s_m)} \quad (1.91)$$

From this inequality, we can deduce that if there exists $t' \in [t_k^n; \min(t_k^n + \Delta t_n, T)]$ such that $\frac{K_{\gamma^2 s}(t' - t_k^n)}{\frac{1}{N} \sum_{i=1}^N \gamma_i S_i \log(S_i/s_m)} = \varepsilon$ then we have that $C_{fc}(t_k^n, t') \leq C_{fc}(t_k^n, t_{k+1}^n)$. Otherwise, if such time t' does not exist, then the closest value to the tolerance ε reached by $t \mapsto C_{fc}(t_k^n, t)$ is at the right boundary of the interval, i.e., either at $t_k^n + \Delta t_n$ or at T . Hence, we have the following minimal value for the time increment.

$$\forall k \in \mathbb{N}, \quad t_{k+1}^n - t_k^n \geq \min \left(\frac{\varepsilon}{NK_{\gamma^2 s}} \sum_{i=1}^N \gamma_i S_i \log(S_i/s_m), \Delta t_n, T - t_k^n \right) \quad (1.92)$$

Let us note $\overline{\Delta t} = \frac{\varepsilon}{NK_{\gamma^2 s}} \sum_{i=1}^N \gamma_i S_i \log(S_i/s_m)$ and $t_\infty^n = \lim_{k \rightarrow \infty} t_k^n$, which exists since the sequence $(t_k^n)_{k \in \mathbb{N}}$ is non-decreasing and bounded. If $t_\infty^n < T$, then there exists k_0 such that $\forall k \geq k_0$

$$t_\infty^n - t_k^n \leq \frac{\min(\overline{\Delta t}, \Delta t_n, T - t_\infty^n)}{2} \quad (1.93)$$

It follows that

$$\begin{aligned} t_{k+1}^n - t_\infty^n &= t_{k+1}^n - t_k^n + t_k^n - t_\infty^n \geq \min(\overline{\Delta t}, \Delta t_n, T - t_k^n) - \frac{\min(\overline{\Delta t}, \Delta t_n, T - t_\infty^n)}{2} \\ &\geq \frac{\min(\overline{\Delta t}, \Delta t_n, T - t_\infty^n)}{2} > 0 \end{aligned} \quad (1.94)$$

So, in particular, $t_{k_0+1}^n > t_\infty^n$, which is in contradiction with the monotonicity of the sequence $(t_k^n)_{k \in \mathbb{N}}$. As a consequence, the limit is $t_\infty^n = T$. We can then find $k_1 \in \mathbb{N}$ such that $T - t_{k_1}^n \leq \frac{\min(\overline{\Delta t}, \Delta t_n)}{2}$, leading to $t_{k_1+1}^n = T$. Hence, the sequence is stationary, and it is possible to build from this sequence a subdivision of the interval $[0; T]$ by removing all the terms that are equal to T . The sequence of functions $(\tilde{s}_{1:N}^n)_{n \in \mathbb{N}^*}$, having each of its term associated with a subdivision extracted from $(t_k^n)_{k \in \mathbb{N}}$, converges to $s_{1:N}$ by proposition (1.4). \square

Algorithm 1 and the one studied in corollary 1.2 exhibits some differences, which aims at alleviating the computation cost of the adaptive algorithm.

1. the sup is removed from the expression of C_{fc} in (1.86); without the sup the function is not necessarily increasing, which precludes the use of inequality (1.92) in the previous proof.

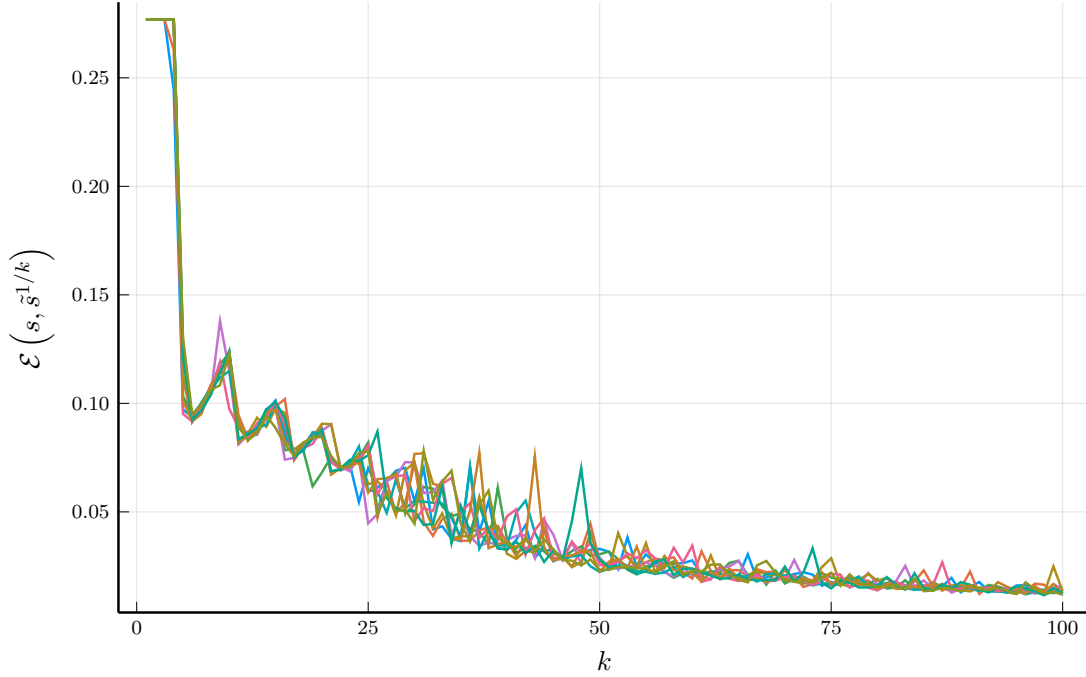


Figure 1.10: Evolution of the consistency error $\mathcal{E}(s, \tilde{s}^{1/k})$ defined in equation (1.95) with the tolerance $\varepsilon_k = 1/k$ and for 10 independent simulations at each value of k . The size of all random subpopulations is $N_{sp} = 10$. All simulations are performed with the same initial configuration $z_{1:N}^0$, generated from distribution μ_0 with the values of parameters in table 1.1, and the competition parameters in table 1.2, for a population of $N = 50$ individuals. The reference solution s is the one obtained with the numerical solver described in subsection 1.4.5 with a relative tolerance of 10^{-2} .

2. The computation of C_{fc} is done over a random sub-population, not on the entire population; the method is, therefore, stochastic if $N_{sp} < N$. It can be proved that consistency still holds in that case.
3. The minimisation procedure in (1.88) is not carried over a small segment $[t; t + \Delta t_n]$ but over the whole time domain $[t; T]$. This extension may constitute a source of bias in the convergence of $(\tilde{s}_n)_{n \in \mathbb{N}^*}$, as the consistency is mainly based on the fact that the time step of the subdivision can be chosen arbitrarily small.

As a consequence, the consistency of algorithm 1 concerns rather a family of functions $(\tilde{s}_{1:N}^\varepsilon)$ indexed by the tolerance ε , and it is based on the postulate that

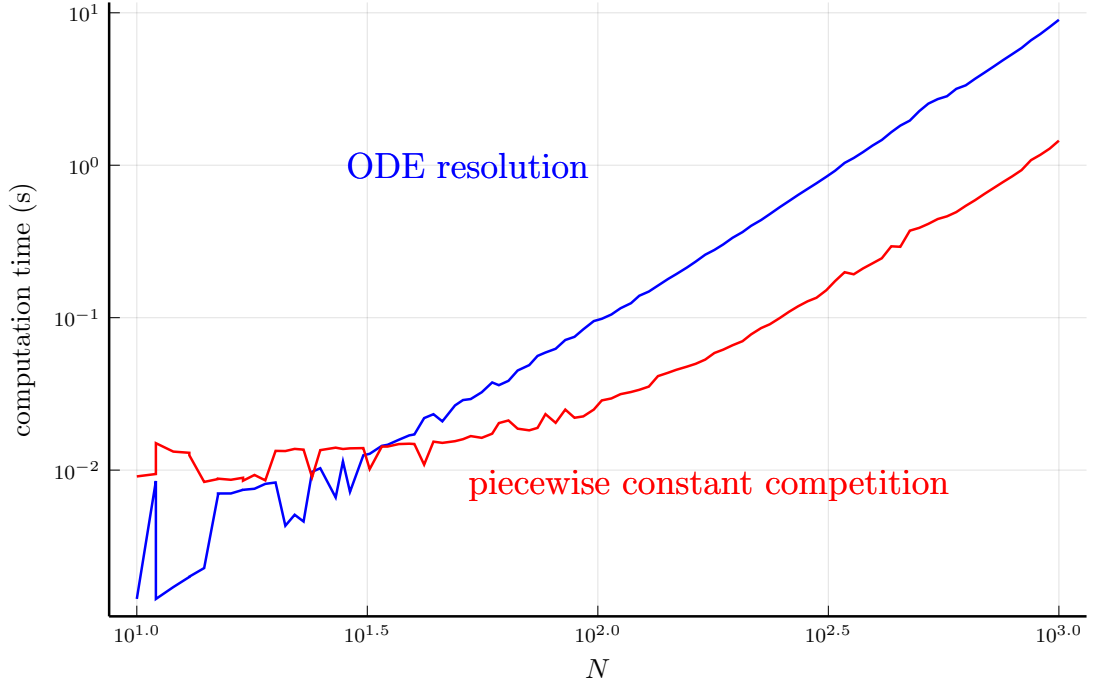


Figure 1.11: Comparison of the computation time used to solve the Schneider system (1.50) over the interval $[0; 1]$ for different population sizes N using a numerical ODE solver^{35,40} (ODE resolution) in the configuration described in subsection 1.4.5, and using the method described in algorithm 1 (piecewise constant competition) with the configuration $\varepsilon = 1/50$ and $N_{sp} = 10$.

the smaller ε is, the smaller the time step is. Despite these differences, figure 1.10 shows that the consistency of the method corresponding to algorithm 1 seems preserved, i.e.

$$\mathcal{E}(s_{1:N}, \tilde{s}_{1:N}^\varepsilon) = \frac{1}{N} \sum_{i=1}^N \sup_{0 \leq t \leq T} |s_i(t) - \tilde{s}_i^\varepsilon(t)| \xrightarrow{\varepsilon \rightarrow 0} 0 \quad (1.95)$$

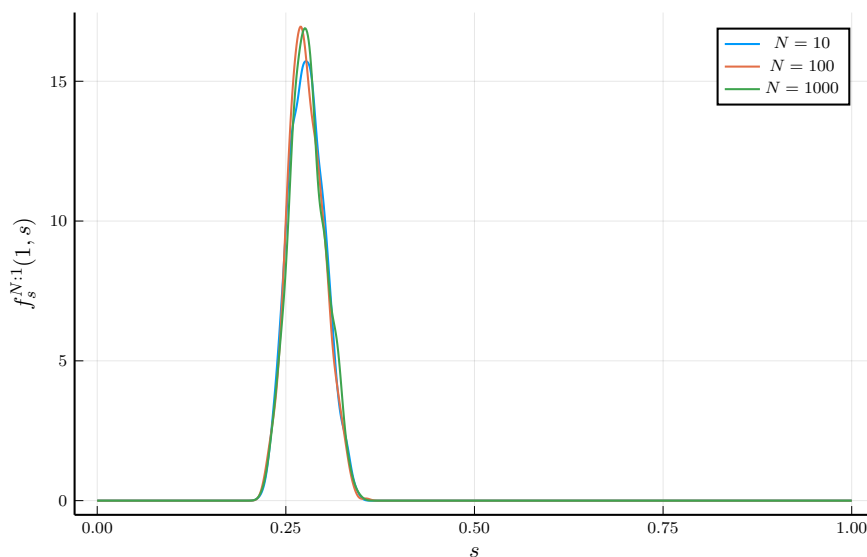
For $\varepsilon = 1/50$, the consistency error is around $\mathcal{E}(s_{1:N}, \tilde{s}_{1:N}^{1/50}) \approx 2 \times 10^{-2}$, which can be considered as an acceptable level of error. We keep this configuration of algorithm 1, i.e. $\varepsilon = 1/50$ and $N_{sp} = 10$ to simulate populations of increasing sizes N , in order to assess the gain in computation efficiency induced by the piecewise constant competition approximation.

Figure 1.11 shows that the computation time is reduced by a factor 10 for large populations ($N > 300$) when the simulation is performed with algorithm 1, which

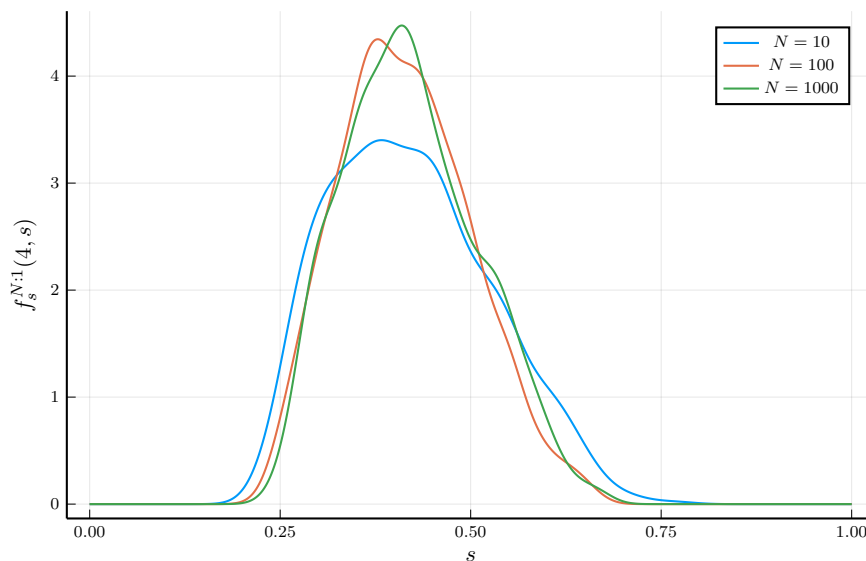
is a significant improvement. The computation time evolution is as expected in N^2 asymptotically due to the pairwise interaction creating the system dynamics. The methodology of approximating competition indices as a piecewise constant process is therefore adopted to approximate the marginal density of the size s for large populations.

Visualization of the marginal density of s

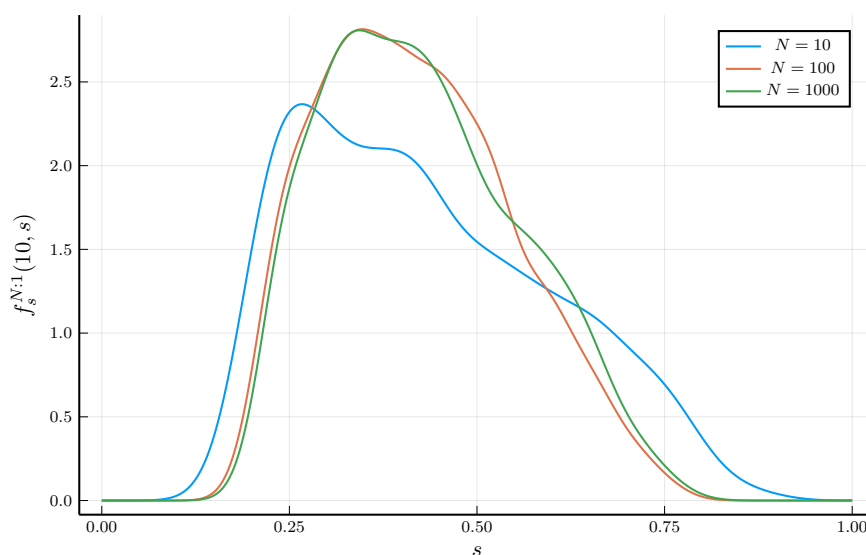
In this subsection, we use the methodology derived in subsection 1.4.5 to alleviate the computation of the density of the distribution $\mu_s^{N:1}[t]$ for large values of N . Figure 1.12 shows that there is very little difference between the densities corresponding to populations of sizes $N = 100$ and $N = 1\,000$, which may indicate that the distribution has a limit with respect to parameter N . This question is dealt with in detail in chapter 3. We can notice that the separability of the densities depends on the observation time t : the marginal density at $t = 1$ is roughly the same for all N , whereas density $f_s^{10:1}(10, \cdot)$ can be clearly distinguished from densities $f_s^{100:1}(10, \cdot)$ and $f_s^{1000:1}(10, \cdot)$ at $t = 10$. Figure 1.12 shows that the Schneider system exhibits a small population behaviour, where the size N may have an influence on the dynamics, and a large population behaviour, where the parameter N does not have a significant impact on individual trajectories. The problem of identifiability of parameter N is further discussed in chapter 2.



(a) $t = 1$



(b) $t = 4$



(c) $t = 10$

Figure 1.12: Kernel density estimation of the marginal density $s \mapsto f_s^{N:1}(t, s)$ from 1000 independent simulations of the Schneider system using algorithm 1 with $\varepsilon = 1/50$ and $N_{sp} = 10$. The shapes of the densities are compared for different values of the population size N . The values of the initial distribution parameters are in table 1.1. The values of the competition parameters are in table 1.2. The Julia package used for density estimation is `KernelEstimation.jl`, where it can be specified that the support of density $f_s^{N:1}(t, \cdot)$ is included within the segment $[s_m(s_0/s_m)^{\exp(-\gamma_M t)}; S_M(s_0/S_M)^{\exp(-\gamma_M t)}]$.

1.5 Conclusion

Chapter 1 introduced the notion of *symmetric and heterogeneous population* that can be applied to many systems where individuals are in interaction. Particular attention has been paid to a specific class of population models, which are necessarily symmetric because of their dependence with respect to the empirical population measure. This specific expression led us to develop simulation methodologies based on piecewise polynomial approximations of the empirical measure's trajectories. These simulation methods allow us to distinguish a level of population dynamics and a level of individual dynamics. They can be adapted to obtain numerical schemes of an arbitrarily large order (but of increasing computational cost), where the order is directly related to the degree of the polynomial approximation. This assertion has only been verified on the Schneider system, but it seems relatively straightforward to generalize this approach to other equally smooth systems (in chapter 2, this numerical scheme is applied to the Spring Cloud system). However, these methods' efficiency is enhanced when isolating in the dynamics population statistics summarizing the interaction between each individual and the population: the barycentre in the Spring Cloud system, the competition potential in the Schneider system, or the communication coefficients in the Cucker-Smale system. However, such statistics are more difficult to find for less smooth systems, e.g., Kuramoto-Vicsek, and an explicit Euler method seems to be more efficient in this case.

The consistency of these methodologies related to symmetric population deserves to be further investigated. On the theoretical plan, it is necessary to write a proof of consistency at a sufficiently generic level to be applied to a broader class of systems and compare it with state-of-the-art numerical methods. A parallel implementation should be carried out on the numerical plan to take full advantage of the partially distributed aspect of the numerical scheme, which allows, at least over a more or less short period, the individuals of a population in interaction to evolve in parallel. Since the complexity of the scheme with respect to the size of the population N is of the same order as an explicit Euler scheme (linear for Spring Cloud, quadratic for Schneider, for an order one scheme), this scheme may not present any computational advantage in situations where the time-step is too small. Favorable situations seem to be related to the temporal stability of this population statistic mentioned above. Nevertheless, the distinction between the population scale and the individual scale provides a unified framework used in chapter 3 to simulate the system's mean-field limit.

Bibliography

- [1] Baey, C. “Modelling inter-individual variability in plant growth models and model selection for prediction”. Theses. Ecole Centrale Paris, Feb. 2014.
- [2] Baey, C., Didier, A., Lemaire, S., Maupas, F., and Cournède, P.-H. “Modelling the interindividual variability of organogenesis in sugar beet populations using a hierarchical segmented model”. In: *Ecological Modelling* vol. 263 (2013), pp. 56–63.
- [3] Berger, U., Piou, C., Schiffrers, K., and Grimm, V. “Competition among plants: concepts, individual-based modelling approaches, and a proposal for a future research strategy”. In: *Perspectives in Plant Ecology, Evolution and Systematics* vol. 9, no. 3-4 (2008), pp. 121–135.
- [4] Beyer, R., Etard, O., Cournède, P. H., and Laurent-Gengoux, P. “Modeling spatial competition for light in plant populations with the porous medium equation”. In: *Journal of Mathematical Biology* vol. 70, no. 3 (2014), pp. 533–547.
- [5] Bolley, F., Canizo, J. A., and Carrillo, J. A. “Stochastic mean-field limit: non-Lipschitz forces and swarming”. In: *Mathematical Models and Methods in Applied Sciences* vol. 21, no. 11 (2011), pp. 2179–2210.
- [6] Brent, R. P. *Algorithms for minimization without derivatives*. Courier Corporation, 1973.
- [7] Clark, B. and Bullock, S. “Shedding light on plant competition: modelling the influence of plant morphology on light capture (and vice versa)”. In: *Journal of theoretical biology* vol. 244, no. 2 (2007), pp. 208–217.
- [8] Claverie, E. “Modélisation de la température du sol avec un bilan d’énergie, application à la prédiction de l’émergence du maïs (*Zea mays*)”. PhD thesis. Paris Saclay, 2018.
- [9] Cornell, S. J., Suprunenko, Y. F., Finkelshtein, D., Somervuo, P., and Ovaskainen, O. “A unified framework for analysis of individual-based models in ecology and beyond”. In: *Nature communications* vol. 10, no. 1 (2019), pp. 1–14.

-
- [10] Cournède, P.-H., Letort, V., Mathieu, A., Kang, M. Z., Lemaire, S., Trevezas, S., Houllier, F., and De Reffye, P. “Some parameter estimation issues in functional-structural plant modelling”. In: *Mathematical Modelling of Natural Phenomena* vol. 6, no. 2 (2011), pp. 133–159.
- [11] Cournède, P.-H., Guyard, T., Bayol, B., Griffon, S., De Coligny, F., Borianne, P., Jaeger, M., and De Reffye, P. “A forest growth simulator based on functional-structural modelling of individual trees”. In: *2009 Third International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications*. IEEE. 2009, pp. 34–41.
- [12] Cournède, P.-H., Mathieu, A., Houllier, F., Barthélémy, D., and De Reffye, P. “Computing competition for light in the GREENLAB model of plant growth: a contribution to the study of the effects of density on resource acquisition and architectural development”. In: *Annals of Botany* vol. 101, no. 8 (2008), pp. 1207–1219.
- [13] Cucker, F. and Smale, S. “On the mathematics of emergence”. In: *Japanese Journal of Mathematics* vol. 2 (2007), pp. 197–227.
- [14] Davidian, M. and Giltinan, D. M. “Some Simple Methods for Estimating Intraindividual Variability in Nonlinear Mixed Effects Models”. In: *Biometrics* vol. 49, no. 1 (1993), pp. 59–73.
- [15] DeAngelis, D. L. and Grimm, V. “Individual-based models in ecology after four decades”. In: *F1000prime reports* vol. 6 (2014).
- [16] Degond, P., Dimarco, G., and Mac, T. B. N. “Hydrodynamics of the Kuramoto–Vicsek model of rotating self-propelled particles”. In: *Mathematical Models and Methods in Applied Sciences* vol. 24, no. 02 (2014), pp. 277–325.
- [17] Della Noce, A., Carrier, M., and Cournède, P.-H. “Optimal control of non-smooth greenhouse models”. In: *International Symposium on Advanced Technologies and Management for Innovative Greenhouses: GreenSys2019 1296*. 2019, pp. 125–132.
- [18] Eloy, C., Fournier, M., Lacoïnte, A., and Moullia, B. “Wind loads and competition for light sculpt trees into self-similar structures”. In: *Nature communications* vol. 8, no. 1 (2017), pp. 1–12.
- [19] Feng, L., Mailhol, J.-C., Rey, H., Griffon, S., Auclair, D., and De Reffye, P. “Comparing an empirical crop model with a functional structural plant model to account for individual variability”. In: *European journal of agronomy* vol. 53 (2014), pp. 16–27.
- [20] Fisher, R. A. “The correlation between relatives on the supposition of Mendelian inheritance.” In: *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* vol. 52, no. 2 (1919), pp. 399–433.

-
- [21] Golse, F. “On the Dynamics of Large Particle Systems in the Mean Field Limit”. In: *arXiv preprint arXiv:1301.5494* (2013), pp. 1–144.
- [22] Harvey, D. G., Fletcher, A. G., Osborne, J. M., and Pitt-Francis, J. “A parallel implementation of an off-lattice individual-based model of multicellular populations”. In: *Computer Physics Communications* vol. 192 (2015), pp. 130–137.
- [23] Hemmerling, R., Kniemeyer, O., Lanwert, D., Kurth, W., and Buck-Sorlin, G. “The rule-based language XL and the modelling environment GroIMP illustrated with simulated tree competition”. In: *Functional plant biology* vol. 35, no. 10 (2008), pp. 739–750.
- [24] Le Gland, F. and Oudjane, N. “Stability and uniform approximation of nonlinear filters using the hilbert metric and application to particle filters”. In: *Annals of Applied Probability* vol. 14, no. 1 (2004), pp. 144–187.
- [25] Letort, V., Mahe, P., Cournède, P.-H., De Reffye, P., and Courtois, B. “Quantitative genetics and functional–structural plant growth models: simulation of quantitative trait loci detection for model parameters and application to potential yield optimization”. In: *Annals of botany* vol. 101, no. 8 (2008), pp. 1243–1254.
- [26] Li, B., Taniguchi, D., Gedara, J. P., Gogulancea, V., Gonzalez-Cabaleiro, R., Chen, J., McGough, A. S., Ofiteru, I. D., Curtis, T. P., and Zuliani, P. “NUFEB: A massively parallel simulator for individual-based modelling of microbial communities”. In: *PLoS computational biology* vol. 15, no. 12 (2019), e1007125.
- [27] Liu, W., Tollenaar, M., Stewart, G., and Deen, W. “Response of corn grain yield to spatial and temporal variability in emergence”. In: *Crop Science* vol. 44, no. 3 (2004), pp. 847–854.
- [28] Lv, Q., Schneider, M. K., and Pitchford, J. W. “Individualism in plant populations: using stochastic differential equations to model individual neighbourhood-dependent plant growth”. In: *Theoretical population biology* vol. 74, no. 1 (2008), pp. 74–83.
- [29] Mischler, S. “Introduction aux limites de champ moyen pour des systèmes de particules”. In: (Feb. 2011).
- [30] Mogensen, P. K. and Riseth, A. N. “Optim: A mathematical optimization package for Julia”. In: *Journal of Open Source Software* vol. 3, no. 24 (2018).
- [31] Nakagawa, Y., Yokozawa, M., and Hara, T. “Competition among plants can lead to an increase in aggregation of smaller plants around larger ones”. In: *Ecological Modelling* vol. 301 (2015), pp. 41–53.

-
- [32] Paine, C. T., Marthews, T. R., Vogt, D. R., Purves, D., Rees, M., Hector, A., and Turnbull, L. A. “How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists”. In: *Methods in Ecology and Evolution* vol. 3, no. 2 (2012), pp. 245–256.
- [33] Patra, D., Sachdev, M., and Subbiah, B. “15 N studies on the transfer of legume-fixed nitrogen to associated cereals in intercropping systems”. In: *Biology and fertility of soils* vol. 2, no. 3 (1986), pp. 165–171.
- [34] Plimpton, S. *Fast parallel algorithms for short-range molecular dynamics*. Tech. rep. Sandia National Labs., Albuquerque, NM (United States), 1993.
- [35] Rackauckas, C. and Nie, Q. “DifferentialEquations.jl—a performant and feature-rich ecosystem for solving differential equations in julia”. In: *Journal of Open Research Software* vol. 5, no. 1 (2017).
- [36] Schneider, M. K., Law, R., and Illian, J. B. “Quantification of Neighbourhood-Dependent Plant Growth by Bayesian Hierarchical Modelling”. In: *Journal of Ecology* vol. 94, no. 2 (2006), pp. 310–321.
- [37] Sievänen, R., Perttunen, J., Nikinmaa, E., and Kaitaniemi, P. “Toward extension of a single tree functional–structural model of Scots pine to stand level: effect of the canopy of randomly distributed, identical trees on development of tree structure”. In: *Functional Plant Biology* vol. 35, no. 10 (2008), pp. 964–975.
- [38] Strand, J. “Random ordinary differential equations”. In: *Journal of Differential Equations* vol. 7, no. 3 (1970), pp. 538–553.
- [39] Trevezas, S. and Cournède, P.-H. “A sequential Monte Carlo approach for MLE in a plant growth model”. In: *Journal of Agricultural, Biological, and Environmental Statistics* vol. 18, no. 2 (2013), pp. 250–270.
- [40] Tsitouras, C. “Runge–Kutta pairs of order 5 (4) satisfying only the first column simplifying assumption”. In: *Computers & Mathematics with Applications* vol. 62, no. 2 (2011), pp. 770–775.
- [41] Weigelt, A. and Jolliffe, P. “Indices of plant competition”. In: *Journal of ecology* (2003), pp. 707–720.

Chapter 2

Uncertainties in symmetric population systems

Statistical uniformity is by no means a harmless scientific ideal; it is the no longer secret political ideal of a society which, entirely submerged in the routine of everyday living, is at peace with the scientific outlook inherent in its existence.

–Hannah Arendt, THE HUMAN CONDITION (1958)

2.1 Introduction

Symmetric population models, as defined in the previous chapter, have been used extensively to model biological systems of great complexity. We have already mentioned the Cucker-Smale model (Cucker and Smale, 2007⁸) or Kuramoto-Vicsek model (Degond et al., 2014⁹), which were applied to collective motions of social animals. These biological populations are complex systems, since their elementary building blocks, i.e., the individuals that constitute them are themselves complex systems that still escape any theory or formalism¹. Individuals interact with each other, but they also interact with the environment outside the system, which is seldom closed. As a consequence, any population model that is to be compared with experimental data has to make a series of assumptions to account for the sources of uncertainty that may disturb the primary mechanism under investigation. When experimental data are analysed at a global scale, one can assume that uncertainties at the microscopic scale can be neglected when individual data are aggregated at a macroscopic scale. In Ballerini et al. (2008),¹ a simple alignment model enabled to evidence the topological nature of interactions within a swarm of starlings. Interactions are said to be topological when the behaviour of an individual is influenced by its k nearest neighbours, in contrast with a geometrical

¹And maybe it should be left that way.

influence coming from its conspecifics located in a disk of a given radius. The model selected for individual motion only represents the alignment phenomenon, without any other source of uncertainty, and it is left to the experimental data to determine whether interactions are topological or geometrical. Neglecting microscopic uncertainty might be justified by the large size of the population observed in the dataset, and by the remarkable accuracy of the experimental protocol used for data collection (Ballerini et al., 2008²). Mann (2011)¹⁹ noticed that many interaction rules, although having singularly dissimilar expressions, lead to the emergence of the same structure or patterns in the collective motion, such as global alignments or rotating mills (studied qualitatively and numerically in Vecil et al., 2013²²). This observation led the author to put forward the thesis that in order to precisely identify the parameters of a specific model, or to decide between two competing models, only observations made at individual scale and during a transitory regime preceding a state of dynamic equilibrium, with potentially universal properties, would be relevant. The author suggests a Bayesian methodology to discriminate between topological or geometrical interaction laws, based on the observation of individual trajectories. Uncertainties on the microscopic dynamics are represented as Gaussian angular perturbations. The formalism of adding stochastic perturbations to the state derivative is a standard modelling practice in the literature (Bolley et al., 2011,⁴ Degond et al., 2014,⁹ Lv et al, 2008¹⁸) to account for the potentially unknown mechanism at the level of the individuals. A notable exception is made in Carrillo et al. (2017)⁶ where the uncertainties on the trajectories result from the uncertainty on parameters driving the interaction force.

Nevertheless, the uncertainties related to the shortcomings of the experimental protocol seem to be overlooked in the literature on statistical inference of kinetic systems and interacting population models in general. The complexity of the dynamics, the potentially large size of the population, makes an exact knowledge of the system out of the reach of any data collection procedure. In the case of symmetric and heterogeneous population models, an exact knowledge of the system presupposes the data of all the initial conditions and individual characteristics, as well as a perfect knowledge of the transition function. Often, in many works, only the parameters to be estimated are assumed to be unknown, while the rest of the variables needed to simulate the system are fixed. For instance, in Bialek et al. (2012),³ the velocities of all the individuals in the population are assumed to be measured precisely at a high frequency. In Bongini et al. (2017),⁵ the interaction kernel is learnt using a non-parametric method based on the exact trajectories of the whole system. A similar type of data is required for the methodology used in Lu et al. (2019).¹⁷ According to the context, such accuracy in observations might be too costly to be achieved. It is thus necessary to include in the model latent variables to complete the partial knowledge brought by the data (Dempster et

al., 1977¹⁰). Resorting to hierarchical models seem to be unavoidable in the case of heterogeneous population models, where the individual characteristics are not directly measurable, but only accessible through their influence on the overall dynamics (Schneider et al., 2006²⁰). In the Bayesian setting, the uncertainties related to these latent variables naturally propagate into uncertainties in the estimation of the parameters. In the case of symmetric population models, we can enumerate the sources of uncertainty exhaustively:

- the uncertainty on the population size: if the experimental protocol only concerns a subgroup of individuals, with very few information on the rest of the population, the interaction network defining the dynamics cannot be fully determined. It is worth noting that this uncertainty only affects parametric identification if individuals interact.
- the uncertainty on the initial configuration of the population: this may be related either to the initial conditions or to the individual characteristics.
- the uncertainty on individual identities: this uncertainty arises especially when individuals are hard to distinguish and when the population is observed with a low frequency, and it leads to difficulty in the reconstitution of individual trajectories.
- the uncertainty due to the inexact integration of the system: in practice, the simulation is carried via a numerical solver, whose deviation from the real trajectories may increase over time. This uncertainty can hardly be neglected in the case of large populations, for which accurate numerical integrations have prohibitive computational costs.

The chapter is organised as follows: section 2.2 introduces the general principle of Bayesian inference, and the next sections elaborates on the modelling of the different sources of uncertainties, along with their consequences on parametric identification. For each source of uncertainty, an illustration is made on different configurations of Spring Cloud model, introduced in example 1.6. This linear system enables exact visualizations of parameters uncertainties, without having to resort to numerical methods for the inference. Moreover, it provides some insights on the difficulties that are encountered when dealing with nonlinear systems. Besides, the different sources of uncertainty are considered separately: when one aspect of the model is assumed to be unknown, the rest of potentially latent variables are assumed to be known. This type of hypothetical situation can find a practical justification in Gibbs' sampler inference method (Geman and Geman, 1993¹¹).

2.2 Bayesian inference

The general principle of Bayesian inference¹³ can be understood by the description of its typical workflow. Initially, we want to carry out an experiment or a data collection to identify the value of an unknown parameter ξ . Before the experiment, i.e., before any data is acquired, we have a more or less clear idea of the possible values that ξ can take. The assumption on the possible values of ξ is formalised as a probability distribution \mathbb{P}_ξ , called the prior distribution, assigning to each possible value a certain probability, which can be interpreted as a quantification of the plausibility of that value. Besides, we have built a model explaining the relation between the collected observations x and the parameter ξ . Such a relation is represented by a collection of probability distributions $\mathbb{P}_{x|\xi}$, indexed by ξ , and called the likelihood distribution. The observation of x may lead to favouring specific values of ξ . The modification of the plausibility landscape of ξ by the observation is formalised by the posterior distribution $\mathbb{P}_{\xi|x}$, which is a conditional distribution (see definition 2.2), and whose expression is given by Bayes' formula (see theorem 2.1).

Definition 2.1. Transition kernel and Markov kernel (from Klenke, 2008,¹⁵ definition 8.24)

Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ and $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ two measurable spaces. A function $\nu : \mathcal{X} \times \mathcal{B}_\mathcal{Y} \rightarrow \mathbb{R}_+$ is said to be a transition kernel if:

1. for all set $B \in \mathcal{B}_\mathcal{Y}$, the function $x \in \mathcal{X} \mapsto \nu(x, B)$ is measurable;
2. for all $x \in \mathcal{X}$, the function $B \in \mathcal{B}_\mathcal{Y} \mapsto \nu(x, B)$ is a σ -finite measure.

In addition, if a transition kernel is such that $\forall x \in \mathcal{X}$, $\nu(x, \cdot)$ is a probability transition, then the transition kernel is said to be a Markov kernel.

The set of transition kernels associating an element of \mathcal{X} to a σ -finite measure over $\mathcal{B}_\mathcal{Y}$ is denoted by $\mathcal{T}(\mathcal{X} \rightarrow \mathcal{B}_\mathcal{Y})$.

Definition 2.2. Conditional distribution

Let $\mathbb{P} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ and $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ being two measurable spaces. Let \mathbb{P}_y be the y -marginal distribution:

$$\forall h \in \mathcal{M}_b(\mathcal{Y} \rightarrow \mathbb{R}), \quad \mathbb{E}_{\mathbb{P}_y}[h(y)] = \int_{\mathcal{X} \times \mathcal{Y}} h(y) \mathbb{P}(dx, dy) \quad (2.1)$$

A transition kernel $\nu \in \mathcal{T}(\mathcal{Y} \rightarrow \mathcal{B}_\mathcal{X})$ is said to be a $x \mid y$ conditional distribution if it satisfies the following relation

$$\forall h \in \mathcal{M}_b(\mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}), \quad \mathbb{E}_{\mathbb{P}}[h(x, y)] = \int_{\mathcal{Y}} \int_{\mathcal{X}} h(x, y) \nu(y, dx) \mathbb{P}_y(dy) \quad (2.2)$$

It is important to notice in definition (2.2) that the conditional distribution extracted from a given joint distribution may not be uniquely defined. In particular, if $\nu \in \mathcal{T}(\mathcal{Y} \rightarrow \mathcal{B}_y)$ is a conditional distribution, then any other ν' such that $\mathbb{P}_y(\nu'(y, \cdot) = \nu(y, \cdot)) = 1$ is also a $x | y$ conditional distribution. In other words, a conditional distribution has constrained values over the support of the marginal distribution of the remaining variables; elsewhere, it can take any value.

Theorem 2.1. Bayes' formula (from Shao, 2003,²¹ theorem 4.1)

Let $(\mathcal{X}, \mathcal{B}_x)$ and $(\mathcal{Y}, \mathcal{B}_y)$ two measurable spaces. Let ν_x and ν_y be two σ -finite measures on \mathcal{X} and \mathcal{Y} respectively. Let $\mathbb{P}_x \in \mathcal{P}(\mathcal{X})$ be a probability measure absolutely continuous with respect to ν_x , with $\pi : x \in \mathcal{X} \mapsto \frac{\partial \mathbb{P}_x}{\partial \nu_x}(x) \in \mathbb{R}_+$. Let $(x, B) \in \mathcal{X} \times \mathcal{B}_y \mapsto \mathbb{P}_{y|x}(B|x)$ a Markov kernel in $\mathcal{T}(\mathcal{X} \rightarrow \mathcal{B}_y)$ such that for all $x \in \mathcal{X}$, the probability measure $\mathbb{P}_{y|x}(\cdot | x)$ is absolutely continuous with respect to ν_y , with $(x, y) \in \mathcal{X} \times \mathcal{Y} \mapsto f(y|x) = \frac{\partial \mathbb{P}_{y|x}(\cdot | x)}{\partial \nu_y}(y)$. We assume that the function $(x, y) \in (\mathcal{X} \times \mathcal{Y}, \sigma(\mathcal{B}_x \times \mathcal{B}_y)) \mapsto f(y|x) \in (\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$ is measurable. Then:

1. the distribution $\mathbb{P}_y : B \in \mathcal{B}_y \mapsto \int_{\mathcal{X}} \mathbb{P}_{y|x}(B|x) \mathbb{P}_x(dx)$ is absolutely continuous with respect to the measure ν_y . Let $g_y : y \in \mathcal{Y} \mapsto \frac{\partial \mathbb{P}_y}{\partial \nu_y}(y) = \int_{\mathcal{X}} f(y|x) \pi(x) \nu_x(dx)$.

2. the transition kernel

$$(y, A) \in \mathcal{Y} \times \mathcal{B}_x \mapsto \mathbb{P}_{x|y}(A | y) = \begin{cases} 0 & \text{if } g_y(y) = 0 \\ \frac{1}{g_y(y)} \int_A f(y|x) \pi(x) \nu_x(dx) & \text{otherwise} \end{cases} \quad (2.3)$$

is a $x | y$ conditional distribution of the distribution $\mathbb{P}_{x,y}$ defined by its action on measurable and bounded functions

$$\forall h \in \mathcal{M}_b(\mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}), \quad \mathbb{E}_{\mathbb{P}_{x,y}}[h(x, y)] = \int_{\mathcal{X}} \int_{\mathcal{Y}} h(x, y) \mathbb{P}_{y|x}(dy|x) \mathbb{P}_x(dx) \quad (2.4)$$

We have now all the elements to define what a Bayesian inference problem is. Let $\mathbb{P}_\xi \in \mathcal{P}(\Xi)$ be a probability distribution over a set of parameters Ξ . This is the prior distribution, modelling the assumptions and the prior knowledge we have on the system under study. Let $\xi \in \Xi \mapsto \mathbb{P}_{x|\xi}(\cdot | \xi) \in \mathcal{P}(\mathcal{X})$ a Markov kernel, i.e., the likelihood distributions defined over the space of observations \mathcal{X} , giving the probability for a model parametrised by ξ to generate the observation

x . The Bayesian inference problem consists of determining the posterior distribution $\mathbb{P}_{\xi|x}(\cdot | x)$ given by equation (2.3) at the specific point x constituting the data set of observations. In most cases, the posterior distribution cannot be exactly determined, mainly because of the intractability of the integral defining the marginal distribution $\mathbb{P}_x(dx) = \int_{\Xi} \mathbb{P}_{x|\xi}(dx | \xi) \mathbb{P}_{\xi}(d\xi)$, also called the evidence in the Bayesian terminology (term g_y in equation 2.3).

2.3 Application to a symmetric system: Spring Cloud model with identical particles

Bayesian setting is particularly useful when dealing with models where the parameters are difficult to identify. Let us come back to the example 1.6 of the model Spring Cloud. To simplify the system, let us assume for now that all the particles have the same parameter κ and the same mass m . Our goal is to identify κ and m from the observation of the trajectories of the different particles in the population. Prior knowledge on the system tells us that the unknown variables of our problem are within known intervals $\kappa \in [\kappa_{\min}; \kappa_{\max}]$ and that $m \in [m_{\min}; m_{\max}]$, so a natural choice for prior distribution would be a uniform distribution over $[\kappa_{\min}; \kappa_{\max}] \times [m_{\min}; m_{\max}]$ of density

$$p_{\kappa,m} : (\kappa, m) \in \mathbb{R}^2 \mapsto \frac{\mathbb{I}\{\kappa_{\min} \leq \kappa \leq \kappa_{\max}\} \mathbb{I}\{m_{\min} \leq m \leq m_{\max}\}}{(\kappa_{\max} - \kappa_{\min})(m_{\max} - m_{\min})} \quad (2.5)$$

When the population system is homogeneous, i.e., when all the particles have the same characteristics, the equation of motion (1.38) becomes

$$\begin{aligned} \forall i \in \llbracket 1; N \rrbracket, \quad \frac{d^2 x_i}{dt^2}(t) &= -\frac{N\kappa^2}{m} x_i(t) + \frac{\kappa^2}{m} \sum_{j=1}^N x_j(t) \\ \text{Besides } \frac{d^2}{dt^2} \sum_{j=1}^N x_j(t) &= \frac{\kappa^2}{m} \sum_{j=1}^N \sum_{k=1}^N (x_k(t) - x_j(t)) = 0 \\ \text{so } \sum_{j=1}^N x_j(t) &= \sum_{j=1}^N x_j^0 + t \sum_{j=1}^N v_j^0 \end{aligned} \quad (2.6)$$

We obtain that the trajectory of each particle is given by

$$\forall i \in \llbracket 1; N \rrbracket, \quad x_i(t) = \left(x_i^0 - \frac{1}{N} \sum_{j=1}^N x_j^0 \right) \cos(\omega_N t) + \frac{1}{\omega_N} \left(v_i^0 - \frac{1}{N} \sum_{j=1}^N v_j^0 \right) \sin(\omega_N t) + \frac{1}{N} \sum_{j=1}^N x_j^0 + \frac{t}{N} \sum_{j=1}^N v_j^0 \quad \text{with } \omega_N = \kappa \sqrt{\frac{N}{m}} \quad (2.7)$$

The trajectories of the particles are elliptic around the barycenter of the cloud, following a uniform and rectilinear motion given by its initial speed. It can be noticed in equation (2.7) that the size N of the population has a direct influence on the pulsation ω_N of the trajectories. The expression of the trajectory also shows that, for any positive factor $\lambda > 0$, the configuration of parameters $(\lambda\kappa, \lambda^2 m)$ would lead to the exact same trajectories as the configuration (κ, m) . We are in the case where the model is not identifiable with respect to the parameters (κ, m) , in the sense that $\mathbb{P}_{x|\kappa, m} = \mathbb{P}_{x|\kappa', m'}$ does not imply that $(\kappa, m) = (\kappa', m')$.

A common choice to model observation errors is the normal or Gaussian distribution. For convenience, we do not distinguish *regular* and *degenerate* Gaussian distribution.

Lemma 2.1. Square root of a positive semi-definite matrix (from Horn and Johnson, 2012,¹⁴ theorem 7.2.6)

Let $A \in \mathcal{M}_d(\mathbb{R})$ be a positive semi-definite matrix. Then there exists a unique $B \in \mathcal{M}_d(\mathbb{R})$ positive semi-definite such that $B^2 = A$. We then use the notation $B = A^{1/2}$.

Definition 2.3. Gaussian distribution

The standard normal distribution over \mathbb{R}^d is the probability measure $\mathcal{N}(0, \mathbf{I}_d)$ absolutely continuous with respect to the Lebesgue measure, whose density is

$$\forall x \in \mathbb{R}^d, \quad \frac{\partial \mathcal{N}(0, \mathbf{I}_d)}{\partial \lambda}(x) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{|x|^2}{2}\right) \quad (2.8)$$

where $|x|$ is the canonical Euclidean norm over \mathbb{R}^d . For any $m \in \mathbb{R}^d$, and for any symmetric matrix $\Sigma \in \mathcal{M}_d(\mathbb{R})$ positive semi-definite, the distribution $\mathcal{N}(m, \Sigma)$ is the law of $m + \Sigma^* u$, for any $u \sim \mathcal{N}(0, \mathbf{I}_d)$ and for any square root Σ^* of the matrix Σ , i.e., satisfying $(\Sigma^*)^T \Sigma^* = \Sigma$.

In particular, any Dirac distribution is Gaussian $\delta_m = \mathcal{N}(m, 0)$ and when the covariance matrix Σ is positive definite, the distribution $\mathcal{N}(m, \Sigma)$ is absolutely

continuous with respect to the Lebesgue measure, with the density

$$\forall x \in \mathbb{R}^d, \quad \frac{\partial \mathcal{N}(m, \Sigma)}{\partial \lambda}(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x - m)^\top \Sigma^{-1}(x - m)\right) \quad (2.9)$$

Let us assume that we are able to observe all the initial positions of the particles at time $t = 0$. We observe also the positions of all the particles at times t_1, \dots, t_M , but this time with some observation error modelled by a bivariate Gaussian variable of covariance matrix $\sigma^2 I_2$, where $\sigma > 0$ is known. The observation model, explaining the relation between the parameters (κ, m) and the data x can be formulated as:

$$\forall i \in \llbracket 1; N \rrbracket, \forall j \in \llbracket 1; M \rrbracket, x_{ij} = x_i(t_j, \kappa, m) + \epsilon_{ij} \text{ where } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2 I_2) \quad (2.10)$$

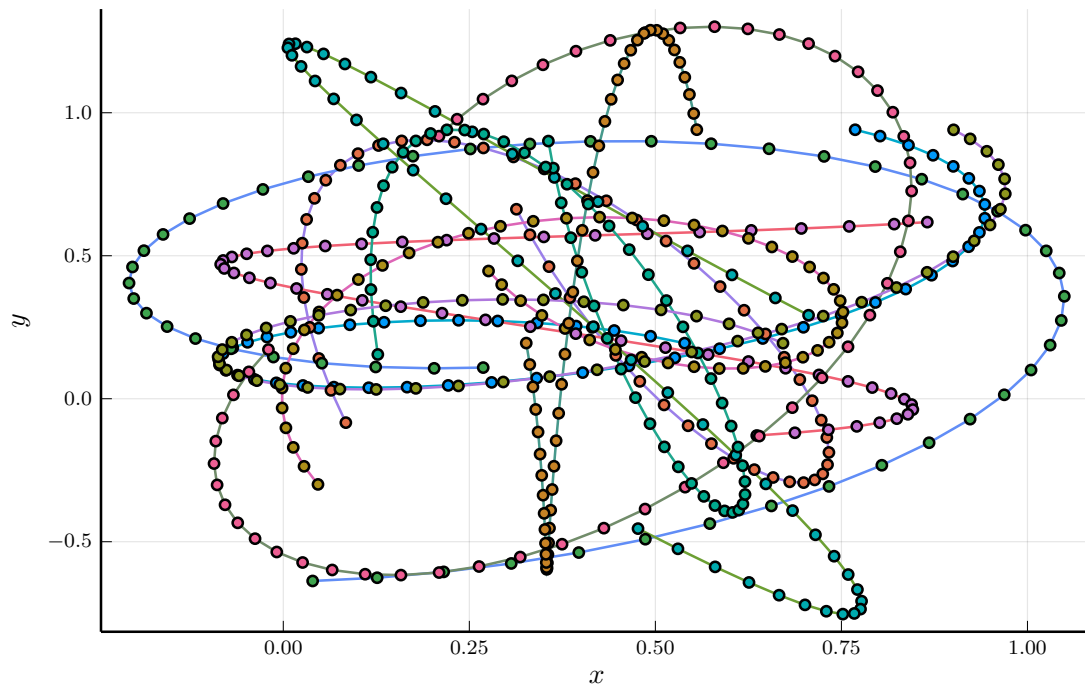
where $x_i(t, \kappa, m)$ is the trajectory of the particle i given by equation (2.7). It follows that the density of the likelihood distribution is, for any $(\kappa, m) \in [\kappa_{\min}; \kappa_{\max}] \times [m_{\min}; m_{\max}]$.

$$\forall x \in \mathbb{R}^{2Nm}, \quad p_{x|\kappa, m}(x | \kappa, m) = \frac{1}{(2\pi\sigma^2)^{NM}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{j=1}^M |x_{ij} - x_i(t_j, \kappa, m)|^2\right) \quad (2.11)$$

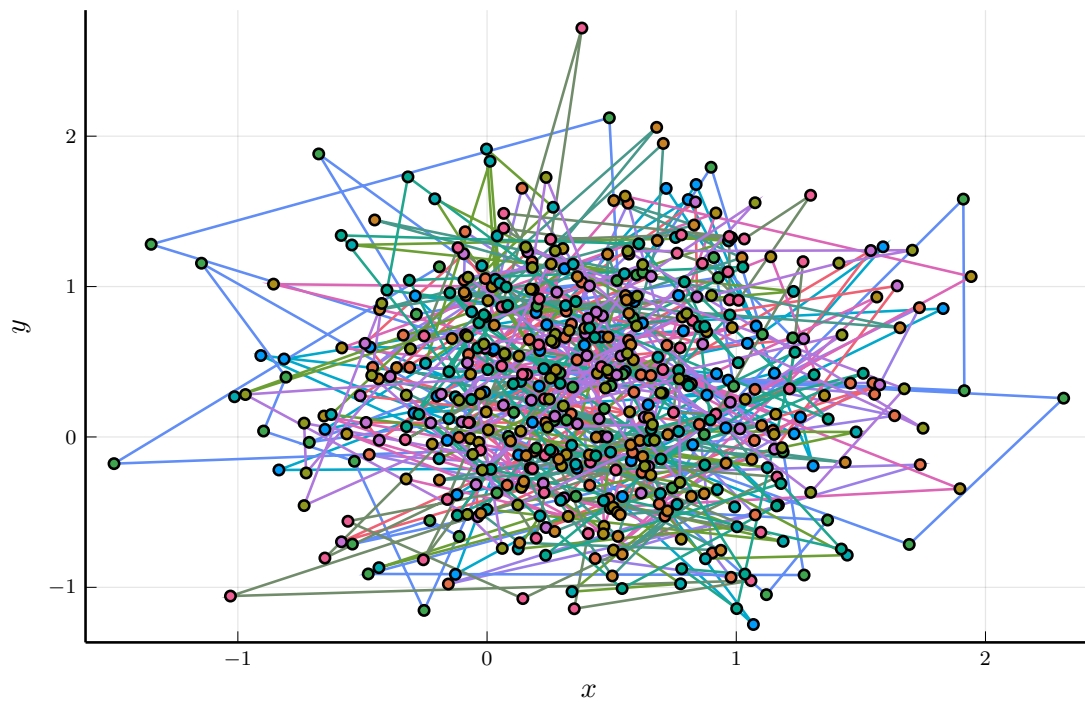
From Bayes' formula (theorem 2.1), the posterior distribution $\mathbb{P}_{\kappa, m|x}$ has a probability density function $p_{\kappa, m|x}$, known up to a normalisation factor.

$$\begin{aligned} p_{\kappa, m|x}(\kappa, m | x) &\propto p_{x|\kappa, m}(x | \kappa, m) p_{\kappa, m}(\kappa, m) \\ &\propto \begin{cases} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{j=1}^M |x_{ij} - x_i(t_j, \kappa, m)|^2\right) & \text{if } (\kappa, m) \in [\kappa_{\min}; \kappa_{\max}] \times [m_{\min}; m_{\max}] \\ 0 & \text{otherwise} \end{cases} \\ &\propto \tilde{p}(\kappa, m | x) \end{aligned} \quad (2.12)$$

We cannot compute the integral $\int_{\kappa_{\min}}^{\kappa_{\max}} \int_{m_{\min}}^{m_{\max}} \tilde{p}(\kappa, m) dm d\kappa$ exactly, but we can represent the surface $(\kappa, m) \mapsto \tilde{p}(\kappa, m | x)$ to see how the data x refines the knowledge modelled by the prior distribution of equation (2.5). Let us consider a data set x corresponding to the observations of a population of 10 particles, having their initial positions in $[0; 1]^2$ and their initial velocities in $[-1; 1]^2$. Their trajectories are generated using the parameters $\kappa_0 = 1$ and $m_0 = 10$, and they are observed over a timeline constituted of 50 regularly space time steps between $t = 0$ and $t = 2\pi$, which corresponds to a rotation period in the barycentric referential. The observation error $\sigma = 0.5$ is chosen very large to show the robustness of the inference. Figure 2.1 compares the trajectories



(a) $\sigma = 0$



(b) $\sigma = 0.5$

Figure 2.1: (a) Representation of the trajectories without observation error (equation 2.7). (b) Scatter plot of the data set \mathcal{Y}_9 (with a large observation error, equation 2.10)

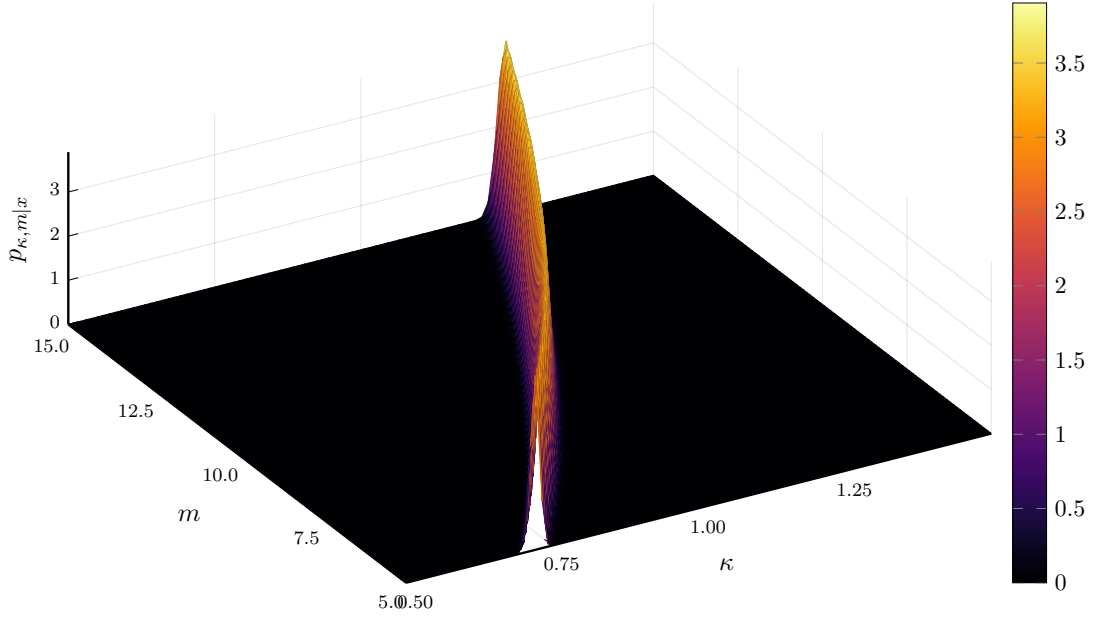


Figure 2.2: Surface $(\kappa, m) \in [\kappa_{\min}; \kappa_{\max}] \times [m_{\min}; m_{\max}] \mapsto \frac{\tilde{p}(\kappa, m | x)}{\tilde{p}(\kappa_0, m_0 | x)}$ for $\sigma = 0.5$, where $\tilde{p}(\kappa, m | x)$ is defined by the equation (2.12), and $(\kappa_0, m_0) = (1, 10)$.

of the individual particles with their noisy observations. We can notice in figure (b) that the error completely break the structure of the motion, that clearly appears in (a). However, despite a standard deviation of $\sigma = 0.5$, figure 2.2 shows that the posterior distribution is concentrated near (but not on) the curve $\left\{ (\kappa, m) \in [\kappa_{\min}; \kappa_{\max}] \times [m_{\min}; m_{\max}] \mid \kappa \sqrt{\frac{N}{m}} = \kappa_0 \sqrt{\frac{N}{m_0}} = 1 \right\}$. For a fixed M , if σ decreases, the posterior distribution converges to the uniform distribution over this curve. The same convergence is observed when σ is fixed and when M , the number of observations per individuals, increases. This last convergence in distribution is yet to be proved.

2.4 Undifferentiated population

Let us come back to the example of subsection 2.3 and let us relax the assumption on our ability to track the different individuals. At a given time, we observe a cloud of points, and, when considering a point at a time t_1 and another point at a time t_2 , we are not able to say whether these points correspond to the different positions of the same particle, or they do not represent the same particle. We

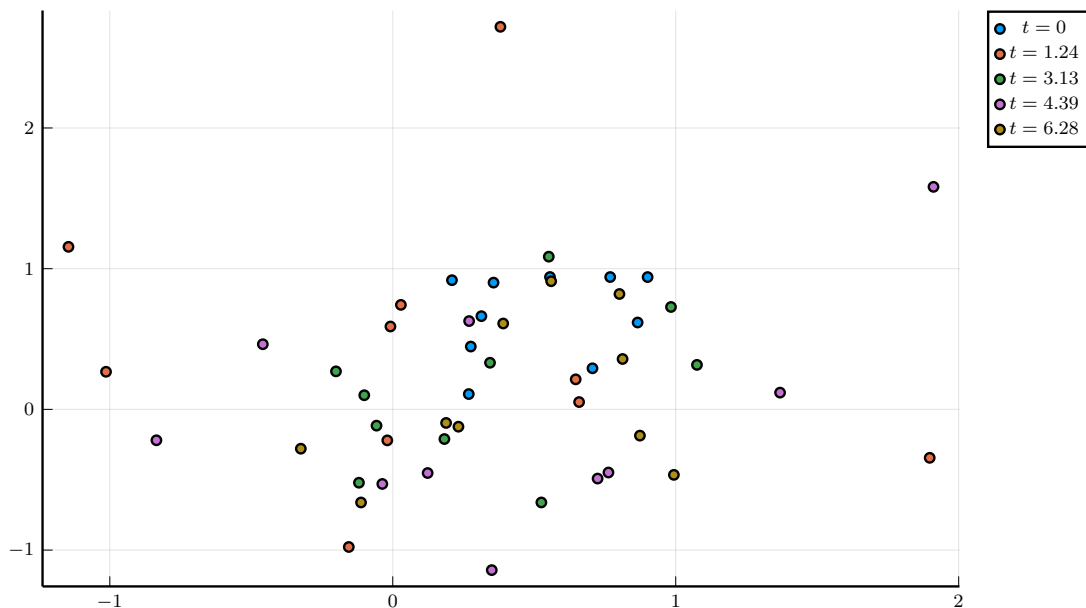


Figure 2.3: Scatter plots of the population of particles at different observation times. Data is generated by the Spring Cloud model (equation 2.10) for a population of 10 particles.

do not have individual observations anymore; we only have observations of the whole population. Figure 2.3 shows a visualization of the data set, consisting of 10 particles following the same observation model as in equation (2.10), with a superimposition of the cloud of points at different observation times. Such a situation may occur when studying a population composed of individuals hard to distinguish, which is the case for the data collected and studied in Ballerini et al. (2008)¹ displaying the evolution of a swarm of starlings. The absence of individual trajectories in the data set constitutes a significant loss of information, impacting the observation model itself. In the case of subsection 2.3, a square loss measures the fit between the model and the data, appearing within the expression of the likelihood function (equation 2.11). This loss function, adapted to the comparison between two lists of points, corresponds to a Gaussian model on the observation errors (equation 2.10). In the undifferentiated case, the pairing between model points and observation points is no longer part of the problem. Let us introduce an arbitrary indexing of the observation points so that the data have the form of a collection of vector $x = (x_{ij})_{\substack{1 \leq i \leq N \\ 1 \leq j \leq M}} \in \mathbb{R}^{2Nm}$ as previously. As this indexing is arbitrary, the model generating the data formulates it as a random permutation,

chosen uniformly in \mathfrak{S}_N , the set of permutations of $\llbracket 1; N \rrbracket$.

$$\begin{aligned} \forall j \in \llbracket 1; M \rrbracket, \tau_j &\sim \mathcal{U}(\mathfrak{S}_N) \text{ defined by } \forall \sigma \in \mathfrak{S}_N, \mathbb{P}\{\tau \sim \mathcal{U}(\mathfrak{S}_N) \mid \tau = \sigma\} = \frac{1}{N!} \\ \forall i \in \llbracket 1; N \rrbracket, x_{ij} &= x_{\tau_j(i)}(t_j, \omega) + \epsilon_{ij} \text{ with } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2 I_2) \end{aligned} \quad (2.13)$$

As we only focus on the uncertainty due to individual tracking, we parametrise the problem by $\omega = \kappa \sqrt{N/m}$. This new model does not have any structural identification problem anymore. The observation model 2.13 implies the following expression for the likelihood distribution.

$$\forall x \in \mathbb{R}^{2NM}, \quad p_{x|\omega}(x \mid \omega) = \frac{1}{(2\pi\sigma^2)^{NM}} \prod_{j=1}^M \left(\frac{1}{N!} \sum_{\tau \in \mathfrak{S}_N} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^N |x_{ij} - x_{\tau(i)}(t_j, \omega)|^2 \right) \right) \quad (2.14)$$

The posterior distribution is then the distribution having a density proportional to $\tilde{p}_{\omega|x}$ where $\tilde{p}_{\omega|x}$ is defined by

$$\tilde{p}_{\omega|x}(\omega \mid x) = \prod_{j=1}^M \left(\frac{1}{N!} \sum_{\tau \in \mathfrak{S}_N} \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^N |x_{ij} - x_{\tau(i)}(t_j, \omega)|^2 \right) \right) \times \mathbb{I}\{\omega_{\min} \leq \omega \leq \omega_{\max}\} \quad (2.15)$$

In the above expression, the prior on the parameter ω is chosen to be the uniform distribution on the interval $[\omega_{\min}; \omega_{\max}]$, with $0 < \omega_{\min} < \omega_{\max}$. We represent in figure 2.4 the posterior distribution in equation (2.15) in the case of a population of 5 particles and for data generated with the parameter $\omega_0 = 0.707$. The choice of a population with few particles is justified by the computational cost involved in averaging over the set of permutations \mathfrak{S}_N when N is higher than 10. We can notice that the uncertainty due to the absence of individual tracking blurs the knowledge of the pulsation ω . The region corresponding to the posterior's highest mode is wider in the undifferentiated case ($\omega \mid x$ has a larger variance) than in the differentiated case. We can even see a second mode's appearance in the undifferentiated posterior graph, located around $\omega = 1.25$, corresponding to a faster motion of the particles. This last scenario emerges in the posterior because of the low frequency of observations (10 observations over $[0; 2\pi]$), which does not prevent to exclude this case.

2.5 Uncertainty on the initial condition

In the previous example, we have assumed an exact knowledge of the initial configuration of the population. This assumption is unrealistic as there is no specific

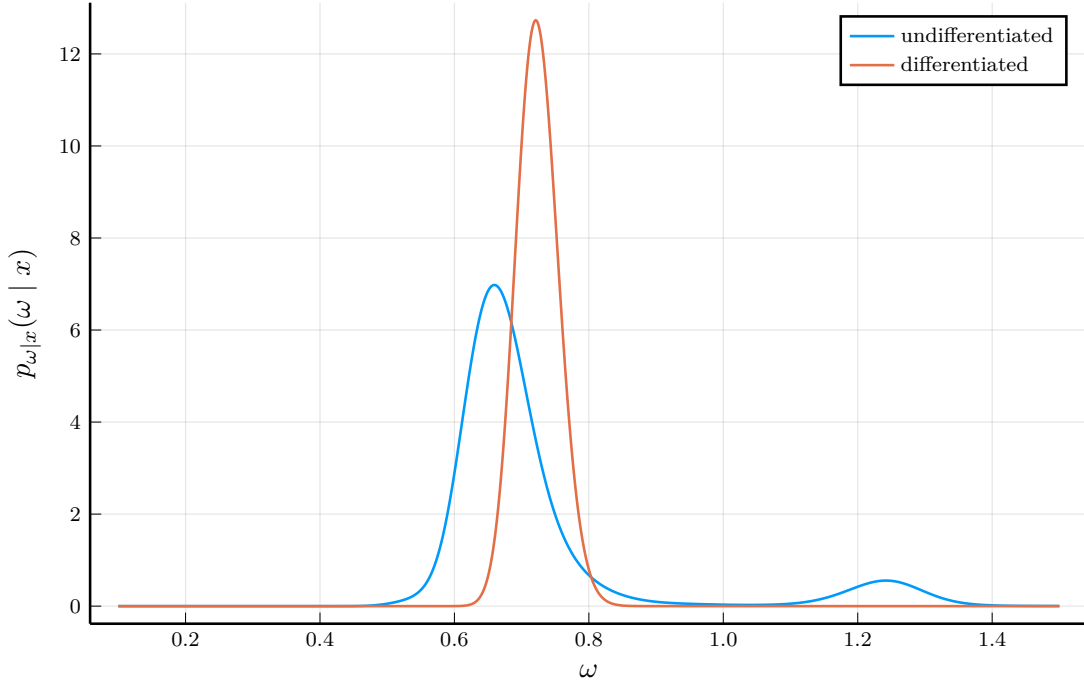


Figure 2.4: Comparison of the posterior densities in two cases: when the individuals are exactly tracked (red line) and when the individual trajectories are unknown (blue line)

reason to believe that time $t = 0$ has such a singularity. What are the consequences we can observe on the posterior distribution if we only assume a partial knowledge of the initial conditions? Let $x \in \mathbb{R}^{2N(m+1)}$ be our data set on Spring Cloud system (equation 2.7), containing observations from the individual trajectories (with distinguished individuals) and the observation of the initial position, which is corrupted by the same observation error as in equation (2.10). The initial velocities of the particles are not observed. The initial configuration of the population is assumed to be generated from a known probability distribution $\mu_0 \in \mathcal{P}(\mathbb{R}^4)$ such that $(x_i^0, v_i^0)_{1 \leq i \leq N} \sim \mu_0^{\otimes N}$. As an illustrative example, we can consider μ_0 to be a Gaussian distribution.

$$\forall i \in \llbracket 1; N \rrbracket, \quad x_i^0 \sim \mathcal{N}(\mu_x, \Sigma_x), \quad v_i^0 \sim \mathcal{N}(0, \Sigma_v) \quad (2.16)$$

We can write the observation model conditionally to the initial configuration $(x_i^0, v_i^0)_{1 \leq i \leq N} \sim \mu_0^{\otimes N}$

$$\forall (i, j) \in \llbracket 1; N \rrbracket \times \llbracket 0; M \rrbracket, \quad x_{ij} = x_i(t_j, \omega, x_{1:N}^0, v_{1:N}^0) + \epsilon_{ij} \text{ where } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2 I_2) \quad (2.17)$$

and $x_i(t_j, \omega, x_{1:N}^0, v_{1:N}^0)$ is given by equation (2.7) and $t_0 = 0$.

$$\begin{aligned}
 p_{x|\omega}(x \mid \omega) &= \frac{1}{(2\pi\sigma^2)^{N(M+1)}} \int_{\mathbb{R}^{4N}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{j=0}^M |x_{ij} - x_i(t_j, \omega, x_{1:N}^0, v_{1:N}^0)|^2\right) \\
 &\times \mu_0^{\otimes N}(dx_{1:N}^0, dv_{1:N}^0) \\
 \text{where } \mu_0^{\otimes N}(dx_{1:N}^0, dv_{1:N}^0) &= \frac{\lambda(dx_{1:N}^0, dv_{1:N}^0)}{(2\pi)^{2N}(\det(\Sigma_x \Sigma_v))^{N/2}} \prod_{i=1}^N \exp\left(-\frac{1}{2}(x_i^0 - \mu_x)^\top \Sigma_x^{-1}(x_i^0 - \mu_x)\right) \\
 &\times \exp\left(-\frac{1}{2}(v_i^0)^\top \Sigma_v^{-1}v_i^0\right)
 \end{aligned} \tag{2.18}$$

The integral over \mathbb{R}^{4N} figuring in the expression (2.18) can be derived analytically by using the formula given in the following lemma.

Lemma 2.2. Expectation of the exponential of a quadratic form with respect to a normal distribution²

Let $H \in \mathcal{M}_d(\mathbb{R})$ be a symmetric matrix positive semi-definite and $y \in \mathbb{R}^d$. Let $\mathcal{N}(m, \Sigma)$ be a normal distribution. Then

$$\begin{aligned}
 &\int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}(x^\top Hx - 2y^\top x)\right) \mathcal{N}(m, \Sigma)(dx) \\
 &= \frac{\exp\left(-\frac{1}{2}(m^\top(Hm - 2y) - (y - Hm)^\top \Sigma^{1/2}(\mathbf{I}_d + \Sigma^{1/2}H\Sigma^{1/2})^{-1}\Sigma^{1/2}(y - Hm))\right)}{\sqrt{\det(\mathbf{I}_d + \Sigma^{1/2}H\Sigma^{1/2})}}
 \end{aligned} \tag{2.19}$$

We can rewrite the integrand in (2.18) as an expectation with respect to some normal distribution defined over \mathbb{R}^{4N} .

$$\begin{aligned}
 &\int_{\mathbb{R}^{4N}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{j=0}^M |x_{ij} - x_i(t_j, \omega, x_{1:N}^0, v_{1:N}^0)|^2\right) \mu_0^{\otimes N}(dx_{1:N}^0, dv_{1:N}^0) \\
 &= \int_{\mathbb{R}^{4N}} \exp\left(-\frac{1}{2\sigma^2} ((X_{1:N}^0)^\top C(\omega)X_{1:N}^0 - 2Y(\omega, x)^\top X_{1:N}^0 + z(x))\right) \mathcal{N}(m_X, \Sigma_X)(dX_{1:N}^0)
 \end{aligned} \tag{2.20}$$

The expressions of matrices $C(\omega), \Sigma_X \in \mathcal{M}_{4N}(\mathbb{R})$, of the vectors $Y(\omega, x), m_X \in \mathbb{R}^{4N}$ and of the scalar $z(x)$ are given in the appendix, page 85. The direct application of the formula of lemma 2.2 gives a closed-form expression for the likelihood

²The proof of lemma 2.2 can be found in the appendix, page 85.

density for all $x \in \mathbb{R}^{2N(m+1)}$

$$\begin{aligned}
 p_{x|\omega}(x | \omega) &= \frac{1}{(2\pi\sigma^2)^{N(m+1)} \sqrt{\det \left(\mathbf{I}_{4N} + \frac{\Sigma_X C(\omega)}{\sigma^2} \right)}} \\
 &\times \exp \left(-\frac{1}{2\sigma^2} \left(z(x) + m_X^\top C(\omega) m_X - 2Y(\omega, x)^\top m_X \right. \right. \\
 &\quad \left. \left. - \frac{1}{\sigma^2} (Y(\omega, x) - C(\omega) m_X)^\top \left(\Sigma_X^{-1} + \frac{C(\omega)}{\sigma^2} \right)^{-1} (Y(\omega, x) - C(\omega) m_X) \right) \right)
 \end{aligned} \tag{2.21}$$

By applying the Bayes formula, we can obtain that the posterior distribution of the pulsation ω , known up to a multiplicative factor.

$$\begin{aligned}
 p_{\omega|x}(\omega | x) &\propto \frac{p_\omega(\omega)}{\sqrt{\det \left(\mathbf{I}_{4N} + \frac{\Sigma_X C(\omega)}{\sigma^2} \right)}} \exp \left(-\frac{1}{2\sigma^2} \left(m_X^\top C(\omega) m_X - 2Y(\omega, x)^\top m_X \right. \right. \\
 &\quad \left. \left. - \frac{1}{\sigma^2} (Y(\omega, x) - C(\omega) m_X)^\top \left(\Sigma_X^{-1} + \frac{C(\omega)}{\sigma^2} \right)^{-1} (Y(\omega, x) - C(\omega) m_X) \right) \right)
 \end{aligned} \tag{2.22}$$

It is worth noting that the latent initial conditions have a Gaussian posterior distribution conditionally to x and ω , thanks to the conjugacy property of the normal distribution.

$$X_{1:N}^0 | x, \omega \sim \mathcal{N} \left[\left(\Sigma_X^{-1} + \frac{C(\omega)}{\sigma^2} \right)^{-1} \left(\Sigma_X^{-1} m_X + \frac{Y(\omega, x)}{\sigma^2} \right), \left(\Sigma_X^{-1} + \frac{C(\omega)}{\sigma^2} \right)^{-1} \right] \tag{2.23}$$

For small observation standard deviation σ , a high weight is given to the observations, whereas for large σ , the prior mean m_X has a higher weight. As individuals are distinguished and labelled in the data set x , the posterior distribution of $X_{1:N}^0 | x, \omega$ is not symmetric, i.e. a permutation of the components of vector $X_{1:N}^0$ may not have the same distribution as the original vector.

We can notice on figure 2.5 that a partial knowledge of the initial condition does not change the structure of the posterior distribution fundamentally. The density remains unimodal, maxima are reached precisely at the same pulsation for the cases with exact or partial knowledge, even when the standard deviation σ of the observation is quite large. Of course, the variance of $\omega | x$ is larger when $X_{1:N}^0$ is not known for sure. A third situation interesting for the variance of $\omega | x$ is when the state of the whole population is known exactly at two different times.

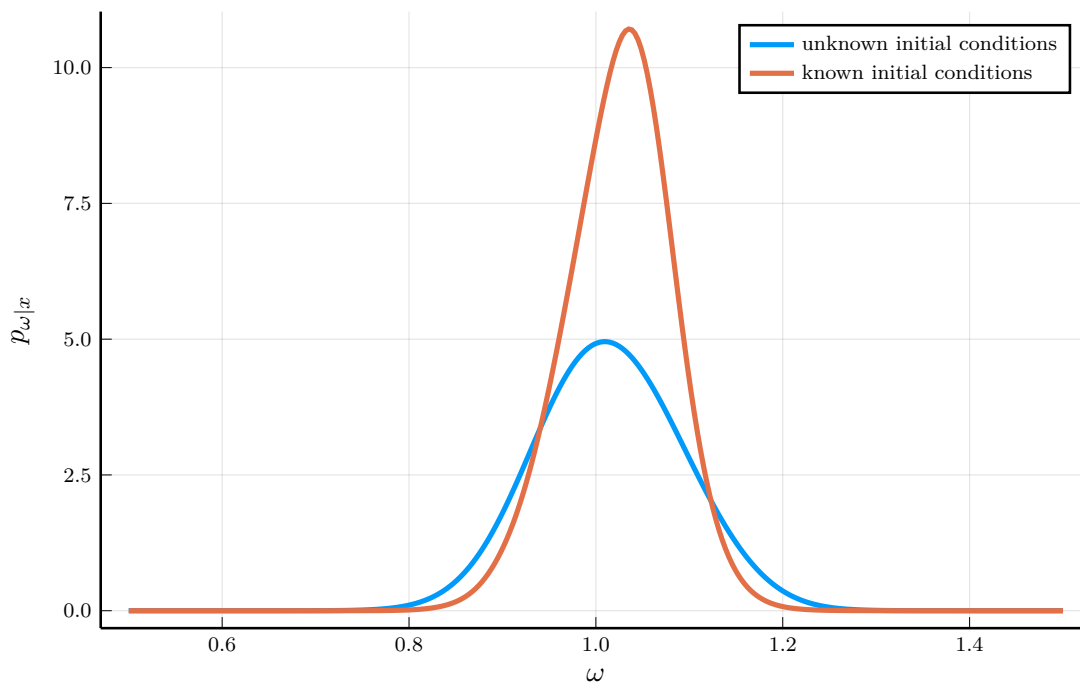


Figure 2.5: Comparison of the posterior densities for unknown and known initial conditions (details on page 87).

By unicity of the trajectories, $\omega | x$ would then be a constant equal to ω_0 , and its variance would be zero. In the specific case of Spring Cloud, determining $p(\omega | x)$ is significantly simplified by the analytical integral over the latent variables (equation 2.20) and by the low dimensionality of the parameter space. In practice, however, the expectation with respect to individual latent variables, such as heterogeneities, does not have a closed-form expression.

2.6 Uncertainty on the population size

In the case of large populations, the exact number of individuals may not be known for sure. This situation typically appears when studying plants in a large crop field. In that case, we cannot have an exact knowledge of the number of individuals in interaction with a given individual. In the model where N figures in the equation of motion, we need to introduce it as a latent variable of the problem, with a specific prior distribution representing the eventual prior knowledge on the range of this variable. The prior distribution is absolutely continuous with respect to the

counting measure $\sum_{n \in \mathbb{N}} \delta_n$ and its density $p_N : \mathbb{N} \rightarrow \mathbb{R}_+$ is such that $\sum_{n=0}^{+\infty} p_N(n) = 1$.

Let us come back to the Spring Cloud example from subsection 2.3 and let us assume that we have a data set $x \in \mathbb{R}^{2N_0m}$, composed of N_0 individual trajectories observed at m observation times. We could have also considered a data set with undifferentiated individuals, but let us focus on the uncertainty due to the population size only. Thus, as N_0 distinct individuals have been observed, it seems logical to assume that the support of the prior p_N is included within $\llbracket N_0; +\infty \llbracket$, i.e., for all $N \leq N_0$, $p_N(N) = 0$. Conditionally to N , the observed individuals evolve in a population of size N , and we do not have any information on the remainder $N - N_0$ individuals. The observations are explained by the following hierarchical model.

$$\begin{aligned}
 & N \sim p_N \text{ prior on the population size} \\
 & \omega \mid N \sim \mathcal{U}([\omega_{\min}\sqrt{N}; \omega_{\max}\sqrt{N}]) \text{ prior on the parameter (the pulsation)} \\
 & x_{N_0+1:N}^0 \sim \mu_0^{\otimes N-N_0} \text{ initial condition of the unobserved individuals} \\
 & \forall i \in \llbracket 1; N_0 \rrbracket, \forall j \in \llbracket 1; M \rrbracket, \quad x_{ij} = x_i(t_j, \omega, x_{1:N}^0, v_{1:N}^0) + \epsilon_{ij} \text{ observation error} \\
 & \text{with } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_2)
 \end{aligned} \tag{2.24}$$

In equation (2.24), the initial conditions of the first N_0 individuals are known exactly, i.e. $x_{1:N_0}^0$ and $v_{1:N_0}^0$ are known, and the remainder $N - N_0$ individuals are sampled from the initial distribution μ_0 . The joint distribution associated with this situation is given by the density $p_{x,\omega}$.

$$\begin{aligned}
 p_{x,\omega}(x, \omega) &= \frac{1}{(2\pi\sigma^2)^{N_0m}} \sum_{N=N_0+1}^{+\infty} p_N(N) p_{\omega|N}(\omega \mid N) \\
 &\times \int_{\mathbb{R}^{4(N-N_0)}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N_0} \sum_{j=1}^M |x_{ij} - x_i(t_j, \omega, x_{1:N}^0, v_{1:N}^0)|^2\right) \mu_0^{\otimes N-N_0}(dx_{N_0+1:N}^0, dv_{N_0+1:N}^0)
 \end{aligned}$$

$$\text{where } p_{\omega|N}(\omega \mid N) = \frac{\mathbb{I}\{\omega_{\min} \leq \omega/\sqrt{N} \leq \omega_{\max}\}}{(\omega_{\max} - \omega_{\min})\sqrt{N}} \tag{2.25}$$

The expectation with respect to the initial conditions can be computed using the formula of lemma 2.2. After application of the Bayes' formula, we obtain that the

posterior density of $\omega \mid x$ is proportional to the following expression.

$$\begin{aligned}
 p_{\omega|x}(\omega \mid x) &\propto \sum_{N=N_0+1}^{+\infty} \frac{p_N(N)p_{\omega|N}(\omega \mid N)}{\sqrt{\det \left(\mathbf{I}_{4(N-N_0)} + \frac{\Sigma_X^{N_0,N} C_{N_0}^N(\omega)}{\sigma^2} \right)}} \\
 &\times \exp \left(-\frac{1}{2\sigma^2} \left(z_N(x, \omega) + (m_X^{N_0,N})^\top C_{N_0}^N(\omega) m_X^{N_0,N} - 2Y_N(x, \omega)^\top m_X^{N_0,N} \right. \right. \\
 &\left. \left. - \frac{1}{\sigma^2} (Y_N(x, \omega) - C_{N_0}^N(\omega) m_X^{N_0,N})^\top \left((\Sigma_X^{N_0,N})^{-1} + \frac{C_{N_0}^N(\omega)}{\sigma^2} \right)^{-1} (Y_N(x, \omega) - C_{N_0}^N(\omega) m_X^{N_0,N}) \right) \right) \\
 &\propto \sum_{N=N_0+1}^{+\infty} p_N(N) \tilde{p}_{\omega|N,x}(\omega \mid N, x)
 \end{aligned} \tag{2.26}$$

The expressions of the scalar $z_N(x, \omega)$, the vectors $Y_N(x, \omega), m_X^{N_0,N} \in \mathbb{R}^{4(N-N_0)}$ and the matrices $C_{N_0}^N(\omega), \Sigma_X^{N_0,N} \in \mathcal{M}_{4(N-N_0)}(\mathbb{R})$ are given page 87. The graph of the unnormalised posterior is given in figure 2.6. We can notice that the density is not smooth, as it is the sum of piecewise continuous functions. The density is unimodal and the mode of the density is close to the pulsation having generated the data. The density may become discontinuous for small length of the prior interval $[\omega_{\min}; \omega_{\max}]$.

The computation of the posterior distribution $\tilde{p}_{\omega|x}$ is simplified by the fact that the latent variables have a Gaussian conditional distribution with respect to the variables ω, N, x , which has an expression similar to the Gaussian distribution in equation (2.23).

$$\begin{aligned}
 \mathbb{E}[X_{N_0+1:N}^0 \mid N, \omega, x] &= \left((\Sigma_X^{N_0,N})^{-1} + \frac{C_{N_0}^N(\omega)}{\sigma^2} \right)^{-1} \left((\Sigma_X^{N_0,N})^{-1} m_X^{N_0,N} + \frac{Y_N(x, \omega)}{\sigma^2} \right) \\
 \text{Cov}(X_{N_0+1:N}^0 \mid N, \omega, x) &= \left((\Sigma_X^{N_0,N})^{-1} + \frac{C_{N_0}^N(\omega)}{\sigma^2} \right)^{-1}
 \end{aligned} \tag{2.27}$$

We can notice in the construction of the matrix $C_{N_0}^N(\omega)$ on page 87 that the conditional distribution of the latent variables $X_{N_0+1:N}^0 \mid N, \omega, x$ is symmetrical, i.e., for any permutation $\sigma \in \mathfrak{S}([N_0+1; N])$, the vectors $X_{N_0+1:N}$ and $X_{\sigma(N_0+1:N)}^0$ have the same distribution. This symmetry of the system is here preserved at the level of the latent variables, as the unobserved individuals are not assigned to specific labels.

We may also use the expression in equation (2.26) to derive the distribution of $N \mid x$, to see whether the observations can help us *guess* the total number of

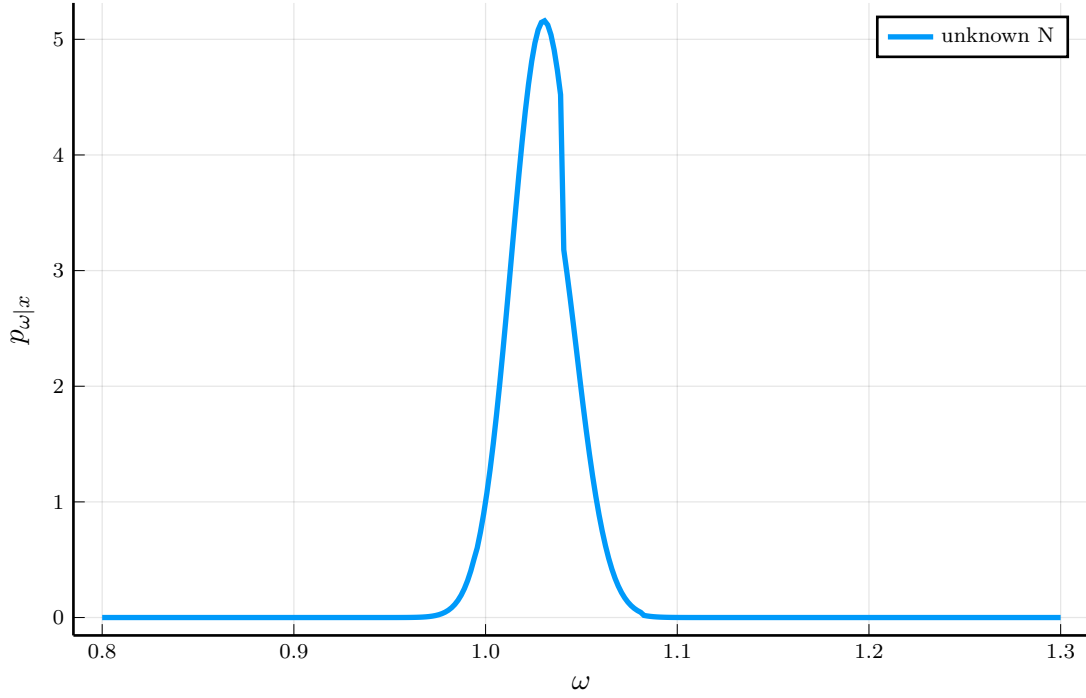


Figure 2.6: Unnormalised posterior distribution $\tilde{p}_{\omega|x}$ when the size of the population is unknown (details on page 89). The true value of the parameter for the data generation is $\omega = 1$.

individuals in the population.

$$\forall N \geq N_0 + 1, \quad p_{N|x}(N | x) \propto p_N(N) \int_0^{+\infty} \tilde{p}_{\omega|N,x}(\omega|N, x) d\omega \quad (2.28)$$

The distribution of the variable $N|x$ is represented in the figure 2.7. The mode of this distribution is reached near the true size of the population.

2.7 Uncertainty due to the inexact simulation of the system

Let us consider a more general setting for the Spring Cloud system, allowing the particles to have respective values for parameters (κ, m) . In this heterogeneous

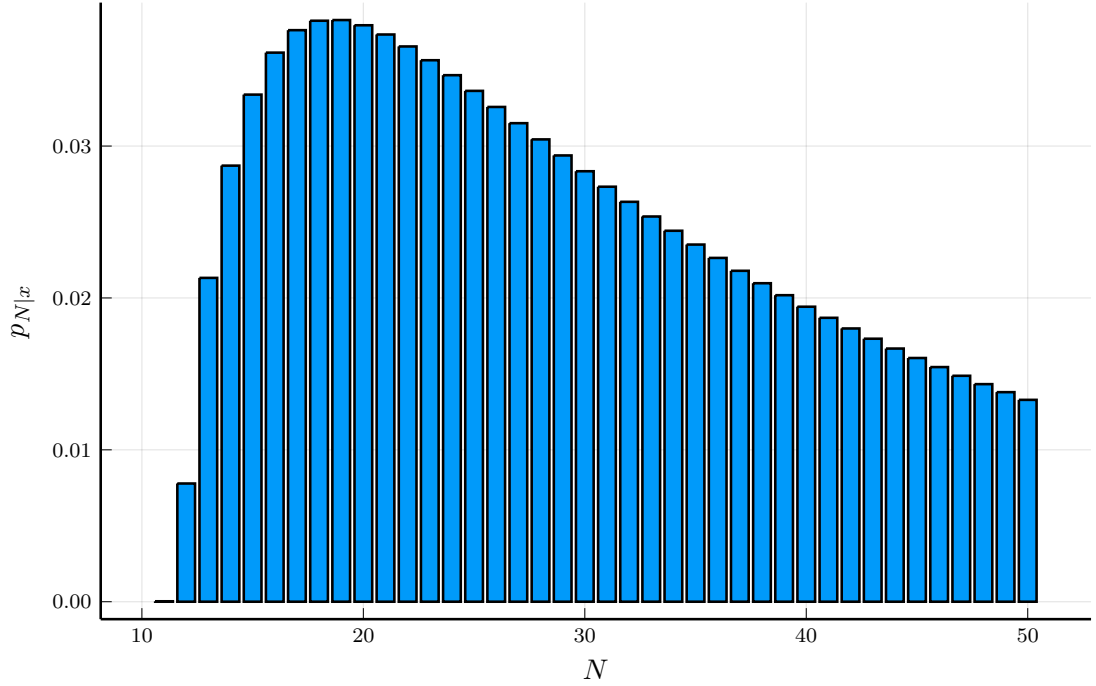


Figure 2.7: Density $p(N|x)$ (with respect to the counting measure of the population size conditionally to the observations (details on page 89)). The size of the population used for data generation is $N = 25$.

configuration, the expression of the trajectories

$$\forall i \in \llbracket 1; N \rrbracket, \quad x_i(t) = \left(x_i^0 - \frac{1}{N} \sum_{j=1}^N x_j^0 \right) \cos(\omega_N t) + \frac{1}{\omega_N} \left(v_i^0 - \frac{1}{N} \sum_{j=1}^N v_j^0 \right) \sin(\omega_N t) + \frac{1}{N} \sum_{j=1}^N x_j^0 + \frac{t}{N} \sum_{j=1}^N v_j^0 \quad \text{with } \omega_N = \kappa \sqrt{\frac{N}{m}} \quad (2.29)$$

is no longer valid. The equation of the motion of the particle i in a population of size N is monitored by the system of equation

$$\frac{d^2 x_i}{dt^2}(t) = -\frac{\kappa_i}{m_i} \left(\sum_{j=1}^N \kappa_j \right) x_i(t) + \frac{\kappa_i}{m_i} \sum_{j=1}^N \kappa_j x_j(t) \quad (2.30)$$

In this section, we consider the situation where we want to infer the values of the parameters κ_1 associated with the particle 1, provided that we have observations

of the whole population at different times, and an exact knowledge of the initial condition of the system, and the parameters $\kappa_2, \dots, \kappa_N$ and m_1, \dots, m_N . We are in fact in a situation quite similar to the one described in section 2.3, except that the population is heterogeneous. In this section, we assume that we are not able to solve exactly the system (1.39), which is not true, since the system is linear. We are only able to simulate the dynamics using a consistent numerical method. This rather unrealistic situation is here to illustrate a methodology to model the uncertainty introduced by the discrete integration, which is unavoidable when the system is nonlinear and does not have an analytical solution. Let us first set the notations concerning the true dynamics of the Spring Cloud system. The matrix associated with the differential system (1.38) is denoted by $C_N^{\kappa, m} \in \mathcal{M}_{4N}(\mathbb{R})$, and can be expressed with the parameters $\kappa_{1:N}$ and $m_{1:N}$.

$$\begin{aligned} \frac{dX_{1:N}(t)}{dt} &= C_N^{\kappa, m} X_{1:N}(t) \\ \text{where } X_{1:N} &= (x_1^1, \dots, x_N^1, x_1^2, \dots, x_N^2, v_1^1, \dots, v_N^1, v_1^2, \dots, v_N^2) \in \mathbb{R}^{4N} \end{aligned} \quad (2.31)$$

$$\begin{aligned} C_N^{\kappa, m} &= \begin{pmatrix} 0 & 0 & I_N & 0 \\ 0 & 0 & 0 & I_N \\ B_N^{\kappa, m} & 0 & 0 & 0 \\ 0 & B_N^{\kappa, m} & 0 & 0 \end{pmatrix} \in \mathcal{M}_{4N}(\mathbb{R}) \\ B_N^{\kappa, m} &= \begin{pmatrix} \frac{\kappa_1}{m_1} \\ \vdots \\ \frac{\kappa_N}{m_N} \end{pmatrix} (\kappa_1 \quad \dots \quad \kappa_N) - \left(\sum_{i=1}^N \kappa_i \right) \begin{pmatrix} \frac{\kappa_1}{m_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{\kappa_N}{m_N} \end{pmatrix} \in \mathcal{M}_N(\mathbb{R}) \end{aligned} \quad (2.32)$$

The exact solution of the system is obtained by computing the exponential of the matrix $C_N^{\kappa, m}$.

$$X_{1:N}(t) = \exp(tC_N^{\kappa, m})X_{1:N}^0 \quad (2.33)$$

The exponential of $C_N^{\kappa, m}$ does not have a closed-form expression for $N \geq 3$. Realisations of exact trajectories associated with a heterogeneous population of 10 Spring Cloud particles is represented in figure 2.8. In the homogeneous case, the barycentre x_κ has a uniform rectilinear motion (see equation 2.7), whereas in the heterogeneous case, the motion has a more complex structure. Recall that the barycentre x_κ constitutes the summary of the population effect on each individual, thanks to equation (1.39). At the level of the particles, the trajectories are no longer ellipses.

If the observation of the system is made with a Gaussian error, as in (2.10) the posterior distribution of parameter κ_1 has an expression quite similar to equation

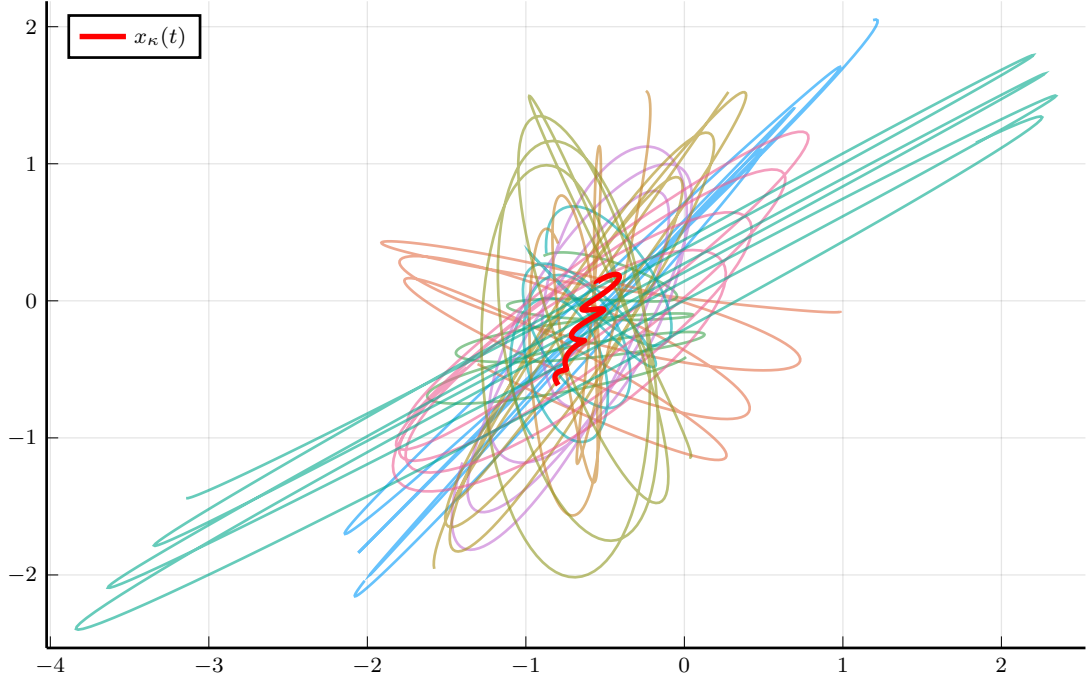


Figure 2.8: Exact trajectories of Spring Cloud system for a population of size $N = 10$ individuals (detailed configuration of the graph on page 90). The trajectory of the attractivity barycentre $x_\kappa = \frac{\sum_{j=1}^N \kappa_j x_j}{\sum_{j=1}^N \kappa_j}$ is represented in red.

of the homogeneous case.

$$p_{\kappa_1|x}(\kappa_1 | x) \propto p_{\kappa_1}(\kappa_1) \exp \left(-\frac{1}{2\sigma^2} \sum_{j=1}^M |x_j - O_x \exp(t_j C_N^{\kappa, m}(\kappa_1)) X_{1:N}^0|^2 \right) \quad (2.34)$$

where $O_x = \begin{pmatrix} I_N & 0 & 0 & 0 \\ 0 & I_N & 0 & 0 \end{pmatrix} \in \mathcal{M}_{2N, 4N}(\mathbb{R})$

and x_j gathers the observed positions at time t_j

2.7.1 Deterministic integrator

Let us assume that, in an analogy with the nonlinear case, we are not able to apply formula (2.33). Instead, we simulate the system using the numerical method described in example 1.7, consisting in approximating the trajectory of the population empirical measure by a piecewise constant function of time. The observation

period $[0; T]$ (with $T = t_m$) is subdivided into intervals $[\tau_0; \tau_1], [\tau_1; \tau_2], \dots, [\tau_{M-1}; \tau_M]$ with $\tau_0 = 0$ and $\tau_M = T$. The numerical method returns a dense estimate $\hat{X}_{1:N}(t)$ of the true solution $X_{1:N}(t)$, given by an inductive equation.

$$\begin{aligned} \forall i \in \llbracket 1; N \rrbracket, \quad \hat{x}_i^0 &= x_i^0, \quad \hat{v}_i^0 = v_i^0 \\ \forall n \in \llbracket 0; M-1 \rrbracket, \quad \hat{x}_\kappa^n &= \frac{\sum_{j=1}^N \kappa_j \hat{x}_j^n}{\sum_{j=1}^N \kappa_j} \\ \forall t \in [0; T], \quad \hat{x}_i^n(t) &= \hat{x}_\kappa^n (1 - \cos(\omega_i(t - \tau_n))) + \hat{x}_i^n \cos(\omega_i(t - \tau_n)) \\ &+ \frac{\hat{v}_i^n}{\omega_i} \sin(\omega_i(t - \tau_n)) \\ \text{where } \omega_i &= \sqrt{\frac{\kappa_i \sum_{j=1}^N \kappa_j}{m_i}} \end{aligned}$$

$$\begin{aligned} \hat{v}_i^n(t) &= \frac{d\hat{x}_i^n}{dt}(t) \\ \hat{x}_i^{n+1} &= \hat{x}_i^n(\tau_{n+1}), \quad \hat{v}_i^{n+1} = \hat{v}_i^n(\tau_{n+1}) \\ \hat{X}_{1:N} : t \in [0; T] &\mapsto \sum_{n=0}^{M-1} \mathbb{I}\{\tau_n < t < \tau_{n+1}\} \begin{pmatrix} \hat{x}_{1:N}^n(t) \\ \hat{v}_{1:N}^n(t) \end{pmatrix} + \sum_{n=0}^M \mathbb{I}\{t = \tau_n\} \begin{pmatrix} \hat{x}_{1:N}^n \\ \hat{v}_{1:N}^n \end{pmatrix} \end{aligned} \quad (2.35)$$

The observation model associated with the numerical method replaces the actual state of the population by its numerical approximation. Such type of observation model is very often used when doing inference on a nonlinear system of differential equations.

$$\forall i \in \llbracket 1; N \rrbracket, \forall j \in \llbracket 1; M \rrbracket, \quad x_{ij} = \hat{x}_i(t_j, \kappa_1) + \epsilon_{ij} \text{ with } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_2) \quad (2.36)$$

The problem underlined by Conrad et al. (2015)⁷ is that such observation model completely neglects the error introduced by the numerical method. This approximation may be valid if the step of the subdivision is sufficiently small, otherwise it may introduce a systemic bias in the estimation of the parameters. In our case, we can quantify the accumulation of the error over time by estimating the truncation error of the numerical scheme.

Proposition 2.1. Truncation error of the numerical scheme³

Let $x_{1:N}^0, v_{1:N}^0, \kappa_{1:N}, m_{1:N}$ be an initial configuration of the system (1.38). Let $\hat{X}_{1:N}$ be the function defined by

$$\hat{X}_{1:N}(t) = \begin{pmatrix} \hat{x}_{1:N}^0(t) \\ \hat{v}_{1:N}^0(t) \end{pmatrix} \quad (2.37)$$

³The proof of proposition 2.1 can be found on page 91

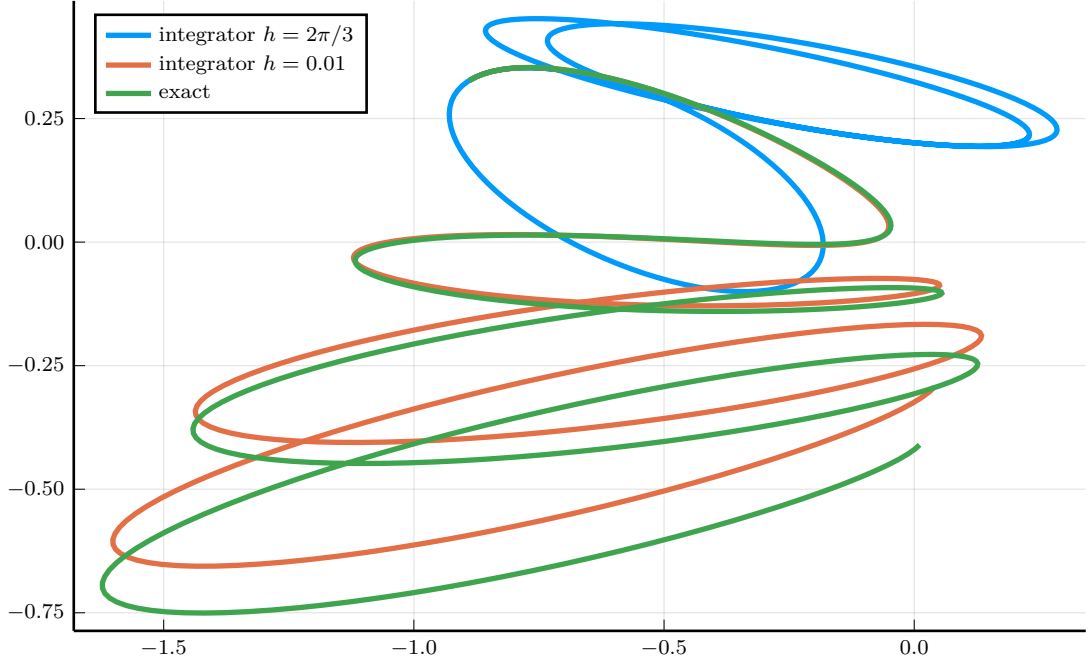


Figure 2.9: Comparison of the trajectories obtained by numerical integration with a large time step $h = 2\pi/3$ and small time step $h = 0.01$ for a given particle. The exact trajectory of the particle is represented in green. Detailed configuration of the graph on page 90.

where $\hat{x}_{1:N}^0, \hat{v}_{1:N}^0$ are defined in equation (2.35). Let $X_{1:N}$ be the solution of the system (1.38). Then for any norm $|\cdot|$ over \mathbb{R}^{4N} , there exists a constant $K > 0$ such that

$$\forall t \in [0; T], \quad |X_{1:N}(t) - \hat{X}_{1:N}(t)| \leq Kt^2 \quad (2.38)$$

Proposition 2.1 shows that the order of the numerical scheme (2.35) is 1. To prove proposition (2.1), the conservation of the energy of the system is helpful to obtain bounds on the motion of the particles (see lemmata 2.3 and 2.4). Similarly as in the case of the Schneider system (see proposition 1.4), we can improve the quality of the approximation by considering piecewise linear trajectory of the barycentre, leading to a second order scheme.

$$\forall t \in [\tau_n; \tau_{n+1}), \quad \hat{x}_\kappa^n(t) = \frac{1}{\sum_{j=1}^N \kappa_j} \sum_{j=1}^N \kappa_j (\hat{x}_j^n + \hat{v}_j^n(t - \tau_n)) \quad (2.39)$$

To obtain an order 3, we need to compute the acceleration of the barycentre, which

is given by

$$a_\kappa(t) = \sum_{j=1}^N \frac{\kappa_j^2}{m_j} (x_\kappa(t) - x_j(t)) \quad (2.40)$$

For the sake of simplicity, we consider a piecewise constant barycentre trajectory in what follows.

The trajectories obtained by numerical integration, for different time steps are displayed on figure 2.9. We can notice that the numerical integrator fails to reproduce the exact trajectories when the time step h is too large, but the simulated trajectories remain stable. With an explicit Euler scheme, one could expect a divergence to occur with such a level of discretisation. The numerically simulated particles, even with rough h , seem to follow continuously differentiable paths, although the continuity of velocities is not imposed at τ_n in the definition of the scheme.

The accumulation of errors due to numerical integration can distort the Bayesian inference by creating a discrepancy between the exact a posteriori density and the density approximated by the numerical method. Such discrepancy appears on figure 2.10. In the case of a large time step, the observation model (2.36) is not relevant for observation times far from the origin $t = 0$, as the numerical error can significantly exceed the standard error σ . Therefore, the approximated posterior density is concentrated in a region disjointed from the relevant region, i.e., the region with high probability for the exact posterior density. In the case of a small timestep, the true density is well approximated, but such a level of precision can be challenging to achieve, for example, in situations where the population size is huge. In practice, Bayesian inference needs a significant amount of model simulations. To obtain a sort of trade-off between time-consuming accuracy and efficient but rough simulations of the system, we introduce in the next subsection a model for the numerical error, enabling to assign time-increasing variances to observations.

2.7.2 Stochastic integrator

We give here some general definitions related to Gaussian processes, that are used in our case to model the uncertainty introduced by the deterministic integrator.

Definition 2.4. positive definite function

Let $\Sigma : \mathcal{X}^2 \rightarrow \mathcal{M}_d(\mathbb{R})$ a function taking values in the space of matrices. The function is said to be positive definite if

$$\forall n \in \mathbb{N}^*, \forall x_1, \dots, x_n \in \mathcal{X}, \forall y_1, \dots, y_n \in \mathbb{R}^d, \quad \sum_{i=1}^n \sum_{j=1}^n y_i^T \Sigma(x_i, x_j) y_j \geq 0 \quad (2.41)$$

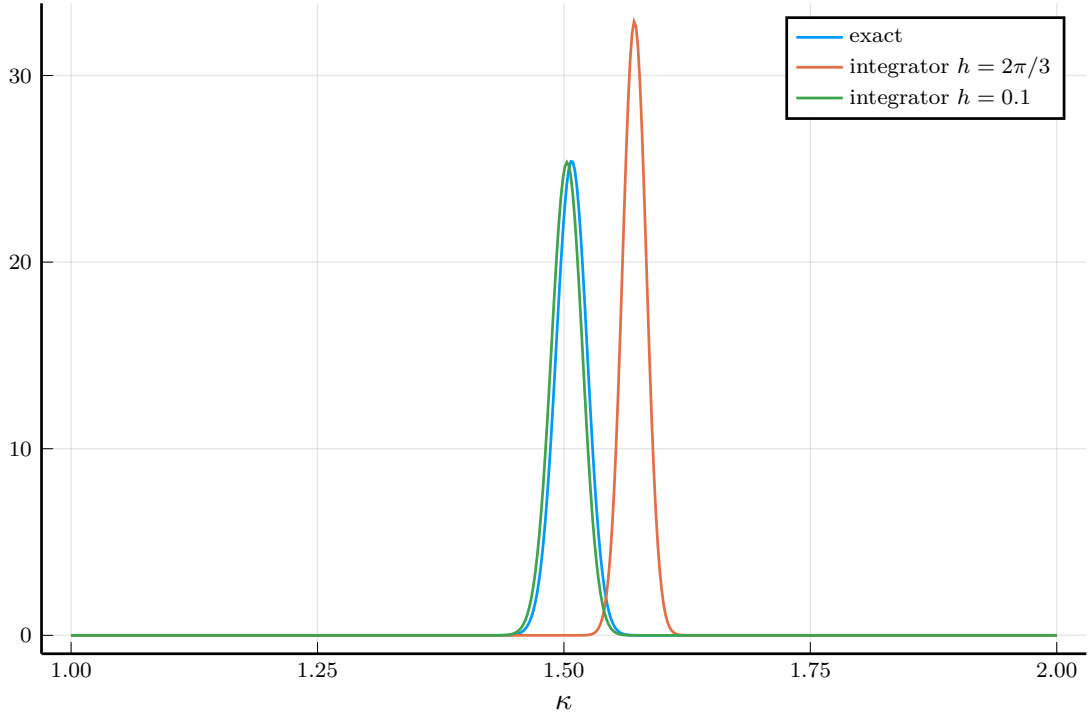


Figure 2.10: Comparison between the exact posterior density and the approximated densities obtained with the numerical method (2.35) for different time steps. Detailed configuration of the graph on pages 90 and 93.

Definition 2.5. continuous Gaussian process

Let $X : (\Omega, \mathcal{F}) \rightarrow (\mathcal{C}(\mathcal{K}, \mathbb{R}^d), \mathcal{B}(\mathcal{C}(\mathcal{K}, \mathbb{R}^d)))$ a random variable taking values in the space of continuous functions defined over a compact metric space \mathcal{K} . The space $\mathcal{C}(\mathcal{K}, \mathbb{R}^d)$ is endowed with the uniform metric. Let $m_X \in \mathcal{C}(\mathcal{K}, \mathbb{R}^d)$ and $\Sigma_X \in \mathcal{C}(\mathcal{K}^2, \mathcal{M}_d(\mathbb{R}))$ a positive definite function. Then X is said to be a Gaussian process of mean function m_X and of covariance function Σ_X if for all $n \in \mathbb{N}^*$ and for all $x_1, \dots, x_n \in \mathcal{X}$, the random vector $\omega \in \Omega \mapsto (X(\omega, x_1), \dots, X(\omega, x_n)) \in \mathbb{R}^{dn}$ is distributed according to a Gaussian distribution $\mathcal{N}(m_X(x_{1:N}), \Sigma_X(x_{1:N}))$, where $m_X(x_{1:N}) = (m_X(x_1), \dots, m_X(x_n)) \in \mathbb{R}^{dn}$ and $\Sigma_X(x_{1:n}) = (\Sigma_X(x_i, x_j))_{1 \leq i, j \leq n} \in \mathcal{M}_{dn}(\mathbb{R})$.

A Gaussian process defines a distribution over mappings from a set \mathcal{Y} to a vector space \mathbb{R}^d . In this section, we consider time-indexed Gaussian processes to model the approximation error of the integrator, following a methodology described in Conrad et al. (2015).⁷ We consider the following process, obtained by adding a collection of Gaussian process to the deterministic integrator, defined in equation

(2.35). The approximation method returns a stochastic process $\tilde{X}_{1:N}$ defined over $[0; T]$.

$$\begin{aligned}
 \forall i \in \llbracket 1; N \rrbracket, \quad & \tilde{x}_i^0 = x_i^0, \quad \tilde{v}_i^0 = v_i^0 \\
 \forall n \in \llbracket 0; M-1 \rrbracket, \quad & \tilde{x}_\kappa^n = \frac{\sum_{j=1}^N \kappa_j \tilde{x}_j^n}{\sum_{j=1}^N \kappa_j} \\
 \forall t \in [0; T], \quad & \tilde{x}_i^n(t) = \tilde{x}_\kappa^n (1 - \cos(\omega_i(t - \tau_n))) + \tilde{x}_i^n \cos(\omega_i(t - \tau_n)) \\
 & + \frac{\tilde{v}_i^n}{\omega_i} \sin(\omega_i(t - \tau_n)) + \xi_i^{x,n}(t - \tau_n)
 \end{aligned} \tag{2.42}$$

where $\xi_i^{x,n} : \Omega \mapsto \mathcal{C}([0; T], \mathbb{R}^2)$ is a Gaussian process.

$$\begin{aligned}
 \tilde{v}_i^n(t) &= \frac{d\tilde{x}_i^n}{dt}(t) \\
 \tilde{x}_i^{n+1} &= \tilde{x}_i^n(\tau_{n+1}), \quad \tilde{v}_i^{n+1} = \tilde{v}_i^n(\tau_{n+1})
 \end{aligned}$$

$$\tilde{X}_{1:N} : t \in [0; T] \mapsto \sum_{n=0}^{M-1} \mathbb{I}\{\tau_n < t < \tau_{n+1}\} \begin{pmatrix} \tilde{x}_{1:N}^n(t) \\ \tilde{v}_{1:N}^n(t) \end{pmatrix} + \sum_{n=0}^M \mathbb{I}\{t = \tau_n\} \begin{pmatrix} \tilde{x}_{1:N}^n \\ \tilde{v}_{1:N}^n \end{pmatrix} \tag{2.43}$$

The collection of Gaussian processes $(\xi_i^{x,n})_{\substack{0 \leq n \leq M-1 \\ 1 \leq i \leq N}}$ are chosen to be independent from one individual to another, and from one subdivision to another. The processes must satisfy a series of conditions in order to be consistent with the analysis of the deterministic integrator.

1. The variance of the processes must increase with time. At the initial time, since the initial condition is known, we must have $\xi_i^{x,0}(0) = 0$ for all i . However, we expect the variance of the process $\tilde{X}_{1:N}(t)$ to increase over the first subdivision $[0; \tau_1]$ with the same order of magnitude as t^2 .
2. The processes $\xi_i^{x,n}$ must be smooth enough, so that their derivatives with respect to time are defined in the classical sense. These derivatives appear when computing the velocities \tilde{v}_i^n .
3. The resulting stochastic integrator must be consistent in some sense. The distribution of the process $\tilde{X}_{1:N}$ must concentrate on the true solution when the time step of the subdivision gets close to zero.

The regularity of a Gaussian process is monitored by the smoothness of its mean and covariance. As we want the integrator to be consistent, and as the deterministic integrator is known to be consistent, we choose a null mean function for all the processes. Concerning the existence of the derivative of ξ^x , we use the following result.

Theorem 2.2. Derivative of a Gaussian process (from Le Gall, 2013,¹⁶ exercise 1.1)

Let $(X(t))_{t \in [a;b]}$ be a real-indexed Gaussian process of zero mean function and of covariance function $K_X : [a; b]^2 \rightarrow \mathcal{M}_d(\mathbb{R})$ being twice continuously differentiable over $[a; b]^2$. Then for any $t \in [a; b]$, the limit $X'(t) = \lim_{\delta \rightarrow 0} \frac{X(t + \delta) - X(t)}{\delta}$ exists in L^2 , i.e.,

$$\mathbb{E} \left| \frac{X(t+s) - X(t)}{s} - \frac{X(t+u) - X(t)}{u} \right|^2 \xrightarrow{s,u \rightarrow 0} 0 \quad (2.44)$$

and the process $(X(t), X'(t))$ is Gaussian, of zero mean function and of covariance function $K_{X,X'} : [a; b]^2 \rightarrow \mathcal{M}_{2d}(\mathbb{R})$ defined by

$$\forall (t_1, t_2) \in [a; b]^2, \quad K_{X,X'}(t_1, t_2) = \begin{pmatrix} K_X(t_1, t_2) & \frac{\partial K_X}{\partial t_2}(t_1, t_2) \\ \frac{\partial K_X}{\partial t_1}(t_1, t_2) & \frac{\partial^2 K_X}{\partial t_1 \partial t_2}(t_1, t_2) \end{pmatrix} \quad (2.45)$$

The expression of the covariance function in (2.45) is natural if you take for granted the existence of the derivative of the Gaussian process.

$$\text{Cov} \left(X(t_1), \frac{X(t_2 + \delta) - X(t_2)}{\delta} \right) = \frac{K_X(t_1, t_2 + \delta) - K_X(t_1, t_2)}{\delta} \quad (2.46)$$

In the case of the Spring Cloud system, we choose a Gaussian process ξ^x having a covariance function of the form

$$\forall t_1, t_2 \in \mathbb{R}_+, K_\xi^h(t_1, t_2) = \sigma_\xi (t_1 t_2)^\beta \exp \left(-\frac{(t_1 - t_2)^2}{2h^2} \right) \mathbf{I}_2 \in \mathcal{M}_2(\mathbb{R}) \quad (2.47)$$

where $\sigma_\xi, \beta, h > 0$. The condition (2.41) reflects the fact that the variance of the random variable $\sum_{i=1}^n y_i^\top \xi^x(t_i)$ must be non-negative. The positive definiteness of

K_ξ^h is a consequence of the positive definiteness of the radial basis kernel and of the positive definiteness of the kernel $(t_1, t_2) \mapsto (t_1 t_2)^\beta$. Notice that the covariance function is not stationary, and that $K_\xi^h(0, t) = 0$, ensuring the absence of uncertainty at $t = 0$. The meta-parameters σ_ξ, β, h are chosen in such a way that the process ξ^x is a plausible representation of the error evolution. h is the typical time scale of the correlation, which is chosen to be equal to the time step of the subdivision $h = \max\{\tau_{n+1} - \tau_n, n \in \llbracket 0; M-1 \rrbracket\}$. There is no correlation between the two components of the process ξ^x , as there are no coupling in the dynamics of the components of vector x , and the covariance matrix is diagonal. The exponent β monitors the rate of increase of the error. The following theorem, establishing the consistency of the stochastic integrator leads to a specific choice for the value of β .

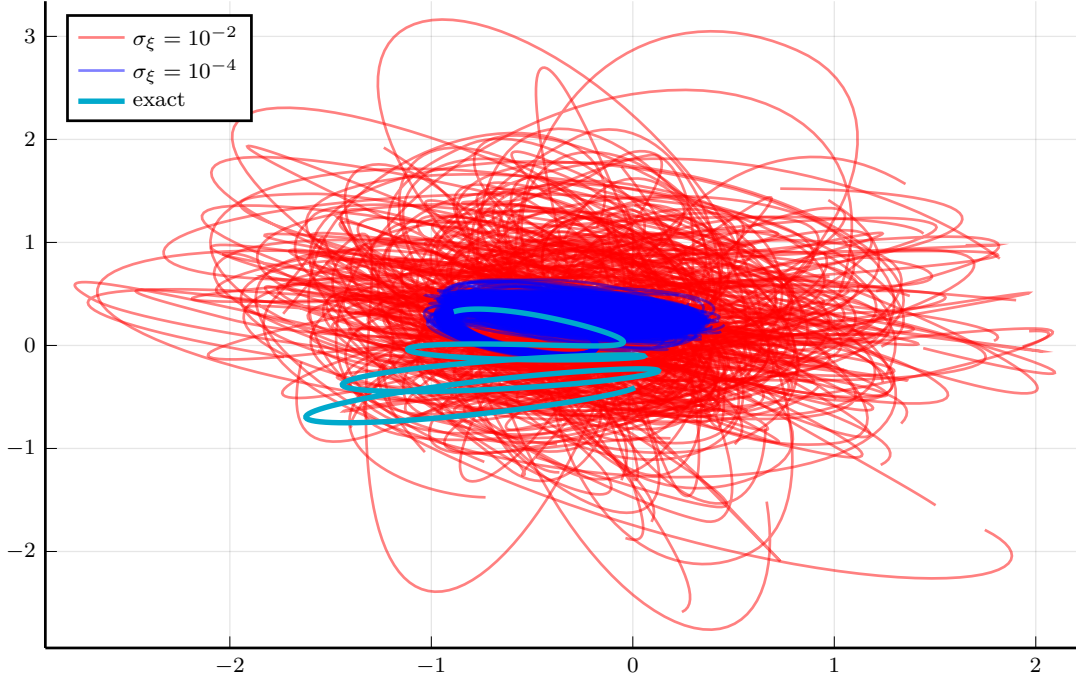


Figure 2.11: Comparison of trajectories samples for two different amplitudes of variance σ_ξ , along with the associated exact trajectory of the system (detailed configuration of the graph on page 94)

Theorem 2.3. (adapted from Conrad et al., 2015,⁷ theorem 2.2)

Let $X_{1:N} : [0; T] \rightarrow \mathbb{R}^{4N}$ be the solution of the system (1.50) and let $\tilde{X}_{1:N} : [0; T] \rightarrow \mathbb{R}^{4N}$ be the output of the stochastic integrator (2.35), associated with the Gaussian process of covariance function K_ξ^h . Then for any norm over \mathbb{R}^{4N} , there exist constants $\rho_1, \rho_2 > 0$ such that

$$\begin{aligned} \sup_{0 \leq n \leq M} \mathbb{E} \left| X_{1:N}(\tau_n) - \tilde{X}_{1:N}(\tau_n) \right|^2 &\leq \rho_1 h^{\min\{2, 2\beta-3\}} \\ \sup_{0 \leq t \leq T} \mathbb{E} \left| X_{1:N}(t) - \tilde{X}_{1:N}(t) \right| &\leq \rho_2 h^{\min\{1, (2\beta-3)/2\}} \end{aligned} \quad (2.48)$$

It follows that the stochastic trajectories converge to the true trajectories in L^1 as the subdivision gets thinner. If $\beta = 5/2$, then the rate of convergence is equal to the one obtained by the deterministic integration (2.35). More generally, if the numerical method is of order p , then the choice of the exponent should be $\beta = (2p + 3)/2$.

So β exponent monitors the local error, at the neighbourhood of the times in the subdivision. As for σ_ξ , it monitors the error of the scheme from a more global

perspective. We can see on figure 2.11 the graphical influence of the variance amplitude σ_ξ . This graph shows the necessity to calibrate parameter σ_ξ if we want the process ξ^x to be a relevant model for the numerical error. A value of σ_ξ too small leads to a biased support of the stochastic solutions that remain concentrated around the approximated deterministic trajectories. On the contrary, a value of σ_ξ too large may destructure the dynamics, and result in simulations that are not sufficiently informative for parameter inference. A methodology for the calibration of σ_ξ is proposed in Conrad et al. (2015).⁷ It consists first in estimating the numerical errors at the level of the points in the subdivision. The estimation of these numerical errors is made by comparing a deterministic trajectory with a reference solution, obtained with a thinner timestep, computed once. The parameter σ_ξ is then chosen so that the empirical variance of the stochastic solutions matches with these error estimates. The metric chosen for this matching is the Bhattacharyya distance.

Let $t \in [0; T] \mapsto \hat{X}_{1:N}^h(t)$ be the deterministic trajectory for some time step h and let $t \in [0; T] \mapsto \hat{X}_{1:N}^{h_{ref}}(t)$ be a reference trajectory obtained with a thinner time step $h_{ref} < h$. The estimation of the error at the level of the subdivision is then

$$\forall n \in \llbracket 1; M \rrbracket, \quad e_n^2 = \frac{1}{N} |\hat{X}_{1:N}^h(\tau_n) - \hat{X}_{1:N}^{h_{ref}}(\tau_n)|^2 \quad (2.49)$$

We consider a normal distribution that is supposed to represent a kind of *ideal* error distribution, that we would like to reproduce using the process ξ . This normal distribution is centered at the the current deterministic trajectory and has a diagonal covariance containing the error at the subdivision points.

$$\nu_t = \bigotimes_{n=1}^M \mathcal{N}(\hat{X}_{1:N}^h(\tau_n), e_n^2 \mathbf{I}_{4N}) \quad (2.50)$$

We then approximate the distribution of the stochastic integrator by the following Gaussian distribution, whose characteristics are computed using a sample of the stochastic solutions $(\tilde{X}_{1:N}^{(k)}(t))_{1 \leq k \leq N_{mc}}$.

$$\begin{aligned} \tilde{\mu}_{N_{mc}}(\sigma_\xi) &= \bigotimes_{n=1}^M \mathcal{N}(\tilde{m}_{N_{mc}}(\tau_n, \sigma_\xi), \tilde{\sigma}_{N_{mc}}(\tau_n, \sigma_\xi)^2 \mathbf{I}_{4N}) \\ \text{where } \tilde{m}_{N_{mc}}(\tau_n, \sigma_\xi) &= \frac{1}{N_{mc}} \sum_{k=1}^{N_{mc}} \tilde{X}_{1:N}^{(k)}(\tau_n, \sigma_\xi) \\ \tilde{\sigma}_{N_{mc}}(\tau_n, \sigma_\xi)^2 &= \frac{1}{N N_{mc}} \sum_{k=1}^{N_{mc}} |\tilde{X}_{1:N}^{(k)}(\tau_n, \sigma_\xi) - \tilde{m}_{N_{mc}}(\tau_n, \sigma_\xi)|^2 \end{aligned} \quad (2.51)$$

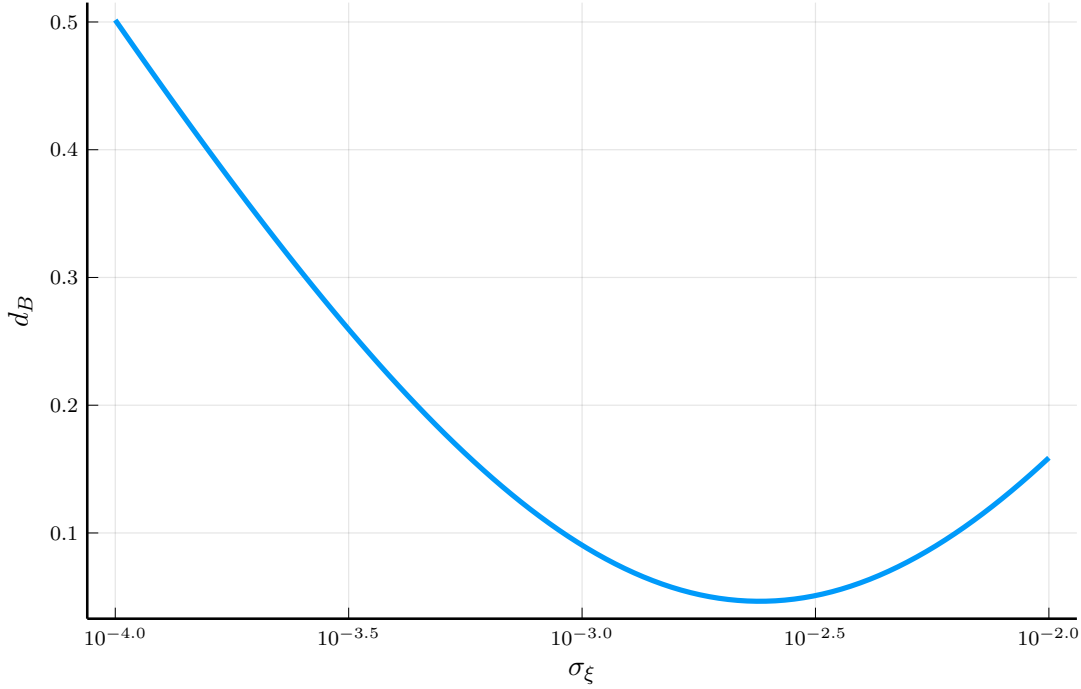


Figure 2.12: Variation of the Bhattacharyya distance according to the amplitude of variance σ_ξ (equation 2.53). The configuration used is detailed on page 94.

It is worth noting that the distribution of the stochastic integrator is indeed Gaussian, as we are dealing with a linear system. The true characteristics of this distribution are derived in the appendix, on page 94. We consider this approximation here to simplify the calibration of σ_ξ , and the true distribution of the stochastic integrator is used later on for the inference.

The amplitude of variance σ_ξ chosen for the process ξ is the one minimising the Bhattacharyya distance between distribution (2.50) and (2.51).

$$\sigma_\xi^* = \operatorname{argmin}_{\sigma_\xi > 0} d_B(\nu_t, \tilde{\mu}_{N_{mc}}(\sigma_\xi)) \quad (2.52)$$

$$d_B(\nu_t, \tilde{\mu}_{N_{mc}}(\sigma_\xi)) = \frac{1}{M} \sum_{n=1}^M \left(\frac{|\hat{X}_{1:N}(\tau_n) - \tilde{m}_{N_{mc}}(\tau_n, \sigma_\xi)|^2}{4N(e_n^2 + \tilde{\sigma}_{N_{mc}}(\tau_n, \sigma_\xi)^2)} + \frac{1}{2} \log \left(\frac{e_n^2 + \tilde{\sigma}_{N_{mc}}(\tau_n, \sigma_\xi)^2}{2e_n \tilde{\sigma}_{N_{mc}}(\tau_n, \sigma_\xi)} \right) \right) \quad (2.53)$$

We can see on figure 2.12 that the minimal distance is reached for σ_ξ around 2.36×10^{-3} , a value which is quite stable even when changing the sample and the initial configuration of the population.

After calibration of σ_ξ , we can consider that the stochastic integrator takes well into account the numerical error introduced by the scheme. The observation model associated with the stochastic integrator is obtained by replacing the true position of the particles by the stochastic trajectories.

$$x_{ij} = \tilde{x}_i(t_j, \kappa_1) + \epsilon_{ij} \text{ with } \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_2) \quad (2.54)$$

In comparison with the observation model (2.36), the accumulation of the error of the numerical scheme is taken into account in the variance of the Gaussian process $\xi^{x,n}$. The likelihood distribution of the data is expressed as an expectation with respect to the distribution $\tilde{\mu}_{\tilde{x}}[\kappa_1]$ of $(\tilde{x}_i(t_j, \kappa_1))_{\substack{1 \leq i \leq N \\ 1 \leq j \leq M}}$.

$$p_{x|\kappa_1}(x | \kappa_1) = \frac{1}{(2\pi\sigma^2)^{Nm}} \int_{\mathbb{R}^{2Nm}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N \sum_{j=1}^M |x_{ij} - \tilde{x}_{ij}|^2\right) \mu_{\tilde{x}}[\kappa_1](d\tilde{x}) \quad (2.55)$$

Due to the linearity of the system (1.38), the distribution $\mu_{\tilde{x}}[\kappa_1]$ is Gaussian, having the deterministic trajectory $\hat{x}_{N,m}(\kappa_1) = (\hat{x}_i(t_j, \kappa_1))_{\substack{1 \leq i \leq N \\ 1 \leq j \leq M}}$ as mean and $\tilde{\Sigma}_{N,m}(\kappa_1)$ as covariance matrix. The components of the matrix are derived in the appendix on page 94.

By applying formula (2.2) to the integral in (2.55) and Bayes' formula, we can obtain an expression of the posterior distribution of the parameter κ_1 .

$$\begin{aligned} & \tilde{p}_{\kappa_1|x}(\kappa_1 | x) \\ & \propto \frac{p_{\kappa_1}(\kappa_1) \exp\left(-\frac{1}{2}(\hat{x}_{N,m}(\kappa_1) - x)^\top (\sigma^2 \mathbf{I}_{2Nm} + \tilde{\Sigma}(\kappa_1))^{-1} (\hat{x}_{N,m}(\kappa_1) - x)\right)}{\sqrt{\det(\sigma^2 \mathbf{I}_{2Nm} + \tilde{\Sigma}_{N,m}(\kappa_1))}} \end{aligned} \quad (2.56)$$

As expected, the covariance matrices associated with the two sources of error add up because of their mutual independence. We can see on figure 2.13 that the posterior density $\tilde{p}_{\kappa_1|x}$ achieves a good trade-off between the exact density, which is out of reach in practice, and the density associated to the deterministic integrator, that can be biased by a too rough time-discretisation.

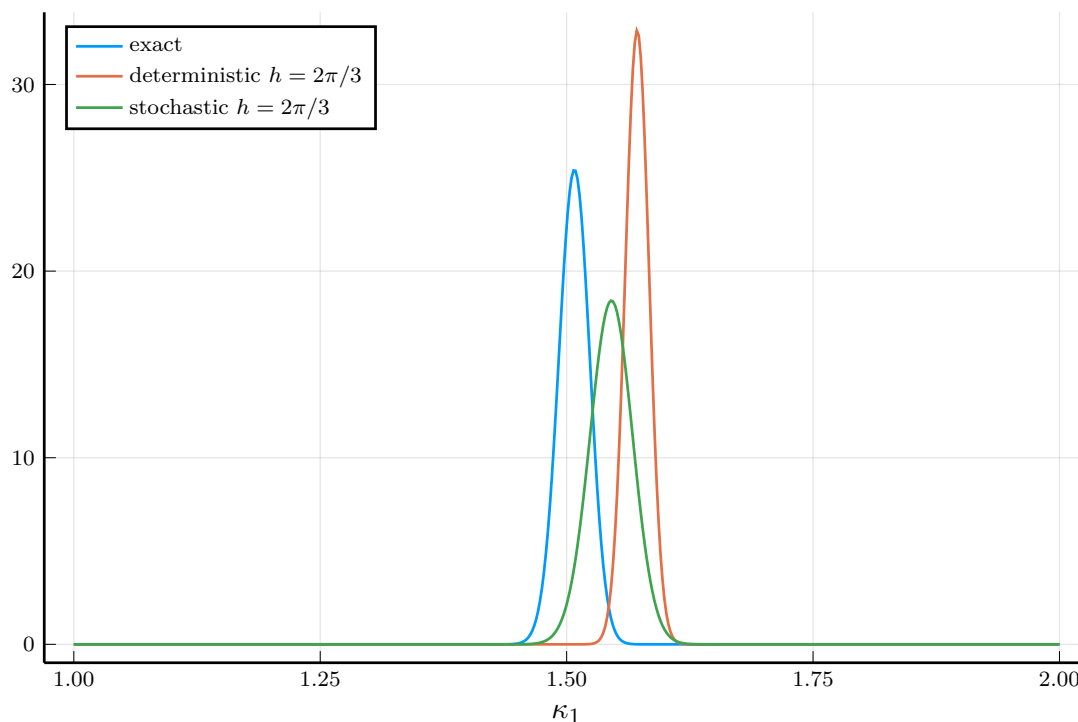


Figure 2.13: Comparison of the posterior densities associated with the exact trajectories, the deterministic and stochastic integrators.

2.8 Conclusion

Chapter 2 presents models of the different sources of uncertainty appearing on the inference problems related to symmetric and heterogeneous populations. We have illustrated these models on the Spring Cloud system, but it is clear that they are formulated identically for any other symmetric systems. In non-linear systems, posterior distributions are more challenging to access. Therefore, the central perspective of this chapter is to apply the research on Monte-Carlo algorithms, and more generally on numerical Bayesian inference, to develop a method adapted to the approximation of the posterior distribution. Within these numerical inference methods, the principal limiting factor lies in the ease of simulating the system, or at least a good approximation of it. Therefore, this factor depends on the consistency results of the numerical method mentioned in chapter 1, particularly in the choice of the time step, which must be selected optimally to guarantee the quality of the approximation and the computational efficiency. Besides, the complexity of some sources of uncertainty, especially the uncertainty on the size N of the popula-

tion, opens up to inference problems that are particularly difficult in the non-linear case, especially when the prior support of the variable N is infinite. This problem is treated in the literature by methods such as reversible-jump Markov Chain Monte-Carlo (Green, 1995¹²), and it still has many unresolved questions. It seems a too high price to pay, mainly because the size N is only a secondary importance variable. Chapter 3, therefore, aims to simplify these inference problems.

2.9 Appendix of chapter 2

2.9.1 Uncertainty on the initial condition

proof of lemma 2.2

Proof.

$$\begin{aligned}
 E &= \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}(x^\top Hx - 2y^\top x)\right) \mathcal{N}(m, \Sigma)(dx) \\
 &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left[-\frac{1}{2}((m + \Sigma^{1/2}u)^\top H(m + \Sigma^{1/2}u) - 2y^\top(m + \Sigma^{1/2}u))\right] \\
 &\times \exp\left(-\frac{1}{2}u^\top u\right) \lambda(du) \tag{2.57} \\
 &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}(u^\top(\mathbf{I}_d + \Sigma^{1/2}H\Sigma^{1/2})u - 2(y - Hm)^\top \Sigma^{1/2}u \right. \\
 &\quad \left. + m^\top Hm - 2y^\top m)\right) \lambda(du)
 \end{aligned}$$

Let us introduce the vector $\tilde{m} \in \mathbb{R}^d$ defined by

$$\tilde{m} = (\mathbf{I}_d + \Sigma^{1/2}H\Sigma^{1/2})^{-1} \Sigma^{1/2}(y - Hm) \tag{2.58}$$

Note that the inverse of the matrix $\mathbf{I}_d + \Sigma^{1/2}H\Sigma^{1/2}$ is well defined since the matrix $\Sigma^{1/2}H\Sigma^{1/2}$ is positive semi-definite, so -1 cannot be an eigenvalue of the matrix. Then using the fact that

$$\int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}(u - \tilde{m})^\top(\mathbf{I}_d + \Sigma^{1/2}H\Sigma^{1/2})(u - \tilde{m})\right) \lambda(du) = \frac{(2\pi)^{d/2}}{\sqrt{\det(\mathbf{I}_d + \Sigma^{1/2}H\Sigma^{1/2})}} \tag{2.59}$$

we obtain that

$$E = \frac{\exp\left(-\frac{1}{2}(m^\top Hm - 2y^\top m - \tilde{m}^\top(\mathbf{I}_d + \Sigma^{1/2}H\Sigma^{1/2})\tilde{m})\right)}{\sqrt{\det(\mathbf{I}_d + \Sigma^{1/2}H\Sigma^{1/2})}} \tag{2.60}$$

which leads to the result. □

Expression of the terms in equation (2.20) Let us first define the structure of the vector $X_{1:N}^0 \in \mathbb{R}^{4N}$ containing all the initial conditions. For any initial conditions (x_i^0, v_i^0) , the components of the two vectors are denoted by $x_i^0 = (x_i^{0,1}, x_i^{0,2})$ and $v_i^0 = (v_i^{0,1}, v_i^{0,2})$.

$$X_{1:N}^0 = (x_1^{0,1}, \dots, x_N^{0,1}, x_1^{0,2}, \dots, x_N^{0,2}, v_1^{0,1}, \dots, v_N^{0,1}, v_1^{0,2}, \dots, v_N^{0,2}) \tag{2.61}$$

Then we express the trajectories of the system as a linear function of the initial conditions.

$$\begin{aligned} \forall (i, j) \in \llbracket 1; N \rrbracket \times \llbracket 0; M \rrbracket, \quad x_i(t_j, \omega, x_{1:N}^0, v_{1:N}^0) &= \sum_{k=1}^N a_{ij}^k(\omega) x_k^0 + \sum_{k=1}^N b_{ij}^k(\omega) v_k^0 \\ \text{where } \forall k \in \llbracket 1; N \rrbracket, \quad a_{ij}^k(\omega) &= \left(\mathbb{I}\{k = i\} - \frac{1}{N} \right) \cos(\omega t_j) + \frac{1}{N} \\ b_{ij}^k(\omega) &= \left(\mathbb{I}\{k = i\} - \frac{1}{N} \right) \frac{\sin(\omega t_j)}{\omega} + \frac{t_j}{N} \end{aligned} \quad (2.62)$$

These can be put into a matrix form.

$$\begin{aligned} x_i(t_j, \omega, x_{1:N}^0, v_{1:N}^0) &= C_{ij}(\omega) X_{1:N}^0 \\ \text{where } C_{ij}(\omega) &= \begin{pmatrix} A_{ij}(\omega) & B_{ij}(\omega) \end{pmatrix} \in \mathcal{M}_{2,4N}(\mathbb{R}) \\ A_{ij}(\omega) &= \begin{pmatrix} a_{ij}^1(\omega) & \dots & a_{ij}^N(\omega) & 0 & \dots & 0 \\ 0 & \dots & 0 & a_{ij}^1(\omega) & \dots & a_{ij}^N(\omega) \end{pmatrix} \in \mathcal{M}_{2,2N}(\mathbb{R}) \\ B_{ij}(\omega) &= \begin{pmatrix} b_{ij}^1(\omega) & \dots & b_{ij}^N(\omega) & 0 & \dots & 0 \\ 0 & \dots & 0 & b_{ij}^1(\omega) & \dots & b_{ij}^N(\omega) \end{pmatrix} \in \mathcal{M}_{2,2N}(\mathbb{R}) \end{aligned} \quad (2.63)$$

The expressions of the matrix $C(\omega)$, of vector $Y(\omega, x)$ and of the scalar $z(x)$ are then given by

$$\begin{aligned} C(\omega) &= \sum_{i=1}^N \sum_{j=0}^M C_{ij}(\omega)^\top C_{ij}(\omega) \\ Y(\omega, x) &= \sum_{i=1}^N \sum_{j=0}^M C_{ij}(\omega)^\top x_{ij} \\ z(x) &= \sum_{i=1}^N \sum_{j=0}^M x_{ij}^\top x_{ij} \end{aligned} \quad (2.64)$$

We derive an expression of the mean and the covariance of the Gaussian variable $X_{1:N}^0$. Let us introduce the components of the mean position $\mu_x = (\mu_x^1, \mu_x^2)$ and the matrices $O_x^i, O_v^i \in \mathcal{M}_{2,4N}(\mathbb{R})$ such that for all $i \in \llbracket 1; N \rrbracket$ we have $O_x^i X_{1:N}^0 = x_i^0$ and $O_v^i X_{1:N}^0 = v_i^0$.

$$\begin{aligned} m_X &= \begin{pmatrix} \mu_x^1 \mathbb{1}_N \\ \mu_x^2 \mathbb{1}_N \\ \mu_v^1 \mathbb{1}_N \\ \mu_v^2 \mathbb{1}_N \end{pmatrix} \in \mathbb{R}^{4N} \quad \text{where } \mathbb{1}_N = (1, \dots, 1) \in \mathbb{R}^N \\ \Sigma_X &= \left(\sum_{i=1}^N (O_x^i)^\top \Sigma_x^{-1} O_x^i + (O_v^i)^\top \Sigma_v^{-1} O_v^i \right)^{-1} \end{aligned} \quad (2.65)$$

Configuration of the graph 2.5 : The observations, associated with the case where initial conditions are partially observed, are generated with the following configuration.

mean initial position and speed: μ_x, μ_v	0
covariance of μ_0 : Σ_x, Σ_v	I_2
random number generator (initial condition and observation noise)	MersenneTwister(123) Julia package <code>Random.jl</code>
observation timeline $(t_j)_{0 \leq j \leq M}$	$(\frac{2j\pi}{5})_{0 \leq j < 4}$
true value for pulsation: ω_0	1.
prior parameters: $(\omega_{\min}, \omega_{\max})$	(0.5, 1.5)
population size N	10
observation standard deviation: σ	1.

Table 2.1: Configuration of the Spring Cloud system used for the generation of the graph 2.5

The curve of label *unknown initial condition* is obtained as follows:

1. we take the logarithm of the expression of the unnormalised posterior in equation (2.22). We denote this function of ω by $\ell(\omega | x)$.
2. We consider the following integral

$$\nu = \int_{\omega_{\min}}^{\omega_{\max}} \exp(\ell(\omega | x) - \ell(\omega_0 | x)) d\omega \quad (2.66)$$

3. We compute the integral numerically using the Gauss-Konrod quadrature formula implemented in the Julia package `QuadGK.jl`, with a relative tolerance of 10^{-8} .
4. The estimation of the posterior density in equation (2.22) is then $\omega \in [\omega_{\min}; \omega_{\max}] \mapsto \frac{\exp(\ell(\omega | x) - \ell(\omega_0 | x))}{\nu}$

For the curve with label *known initial conditions*, we use the same process to represented the curve with label *differentiated* in figure 2.4.

2.9.2 Uncertainty on the population size

Expressions of the terms in the likelihood (equation 2.26) For all $N \geq N_0$, the latent variables are denoted by $X_{N_0+1:N}^0$. The initial conditions for the N_0

observed particles are gathered in a constant vector $X_{1:N_0}^0$. Both vectors have the same structure as in equation (2.61). The individual trajectories can be expressed as an affine function of the latent variables.

$$\forall (i, j) \in \llbracket 1; N \rrbracket \times \llbracket 1; M \rrbracket, x_i^N(t_j, X_{1:N}^0) = C_{ij}^{N_0, N}(\omega) X_{1:N_0}^0 + C_j^{N_0, N}(\omega) X_{N_0+1:N}^0 \quad (2.67)$$

where $C_{ij}^{N_0, N}(\omega) = (A_{ij}^{N_0, N}(\omega) \ B_j^{N_0, N}(\omega)) \in \mathcal{M}_{2, 4N_0}(\mathbb{R})$ and $C_j^{N_0, N}(\omega) = (A_j^{N_0, N}(\omega) \ B_j^{N_0, N}(\omega)) \in \mathcal{M}_{2, 4(N-N_0)}(\mathbb{R})$. The coefficients of the A, B matrices have the same expressions as in equation (2.62).

$$\begin{aligned} \forall k \in \llbracket 1; N_0 \rrbracket, a_{ij}^{k, N}(\omega) &= \left(\mathbb{I}\{k = i\} - \frac{1}{N} \right) \cos(\omega t_j) + \frac{1}{N} \\ a_j^N(\omega) &= \frac{1}{N} (1 - \cos(\omega t_j)) \\ b_{ij}^{k, N}(\omega) &= \left(\mathbb{I}\{k = i\} - \frac{1}{N} \right) \frac{\sin(\omega t_j)}{\omega} + \frac{t_j}{N} \\ b_j^N(\omega) &= \frac{1}{N} \left(t_j - \frac{\sin(\omega t_j)}{\omega} \right) \\ A_{ij}^{N_0, N}(\omega) &= \begin{pmatrix} a_{ij}^{1, N}(\omega) & \dots & a_{ij}^{N_0, N}(\omega) & 0 & \dots & 0 \\ 0 & \dots & 0 & a_{ij}^{1, N}(\omega) & \dots & a_{ij}^{N_0, N}(\omega) \end{pmatrix} \in \mathcal{M}_{2, 2N_0}(\mathbb{R}) \\ A_j^{N_0, N}(\omega) &= a_j^N(\omega) \begin{pmatrix} \mathbb{1}_{N-N_0}^\top & 0 \\ 0 & \mathbb{1}_{N-N_0}^\top \end{pmatrix} \in \mathcal{M}_{2, 2(N-N_0)}(\mathbb{R}) \\ B_{ij}^{N_0, N}(\omega) &= \begin{pmatrix} b_{ij}^{1, N}(\omega) & \dots & b_{ij}^{N_0, N}(\omega) & 0 & \dots & 0 \\ 0 & \dots & 0 & b_{ij}^{1, N}(\omega) & \dots & b_{ij}^{N_0, N}(\omega) \end{pmatrix} \in \mathcal{M}_{2, 2N_0}(\mathbb{R}) \\ B_j^{N_0, N}(\omega) &= b_j^N(\omega) \begin{pmatrix} \mathbb{1}_{N-N_0}^\top & 0 \\ 0 & \mathbb{1}_{N-N_0}^\top \end{pmatrix} \in \mathcal{M}_{2, 2(N-N_0)}(\mathbb{R}) \end{aligned} \quad (2.68)$$

The expressions of the terms in the likelihood (2.26) are then

$$\begin{aligned} z_N(x, \omega) &= \sum_{i=1}^{N_0} \sum_{j=1}^M |x_{ij} - C_{ij}^{N_0, N}(\omega) X_{1:N_0}^0|^2 \\ C_{N_0}^N(\omega) &= N_0 \sum_{j=1}^M C_j^{N_0, N}(\omega)^\top C_j^{N_0, N}(\omega) \\ Y_N(x, \omega) &= \sum_{j=1}^M C_j^{N_0, N}(\omega)^\top \sum_{i=1}^{N_0} \left(x_{ij} - C_{ij}^{N_0, N}(\omega) X_{1:N_0}^0 \right) \end{aligned} \quad (2.69)$$

$$\begin{aligned}
 m_X^{N_0, N} &= \begin{pmatrix} \mu_x^1 \mathbb{1}_{N-N_0} \\ \mu_x^2 \mathbb{1}_{N-N_0} \\ \mu_v^1 \mathbb{1}_{N-N_0} \\ \mu_v^2 \mathbb{1}_{N-N_0} \end{pmatrix} \in \mathbb{R}^{4(N-N_0)} \\
 \Sigma_X^{N_0, N} &= \left(\sum_{i=1}^{N-N_0} (O_x^{i, N-N_0})^\top \Sigma_x^{-1} O_x^{i, N-N_0} + (O_v^{i, N-N_0})^\top \Sigma_v^{-1} O_v^{i, N-N_0} \right)^{-1}
 \end{aligned} \tag{2.70}$$

where the matrices $O_x^{i, N-N_0}, O_v^{i, N-N_0} \in \mathcal{M}_{2,4(N-N_0)}(\mathbb{R})$ are such that for all $i \in \llbracket 1; N - N_0 \rrbracket$, $O_x^i X_{N_0+1:N}^0 = x_{N_0+i}^0$ and $O_v^i X_{N_0+1:N}^0 = v_{N_0+i}^0$.

Configuration for the figures 2.6 and 2.7 The observations, associated with the case where the population size is uncertain, are generated with the following configuration.

means of μ_0 μ_x, μ_v	0
covariances of μ_0 : Σ_x, Σ_v	I_2
random number generator (initial condition and observation noise)	MersenneTwister(123) (Julia package <code>Random.jl</code>)
observation timeline	$\left(\frac{2j\pi}{10} \right)_{1 \leq j \leq 10}$
true size of the population: N	25
number of observed particles N_0	10
prior parameters of p_N : (N_{\min}, N_{\max})	(11,50)
true pulsation ω	1
prior parameters of $p_{\omega N}$: $(\omega_{\min}, \omega_{\max})$	(1/10, 3/10)
observations error σ	1

The prior p_N on the population size is chosen with a finite support $\llbracket N_{\min}; N_{\max} \rrbracket$ with $N_{\min} \geq N_0 + 1$. In our case, we have chosen a uniform distribution over the discrete interval.

$$\forall N \in \llbracket N_{\min}; N_{\max} \rrbracket, \quad p_N(N) = \frac{1}{N_{\max} - N_{\min} + 1} \tag{2.71}$$

The unnormalised posterior density $\tilde{p}_{\omega|x}$, represented in graph 2.6, is computed as follows:

1. for any $\omega \in [\omega_{\min} \sqrt{N_{\min}}; \omega_{\max} \sqrt{N_{\max}}]$, we compute the vector containing the logarithms of the quantities $\tilde{p}_{\omega|N,x}(\omega|N, x)$ for all $N \in \llbracket N_{\min}; N_{\max} \rrbracket$.

$$\ell_{N_{\min}:N_{\max}}(\omega) = \left(\log(\tilde{p}_{\omega|N,x}(\omega|N, x)) \right)_{N_{\min} \leq N \leq N_{\max}} \tag{2.72}$$

2. To avoid underflow, the logarithm of $\tilde{p}_{\omega|x}(\omega|x)$ is computed as follows

$$\begin{aligned} \ell_{\max}(\omega) &= \max_{N_{\min} \leq N \leq N_{\max}} [\log(\tilde{p}_{\omega|N,x}(\omega|N, x) + \log(p_N(N)))] \\ \log(\tilde{p}_{\omega|x}(\omega|x)) &= \ell_{\max}(\omega) + \log \left(\sum_{N=N_{\min}}^{N_{\max}} \exp(\log(p_N(N)\tilde{p}_{\omega|N,x}(\omega|N, x)) - \ell_{\max}(\omega)) \right) \end{aligned} \quad (2.73)$$

3. in figure 2.6, the function $\omega \mapsto \exp(\log(\tilde{p}(\omega|x)) - \log(\tilde{p}(\omega_0|x)))$ with ω_0 is the value of the pulsation having generated the observation data, i.e. $\omega_0 = 1$.

In figure 2.7, we compute for all $N \in \llbracket N_{\min}; N_{\max} \rrbracket$ the integral

$$\nu_N(x) = p_N(N) \int_{\omega_{\min}\sqrt{N}}^{\omega_{\max}\sqrt{N}} \tilde{p}_{\omega|N,x}(\omega|N, x) d\omega \quad (2.74)$$

using Gauss-Konrod method, implemented in the package `QuadGK.jl`, with a relative tolerance of 10^{-8} . The estimation of the probability $p_{N|x}(N|x)$ is then

$$p_{N|x}(N|x) = \frac{\nu_N(x)}{\sum_{N'=N_{\min}}^{N_{\max}} \nu_{N'}(x)} \quad (2.75)$$

2.9.3 Uncertainty due to the inexact simulation of the system

Initial configuration of Spring Cloud system (figure 2.8) The initial configuration of Spring Cloud heterogeneous is sampled from the distribution μ_0 having the following parameters.

initial positions and speeds μ_0^x, μ_0^v	$\mathcal{N}(0, I_2)$
attractivities and masses of particles μ_0^κ, μ_0^m	$\mathcal{U}([1; 2])$
population size N	10
random number generator	MersenneTwister(123)
true value of parameter κ_1	1.5
observation interval	$[0; 2\pi]$

Table 2.2: Parameters monitoring the initial configuration of Spring Cloud heterogeneous

Proof of proposition 2.1

Lemma 2.3. Conservation of the energy

Let $x_{1:N}, v_{1:N}$ be the solution of the system (1.38). Then the Hamiltonian of the system

$$H : t \in [0; T] \mapsto \frac{1}{2} \sum_{i=1}^N m_i |v_i(t)|^2 + \frac{1}{2} \sum_{i < j} \kappa_i \kappa_j |x_i(t) - x_j(t)|^2 \quad (2.76)$$

is constant over time.

Proof. Let $t \in [0; T]$

$$\begin{aligned} \frac{dH}{dt}(t) &= \sum_{i=1}^N \kappa_i v_i(t) \cdot \sum_{j=1}^N \kappa_j (x_j(t) - x_i(t)) + \sum_{i < j} \kappa_i \kappa_j (x_i(t) - x_j(t)) \cdot (v_i(t) - v_j(t)) \\ &= \sum_{i < j} \kappa_i \kappa_j [v_i(t) \cdot (x_j(t) - x_i(t)) + v_j(t) \cdot (x_i(t) - x_j(t)) + (x_i(t) - x_j(t)) \cdot (v_i(t) - v_j(t))] \\ &= 0 \end{aligned} \quad (2.77)$$

□

A consequence of the energy conservation is that each particle remains at a bounded distance from the barycentre of the cloud.

Lemma 2.4. Bound on the distance to the barycentre

Let $x_{1:N}$ be a solution of the system (1.38), let $x_\kappa = \frac{\sum_{j=1}^N \kappa_j x_j}{\sum_{j=1}^N \kappa_j}$ be the barycentre of the cloud, and let H be the Hamiltonian of the system. Then for all time $t \in [0; T]$, we have

$$\sum_{j=1}^N \kappa_j |x_\kappa(t) - x_j(t)|^2 \leq \frac{4H}{\sum_{j=1}^N \kappa_j} \quad (2.78)$$

Proof. By conservation of the energy (lemma 2.3), we have for all time $t \in [0; T]$

$$\begin{aligned} \sum_{i < j} \kappa_i \kappa_j |x_i(t) - x_j(t)|^2 &\leq 2H \\ \sum_{i=1}^N \sum_{j=1}^N \kappa_i \kappa_j |x_i(t) - x_j(t)|^2 &\leq 4H \end{aligned} \quad (2.79)$$

By Leibniz's inequality, we have for all $i \in \llbracket 1; N \rrbracket$

$$\sum_{j=1}^N \kappa_j |x_\kappa(t) - x_j(t)|^2 \leq \sum_{j=1}^N \kappa_j |x_i(t) - x_j(t)|^2 \quad (2.80)$$

$$\left(\sum_{i=1}^N \kappa_i \right) \sum_{j=1}^N \kappa_j |x_\kappa(t) - x_j(t)|^2 \leq 4H \quad (2.81)$$

□

We have now all the elements to estimate the truncation error of the numerical scheme.

Proof. (of proposition 2.1) Let us consider the following norm over \mathbb{R}^4 , defined by $|x, v| = |x| + \tau_* |v|$ where $|x|, |v|$ is the standard Euclidean norm over \mathbb{R}^2 and $\tau_* > 0$ is a characteristic time. Then for $t \in [0; T]$, we have for all $i \in \llbracket 1; N \rrbracket$

$$\begin{aligned} |x_i(t) - \hat{x}_i(t)| &\leq \int_0^t |v_i(s) - \hat{v}_i(s)| ds \\ |v_i(t) - \hat{v}_i(t)| &\leq \left(\frac{\kappa_i}{m_i} \sum_{j=1}^N \kappa_j \right) \int_0^t [|x_i(s) - \hat{x}_i(s)| + |x_\kappa(s) - x_\kappa^0|] ds \\ |x_\kappa(t) - x_\kappa^0| &\leq \int_0^t \left| \frac{dx_\kappa}{ds}(s) \right| ds = \int_0^t |v_\kappa(s)| ds \\ \frac{dv_\kappa}{dt}(t) &= \sum_{j=1}^N \frac{\kappa_j^2}{m_j} (x_\kappa(t) - x_j(t)) \\ |v_\kappa(t)| &\leq |v_\kappa^0| + \sum_{j=1}^N \frac{\kappa_j^2}{m_j} \int_0^t |x_\kappa(s) - x_j(s)| ds \end{aligned} \quad (2.82)$$

By Cauchy-Schwarz inequality, we have

$$\sum_{j=1}^N \frac{\kappa_j^2}{m_j} |x_\kappa(t) - x_j(t)| \leq \sqrt{\sum_{j=1}^N \frac{\kappa_j^3}{m_j^2}} \times \sqrt{\sum_{j=1}^N \kappa_j |x_\kappa(t) - x_j(t)|^2} \quad (2.83)$$

By lemma 2.4, we have

$$\sum_{j=1}^N \frac{\kappa_j^2}{m_j} |x_\kappa(t) - x_j(t)| \leq 2 \sqrt{\frac{H \sum_{j=1}^N \kappa_j^3 / m_j^2}{\sum_{j=1}^N \kappa_j}} = A_\kappa \quad (2.84)$$

The last inequality can be seen as bound on the acceleration of the barycentre. It implies a bound on the motion of the barycenter.

$$|x_\kappa(t) - x_\kappa^0| \leq |v_\kappa^0| t + A_\kappa \frac{t^2}{2} \quad (2.85)$$

$$\begin{aligned}
 |x_i(t) - \hat{x}_i(t)| + \tau_* |v_i(t) - \hat{v}_i(t)| &\leq \int_0^t [|v_i(s) - \hat{v}_i(s)| + \tau_* \omega_i^2 |x_i(s) - \hat{x}_i(s)|] ds \\
 &+ \tau_* \omega_i^2 |v_\kappa^0| \frac{t^2}{2} + \tau_* \omega_i^2 A_\kappa \frac{t^3}{6} \\
 |X_i(t) - \hat{X}_i(t)| &\leq \max \left\{ \frac{1}{\tau_*}, \tau_* \omega_i^2 \right\} \int_0^t |X_i(s) - \hat{X}_i(s)| ds + \tau_* \omega_i^2 |v_\kappa^0| \frac{t^2}{2} + \tau_* \omega_i^2 A_\kappa \frac{t^3}{6}
 \end{aligned} \tag{2.86}$$

By Grönwall's inequality, the upper bound of $|X_i(t) - \hat{X}_i(t)|$ is solution of a linear differential equation.

$$\begin{cases} \forall t \in [0; T], & \frac{dy}{dt}(t) = \frac{y(t)}{\tau_i} + B_i t + C_i t^2 \\ y(0) = 0 \end{cases} \tag{2.87}$$

$$\text{where } \frac{1}{\tau_i} = \max \left\{ \frac{1}{\tau_*}, \tau_* \omega_i \right\}, \quad B_i = \tau_* \omega_i^2 |v_\kappa^0|, \quad C_i = \frac{\tau_* \omega_i^2 A_\kappa}{2}$$

The solution of the differential equation (3.51) is

$$\forall t \in [0; T], \quad y(t) = \tau_i [B_i ((e^{t/\tau_i} - 1) - t) + C_i ((e^{t/\tau_i} - 1) \tau_i^2 - 2\tau_i t - t^2)] \tag{2.88}$$

By Taylor's inequality, we have that, since $y(0) = 0$ and $\frac{dy}{dt}(0) = 0$,

$$\begin{aligned}
 \forall t \in [0; T], \quad y(t) &\leq \max \left\{ \left| \frac{d^2 y}{dt^2}(t) \right|, 0 \leq t \leq T \right\} \frac{t^2}{2} \\
 \frac{d^2 y}{dt^2}(t) &= B_i e^{t/\tau_i} + 2C_i \tau_i (e^{t/\tau_i} - 1) \\
 y(t) &\leq (B_i e^{T/\tau_i} + 2C_i \tau_i (e^{T/\tau_i} - 1)) \frac{t^2}{2}
 \end{aligned} \tag{2.89}$$

By defining over \mathbb{R}^{4N} the norm $|X_{1:N}| = \sum_{i=1}^N |X_i|$, we obtain the estimate of the truncation error. □

Configuration for the data generation (figure 2.10) We consider the Spring Cloud system with an initial configuration described in table 2.2. The parameters monitoring the generation of the data are enumerated in the following table.

true value of parameters κ_1	1.5
observation deviation σ	0.4
observation timeline $(t_j)_{1 \leq j \leq M}$	$\left(\frac{2\pi j}{10} \right)_{1 \leq j \leq 10}$

time step h	$2\pi/3$
sample size N_{mc}	100
random number generator	MersenneTwister(123)

Table 2.3: Configuration of the graph 2.11. The initial configuration of the system is given on table 2.2.

time step used for the reference solution h_{ref}	10^{-3}
time step of the integrator h	$2\pi/3$
sample size N_{mc}	100

Table 2.4: Configuration used for the figure 2.12. The seed of the random number generator is reset to the same value MersenneTwister(123) for each σ_ξ to ensure smooth variations of the Bhattacharyya distance.

Characteristics of the Gaussian distribution associated with the stochastic integrator Let us identify the characteristics of the distribution $\mu_{\tilde{x}}[\kappa_1]$ by first considering it as a marginal of the joint distribution $\mu_{\tilde{x}}^{N,K}$ of the variables $(\tilde{x}_i(\tilde{\tau}_1), \tilde{v}_i(\tilde{\tau}_1), \dots, \tilde{x}_i(\tilde{\tau}_K), \tilde{v}_i(\tilde{\tau}_K))_{1 \leq i \leq N}$ where $\{\tilde{\tau}_1, \dots, \tilde{\tau}_K\} = \{\tau_1, \dots, \tau_M\} \cup \{t_1, \dots, t_m\}$.

Let us first identify the mean of the stochastic integrator $\tilde{m}_{1:N}(t) = \mathbb{E} \left[\tilde{X}_{1:N}(t) \right]$. The stochastic integrator at the population level can be expressed in the following matrix form

$$\begin{aligned}
 \tilde{X}_{1:N}^0 &= X_{1:N}^0 \\
 \forall n \in \llbracket 0; M-1 \rrbracket, \forall t \in [\tau_n; \tau_{n+1}), \quad \tilde{X}_{1:N}(t) &= A(t - \tau_n) \tilde{X}_{1:N}^n + \xi_{1:N}^n(t - \tau_n) \\
 \tilde{X}_{1:N}^{n+1} &= A(\tau_{n+1} - \tau_n) \tilde{X}_{1:N}^n + \xi_{1:N}^n(\tau_{n+1} - \tau_n)
 \end{aligned} \tag{2.90}$$

The coefficients of the matrix A are defined by the linear relation in equation

(2.43).

$$\begin{aligned}
 \tilde{X}_{1:N}(t) &= (\tilde{x}_1^1(t) \ \dots \ \tilde{x}_N^1(t) \ \tilde{x}_1^2(t) \ \dots \ \tilde{x}_N^2(t) \ \tilde{v}_1^1(t) \ \dots)^\top \in \mathcal{M}_{4N,1}(\mathbb{R}) \\
 \forall t \in \mathbb{R}_+, \forall (i, j) \in \llbracket 1; N \rrbracket^2, \quad a_{ij}(t) &= \frac{\kappa_j(1 - \cos(\omega_i t))}{\sum_{\ell=1}^N \kappa_\ell} + \mathbb{I}\{i = j\} \cos(\omega_i t) \\
 a'_{ij}(t) &= \frac{\kappa_j \omega_i}{\sum_{\ell=1}^N \kappa_\ell} \sin(\omega_i t) - \mathbb{I}\{i = j\} \omega_i \sin(\omega_i t) \\
 b_i(t) &= \frac{\sin(\omega_i t)}{\omega_i} \\
 b'_i(t) &= \cos(\omega_i t) \\
 a(t) &= (a_{ij}(t))_{1 \leq i, j \leq N} \\
 d_b(t) &= \text{diag}((b_i(t))_{1 \leq i \leq N}) \\
 A(t) &= \begin{pmatrix} a(t) & 0 & d_b(t) & 0 \\ 0 & a(t) & 0 & d_b(t) \\ a'(t) & 0 & d'_b(t) & 0 \\ 0 & a'(t) & 0 & d'_b(t) \end{pmatrix} \in \mathcal{M}_{4N}(\mathbb{R})
 \end{aligned} \tag{2.91}$$

The evolution of the mean function follows exactly the same inductive equation as the deterministic integrator, and therefore the mean is equal to the deterministic trajectory.

$$\begin{aligned}
 \tilde{m}_{1:N}^0 &= X_{1:N}^0 \\
 \forall n \in \llbracket 0; M-1 \rrbracket, \forall t \in [\tau_n; \tau_{n+1}), \quad \tilde{m}_{1:N}(t) &= A(t - \tau_n) \tilde{m}_{1:N}^n \\
 \tilde{m}_{1:N}^{n+1} &= A(\tau_{n+1} - \tau_n) \tilde{m}_{1:N}^n
 \end{aligned} \tag{2.92}$$

Concerning the covariance of the process, we want to establish an expression of $\text{Cov}(\tilde{X}_{1:N}(t_1), \tilde{X}_{1:N}(t_2))$ for $t_1, t_2 \in [0; T]$. For this purpose, we introduce the covariance matrix of the uncertainty process $\xi_{1:N}$, which is computed using the formula (2.45).

$$\begin{aligned}
 k_{x,x}(t_1, t_2) &= \sigma_\xi(t_1 t_2)^{5/2} \exp\left(-\frac{(t_1 - t_2)^2}{2h^2}\right) \\
 k_{x,v}(t_1, t_2) &= \sigma_\xi t_1 (t_1 t_2)^{3/2} \frac{5h^2 + 2t_2(t_1 - t_2)}{2h^2} \exp\left(-\frac{(t_1 - t_2)^2}{2h^2}\right) \\
 k_{v,v}(t_1, t_2) &= \sigma_\xi (t_1 t_2)^{3/2} \frac{25h^4 - 4t_1 t_2 (t_1 - t_2)^2 - 2h^2(5t_1^2 - 12t_1 t_2 + 5t_2^2)}{4h^4} \exp\left(-\frac{(t_1 - t_2)^2}{2h^2}\right)
 \end{aligned} \tag{2.93}$$

$$\begin{aligned}
 K(t_1, t_2) &= \text{Cov}(\xi_{1:N}^n(t_1), \xi_{1:N}^n(t_2)) \\
 &= \begin{pmatrix} k_{x,x}(t_1, t_2)\mathbf{I}_N & 0_N & k_{x,v}(t_1, t_2)\mathbf{I}_N & 0_N \\ 0_N & k_{x,x}(t_1, t_2)\mathbf{I}_N & 0_N & k_{x,v}(t_1, t_2)\mathbf{I}_N \\ k_{x,v}(t_2, t_1)\mathbf{I}_N & 0_N & k_{v,v}(t_1, t_2)\mathbf{I}_N & 0_N \\ 0_N & k_{x,v}(t_2, t_1)\mathbf{I}_N & 0_N & k_{v,v}(t_1, t_2)\mathbf{I}_N \end{pmatrix} \in \mathcal{M}_{4N}(\mathbb{R})
 \end{aligned} \tag{2.94}$$

Let $t_1 \in [\tau_n; \tau_{n+1})$ and $t_2 \in [\tau_m; \tau_{m+1})$ for $m, n \in \llbracket 0; M-1 \rrbracket$. Let us expand the expression of the covariance.

$$\begin{aligned}
 &\text{Cov}(\tilde{X}_{1:N}(t_1), \tilde{X}_{1:N}(t_2)) \\
 &= \text{Cov}(A(t_1 - \tau_n)\tilde{X}_{1:N}^n + \xi_{1:N}^n(t_1 - \tau_n), A(t_2 - \tau_m)\tilde{X}_{1:N}^M + \xi_{1:N}^M(t_2 - \tau_m)) \\
 &= A(t_1 - \tau_n)C_{n,m}A(t_2 - \tau_m)^\top + K_{n:m}(t_1 - \tau_n)A(t_2 - \tau_m)^\top + A(t_1 - \tau_n)K_{m:n}(t_2 - \tau_m)^\top \\
 &\quad + \mathbb{I}\{m = n\}K(t_1 - \tau_n, t_2 - \tau_n)
 \end{aligned}$$

where $C_{n,m} = \text{Cov}(\tilde{X}_{1:N}^n, \tilde{X}_{1:N}^M)$, $K_{n:m}(t) = \text{Cov}(\xi_{1:N}^n(t), \tilde{X}_{1:N}^M)$

(2.95)

The covariance terms $C_{n,m}$ and $K_{n:m}$ can be determined by inductive equations. By independence of the collection $(\xi_{1:N}^n)_{0 \leq n \leq M-1}$, we have for all $n \in \llbracket 0; M-1 \rrbracket$

$$\begin{aligned}
 \forall k \leq n, K_{n:k}(t) &= 0 \\
 K_{n:n+1}(t) &= K(t - \tau_n, \tau_{n+1} - \tau_n) \\
 \forall m \in \llbracket n+1; M-1 \rrbracket, K_{n:m+1}(t) &= K_{n:m}(t)A(\tau_{m+1} - \tau_m)^\top
 \end{aligned} \tag{2.96}$$

A double induction is necessary to compute all the covariances $C_{n,m}$.

$$\begin{aligned}
 C_{0,0} &= 0 \\
 \forall n, m \in \llbracket 0; M-1 \rrbracket, \forall \epsilon_1, \epsilon_2 \in \{0, 1\}, \\
 C_{n+\epsilon_1, m+\epsilon_2} &= A(\epsilon_1(\tau_{n+1} - \tau_n))C_{n,m}A(\epsilon_2(\tau_{m+1} - \tau_m))^\top \\
 &\quad + K_{n:m}(\epsilon_1(\tau_{n+1} - \tau_n))A(\epsilon_2(\tau_{m+1} - \tau_m)) + A(\epsilon_1(\tau_{n+1} - \tau_n))K_{m:n}(\epsilon_2(\tau_{m+1} - \tau_m)) \\
 &\quad + \mathbb{I}\{m = n\}K(\tau_{n+1} - \tau_n, \tau_{n+1} - \tau_n)
 \end{aligned} \tag{2.97}$$

We use formulas (2.92) and (2.95) to obtain the distribution of the vector $\tilde{X}_{N,K} = (\tilde{X}_{1:N}(\tilde{\tau}_1)^\top, \dots, \tilde{X}_{1:N}(\tilde{\tau}_K)^\top)^\top \in \mathcal{M}_{4NK,1}(\mathbb{R})$. We note $\tilde{X}_{N,K} = \mathcal{N}(\tilde{m}_{N,K}, \tilde{\Sigma}_{N,K})$. We consider the observation matrix $O_{N,K}$ extracting the observed states from $\tilde{X}_{N,K}$, such that

$$\tilde{x}_{N,m} \sim \mathcal{N}(O_{N,K}\tilde{m}_{N,K}, O_{N,K}^\top\tilde{\Sigma}_{N,K}O_{N,K}) \tag{2.98}$$

Bibliography

- [1] Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., et al. “Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study”. In: *Proceedings of the national academy of sciences* vol. 105, no. 4 (2008), pp. 1232–1237.
- [2] Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Orlandi, A., Parisi, G., Procaccini, A., Viale, M., et al. “Empirical investigation of starling flocks: a benchmark study in collective animal behaviour”. In: *Animal behaviour* vol. 76, no. 1 (2008), pp. 201–215.
- [3] Bialek, W., Cavagna, A., Giardina, I., Mora, T., Silvestri, E., Viale, M., and Walczak, A. M. “Statistical mechanics for natural flocks of birds”. In: *Proceedings of the National Academy of Sciences* vol. 109, no. 13 (2012), pp. 4786–4791.
- [4] Bolley, F., Canizo, J. A., and Carrillo, J. A. “Stochastic mean-field limit: non-Lipschitz forces and swarming”. In: *Mathematical Models and Methods in Applied Sciences* vol. 21, no. 11 (2011), pp. 2179–2210.
- [5] Bongini, M., Fornasier, M., Hansen, M., and Maggioni, M. “Inferring interaction rules from observations of evolutive systems I: The variational approach”. In: *Mathematical Models and Methods in Applied Sciences* vol. 27, no. 5 (2017), pp. 909–951. arXiv: [1602.00342](#).
- [6] Carrillo, J. A., Pareschi, L., and Zanella, M. “Particle based gPC methods for mean-field models of swarming with uncertainty”. In: (2017). arXiv: [1712.01677](#).
- [7] Conrad, P. R., Girolami, M., Särkkä, S., Stuart, A., and Zygalakis, K. “Probability measures for numerical solutions of differential equations”. In: *arXiv preprint arXiv:1506.04592* (2015).
- [8] Cucker, F. and Smale, S. “On the mathematics of emergence”. In: *Japanese Journal of Mathematics* vol. 2 (2007), pp. 197–227.

-
- [9] Degond, P., Dimarco, G., and Mac, T. B. N. “Hydrodynamics of the Kuramoto–Vicsek model of rotating self-propelled particles”. In: *Mathematical Models and Methods in Applied Sciences* vol. 24, no. 02 (2014), pp. 277–325.
- [10] Dempster, A. P., Laird, N. M., and Rubin, D. B. “Maximum likelihood from incomplete data via the EM algorithm”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* vol. 39, no. 1 (1977), pp. 1–22.
- [11] Geman, S. and Geman, D. “Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images”. In: *Journal of Applied Statistics* vol. 20, no. 5-6 (1993), pp. 25–62.
- [12] Green, P. J. “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination”. In: *Biometrika* vol. 82, no. 4 (1995), pp. 711–732.
- [13] Hoang, L. N. *La Formule du Savoir: Une philosophie unifiée du savoir fondée sur le théorème de Bayes*. HORS COLLECTION. EDP sciences, 2018.
- [14] Horn, R. A. and Johnson, C. R. *Matrix analysis*. Cambridge university press, 2012.
- [15] Klenke, A. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2008.
- [16] Le Gall, J.-F. *Mouvement brownien, martingales et calcul stochastique*. Springer, 2013.
- [17] Lu, F., Zhong, M., Tang, S., and Maggioni, M. “Nonparametric inference of interaction laws in systems of agents from trajectory data”. In: *Proceedings of the National Academy of Sciences* vol. 116, no. 29 (2019), pp. 14424–14433.
- [18] Lv, Q., Schneider, M. K., and Pitchford, J. W. “Individualism in plant populations: using stochastic differential equations to model individual neighbourhood-dependent plant growth”. In: *Theoretical population biology* vol. 74, no. 1 (2008), pp. 74–83.
- [19] Mann, R. P. “Bayesian inference for identifying interaction rules in moving animal groups”. In: *PLoS ONE* vol. 6, no. 8 (2011).
- [20] Schneider, M. K., Law, R., and Illian, J. B. “Quantification of Neighbourhood-Dependent Plant Growth by Bayesian Hierarchical Modelling”. In: *Journal of Ecology* vol. 94, no. 2 (2006), pp. 310–321.
- [21] Shao, J. *Mathematical Statistics*. Springer Texts in Statistics. New York, NY: Springer New York, 2003. arXiv: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).

- [22] Vecil, F., Lafitte, P., and Linares, J. R. “A numerical study of attraction/repulsion collective behavior models: 3D particle analyses and 1D kinetic simulations”. In: *Physica D: Nonlinear Phenomena* vol. 260 (2013), pp. 127–144.

Chapter 3

Mean-field approximated inference

Disappear here.

–Bret Easton Ellis, *LESS THAN ZERO* (1985)

3.1 Introduction

The previous chapter highlights the difficulties arising in the context of statistical inference on symmetrical models. The uncertainty about the population's size and the unobserved individuals' states leads to many latent variables. The individuals' interdependence may hinder the scaling up of the inference algorithms for a large population. This chapter considers the asymptotic analysis of symmetrical population models when their sizes tend to infinity. This analysis aims to find an approximation of the population's overall motion when it is crowded enough. The required approximation should be independent of the unknown size N , or at least have a dependency with respect to N much simpler than in the microscopic case. In symmetrical population models, such approximation is inevitably linked to the notion of mean-field distribution. For some systems, it is possible to prove that the empirical measure of population, which is a stochastic measure, depending on the realization of the system's initial configuration, converges to a deterministic probability measure when the population becomes infinitely crowded. This deterministic measure is referred to as the mean-field limit distribution or the mean-field distribution. It can be interpreted from a mesoscopic perspective: it represents the probability of finding at a given time an individual in a given state, and at the same time, it represents the entire population, that is, the source of the interactions governing the trajectories of each individual. Formally, the mean-field limit appears as the solution of a partial differential equation of the hyperbolic

type, whose velocity field is closely related to the underlying microscopic system's transition function.

The mean-field limit, when it is well-defined, seems to solve at the same time the two statistical difficulties mentioned above. Under this regime, the population dynamics do not depend on N , and all the latent variables related to the unobserved individuals are naturally incorporated into this probability distribution representing the entire population. Besides, the interdependence between individuals disappears since trajectories are now only a function of their respective initial conditions and this mean-field distribution following a deterministic time-evolution.

This phenomenon of the evanescence of the correlation between individuals was first noticed in systems derived from statistical physics. Kadanoff (2009)¹⁵ traces the idea back to the work of Weiss (1907)³² on the ferromagnetic properties of some materials. As several formalisms traditionally used in statistical physics have been extended to machine learning problems (Kindermann and Snell, 1980¹⁶), the notion of *mean-field* was integrated into the statistical lexicography, but with a meaning a bit different from that previously defined. In the statistical literature, in general, mean-field inference is a synonym of variational Bayes approximation, which consists in replacing a complex distribution, that is too difficult to handle within the inference process by a tractable and factorized distribution (Smidl and Quinn, 2006²⁵). This methodology is applied both to inference on dynamical systems, as in Vrettas et al. (2015)³¹ or Gorbach et al. (2017),¹³ and to inverse problems without the notion of time, like for instance, image restoration (Marnissi et al., 2017²²). Therefore, the link with the physical phenomenon is in the search for an alternative probability distribution for which the model's components are independent, despite their total interdependence originally. Such a distribution is generally obtained by projecting in the Kullback-Leibler sense the original distribution onto a class of factorized and tractable distributions. However, the gap between the approximation space and the initial space is, up to our knowledge, rarely investigated.

In parallel, the mean-field formalism has been extended to plasma physics by Vlasov (1968)³⁰ and more recently to macroscopic systems constituted by birds, fishes, sheep, ..., including the aforementioned Cucker-Smale model (Carrillo et al., 2010⁴). Mean-field limits can also be found in kinetic equations evolving in phase spaces other than position-velocity, notably in order to model the plasticity of natural neural networks in Perthame et al. (2017),²³ or the formation of opinions in Boudin and Salvarini (2010).³ These systems are generally much smaller in size than the systems studied in statistical physics, and the authors use the mean-field approximation to obtain a qualitative intuition of the macroscopic dynamics of the population.

Up to our knowledge, the mean-field limit, in its native meaning, has never been used for the simplification of inference problems on symmetrical population models, at least in a Bayesian framework. In Bongini et al. (2017),² as well as in Lu et al. 2019)²¹ and Lu et al. (2020),²⁰ the authors suggest non-parametric methods for estimating the interaction kernel using trajectories that are exactly observed over the entire population. The mean-field limit of the microscopic systems is used to check the consistency and the well-posedness of the learning procedure. In other words, the authors try to answer the question: “Does the sequence of estimators formed by this method have the same limit as the microscopic system when the size of the population tends towards infinity ?”

In this chapter, we wish to bridge the existing gap between, on the one hand, a rigorous analysis of the asymptotic behavior of large populations, and on the other hand, the approximations of inference allowing their efficient resolution. Therefore, our problem is to study how symmetric population models’ asymptotic analysis can provide consistent approximations to the Bayesian inference problems we considered in chapter 2.

In section 2, we start by defining the notion of weak solutions for non-local transport equations, and we prove that the dynamics of the symmetric population can be equivalently written as a transport equation on the empirical population measure. Section 3 deals with the existence and uniqueness of the mean-field limit, defined in the case where the velocity field on the population’s empirical measure has a pointwise limit when $N \rightarrow \infty$. In the eventuality where this pointwise limit is not defined, it is still possible in most cases to study the limit obtained after the renormalization of the time and state variables of the system. Section 4 links the propagation of chaos, i.e., the convergence of the population’s empirical measure towards the mean-field distribution and simplifying an inference problem with unknown population size. In particular, we discuss how the mean-field limit can provide consistent approximations in some cases, but we also notice that in other cases, it is necessary to develop further (a bit like in the case of Taylor’s expansion) to have an approximation with a simplified dependence with respect to N . Nevertheless, the mean-field limit remains an essential step in simulating large populations from a macroscopic perspective. The statistical approximations suggested in section 4 are effectively simplifications only if the mean-field transport equation’s solution is numerically attainable. This issue is addressed in section 5, which gives a preliminary work to simulate the mean-field dynamics efficiently in the Spring-Cloud and Schneider systems cases.

3.2 Empirical measure and population distribution dynamics

Within symmetric populations, the labels of the individuals are arbitrary. Therefore the evolution of the system can be described equivalently in the quotient space $\mathcal{Z}^N/\mathfrak{S}_N$ or in $\mathcal{P}_N(\mathcal{Z})$, the set of empirical measures with N Dirac masses (using proposition 1.1). In $\mathcal{P}_N(\mathcal{Z})$, we can describe the dynamics of empirical measures using partial differential equations, satisfied in the weak sense. We need to define a series of notions to characterize the continuity of trajectories taking place in the set of probability measures. In keeping with the literature related to kinetic equations theory (Golse, 2013,¹² Bolley et al., 2011,¹ Lagoutière and Vauchelet, 2017¹⁸), this regularity is characterized in terms of Wasserstein distance.

Definition 3.1. *Wasserstein distance (Villani, 2009,²⁹ definition 6.1)*

Let (\mathcal{Z}, d) be a Polish metric space. For any probability measures $\mu, \nu \in \mathcal{P}(\mathcal{Z})$, we define the set of couplings $\Pi(\mu, \nu)$ as the set of probability measures over \mathcal{Z}^2 having their first marginal equal to μ and their second marginal equal to ν .

$$\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathcal{Z}^2) \mid \mu(dz_1) = \int_{\mathcal{Z}} \pi(dz_1, dz_2), \nu(dz_2) = \int_{\mathcal{Z}} \pi(dz_1, dz_2) \right\} \quad (3.1)$$

The Wasserstein distance of order p is then defined using the coupling minimizing the expected distance to the power $p \geq 1$.

$$\mathcal{W}_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{Z}^2} d(z_1, z_2)^p \pi(dz_1, dz_2) \right)^{1/p} \quad (3.2)$$

The verification of the fact that this quantity satisfies the axioms of a distance can be found in Villani (2009),²⁹ page 106.

Example 3.1. *When restricted to the set of empirical measures having the same number of Dirac masses, the Wasserstein distance can be expressed as the solution of an optimization problem over the set of permutations.*

$$\forall (x_i, y_i)_{1 \leq i \leq N} \in \mathcal{Z}^{2N}, \quad \mathcal{W}_p \left(\frac{1}{N} \sum_{i=1}^N \delta_{x_i}, \frac{1}{N} \sum_{i=1}^N \delta_{y_i} \right) = \left(\frac{1}{N} \inf_{\sigma \in \mathfrak{S}_N} \sum_{i=1}^N d(x_i, y_{\sigma(i)})^p \right)^{1/p} \quad (3.3)$$

The distance is reached for the permutation transforming the cloud of points y into x with minimal displacement. However, this intuition of the Wasserstein distance is not true in general, as there are cases where a transformation T satisfying $T\#\mu = \nu$ (the pushforward image of μ by T is ν) cannot exist. For instance, when $\mu = \delta_x$ and $\nu = \delta_{x_1} + \delta_{x_2}$, because we cannot have $T(x) = x_1$ and $T(x) = x_2$ if $x_1 \neq x_2$.

The Wasserstein distance is associated with a convergence, which is a bit stronger than the weak convergence.

Definition 3.2. Weak convergence, or convergence in distribution

Let \mathcal{Z} be a Polish metric space, $(\mu_n)_{n \in \mathbb{N}}$ a sequence of probability measures in $\mathcal{P}(\mathcal{Z})$, and $\mu \in \mathcal{P}(\mathcal{Z})$. Then we say that the sequence $(\mu_n)_{n \in \mathbb{N}}$ converges to μ in distribution or weakly if for all $\varphi \in \mathcal{C}_b(\mathcal{Z} \rightarrow \mathbb{R})$ continuous and bounded we have

$$\int_{\mathcal{Z}} \varphi(z) \mu_n(dz) \xrightarrow{n \rightarrow \infty} \int_{\mathcal{Z}} \varphi(z) \mu(dz) \quad (3.4)$$

We then note $\mu_n \xrightarrow[n \rightarrow \infty]{w} \mu$.

Definition 3.3. Set of probability having moments of order p

Let (\mathcal{Z}, d) be a Polish metric space, let $z_0 \in \mathcal{Z}$, and let $p \geq 0$. The set $\mathcal{P}_p(\mathcal{Z})$ of probability measures having moments of order p is defined by

$$\mathcal{P}_p(\mathcal{Z}) = \left\{ \mu \in \mathcal{P}(\mathcal{Z}) \mid \int_{\mathcal{Z}} d(z_0, z)^p \mu(dz) < +\infty \right\} \quad (3.5)$$

By triangular inequality this definition is independent on the choice of the origin z_0 .

Theorem 3.1. Convergence associated with the Wasserstein distance (Villani, 2009,²⁹ theorem 6.9)

Let $(\mu_n)_{n \in \mathbb{N}}$ be a sequence of probability measures in $\mathcal{P}_p(\mathcal{Z})$ and $\mu \in \mathcal{P}_p(\mathcal{Z})$ defined over a Polish metric space (\mathcal{Z}, d) . Then the two statements are equivalent:

$$(i) \mu_n \xrightarrow[n \rightarrow \infty]{w} \mu \text{ and } \int_{\mathcal{Z}} d(z_0, z)^p \mu_n(dz) \xrightarrow{n \rightarrow \infty} \int_{\mathcal{Z}} d(z_0, z)^p \mu(dz)$$

and

$$(ii) \mathcal{W}_p(\mu_n, \mu) \xrightarrow[n \rightarrow \infty]{} 0$$

The metric \mathcal{W}_1 is used especially used here, because of its dual formulation, which is particularly convenient for proofs.

Theorem 3.2. Kantorovich-Rubinstein formula (Villani, 2009,²⁹ particular case 5.16)

Let $\mathcal{P}_1(\mathcal{Z})$ be the set of probability measures having first order moments, i.e. for all $\mu \in \mathcal{P}_1(\mathcal{Z})$ and for some $z_0 \in \mathcal{Z}$ (and therefore for any z_0)

$$\int_{\mathcal{Z}} d(z_0, z) \mu(dz) < \infty \quad (3.6)$$

Let $\mathcal{C}_L(\mathcal{Z} \rightarrow \mathbb{R})$ be the space of Lipschitz-continuous functions over \mathcal{Z} and let us denote $\text{Lip}(\phi)$ the Lipschitz constant of any element $\phi \in \mathcal{C}_L(\mathcal{Z} \rightarrow \mathbb{R})$. Then, for any $\mu_1, \mu_2 \in \mathcal{P}_1(\mathcal{Z})$:

$$\mathcal{W}_1(\mu_1, \mu_2) = \sup \left\{ \left| \int_{\mathcal{Z}} \phi(z) \mu_1(dz) - \int_{\mathcal{Z}} \phi(z) \mu_2(dz) \right|, \phi \in \mathcal{C}_L(\mathcal{Z} \rightarrow \mathbb{R}), \text{Lip}(\phi) \leq 1 \right\} \quad (3.7)$$

The space of test functions used to define weak solutions of transport equations is composed of maps differentiable with respect to time and with respect the state variable x .

$$\begin{aligned} \mathcal{D}(\mathbb{R}_+ \times \mathcal{X} \times \Theta) = & \left\{ \varphi \in \mathcal{C}_b^0(\mathbb{R}_+ \times \mathcal{X} \times \Theta \rightarrow \mathbb{R}) \mid \right. \\ & \forall (t, x, \theta) \in \mathbb{R}_+ \times \mathcal{X} \times \Theta, \frac{\partial \varphi}{\partial t}(t, x, \theta) \text{ and } \frac{\partial \varphi}{\partial x}(t, x, \theta) \text{ exists} \\ & \left. \text{and } \sup_{(t, x, \theta) \in \mathbb{R}_+ \times \mathcal{X} \times \Theta} \left| \frac{\partial \varphi}{\partial t}(t, x, \theta) \right| + \left| \frac{\partial \varphi}{\partial x}(t, x, \theta) \right| < \infty \right\} \quad (3.8) \end{aligned}$$

Notice that we do not assume the test function to be differentiable with respect to the individual parameter θ , as it remains constant through time. The dynamics of the empirical measure in symmetric populations are given by the following definition.

Definition 3.4. measure transport equation

Let $\mu_0 \in \mathcal{P}_p(\mathcal{X} \times \Theta)$ for some $p \geq 1$, $t \in \mathbb{R}_+ \mapsto \mu(t) \in \mathcal{P}_p(\mathcal{X} \times \Theta)$ a trajectory within the set of probability measures, and $F : \mathbb{R}_+ \times \mathcal{X} \times \Theta \times \mathcal{P}_p(\mathcal{X} \times \Theta) \rightarrow \mathcal{X}$ a functional. Then we say that $t \mapsto \mu(t)$ is a solution in $\mathcal{P}_p(\mathcal{X} \times \Theta)$ of the transport equation of transition F and of initial condition $(0, \mu_0)$, and we denote

$$\begin{cases} \forall t \geq 0, \frac{\partial \mu}{\partial t}(t, dx, d\theta) + \text{div}_x [\mu(t, dx, d\theta) F(t, x, \theta, \mu(t))] = 0 \\ \mu(0, dx, d\theta) = \mu_0(dx, d\theta) \end{cases} \quad (3.9)$$

if the following statements are true:

1. $t \mapsto \mu(t)$ is continuous with respect to the metric \mathcal{W}_p .
2. the function $(t, x, \theta) \in \mathbb{R}_+ \times \mathcal{X} \times \Theta \mapsto F(t, x, \theta, \mu(t))$ is continuous, and $t \in \mathbb{R}_+ \mapsto \int_{\mathcal{X} \times \Theta} F(t, x, \theta, \mu(t)) \mu(t, dx, d\theta)$ is also continuous.

3. for all $\varphi \in \mathcal{D}(\mathbb{R}_+ \times X \times \Theta)$ we have

$$\begin{aligned} & \int_{\mathcal{X} \times \Theta} \varphi(t, x, \theta) \mu(t, dx, d\theta) - \int_{\mathcal{X} \times \Theta} \varphi(0, x, \theta) \mu_0(dx, d\theta) \\ &= \int_0^t \int_{\mathcal{X} \times \Theta} \left(\frac{\partial \varphi}{\partial s}(s, x, \theta) + \frac{\partial \varphi}{\partial x}(s, x, \theta) \cdot F(s, x, \theta, \mu(s)) \right) \mu(s, dx, d\theta) ds \end{aligned} \quad (3.10)$$

The expression of the partial differential equation (3.9) is the same as the one being satisfied by the density of the measure, if the latter is smooth enough. This statement can be derived by direct integration by part of equation (3.10). We apply this formalism to describe the dynamics of the empirical measure of the population.

Proposition 3.1. Transport of the empirical measure

Let $(x_i^0, \theta_i)_{1 \leq i \leq N}$ be an initial configuration of the symmetric population system

$$\forall i \in \llbracket 1; N \rrbracket, \quad \begin{cases} x_i(0) = x_i^0 \\ \forall t \geq 0, \quad \frac{dx_i}{dt}(t) = h_N \left(t, x_i(t), \theta_i, \frac{1}{N} \sum_{j=1}^N \delta_{(x_j(t), \theta_j)} \right) \end{cases} \quad (3.11)$$

leading to a global and continuously differentiable $t \in \mathbb{R}_+ \mapsto x_{1:N}(t)$. We assume that the function $H_N : (t, x_1, \theta_1, \dots, x_N, \theta_N) \in \mathbb{R}_+ \times (\mathcal{X} \times \Theta)^N \mapsto$

$h_N \left(t, x_1, \theta_1, \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, \theta_i)} \right)$ is continuous. Then the empirical measure of the

population $t \mapsto \hat{\mu}_N(t) = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i(t), \theta_i)}$ is a solution in $\mathcal{P}_1(\mathcal{X} \times \Theta)$ of the transport equation

$$\begin{cases} \hat{\mu}_N(0) = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i^0, \theta_i)} \\ \forall t \geq 0, \quad \frac{\partial \hat{\mu}_N}{\partial t}(t, dx, d\theta) + \operatorname{div}_x(\hat{\mu}_N(t, dx, d\theta) h_N(t, x, \theta, \hat{\mu}_N(t))) = 0 \end{cases} \quad (3.12)$$

Proof. Let $\phi \in \mathcal{C}_L(\mathcal{X} \times \Theta \rightarrow \mathbb{R})$ a Lipschitz function. We have for all $t_1, t_2 \in \mathbb{R}_+$

$$\begin{aligned} & \left| \int_{\mathcal{X} \times \Theta} \phi(x, \theta) \hat{\mu}_N(t_1, dx, d\theta) - \int_{\mathcal{X} \times \Theta} \phi(x, \theta) \hat{\mu}_N(t_2, dx, d\theta) \right| \\ &= \frac{1}{N} \left| \sum_{i=1}^N (\phi(x_i(t_1), \theta) - \phi(x_i(t_2), \theta)) \right| \leq \frac{\operatorname{Lip}(\phi)}{N} \sum_{i=1}^N |x_i(t_1) - x_i(t_2)| \end{aligned} \quad (3.13)$$

Therefore using theorem 3.2

$$\mathcal{W}_1(\hat{\mu}_N(t_1), \hat{\mu}_N(t_2)) \leq \frac{1}{N} \sum_{i=1}^N |x_i(t_1) - x_i(t_2)| \quad (3.14)$$

So the trajectory $t \in \mathbb{R}_+ \mapsto \hat{\mu}_N(t)$ is continuous with respect to \mathcal{W}_1 . Besides, as the trajectories $x_{1:N}$ are continuously differentiable with respect to time, we

clearly have that $(t, x, \theta) \mapsto h_N(t, x, \theta, \hat{\mu}_N(t))$ and $t \mapsto \frac{1}{N} \sum_{i=1}^N h_N(t, x_i(t), \theta_i, \hat{\mu}_N(t))$

are continuous. We also have that

$$\begin{aligned} \frac{d}{dt} \int_{\mathcal{X} \times \Theta} \varphi(t, x, \theta) \hat{\mu}_N(t, dx, d\theta) &= \frac{1}{N} \sum_{i=1}^N \left(\frac{\partial \varphi}{\partial t}(t, x_i(t), \theta_i) + \frac{\partial \varphi}{\partial x}(t, x_i(t), \theta_i) \cdot \frac{dx_i}{dt}(t) \right) \\ &= \int_{\mathcal{X} \times \Theta} \left(\frac{\partial \varphi}{\partial t}(t, x, \theta) + \frac{\partial \varphi}{\partial x}(t, x, \theta) \cdot h_N(t, x, \theta, \hat{\mu}_N(t)) \right) \hat{\mu}_N(t, dx, d\theta) \end{aligned} \quad (3.15)$$

□

The existence of a global solution to the system implies a partial differential equation on the empirical measure of the population. Thus, the equation (3.12) can be seen as a rewriting of the differential system (3.11), but where the individual labels are not taken into account. In the specific case of the Spring Cloud system (example 1.6), the empirical measure satisfies the following partial differential equation

$$\begin{aligned} \frac{\partial \hat{\mu}_N}{\partial t}(t, dx, dv, d\kappa, dm) + v \cdot \frac{\partial \hat{\mu}_N}{\partial x}(t, dx, dv, d\kappa, dm) \\ + \frac{N\kappa}{m} \frac{\partial \hat{\mu}_N}{\partial v}(t, dx, dv, d\kappa, dm) \cdot \int_{\mathbb{R}^3} \kappa'(x' - x) \hat{\mu}_N^{x, \kappa}(t, dx', d\kappa') = 0 \end{aligned} \quad (3.16)$$

We can notice that the transport equation is nonlinear with respect to the measure $\hat{\mu}_N$, although the associated differential system is linear with respect to the state variables. For the Schneider system, the dynamical equation of the empirical measure has a slightly different structure. This dynamical equation can be derived from equations (1.47) and (1.48).

$$\begin{aligned} \frac{\partial \hat{\mu}_N}{\partial t}(t, ds, dx, dS, d\gamma) \\ + \frac{\partial}{\partial s} \left(\hat{\mu}_N(t, ds, dx, dS, d\gamma) \int_{\mathbb{R}_+ \times \mathbb{R}^2} g_N(s, x, S, \gamma, s', x') \hat{\mu}_N^{s, x}(t, ds', dx') \right) = 0 \end{aligned} \quad (3.17)$$

$$\text{where } g_N(s, x, S, \gamma, s', x') = \frac{N}{N-1}g(s, x, S, \gamma, s', x') - \frac{g(s, x, S, \gamma, s, x)}{N-1} \quad (3.18)$$

$$g(s, x, S, \gamma, s', x') = \gamma s(\log(S/s_m)(1 - C(s, s', |x - x'|)) - \log(s/s_m))$$

The time evolution of the empirical measure, starting from a specific configuration of the population must be distinguished from the time evolution of the population distribution $t \in \mathbb{R}_+ \mapsto \mu_{1:N}(t) \in \mathcal{P}((\mathcal{X} \times \Theta)^N)$, which is a deterministic object. $\mu_{1:N}(t)$ is the joint distribution of the states and parameters of the individuals in the population. When the population model is symmetric, this distribution is invariant by permutation. It can be proved that this distribution satisfies also a transport equation, with velocity field expressed with the transition function of the system.

Proposition 3.2. Transport of the population distribution

Let $\mu_0 \in \mathcal{P}_1(\mathcal{X} \times \Theta)$ and $h_N : \mathbb{R}_+ \times \mathcal{X} \times \Theta \times \mathcal{P}_N(\mathcal{X} \times \Theta) \rightarrow \mathcal{X}$ be such that $(t, x_1, \theta_1, \dots, x_N, \theta_N) \mapsto h_N\left(t, x_1, \theta_1, \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, \theta_i)}\right)$ is continuous. It is also assumed that there exists $K > 0$ such that for all $t, x_1, \theta_1, \dots, x_N, \theta_N$, we have

$$h_N\left(t, x_1, \theta_1, \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, \theta_i)}\right) \leq K \left(1 + \sum_{i=1}^N |x_i|\right) \quad (3.19)$$

Then the population distribution $t \in \mathbb{R}_+ \mapsto \mu_{1:N}(t)$ is a solution in $\mathcal{P}_1(\mathcal{X} \times \Theta)$ of the transport equation

$$\begin{aligned} & \frac{\partial \mu_{1:N}}{\partial t}(t, dx_{1:N}, d\theta_{1:N}) \\ & + \sum_{i=1}^N \operatorname{div}_{x_i} \left(\mu_{1:N}(t, dx_{1:N}, d\theta_{1:N}) h_N \left(t, x_i, \theta_i, \frac{1}{N} \sum_{j=1}^N \delta_{(x_j, \theta_j)} \right) \right) = 0 \end{aligned} \quad (3.20)$$

Proof. Let us start by proving that for all $t \geq 0$, $\mu_{1:N}(t) \in \mathcal{P}_1((\mathcal{X} \times \Theta)^N)$. For any initial configuration $(x_{1:N}^0, \theta_{1:N})$ we have

$$\begin{aligned} |x_1(t, x_{1:N}^0, \theta_{1:N})| & \leq |x_1^0| + K \int_0^t \left(1 + \sum_{i=1}^N |x_i(s, x_{1:N}^0, \theta_{1:N})| \right) ds \\ \text{therefore: } \int_{(\mathcal{X} \times \Theta)^N} |x_1(t, x_{1:N}^0, \theta_{1:N})| \mu_0^{\otimes N}(dx_{1:N}, d\theta_{1:N}) & \leq \int_{\mathcal{X} \times \Theta} |x| \mu_0(dx) \quad (3.21) \\ & + K \int_0^t \left(1 + N \int_{(\mathcal{X} \times \Theta)^N} |x_1(s, x_{1:N}^0, \theta_{1:N})| \mu_0^{\otimes N}(dx_{1:N}, d\theta_{1:N}) \right) ds \end{aligned}$$

By Grönwall lemma, we obtain that

$$\int_{(\mathcal{X} \times \Theta)^N} |x_1(t, x_{1:N}^0, \theta_{1:N})| \mu_0^{\otimes N}(dx_{1:N}, d\theta_{1:N}) \leq \frac{1}{N} \left(\left(N \int_{\mathcal{X}} |x| \mu_0^x(dx) + 1 \right) e^{Kt} - 1 \right) \quad (3.22)$$

Let $\phi \in \mathcal{C}_L((\mathcal{X} \times \Theta)^N \rightarrow \mathbb{R})$ a Lipschitz continuous function and let $t_1, t_2 \in \mathbb{R}_+$.

$$\begin{aligned}
 & \left| \int_{(\mathcal{X} \times \Theta)^N} \phi(x_{1:N}, \theta_{1:N}) \mu_{1:N}(t_1, dx_{1:N}, d\theta_{1:N}) - \int_{(\mathcal{X} \times \Theta)^N} \phi(x_{1:N}, \theta_{1:N}) \mu_{1:N}(t_2, dx_{1:N}, d\theta_{1:N}) \right| \\
 & \leq \text{Lip}(\phi) \sum_{i=1}^N \int_{(\mathcal{X} \times \Theta)^N} |x_i(t_1, x_{1:N}^0, \theta_{1:N}) - x_i(t_2, x_{1:N}^0, \theta_{1:N})| \mu_0^{\otimes N}(dx_{1:N}, d\theta_{1:N}) \\
 & \text{thus: } \mathcal{W}_1(\mu_{1:N}(t_1), \mu_{1:N}(t_2)) \\
 & \leq N \int_{(\mathcal{X} \times \Theta)^N} |x_1(t_1, x_{1:N}^0, \theta_{1:N}) - x_1(t_2, x_{1:N}^0, \theta_{1:N})| \mu_0^{\otimes N}(dx_{1:N}, d\theta_{1:N})
 \end{aligned} \tag{3.23}$$

The above upper-bound tends to zero when t_1 tends to t_2 by dominated convergence. It follows that $t \mapsto \mu_{1:N}(t)$ is continuous with respect to the metric \mathcal{W}_1 . The function $(t, x_1, \theta_1, \dots, x_N, \theta_N) \mapsto h_N\left(t, x_1, \theta_1, \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, \theta_i)}\right)$ is continuous by assumption. The function $t \mapsto \int_{(\mathcal{X} \times \Theta)^N} h_N\left(t, x_1, \theta_1, \frac{1}{N} \sum_{i=1}^N \delta_{(x_i, \theta_i)}\right) \mu_{1:N}(t, dx_{1:N}, d\theta_{1:N})$ is continuous by dominated convergence. Finally, we have for all $t \in \mathbb{R}_+$

$$\begin{aligned}
 & \frac{d}{dt} \int_{\mathcal{X}^N \times \Theta^N} \varphi(t, x_{1:N}, \theta_{1:N}) \mu_{1:N}(t, dx_{1:N}, d\theta_{1:N}) = \int_{\mathcal{X}^N \times \Theta^N} \left(\frac{\partial \varphi}{\partial t}(t, x_{1:N}(t, x_{1:N}^0, \theta_{1:N}), \theta_{1:N}) \right. \\
 & \left. + \sum_{i=1}^N \frac{\partial \varphi}{\partial x_i}(t, x_{1:N}(t, x_{1:N}^0, \theta_{1:N}), \theta_{1:N}) \cdot h_N(t, x_i(t, x_{1:N}^0, \theta_{1:N}), \theta_i, \hat{\mu}_N(t, x_{1:N}^0, \theta_{1:N})) \right) \\
 & \times \mu_0^{\otimes N}(dx_{1:N}^0, d\theta_{1:N})
 \end{aligned} \tag{3.24}$$

□

In the case of the Spring Cloud system, when the initial distribution is such that $\int_{\mathbb{R}_+} \kappa \mu_0^\kappa(d\kappa) = +\infty$, the transport equation of the population is not defined. However, if μ_0 has compact support for the variables κ, m , with $m \geq m_{\min} > 0$, and if μ_0 has first order moment with respect to the variable x and v , then the population distribution satisfies the following partial differential equation.

$$\begin{aligned}
 & \frac{\partial \mu_{1:N}}{\partial t}(t, dx_{1:N}, dv_{1:N}, d\kappa_{1:N}, dm_{1:N}) + \sum_{i=1}^N v_i \cdot \frac{\partial \mu_{1:N}}{\partial x_i}(t, dx_{1:N}, dv_{1:N}, d\kappa_{1:N}, dm_{1:N}) \\
 & + \sum_{i=1}^N \sum_{j=1}^N \frac{\kappa_i \kappa_j}{m_i} (x_j - x_i) \cdot \frac{\partial \mu_{1:N}}{\partial v_i}(t, dx_{1:N}, dv_{1:N}, d\kappa_{1:N}, dm_{1:N}) = 0
 \end{aligned} \tag{3.25}$$

Thanks to the linearity of the equation, we can obtain the general expression of the population distribution, by using the relation (2.33).

$$\begin{aligned} & \mu_{1:N}(t, dx_{1:N}, dv_{1:N}, d\kappa_{1:N}, dm_{1:N}) \\ &= \left(\exp(tC(\kappa_{1:N}, m_{1:N})) \# \bigotimes_{i=1}^N \mu_0^{x,v|\kappa,m}(dx_i, dv_i | \kappa_i, m_i) \right) \bigotimes_{i=1}^N \mu_0^{\kappa,m}(d\kappa_i, dm_i) \end{aligned} \quad (3.26)$$

with the notation $f \# \mu$ representing the pushforward measure of μ by the map f .

In particular, if the initial distribution is absolutely continuous with respect to the Lebesgue measure, with the decomposition $\mu_0(dx, dv, d\kappa, dm) = f_0^{x,v}(x, v) f_0^{\kappa,m}(\kappa, m) \lambda(dx, dv, d\kappa, dm)$, then $\mu_{1:N}(t)$ has also a density f .

$$\begin{aligned} f(t, x_{1:N}, v_{1:N}, \kappa_{1:N}, m_{1:N}) &= \prod_{i=1}^N f_0^{x,v} \left(O_x^i e^{-tC(\kappa_{1:N}, m_{1:N})} \begin{pmatrix} x_{1:N} \\ v_{1:N} \end{pmatrix}, O_v^i e^{-tC(\kappa_{1:N}, m_{1:N})} \begin{pmatrix} x_{1:N} \\ v_{1:N} \end{pmatrix} \right) \\ &\quad \times |\det(e^{-tC(\kappa_{1:N}, m_{1:N})})| \prod_{i=1}^N f_0^{\kappa,m}(\kappa_i, m_i) \\ &\quad \text{where } O_x^i \begin{pmatrix} x_{1:N} \\ v_{1:N} \end{pmatrix} = x_i \text{ and } O_v^i \begin{pmatrix} x_{1:N} \\ v_{1:N} \end{pmatrix} = v_i \end{aligned} \quad (3.27)$$

In general, the marginal distribution of the individual parameters remains unchanged, while the distribution of the state variables conditionally to the individual parameters is the pushforward measure of the initial distribution by the semi-group of the system.

3.3 Existence and uniqueness of the mean-field measure

In this section, we study the behaviour of symmetric systems when the number of individuals tends to infinity. The dynamics being determined by the transition functions $h_N : \mathcal{X} \times \Theta \times \mathcal{P}_N(\mathcal{X} \times \Theta) \rightarrow \mathcal{X}$, we shall study the asymptotic properties of this sequence of functions indexed by $N \geq 2$. To simplify the analysis, we shall extend the definition of h_N from $\mathcal{P}_N(\mathcal{X} \times \Theta)$ to some set of probability measures $\mathcal{P}_h(\mathcal{X} \times \Theta)$ such that the value of the function $h_N(t, x, \theta, \mu)$ is well defined for all $N \geq 2$ and all $\mu \in \mathcal{P}_h(\mathcal{X} \times \Theta)$. For instance, in the case of the Schneider system, we shall prove that having $\mu \in \mathcal{P}_1(\mathbb{R}_+^* \times \Theta)$ is enough to have well defined values for the transition function, no matter the size of the population. This extension of domain for transition functions enables to study this sequence of mappings over a fixed domain, not depending on N .

Let us consider the situation where the sequence of transition functions $(h_N)_{N \geq 2}$ admits a pointwise limit when $N \rightarrow \infty$, which is the case, for instance, of the Schneider system, where for all $(s, x, S, \gamma, \mu) \in \mathcal{X} \times \Theta \times \mathcal{P}_1(\mathcal{X} \times \Theta)$

$$\begin{aligned} h_N(s, x, S, \gamma, \mu) &= \int_{\mathbb{R}_+ \times \mathbb{R}^2} \left(\frac{N}{N-1} g(s, x, S, \gamma, s', x') - \frac{g(s, x, S, \gamma, s, x)}{N-1} \right) \mu^{s,x}(ds', dx') \\ h_N(s, x, S, \gamma, \mu) &\xrightarrow{N \rightarrow \infty} \int_{\mathbb{R}_+ \times \mathbb{R}^2} g(s, x, S, \gamma, s', x') \mu^{s,x}(ds', dx') = h(s, x, S, \gamma, \mu) \end{aligned} \quad (3.28)$$

We can then investigate under which conditions the transport equation associated with the limit field h has a unique solution.

$$\begin{cases} \frac{\partial \mu}{\partial t}(t, ds, dS, d\gamma) + \frac{\partial}{\partial s} (\mu(t, ds, dx, dS, d\gamma) h(s, x, S, \gamma, \mu(t))) = 0 \\ \mu(0) = \mu_0 \end{cases} \quad (3.29)$$

In the above equation, μ_0 is some initial population distribution in $\mathcal{P}(\mathcal{X} \times \Theta)$. The proof of existence and uniqueness of the measure solution for different types of transition function is a well-documented question in the literature. We highlight three categories of works for the theoretical resolution of this type of non-local partial differential equation.

- Golse (2013):¹² in this work, the transition function is expressed as the expectation of a globally-Lipschitz continuous field. The solution is obtained by Picard iterations at the level of the characteristic flow equation.
- Bolley et al. (2011):¹ the transition function is not globally Lipschitz continuous, but with variations that are at most quadratic (like in the case of the Cucker-Smale model). The proof follows an iterative scheme, alternating between a linear transport partial differential equation and the associated characteristic flow equation, that converges towards the solution of the non-linear partial differential equation.
- Carrillo et al. (2014),⁵ Lagoutière and Vauchelet (2017):¹⁸ the transition function is expressed as the expectation of a force admitting a finite number of discontinuities. The existence of the solution is obtained by first solving the characteristic flow equation with initial condition being expressed as discrete probability measures, convex combination of Dirac masses.

The structure of the proof suggested here is adapted specifically to the regularity of the Schneider system, but we try to isolate the properties that make each step works. In keeping with Golse (2013),¹² existence and uniqueness are obtained as

a generalization of Cauchy-Lipschitz to differential equations taking place in an Euclidean space, with initial conditions being probability measures.

Let us start by introducing the Banach space in which the proof of existence of the characteristic flow takes place.

Definition 3.5. Space of sublinear functions (Golse, 2013,¹² page 18)

The space of continuous functions over $\mathcal{X} \times \Theta$ that are at most linear is denoted by

$$\mathcal{SL}(\mathcal{X}, \Theta) = \{\xi \in \mathcal{C}^0(\mathcal{X} \times \Theta \rightarrow \mathcal{X}) \mid \sup_{(x,\theta) \in \mathcal{X} \times \Theta} \frac{|\xi(x, \theta)|}{1 + |x|} < \infty\} \quad (3.30)$$

$\mathcal{SL}(\mathcal{X}, \Theta)$ endowed with the norm

$$\xi \in \mathcal{SL}(\mathcal{X}, \Theta) \mapsto \|\xi\|_{\mathcal{SL}} = \sup_{(x,\theta) \in \mathcal{X} \times \Theta} \frac{|\xi(x, \theta)|}{1 + |x|} \quad (3.31)$$

is a Banach space.

Proof. Let $(\xi_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in $\mathcal{SL}(\mathcal{X}, \Theta)$. Then for all (x, θ) , the sequence $(\xi_n(x, \theta))_{n \in \mathbb{N}}$ is Cauchy, so it has a pointwise limit, since $(\mathcal{X}, |\cdot|)$ is complete. Let ξ be the pointwise limit of the sequence.

Let $\epsilon > 0$. There exists $n_0 \in \mathbb{N}$ such that for all $p, q \geq n_0$, we have for all $(x, \theta) \in \mathcal{X} \times \Theta$,

$$\begin{aligned} \frac{|\xi_p(x, \theta) - \xi_q(x, \theta)|}{1 + |x|} &\leq \epsilon \\ \lim_{q \rightarrow \infty} \frac{|\xi_p(x, \theta) - \xi_q(x, \theta)|}{1 + |x|} &\leq \epsilon \\ \frac{|\xi_p(x, \theta) - \xi(x, \theta)|}{1 + |x|} &\leq \epsilon \\ \text{so } \|\xi_n - \xi\|_{\mathcal{SL}} &\xrightarrow[n \rightarrow \infty]{} 0 \end{aligned} \quad (3.32)$$

Besides, ξ is continuous at all points (x, θ) , since we have for all $R > |x|$ and for all $n \geq n_0$,

$$\sup\{|\xi_n(x', \theta') - \xi(x', \theta')|, |x'| \leq R\} \leq \epsilon(1 + R) \quad (3.33)$$

and ξ is sublinear as we have for all (x, θ)

$$\frac{|\xi(x, \theta)|}{1 + |x|} \leq \frac{|\xi_{n_0}(x, \theta)|}{1 + |x|} + \epsilon \leq \|\xi_{n_0}\|_{\mathcal{SL}} + \epsilon \quad (3.34)$$

□

The following lemma gives conditions for the local existence of the characteristic flow associated to the non-local partial differential equation. The flow of a differential equation $y' = f(t, y)$ is the function $y(t, y_0)$ giving the value of the solution at t starting from the initial condition $(0, y_0)$. In the case of transport equation, the characteristic flow is the flow associated with the differential equation $x' = h(t, x)$ where h is the velocity field of the transport equation.

Lemma 3.1. Local existence of the characteristic flow

Let $g : (\mathcal{X} \times \Theta)^2 \rightarrow \mathcal{X}$ be a continuous function such that:

- there exists $K_1 > 0$ verifying for all $(x, \theta, x', \theta') \in (\mathcal{X} \times \Theta)^2$,

$$|g(x, \theta, x', \theta')| \leq K_1(1 + |x| + |x'|) \quad (3.35)$$

- there exists $K_2 > 0$ and $p \in \mathbb{N}$ verifying for all $x_1, x'_1, x_2, x'_2 \in \mathcal{X}$ and $\theta, \theta' \in \Theta$

$$|g(x_1, \theta, x'_1, \theta') - g(x_2, \theta, x'_2, \theta')| \leq K_2(1 + |x'_1|^p + |x'_2|^p)(|x_1 - x_2| + |x'_1 - x'_2|) \quad (3.36)$$

Let $\mu_0 \in \mathcal{P}(\mathcal{X} \times \Theta)$ be a probability measure having moments up to the order $1 + p$ with respect to the variable x . Then for all $R > 1$, there exists $\alpha > 0$ and $x_\alpha \in \mathcal{C}^1([-\alpha; \alpha] \rightarrow \mathcal{S}\mathcal{L}(\mathcal{X}, \Theta))$ such that for all $t \in [-\alpha; \alpha]$ and for all $(x, \theta) \in \mathcal{X} \times \Theta$

$$\begin{aligned} \frac{\partial x_\alpha}{\partial t}(t, x, \theta) &= \int_{\mathcal{X} \times \Theta} g(x_\alpha(t, x, \theta), \theta, x_\alpha(t, x', \theta'), \theta') \mu_0(dx', d\theta') \\ x_\alpha(0, x, \theta) &= x \\ \text{and } \sup_{-\alpha \leq t \leq \alpha} \|x_\alpha(t, \cdot, \cdot)\|_{\mathcal{S}\mathcal{L}} &\leq R \end{aligned} \quad (3.37)$$

Proof. Let $\alpha > 0$. We consider the Banach space $\mathcal{S}\mathcal{L}_\alpha = \mathcal{C}^0([-\alpha; \alpha] \rightarrow \mathcal{S}\mathcal{L}(\mathcal{X}, \Theta))$, endowed with the norm

$$\forall \xi \in \mathcal{S}\mathcal{L}_\alpha, \|\xi\|_{\mathcal{S}\mathcal{L}_\alpha} = \sup_{-\alpha \leq t \leq \alpha} \|\xi(t, \cdot, \cdot)\|_{\mathcal{S}\mathcal{L}} \quad (3.38)$$

We consider the application Φ_α defined over $\mathcal{S}\mathcal{L}_\alpha$. For all $\xi \in \mathcal{S}\mathcal{L}_\alpha$ and $(t, x, \theta) \in [-\alpha; \alpha] \times \mathcal{X} \times \Theta$

$$\Phi_\alpha(\xi, t, x, \theta) = x + \int_0^t \int_{\mathcal{X} \times \Theta} g(\xi(s, x, \theta), \theta, \xi(s, x', \theta'), \theta') \mu_0(dx', d\theta') ds \quad (3.39)$$

Let us first notice that the integral over $\mathcal{X} \times \Theta$ is well-defined, since μ_0 has finite first moments, and

$$\begin{aligned} &\int_{\mathcal{X} \times \Theta} |g(\xi(t, x, \theta), \theta, \xi(t, x', \theta'), \theta')| \mu_0(dx', d\theta') \\ &\leq K_1 \int_{\mathcal{X} \times \Theta} (1 + \|\xi\|_{\mathcal{S}\mathcal{L}_\alpha} (2 + |x| + |x'|)) \mu_0(dx', d\theta') < \infty \end{aligned} \quad (3.40)$$

This inequality also proves the continuity of the integral

$$\int_{\mathcal{X} \times \Theta} g(\xi(t, x, \theta), \theta, \xi(t, x', \theta'), \theta') \mu_0(dx', d\theta')$$

with respect to its arguments (t, x, θ) . Therefore, for any $\xi \in \mathcal{SL}_\alpha$ $\Phi_\alpha(\xi)$ is continuous with respect to (x, θ) and continuously differentiable with respect to t . Inequality (3.40) also proves that $\Phi_\alpha(\xi) \in \mathcal{SL}_\alpha$ for all $\xi \in \mathcal{SL}_\alpha$, and we have

$$\|\Phi_\alpha(\xi)\|_{\mathcal{SL}_\alpha} \leq \max \left(K_1 \alpha \left(1 + \|\xi\|_{\mathcal{SL}_\alpha} \left(2 + \int_{\mathcal{X}} |x| \mu_0^x(dx) \right) \right), 1 + K_1 \alpha \|\xi\|_{\mathcal{SL}_\alpha} \right) \quad (3.41)$$

We consider the ball of radius R in \mathcal{SL}_α , $\mathcal{B}_\alpha(0, R) = \{\xi \in \mathcal{SL}_\alpha \mid \|\xi\|_{\mathcal{SL}_\alpha} \leq R\}$. Then for α small enough, $\Phi_\alpha(\xi) \in \mathcal{B}_\alpha(0, R)$ for any $\xi \in \mathcal{B}_\alpha(0, R)$. The threshold for α has the following expression

$$\alpha \leq \min \left(\frac{R}{K_1(1 + R(2 + \int_{\mathcal{X}} |x| \mu_0^x(dx)))}, \frac{R-1}{K_1 R} \right) \quad (3.42)$$

For any $\xi_1, \xi_2 \in \mathcal{B}_\alpha(0, R)$

$$\|\Phi_\alpha(\xi_1) - \Phi_\alpha(\xi_2)\|_{\mathcal{SL}_\alpha} \leq K_2 \alpha \left(1 + 2R^p \int_{\mathcal{X}} (1 + |x|)^p (2 + |x|) \mu_0(dx) \right) \|\xi_1 - \xi_2\|_{\mathcal{SL}_\alpha} \quad (3.43)$$

From this inequality, we can deduce that if α is chosen small enough, the map Φ_α is a contraction over $\mathcal{B}_\alpha(0, R)$. The threshold for α has the following expression

$$\alpha < \min \left(\frac{R}{K_1(1 + R(2 + \int_{\mathcal{X}} |x| \mu_0^x(dx)))}, \frac{R-1}{K_1 R}, \frac{1}{K_2 \left(1 + 2R^p \int_{\mathcal{X}} (1 + |x|)^p (2 + |x|) \mu_0(dx) \right)} \right) \quad (3.44)$$

For this range of α , Banach fixed-point theorem states that there exists a unique $x_\alpha \in \mathcal{B}_\alpha(0, R)$ such that $\Phi_\alpha(x_\alpha) = x_\alpha$. \square

The next step is to prove the local uniqueness of the solution of the differential equation, which is done exactly as in the case of the classical Cauchy-Lipschitz theorem.

Lemma 3.2. Local uniqueness of the characteristic flow

Let I be an interval of \mathbb{R} containing 0 and $x_I : \mathbb{R}_+ \times \mathcal{X} \times \Theta \rightarrow \mathcal{X}$ satisfy the functional equation, for all $(x, \theta) \in \mathcal{X} \times \Theta$

$$\forall t \in I, \quad \frac{\partial x_I}{\partial t}(t, x, \theta) = \int_{\mathcal{X} \times \Theta} g(x_I(t, x, \theta), \theta, x_I(t, x', \theta'), \theta') \mu_0(dx', d\theta') \quad (3.45)$$

$$x_I(0, x, \theta) = x$$

where $g : (\mathcal{X} \times \Theta)^2 \rightarrow \mathcal{X}$ and μ_0 satisfy the same assumptions as in lemma 3.1. Let J be another interval \mathbb{R} , also containing 0 and included in I . If x_J satisfies the same functional equation than x_I over J , then we have that $x_{I|J} = x_J$.

Proof. Let us consider the interval $T_+ = \{t \in J \mid t \geq 0, \forall s \in [0; t], x_I(s) = x_J(s)\}$. We would like to prove that $t^* = \sup(T_+) = \sup(J)$. Let us prove first that $t^* > 0$. Let $R = \min(2, \sup_{t \in I} \|x_I(t)\|_{\mathcal{S}\mathcal{L}})$. By continuity of the norm $\|\cdot\|_{\mathcal{S}\mathcal{L}}$ and by lemma 3.1, there exists $\alpha_1 > 0$ such that for all $t \in [0; \alpha_1]$, $\|x_I(t)\|_{\mathcal{S}\mathcal{L}} \leq R$, $\|x_J(t)\|_{\mathcal{S}\mathcal{L}} \leq R$, and Φ_{α_1} is a contraction over the ball $\mathcal{B}_{\alpha_1}(0, R)$. By uniqueness of the fixed point, we have that $x_{I|[-\alpha_1; \alpha_1]} = x_{J|[-\alpha_2; \alpha_2]}$. In particular, $t^* \geq \alpha_1 > 0$.

Let us assume that $t^* < \sup(J)$, in particular t^* has a finite value. Then for all $t \in [0; t^*)$, we have $\|x_I(t) - x_J(t)\|_{\mathcal{S}\mathcal{L}} = 0$. By continuity of the norm, we obtain that $x_I(t^*) = x_J(t^*)$. There exists $\delta > 0$ such that for all $t \in [t^*; t^* + \delta]$, $\|x_I(t)\|_{\mathcal{S}\mathcal{L}} \leq \|x_I(t^*)\|_{\mathcal{S}\mathcal{L}} + 1 = R_{t^*}$ and $\|x_J(t)\|_{\mathcal{S}\mathcal{L}} \leq R_{t^*}$. For all $t \in [t^*; t^* + \delta]$, we have the following inequality

$$\|x_I(t) - x_J(t)\|_{\mathcal{S}\mathcal{L}} \leq K_2 \int_{\mathcal{X}} (1 + 2R_{t^*}^p (1 + |x|)^p) (2 + |x|) \mu_0^x(dx) \int_{t^*}^t \|x_I(s) - x_J(s)\|_{\mathcal{S}\mathcal{L}} ds \quad (3.46)$$

By Grönwall lemma, we have for all $t \in [t^*; t^* + \delta]$ that $x_I(t) = x_J(t)$, which is in contradiction with the definition of t^* . Therefore, $t^* = \sup(J)$. We use the same reasoning to prove that

$$\inf\{t \in J \mid t \leq 0, \forall s \in [t; 0], x_I(s) = x_J(s)\} = \inf(J) \quad (3.47)$$

□

Lemma 3.2 introduces a partial order relation amongst the solutions, based on the inclusion. According to Zorn's lemma, the partially ordered set of solutions has at least one maximal solution. To prove that the maximal solution is in fact global, defined over \mathbb{R}_+ , we need to characterize the structure of the semi-group associated with this kind of differential flow. First, it can be proved that the properties of the initial distribution are preserved by the flow.

Lemma 3.3. Moments of the pushforward measure

Let $g : (\mathcal{X} \times \Theta)^2 \rightarrow \mathcal{X}$ and μ_0 satisfy the same assumptions in lemma 3.1. Let (I, x_I) be the maximal solution associated with the differential equation, where I is an interval containing 0. For all $t \in I$, we define the pushforward measure $(x_I(t), \text{Id}_{\Theta})\#\mu_0$ as

$$\forall \varphi \in \mathcal{C}_b(\mathcal{X} \times \Theta), \int_{\mathcal{X} \times \Theta} \varphi(x, \theta) [(x_I(t), \text{Id}_{\Theta})\#\mu_0](dx, d\theta) = \int_{\mathcal{X} \times \Theta} \varphi(x_I(t, x, \theta), \theta) \mu_0(dx, d\theta) \quad (3.48)$$

Then for all $t \in I$ we have that

$$\int_{\mathcal{X} \times \Theta} |x|^{1+p} [(x_I(t), \text{Id}_\Theta) \# \mu_0](dx, d\theta) < \infty \quad (3.49)$$

Proof. Let $t \in I$ and $(x, \theta) \in \mathcal{X} \times \Theta$, we have, using equations (3.35) and (3.45).

$$\begin{aligned} |x_I(t, x, \theta)| &\leq |x| + K_1 \int_0^t \left(1 + |x_I(s, x, \theta)| + \int_{\mathcal{X} \times \Theta} |x_I(s, x', \theta')| \mu_0(dx', d\theta') \right) ds \\ \text{thus: } \int_{\mathcal{X} \times \Theta} |x_I(t, x, \theta)| \mu_0(dx, d\theta) &\leq \\ \int_{\mathcal{X}} |x| \mu_0^x(dx) + K_1 \int_0^t \left(1 + 2 \int_{\mathcal{X} \times \Theta} |x_I(s, x, \theta)| \mu_0(dx, d\theta) \right) ds & \end{aligned} \quad (3.50)$$

By Grönwall lemma, we obtain that for all $t \in I$

$$\int_{\mathcal{X} \times \Theta} |x_I(t, x, \theta)| \mu_0(dx, d\theta) \leq \frac{1}{2} \left(\left(2 \int_{\mathcal{X}} |x| \mu_0(dx) + 1 \right) e^{2K_1 t} - 1 \right) \quad (3.51)$$

We can inject the expression of this upper-bound in the estimation of the state.

$$|x_I(t, x, \theta)| \leq |x| + K_1 \int_0^t \left(1 + |x_I(s, x, \theta)| + \frac{1}{2} \left(\left(2 \int_{\mathcal{X}} |x| \mu_0(dx) + 1 \right) e^{2K_1 s} - 1 \right) \right) ds \quad (3.52)$$

By a second application of Grönwall lemma, we obtain that for all $(t, x, \theta) \in I \times \mathcal{X} \times \Theta$

$$|x_I(t, x, \theta)| \leq \frac{1}{2} \left(e^{K_1 t} \left(\left(2 \int_{\mathcal{X}} |x'| \mu_0^x(dx') + 1 \right) e^{K_1 t} - 2 \int_{\mathcal{X}} |x'| \mu_0^x(dx') + 2|x| \right) - 1 \right) \quad (3.53)$$

This estimation proves that $\int_{\mathcal{X} \times \Theta} |x_I(t, x, \theta)|^{1+p} \mu_0(dx, d\theta) < +\infty$. \square

For any $g : (\mathcal{X} \times \Theta)^2 \rightarrow \mathcal{X}$ and $\mu_0 \in \mathcal{P}_{1+p}(\mathcal{X} \times \Theta)$ satisfying the assumptions of lemma 3.1, we introduce the notation $(t, x, \theta) \rightarrow x_{t_0}(t, x, \theta, \mu_0)$ to specify the dependence with respect to the initial condition (t_0, μ_0) of the maximal solution. This notation enables us to express the structure of the semi-group associated with this kind of differential equation.

Lemma 3.4. Semi-group of the non-local flow

Let $g : (\mathcal{X} \times \Theta)^2 \rightarrow \mathcal{X}$ and $\mu_0 \in \mathcal{P}_{1+p}(\mathcal{X})$ satisfying the same assumptions as in lemma 3.1, and let $(t, x, \theta) \in I \times \mathcal{X} \times \Theta \mapsto x_{t_0}(t, x, \theta, \mu_0)$ be the maximal solution associated with the initial condition (t_0, μ_0) . Then for all $t, t_1 \in I$ we have

$$x_{t_0}(t, x, \theta, \mu_0) = x_{t_1}(t, x_{t_0}(t_1, x, \theta, \mu_0), \theta, (x_{t_0}(t_1), \text{Id}_\Theta) \# \mu_0) \quad (3.54)$$

Proof. Let us use the notation $\xi(t, x, \theta) = x_{t_1}(t, x_{t_0}(t_1, x, \theta, \mu_0), \theta, (x_{t_0}(t_1), \text{Id}_\Theta) \# \mu_0)$. We have that $\xi(t_1, x, \theta) = x_{t_0}(t_1, x, \theta)$ for all $(x, \theta) \in \mathcal{X} \times \Theta$. Besides for all $t \in I$, we have

$$\begin{aligned} & \frac{\partial \xi}{\partial t}(t, x, \theta) \\ &= \int_{\mathcal{X} \times \Theta} g(\xi(t, x, \theta), \theta, x_{t_1}(t, x', \theta', (x_{t_0}(t_1), \text{Id}_\Theta) \# \mu_0), \theta') [(x_{t_0}(t_1), \text{Id}_\Theta) \# \mu_0](dx', d\theta') \\ &= \int_{\mathcal{X} \times \Theta} g(\xi(t, x, \theta), \theta, \xi(t, x', \theta'), \theta') \mu_0(dx', d\theta') \end{aligned} \quad (3.55)$$

By uniqueness of the maximal solution of the characteristic flow equation (see lemma 3.2), we have $\xi(t, \cdot, \cdot) = x_{t_0}(t, \cdot, \cdot)$ for all $t \in I$. \square

The proof that the unique maximal solution is defined over \mathbb{R}_+ is then a consequence of the boundedness of the state variable, that was established by lemma 3.3.

Lemma 3.5. *Globality of the maximal solution*

Let g, μ_0 satisfy the assumptions in lemma 3.1. Let (I, x_I) be the maximal solution associated with the initial condition $(0, \mu_0)$. Then $\mathbb{R}_+ \subset I$.

Proof. Let $t^* = \sup(I)$. Let us assume by contradiction that $t^* < +\infty$. If $t^* \in I$, then we can consider the maximal solution associated with the initial condition $(t^*, (x_I(t^*), \text{Id}_\Theta) \# \mu_0)$, which is defined over some interval J containing the segment $[t^* - \alpha; t^* + \alpha]$ for some $\alpha > 0$. So t^* cannot be in I . From the estimations (3.51) and (3.53), we deduce that, for all $t < t^*$,

$$\begin{aligned} & \int_0^t \left| \frac{\partial x_I}{\partial t}(s, x, \theta) \right| ds \\ & \leq K_1 \int_0^t e^{K_1 s} \left(\left(2 \int_{\mathcal{X}} |x'| \mu_0(dx') + 1 \right) e^{K_1 s} - \int_{\mathcal{X}} |x'| \mu_0(dx') + |x| \right) ds \quad (3.56) \\ & \leq \frac{1}{2} (e^{K_1 t} - 1) \left(\left(2 \int_{\mathcal{X}} |x'| \mu_0(dx') + 1 \right) e^{K_1 t} + 2|x| + 1 \right) \end{aligned}$$

In particular, the integral $\int_0^t \left| \frac{\partial x_I}{\partial t}(s, x, \theta) \right| ds$ is convergent when $t \rightarrow t^*$. We can thus define $\tilde{x}_I(t^*, x, \theta) = x + \int_0^{t^*} \frac{\partial x_I}{\partial t}(t, x, \theta) dt$. Then we can consider the maximal solution associated with the initial condition $(t^*, (\tilde{x}_I(t^*), \text{Id}_\Theta) \# \mu_0)$, which provides once again an extension of the maximal solution, leading to a contradiction. Therefore, $t^* = +\infty$. \square

Now, that the characteristic flow is uniquely defined over \mathbb{R}_+ , the measure solution of the transport equation can be expressed as the pushforward measure of the initial distribution by the characteristic flow. We first prove that the pushforward measure satisfies the transport equation, and afterwards, we prove that this is the unique possible solution.

Lemma 3.6. *The pushforward measure by the non-local flow is a solution of the transport equation*

Let $\mu_0 \in \mathcal{P}_{1+p}(\mathcal{X} \times \Theta)$ be a probability measure and g satisfying the assumptions in lemma 3.1. Let $x_0 : \mathbb{R}_+ \times \mathcal{X} \times \Theta \rightarrow \mathcal{X}$ be the characteristic flow associated with the initial condition $(0, \mu_0)$. Then the measure trajectory $t \mapsto (x_0(t), \text{Id}_\Theta) \# \mu_0$ is a measure solution in $\mathcal{P}_{1+p}(\mathcal{X} \times \Theta)$ of the transport equation

$$\begin{aligned} \forall t \geq 0, \quad & \frac{\partial \mu}{\partial t}(t, dx, d\theta) + \text{div}_x \left(\mu(t, dx, d\theta) \int_{\mathcal{X} \times \Theta} g(x, \theta, x', \theta') \mu(t, dx', d\theta') \right) = 0 \\ & \mu(0, dx, d\theta) = \mu_0(dx, d\theta) \end{aligned} \tag{3.57}$$

Proof. Let $t_1, t_2 \in \mathbb{R}_+$ and let us consider the distribution π_{t_1, t_2} defined over $(\mathcal{X} \times \Theta)^2$ by

$$\pi_{t_1, t_2}(dx_1, d\theta_1, dx_2, d\theta_2) = (x_0(t_1), \text{Id}_\Theta, x_0(t_2), \text{Id}_\Theta) \# (\mu_0(dx_1, d\theta_1) \delta_{(x_1, \theta_1)}(dx_2, d\theta_2)) \tag{3.58}$$

This distribution is the law of the random variable $(x_0(t_1, x, \theta), x_0(t_2, x, \theta))$ with (x, θ) being a random variable of distribution μ_0 . Therefore π_{t_1, t_2} is a coupling of the distributions $\mu(t_1)$ and $\mu(t_2)$, or $\pi_{t_1, t_2} \in \Pi(\mu(t_1), \mu(t_2))$. So

$$\mathcal{W}_{1+p}(\mu(t_1), \mu(t_2))^{1+p} \leq \int_{\mathcal{X} \times \Theta} |x_0(t_1, x, \theta) - x_0(t_2, x, \theta)|^{1+p} \mu_0(dx, d\theta) \tag{3.59}$$

From lemma 3.3 and dominated convergence theorem, we have that the upper-bound tends to 0 when t_1 tends to t_2 . So $t \mapsto \mu(t)$ is continuous for the metric \mathcal{W}_{1+p} . By dominated convergence, we also obtain the continuity of the functions

$$\begin{aligned} (t, x, \theta) & \mapsto \int_{\mathcal{X} \times \Theta} g(x, \theta, x_0(t, x', \theta'), \theta') \mu_0(dx', d\theta') = \int_{\mathcal{X} \times \Theta} g(x, \theta, x', \theta') \mu(t, dx', d\theta') \\ \text{and } t & \mapsto \int_{\mathcal{X} \times \Theta} \int_{\mathcal{X} \times \Theta} g(x_0(t, x, \theta), \theta, x_0(t, x', \theta'), \theta') \mu_0(dx', d\theta') \mu_0(dx, d\theta) \\ & = \int_{\mathcal{X} \times \Theta} \int_{\mathcal{X} \times \Theta} g(x, \theta, x', \theta') \mu(t, dx, d\theta) \mu(t, dx', d\theta') \end{aligned} \tag{3.60}$$

Besides for all test function $\varphi \in \mathcal{D}(\mathbb{R}_+ \times \mathcal{X} \times \Theta)$

$$\begin{aligned} & \frac{d}{dt} \int_{\mathcal{X} \times \Theta} \varphi(t, x, \theta) \mu(t, dx, d\theta) \\ &= \int_{\mathcal{X} \times \Theta} \left(\frac{\partial \varphi}{\partial t}(t, x, \theta) + \frac{\partial \varphi}{\partial x}(t, x, \theta) \cdot \int_{\mathcal{X} \times \Theta} g(x, \theta, x', \theta') \mu(t, dx', d\theta') \right) \mu(t, dx, d\theta) \end{aligned} \quad (3.61)$$

□

To prove the uniqueness of the solution, we need to have additional assumptions on the transition function g to ensure the smoothness of the flow with respect to the initial conditions.

Lemma 3.7. Differentiability of the flow with respect to its initial condition

Let $t \in \mathbb{R}_+ \mapsto \nu(t) \in \mathcal{P}_{1+p}(\mathcal{X} \times \Theta)$ be a continuous trajectory for the metric \mathcal{W}_{1+p} . Let g satisfy the assumptions of lemma 3.1 and an additional regularity assumption: $(x_1, \theta_1, x_2, \theta_2) \in (\mathcal{X} \times \Theta)^2 \mapsto g(x_1, \theta_1, x_2, \theta_2)$ is continuously differentiable with respect to x_1 and there exists $K_3 > 0$ such that for all $x_1, x_2 \in \mathcal{X}$, for all $\theta_1, \theta_2 \in \Theta$

$$\left| \frac{\partial g}{\partial x_1}(x_1, \theta_1, x_2, \theta_2) \right| = \sup_{|x|=1} \left| \frac{\partial g}{\partial x_1}(x_1, \theta_1, x_2, \theta_2) x \right| \leq K_3(1 + |x_2|^p) \quad (3.62)$$

Let us define $x_\nu : \mathbb{R}_+^* \times \mathcal{X} \times \Theta \rightarrow \mathcal{X}$ satisfying the differential equation:

$$\begin{aligned} x_\nu(t_0, t_0, x, \theta) &= x \\ \frac{\partial x_\nu}{\partial t}(t, t_0, x, \theta) &= \int_{\mathcal{X} \times \Theta} g(x_\nu(t, t_0, x, \theta), \theta, x', \theta') \nu(t, dx', d\theta') \end{aligned} \quad (3.63)$$

Then x_ν is globally defined for all $(t, t_0, x, \theta) \in \mathbb{R}_+^2 \times \mathcal{X} \times \Theta$ and is continuously differentiable with respect to t, t_0, x .

Proof. Let us introduce the following notation for the transition function of the differential equation

$$G_\nu : (t, x, \theta) \in \mathbb{R}_+ \times \mathcal{X} \times \Theta \mapsto \int_{\mathcal{X} \times \Theta} g(x, \theta, x', \theta') \nu(t, dx', d\theta') \quad (3.64)$$

which is continuous by dominated convergence. Moreover, we have for all $t \in \mathbb{R}_+$ and for all $x_1, x_2 \in \mathcal{X}$

$$|G(t, x_1, \theta) - G(t, x_2, \theta)| \leq K_2 \left(1 + \int_{\mathcal{X} \times \Theta} |x'|^p \nu(t, dx', d\theta') \right) |x_1 - x_2| \quad (3.65)$$

By continuity of the trajectory $t \mapsto \nu(t)$ for the metric \mathcal{W}_{1+p} , we obtain that the Lipschitz factor is bounded over all segment $[0; T]$ and therefore the solution of the differential equation is globally defined.

The differentiability of the characteristic flow with respect to initial conditions is a straightforward consequence of theorem 2.2.3 in Golse (2013).¹¹ \square

The uniqueness of the solution is then obtained by considering a specific class of test functions, satisfying a conservative transport equation.

Lemma 3.8. Uniqueness of the solution of the transport equation

Let $\mu_0 \in \mathcal{P}_{1+p}(\mathcal{X} \times \Theta)$ and g satisfy the assumptions of lemma 3.1 and assumption (3.62). Let $t \in \mathbb{R}_+ \mapsto \nu(t) \in \mathcal{P}_{1+p}(\mathcal{X} \times \Theta)$ satisfy in $\mathcal{P}_{1+p}(\mathcal{X} \times \Theta)$ the transport equation

$$\begin{aligned} \frac{\partial \nu}{\partial t}(t, dx, d\theta) + \operatorname{div}_x \left(\nu(t, dx, d\theta) \int_{\mathcal{X} \times \Theta} g(x, \theta, x', \theta') \nu(t, dx', d\theta') \right) &= 0 \\ \nu(0) &= \mu_0 \end{aligned} \quad (3.66)$$

Then for all $t \in \mathbb{R}_+$, $\nu(t) = (x_0(t), \operatorname{Id}_\Theta) \# \mu_0$.

Proof. Let $\varphi_0 \in \mathcal{C}_b^0(\mathcal{X} \times \Theta)$ be a continuously differentiable with respect to x and let $\varphi : (t, x, \theta) \in \mathbb{R}_+ \times \mathcal{X} \times \Theta \mapsto \varphi_0(x_\nu(0, t, x, \theta)) \in \mathbb{R}$, where x_ν is the characteristic flow associated with the differential equation (3.63). Then, according to theorem 2.2.4 in Golse (2013),¹¹ φ is such that for all $(t, x, \theta) \in \mathbb{R}_+ \times \mathcal{X} \times \Theta$

$$\frac{\partial \varphi}{\partial t}(t, x, \theta) + \frac{\partial \varphi}{\partial x}(t, x, \theta) \cdot \int_{\mathcal{X} \times \Theta} g(x, \theta, x', \theta') \nu(t, dx', d\theta') = 0 \quad (3.67)$$

So for all $t \in \mathbb{R}_+$, we have

$$\begin{aligned} \int_{\mathcal{X} \times \Theta} \varphi_0(x_\nu(0, t, x, \theta), \theta) \nu(t, dx, d\theta) &= \int_{\mathcal{X} \times \Theta} \varphi_0(x, \theta) \mu_0(dx, d\theta) \\ \int_{\mathcal{X} \times \Theta} \varphi_0(x, \theta) [(x_\nu(0, t), \operatorname{Id}_\Theta) \# \nu(t)](dx, d\theta) &= \int_{\mathcal{X} \times \Theta} \varphi_0(x, \theta) \mu_0(dx, d\theta) \end{aligned} \quad (3.68)$$

As φ_0 is arbitrary, we obtain that $(x_\nu(0, t), \operatorname{Id}_\Theta) \# \nu(t) = \mu_0$ and therefore that $\nu(t) = (x_\nu(t, 0), \operatorname{Id}_\Theta) \# \mu_0$. By uniqueness of the non local flow, we obtain that $\nu(t) = (x_0(t), \operatorname{Id}_\Theta) \# \mu_0$. \square

We can now apply the previous results to prove the existence and uniqueness of the mean field limit distribution associated with the Schneider system. As in the case where N is finite, we have chosen to fix some conditions on the initial distribution in order to ensure that the behaviour of the system remains biologically conceivable, in particular that the average competition remains between 0 and 1.

Corollary 3.1. Existence and uniqueness of the mean-field distribution associated with the Schneider system

Let $s_m > 0, R_M > 0, s_m < S_m < s_m \exp(R_M), \gamma_M > 0$ and $\Theta = \mathbb{R}^2 \times [s_m; s_m \exp(R_M)] \times [0; \gamma_M]$ ($x \in \mathbb{R}^2, S \in [S_m; s_m \exp(R_M)], \gamma \in [0; \gamma_M]$). Let $\mu_0 \in \mathcal{P}(\mathbb{R}_+^* \times \Theta)$ ($s_0 \in \mathbb{R}_+^*$) be a probability distribution satisfying the following properties:

$$\mathbb{P} \{s_0 \sim \mu_0^s \mid s_0^{\min} \leq s_0 \leq s_0^{\max}\} = 1 \quad (3.69)$$

where $s_m < s_0^{\min}$ and s_0^{\max} . Then, there exists a unique measure solution in $\mathcal{P}_2(\mathbb{R}_+^* \times \Theta)$ satisfying for all $t \in \mathbb{R}_+, s \in \mathbb{R}_+^*, x \in \mathbb{R}^2, S \in [s_m; s_m \exp(R_M)], \gamma \in [0; \gamma_M]$

$$\begin{aligned} & \frac{\partial \mu}{\partial t}(t, ds, dx, dS, d\gamma) \\ & + \frac{\partial}{\partial s} \left(\gamma s \left(\log(S/s_m) \left(1 - \int_{\mathbb{R}_+ \times \mathbb{R}^2} C(s, s', |x - x'|) \mu^{s,x}(t, ds', dx') \right) \right. \right. \\ & \quad \left. \left. - \log(s/s_m) \right) \mu(t, ds, dS, d\gamma) \right) = 0 \\ & \mu(0, ds, dx, dS, d\gamma) = \mu_0(ds, dx, dS, d\gamma) \end{aligned} \quad (3.70)$$

where C is defined in equation (1.48). Moreover, we have for all $t \geq 0$ that

$$\begin{aligned} & \mathbb{P} \{(s_t, S) \sim \mu^{s,S}(t) \mid s_m < s_t < S\} = 1 \\ & \mathbb{P} \{(s_t, x, s'_t, x') \sim \mu^{s,x}(t)^{\otimes 2} \mid 0 \leq C(s_t, s'_t, |x - x'|) \leq 1\} = 1 \end{aligned} \quad (3.71)$$

Proof. We consider the function $g_r : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ defined by

$$\begin{aligned} g_r(r_1, x_1, S_1, \gamma_1, r_2, x_2) &= \gamma_1 (\log(S_1/s_m)(1 - C_r(r_1, r_2, |x_1 - x_2|) - r_1) \\ C_r(r_1, r_2, |x_1 - x_2|) &= \frac{r_2}{2R_M(1 + |x_1 - x_2|^2/\sigma_x^2)} \left(1 + \tanh \left(\frac{r_2 - r_1}{\sigma_r} \right) \right) \end{aligned} \quad (3.72)$$

g_r is sublinear with respect to the variables r . For all $r_1, r_2 \in \mathbb{R}$ and for all $x_1, x_2 \in \mathbb{R}^2, S_1 \in [S_m; s_m \exp(R_M)], \gamma_1 \in [0; \gamma_M]$ we have

$$|g_r(r_1, x_1, S_1, \gamma_1, r_2, x_2)| \leq \gamma_M (R_M + |r_1| + |r_2|) \quad (3.73)$$

The partial derivatives are also sublinear with respect to r :

$$\begin{aligned}
 \frac{\partial g_r}{\partial r_1}(r_1, x_1, S_1, \gamma_1, r_2, x_2) &= \gamma_1 \left(\frac{r_2 \log(S_1/s_m)(1 - \tanh^2((r_2 - r_1)/\sigma_r))}{2R_M \sigma_r (1 + |x_1 - x_2|^2/\sigma_x^2)} - 1 \right) \\
 \left| \frac{\partial g_r}{\partial r_1}(r_1, x_1, S_1, \gamma_1, r_1, x') \right| &\leq \gamma_M \left(1 + \frac{|r_2|}{2R_M \sigma_r} \right) \\
 \frac{\partial g_r}{\partial r_2}(r_1, x_1, S_1, \gamma_1, r_2, x_2) &= -\frac{\gamma_1 \sigma_x^2 \log(S_1/s_m)(\sigma_r \tanh(\frac{r_2-r_1}{\sigma_r}) + r_2(1 - \tanh^2(\frac{r_2-r_1}{\sigma_r})) + \sigma_r)}{2R_M \sigma_r (\sigma_x^2 + |x_1 - x_2|^2)} \\
 \left| \frac{\partial g_r}{\partial r_2}(r_1, x_1, S_1, \gamma_1, r_2, x_2) \right| &\leq \gamma_M \left(2 + \frac{|r_2|}{2\sigma_r} \right)
 \end{aligned} \tag{3.74}$$

We deduce from these inequalities an upper-bound on the variations of g_r with respect to variables r . For all $r_1, r'_1, r_2, r'_2 \in \mathbb{R}$ and for all $x, x' \in \mathbb{R}^2, S \in [S_m; s_m \exp(R_M)], \gamma \in [0; \gamma_M]$

$$\begin{aligned}
 |g_r(r_1, x, S, \gamma, r'_1, x') - g_r(r_2, x, S, \gamma, r'_2, x')| &\leq \\
 \gamma_M \left(\left(1 + \frac{|r'_1|}{2R_M \sigma_r} \right) |r_1 - r_2| + \left(2 + \frac{|r'_1| + |r'_2|}{2\sigma_r} \right) |r'_1 - r'_2| \right) &\tag{3.75}
 \end{aligned}$$

We consider the distribution ν_0 obtained by the change of variable $\nu_0 = (\log(\cdot/s_m), \text{Id}_\Theta) \# \mu_0$. Then ν_0 has moments up to any order as it has a compact support. By applying the results from lemma 3.1 to lemma 3.8, we obtain that there exists a unique measure trajectory $t \in \mathbb{R}_+ \mapsto \nu(t) \in \mathcal{P}(\mathbb{R} \times \Theta)$ such that for all $r \in \mathbb{R}, x \in \mathbb{R}^2, S \in [S_m; s_m \exp(R_M)]$ and $\gamma \in [0; \gamma_M]$

$$\begin{aligned}
 \frac{\partial \nu}{\partial t}(t, dr, dx, dS, d\gamma) + \frac{\partial}{\partial r} \left(\nu(t, dr, dx, dS, d\gamma) \int_{\mathbb{R}_+ \times \mathbb{R}^2} g_r(r, x, S, \gamma, r', x') \nu^{r,x}(t, dr', dx') \right) &= 0 \\
 \nu(0) = \nu_0 &\tag{3.76}
 \end{aligned}$$

By doing the reverse change of variable $\mu(t) = (s_m \exp(\cdot), \text{Id}_\Theta) \# \nu(t)$, we obtain the solution of the desired transport equation.

The characteristic flow associated with the measure trajectory satisfies the following non-local differential equation for all $t \in \mathbb{R}_+$ and all s_0, x, S, γ such that $s_m < s_0 < S$ and $(x, S, \gamma) \in \Theta$

$$\begin{aligned}
 \frac{\partial}{\partial t} \log(s(t, s_0, x, S, \gamma)/s_m) &= \gamma \left(\log(S/s_m) \right. \\
 \left. \left(1 - \int_{\mathbb{R}_+ \times \mathbb{R}^2} C(s(t, s_0, x, S, \gamma), s', |x - x'|) \mu^{s,x}(t, ds', dx') \right) - \log(s(t, s_0, x, S, \gamma)/s_m) \right) &\tag{3.77}
 \end{aligned}$$

Let us consider the interval

$$\mathcal{I}_{\mathcal{D}} = \{t \in \mathbb{R}_+ \mid \forall \tau \in [0; t), \forall (s_0, x, S, \gamma) \in [s_0^{\min}; s_0^{\max}] \times \Theta, s_m < s(t, s_0, x, S, \gamma) < S\} \quad (3.78)$$

and $t^* = \sup(\mathcal{I}_{\mathcal{D}})$. By continuity of the characteristic flow, $t^* > 0$. For all $t \in [0; t^*)$ we have

$$\begin{aligned} -\gamma \log(s(t, s_0, x, S, \gamma)/s_m) &\leq \frac{\partial}{\partial t} \log(s(t, s_0, x, S, \gamma)/s_m) \leq \gamma \log(S/s(t, s_0, x, S, \gamma)) \\ s_m(s_0/s_m)^{e^{-\gamma t}} &\leq s(t, s_0, x, S, \gamma) \leq S(s_0/S)^{e^{-\gamma t}} \end{aligned} \quad (3.79)$$

So, if $t^* < +\infty$, we obtain that $s_m < s(t^*, s_0, x, S, \gamma) < S$, which is in contradiction with the definition of t^* . So $t^* = +\infty$. \square

In equation (3.77), the state variable $s(t)$ represents the growth of a plant in an infinitely crowded population. However, we have not proved yet the connection between the evolution of an individual within a finite population and the mean-field flow. That is the topic of the next section. This asymptotic behaviour of the dynamics is well-defined in the case of the Schneider system, as the transition function of the system has a pointwise limit when N tends to ∞ . But what can we say about other symmetric systems, that may not have this asymptotic property?

For instance, in the case of the Spring Cloud system, the transition function has the following expression

$$h_N(x, v, \kappa, m, \mu) = \left(\begin{array}{c} v \\ \frac{N\kappa}{m} \int_{\mathbb{R}^2 \times \mathbb{R}_+} \kappa'(x' - x) \mu^{x, \kappa}(dx', d\kappa') \end{array} \right) \quad (3.80)$$

where μ can be taken in the set of probability distributions satisfying $\int_{\mathbb{R}^2 \times \mathbb{R}_+} \kappa|x|\mu^{x, \kappa}(dx, d\kappa) < \infty$. Here, the factor N causes the divergence of the transition function. We can get rid of it by considering a change of time scale. Let us consider the new time scale $t' \leftarrow t\sqrt{N}$ depending on the size of the population and let us consider new state variables

$$\begin{aligned} \hat{x}_i(t) &= x_i(t/\sqrt{N}) \\ \hat{v}_i(t) &= \frac{v_i(t/\sqrt{N})}{\sqrt{N}} \end{aligned} \quad (3.81)$$

The new system for these state variables is then, for all $i \in \llbracket 1; N \rrbracket$

$$\frac{d^2 \hat{x}_i}{dt^2}(t) = \frac{\kappa_i}{Nm_i} \sum_{j=1}^N \kappa_j (\hat{x}_j(t) - \hat{x}_i(t)) \quad (3.82)$$

and the divergent component of the transition function becomes

$\frac{\kappa}{m} \int_{\mathbb{R}^2 \times \mathbb{R}_+} \kappa'(x' - x) \mu^{x, \kappa}(dx', d\kappa')$ and therefore the new system has a pointwise limit. The exact same methodology as in the case of the Schneider system can be used here to prove the existence and uniqueness of the mean-field distribution.

Corollary 3.2. Existence and uniqueness of the mean-field distribution for the Spring Cloud system

Let $0 < \kappa_{\min} < \kappa_{\max}$, $0 < m_{\min} < m_{\max}$, $\Theta = [\kappa_{\min}; \kappa_{\max}] \times [m_{\min}; m_{\max}]$ and $\mu_0 \in \mathcal{P}_1(\mathbb{R}^4 \times \Theta)$. Then there exists a unique solution in $\mathcal{P}_1(\mathbb{R}^4 \times \Theta)$ of the transport equation

$$\begin{aligned} & \frac{\partial \mu}{\partial t}(t, dx, dv, d\kappa, dm) + v \cdot \frac{\partial \mu}{\partial x}(t, dx, dv, d\kappa, dm) \\ & + \frac{\kappa}{m} \frac{\partial \mu}{\partial v}(t, dx, dv, d\kappa, dm) \cdot \int_{\mathbb{R}^2 \times \mathbb{R}_+} \kappa'(x' - x) \mu^{x, \kappa}(t, dx', d\kappa') = 0 \quad (3.83) \\ & \mu(0) = \mu_0 \end{aligned}$$

This transport equation is associated with the following non-local characteristic flow

$$\begin{aligned} \frac{\partial^2 x}{\partial t^2}(t, x_0, v_0, \kappa, m) &= \frac{\kappa \bar{\kappa}}{m} (x_\kappa(t) - x(t, x_0, v_0, \kappa, m)) \\ \text{where } \bar{\kappa} &= \int_{\mathbb{R}_+} \kappa' \mu_0^\kappa(d\kappa') \text{ and } x_\kappa(t) = \frac{1}{\bar{\kappa}} \int_{\mathbb{R}^2 \times \mathbb{R}_+} \kappa' x' \mu^{x, \kappa}(t, dx', d\kappa') \end{aligned} \quad (3.84)$$

The mean-field characteristic flow in the case of Spring Cloud might be more difficult to interpret than in the case of the Schneider system. Equation (3.84) monitors the motion of a particle in interaction with infinitely many particles, but whose motion is also slowed down with an infinite factor. When the size of the population becomes larger, the particles goes faster but their trajectories tend to have invariant shapes.

A similar phenomenon can be observed for the Cucker-Smale system (see equation 1.19), but with a different time scale. We consider the change of time scale $t \leftarrow Nt$ and the state variables

$$\begin{aligned} \hat{x}_i(t) &= x_i(t/N) \\ \hat{v}_i(t) &= \frac{v_i(t/N)}{N} \end{aligned} \quad (3.85)$$

As proved in Bolley et al (2011),¹ if the initial distribution μ_0 is chosen in $\mathcal{P}_2(\mathbb{R}^4 \times [m_{\min}; m_{\max}])$ with $m_{\min} > 0$, then there exists a unique measure trajectory $t \in$

$\mathbb{R}_+ \mapsto \mu(t) \in \mathcal{P}_2(\mathbb{R}^4 \times [m_{\min}; m_{\max}])$ solution of

$$\begin{aligned} & \frac{\partial \mu}{\partial t}(t, dx, dv, dm) + v \cdot \frac{\partial \mu}{\partial x}(t, dx, dv, dm) \\ & + \frac{H}{m} \frac{\partial}{\partial v} \left(\mu(t, dx, dv, dm) \int_{\mathbb{R}^{2d}} \frac{v' - v}{(1 + |x - x'|/\sigma_x)^\beta} \mu^{x,v}(t, dx', dv') \right) = 0 \end{aligned} \quad (3.86)$$

3.4 Propagation of chaos and mean-field approximation

This section gives the proofs associated with the convergence of the microscopic system, represented by the empirical population measure $\hat{\mu}_N(t)$, to the macroscopic system, represented by the solution of the non-local transport equation, whose existence and uniqueness were established in the previous section. We also examine how the mean-field limit can simplify the inference problems introduced in chapter 2. These proofs are carried out in the specific cases of the Schneider system and of the Spring Cloud system.

The empirical measure's convergence to the mean-field distribution is referred to in the literature as the phenomenon of chaos propagation. The fundamentals of this notion were introduced by Kac (1956)¹⁴ for gas kinetic equations. A comprehensive description of the different aspects of chaos propagation is provided by Sznitman (1991).²⁶ In this context, chaos describes the state of a symmetric system containing an infinite number of independent and identically distributed particles. In particular, the initial configuration of a system sampled from some distribution μ_0 constitutes the simplest example of chaotic systems. It is clear, for any finite population of interacting particles, that the particles' initial independence is no longer valid for any $t > 0$. In some situations, we can nevertheless say that the system is chaotic if the population distribution $\mu_{1:N}(t)$, defined over $\mathcal{P}((\mathcal{X} \times \Theta)^N)$, behaves as a factorized distribution $\mu_{1:N}(t) \approx \mu_\infty(t)^{\otimes N}$ asymptotically when $N \rightarrow \infty$. More formally, a sequence of symmetric distributions is called chaotic if for any $k \in \mathbb{N}^*$, the marginal distribution $\mu_{1:N}^{(k)}(t)$ converges weakly in $\mathcal{P}((\mathcal{X} \times \Theta)^k)$ to some factorized distribution $\mu_\infty(t)^{\otimes k}$. Here, $\mu_{1:N}^{(k)}(t)$ denotes the marginal distribution of any sub-group of k individuals within a population of size N . As stated in proposition 2.2 of Sznitman (1991),²⁶ chaos is equivalent to the convergence of the empirical distribution to some distribution. We can intuitively understand this equivalence for the symmetric empirical system having a transition function of the form $h_N(t, x, \theta, \hat{\mu}_N(t))$. If the transition function h_N converges in some sense towards a function $h(t, x, \theta, \mu(t))$, where $\mu(t)$ is a deterministic probability distribution¹, then within this infinitely crowded population, any couple of

¹oxymoron to be understood in opposition to a random probability measure, such as the

particles x_1, x_2 would evolve according to the equations $\dot{x}_1(t) = h(t, x_1(t), \theta_1, \mu(t))$ and $\dot{x}_2(t) = h(t, x_2(t), \mu(t))$, so the two particles would be independent. This result justifies the idea of *propagation*: the initial independence of the particles is somewhat propagated over time by the dynamics of the macroscopic system. Besides, the marginal distributions of the two particles could only be the limit distribution $\mu(t)^{\otimes 2}$, in keeping with the fact that, in the microscopic case, we can sample two individuals in the population from $\hat{\mu}_N(t)^{\otimes 2}$.

A simple methodology to prove the propagation of chaos, especially suited to models with a certain degree of smoothness, is based on Dobrushin's stability.¹⁰ The steps are clearly explained in Golse (2013).¹² First, one starts from a classical statistical result linking the initial distribution μ_0 with the empirical measure associated with an independent and identically distributed sample from μ_0 . Then, the convergence $\hat{\mu}_N(0) \rightarrow \mu_0$ is propagated over time using consequences of Grönwall lemma to establish that $\hat{\mu}_N(t) \rightarrow \mu(t)$, where $\mu(t)$ is the unique solution of the non-local transport equation of initial condition μ_0 . Let us start by introducing the result concerning the convergence of the initial conditions.

Theorem 3.3. *Fundamental principle of statistics (Varadarajan, 1958²⁸)*
Let $\mu \in \mathcal{P}_p(\mathcal{Z})$ for some Polish metric space (\mathcal{Z}, d) , and let \mathcal{W}_p be the Wasserstein distance associated with d . We use the notation $(z_n)_{n \in \mathbb{N}^} \sim \mu^{\otimes \infty}$ for any sequence of independent and identically distributed variables of distribution μ . Then we have that*

$$\mathbb{P} \left\{ (z_n)_{n \in \mathbb{N}} \sim \mu^{\otimes \infty} \mid \lim_{N \rightarrow \infty} \mathcal{W}_p \left(\frac{1}{N} \sum_{n=1}^N \delta_{z_n}, \mu \right) = 0 \right\} = 1 \quad (3.87)$$

The distance used to quantify the propagation of chaos is the Wasserstein distance of some order. The propagation of chaos enables to prove that $\mathcal{W}_p(\hat{\mu}_N(t), \mu(t)) \rightarrow 0$ almost surely even for empirical measures that are sums of non-independent Dirac distributions (for $t > 0$). This result is possible because the initial convergence is jointly transported by the flows associated with $\hat{\mu}_N(t)$ and $\mu(t)$.

3.4.1 Propagation of chaos in the case of the Schneider system

In the following of this section, we consider the Schneider system. We first prove that the empirical flow associated with $\hat{\mu}_N(t)$ is well-defined in the case of the Schneider system.

Lemma 3.9. *Well-posedness of the empirical flow associated with the Schneider system*

empirical population measure $\hat{\mu}_N(t)$

We use the same assumptions for the initial distribution $\mu_0 \in \mathcal{P}(\mathbb{R}_+^* \times \Theta)$ as in corollary 3.1. Let $z_{1:N}^0 = (s_i^0, x_i, S_i, \gamma_i)_{1 \leq i \leq N} \in ([s_0^{\min}; s_0^{\max}] \times \mathbb{R}^2 \times [S_m; s_m e^{R_M}] \times [0; \gamma_M])^N$ an initial configuration, and let $t \in \mathbb{R}_+ \mapsto \hat{\mu}_N(t)$ be the empirical population measure associated with this initial configuration. Then for any $(s_0, \theta) = (s_0, x, S, \gamma) \in [s_0^{\min}; s_0^{\max}] \times \Theta$, there exists a unique global solution of the differential equation

$$\begin{aligned} \hat{s}_N(0, s_0, \theta) &= s_0 \\ \frac{\partial \hat{s}_N}{\partial t}(t, s_0, \theta) &= \gamma_1 \hat{s}_N(t, s_0, \theta) \left(\log(S_1/s_m) \left(1 - \frac{1}{N-1} \sum_{i=1}^N C(\hat{s}_N(t, \theta_1), s_i(t), |x_1 - x_i|) \right. \right. \\ &\quad \left. \left. + \frac{C(\hat{s}_N(t, s_0, \theta_1), \hat{s}_N(t, \theta_1), 0)}{N-1} \right) - \log\left(\frac{\hat{s}_N(t, s_0, \theta_1)}{s_m}\right) \right) \end{aligned} \quad (3.88)$$

Moreover, we have for all $i \in \llbracket 1; N \rrbracket$: $\hat{s}_N(t, s_i^0, x_i, S_i, \gamma_i) = s_i(t)$, the individual trajectory associated with the initial configuration $z_{1:N}^0$, and for all $(s_0, x, S, \gamma) \in [s_0^{\min}; s_0^{\max}]$ and $t \in \mathbb{R}_+$, we have the bounds

$$s_m^N = s_m \exp\left(-\frac{R_M}{2N-3}\right) \leq \hat{s}_N(t, s, x, S, \gamma) \leq s_m \exp\left(\frac{\log(S/s_m)}{1 - \frac{\log(S/s_m)}{2(N-1)R_M}}\right) = \hat{S}_N(S) \quad (3.89)$$

In particular, the empirical flow \hat{s}_N determines the semi-group associated with the empirical measure of the population.

$$(\hat{s}_N(t), \text{Id}_\Theta) \# \hat{\mu}_N(0, z_{1:N}^0) = \hat{\mu}_N(t, z_{1:N}^0) \quad (3.90)$$

Proof. We consider the function $G_N : \mathbb{R}_+ \times \mathbb{R} \times \Theta \times \mathbb{R}$ defined by

$$\begin{aligned} \forall r \in \mathbb{R}, (x, S, \gamma) \in \Theta, \quad &G_N(t, r, x, S, \gamma) \\ &= \gamma \left(\log(S/s_m) \left(1 - \frac{1}{N-1} \sum_{i=1}^N C_r(r, \log(s_i(t)/s_m), |x - x_i|) \right) \right. \\ &\quad \left. - \left(1 - \frac{\log(S/s_m)}{2(N-1)R_M} \right) r \right) \end{aligned} \quad (3.91)$$

where C_r is defined in equation (3.72). The function G_N is continuously differentiable, in particular locally Lipschitz continuous. Therefore, there exists a unique maximal solution for any initial configuration $(r_0, \theta) \in \mathbb{R} \times \Theta$. Moreover, according to proposition 1.3, we can obtain an inequality enabling to prove that the maximal solution is globally defined.

$$\forall (t, r, \theta) \in \mathbb{R}_+ \times \mathbb{R} \times \Theta, \quad |G_N(t, r, \theta)| \leq \gamma_M (R_M + |r|) \quad (3.92)$$

By the change of variable $r = \log(s/s_m)$, there exists a unique global solution of the differential equation (3.88). For any time $t \geq 0$, we have for all $(s, x, S, \gamma) \in [s_0^{\min}; s_0^{\max}] \times \Theta$

$$\begin{aligned} & \frac{\partial}{\partial t} \log(\hat{s}_N(t, s, x, S, \gamma)/s_m) \\ & \geq \gamma \left(\log(S/s_m) \left(1 - \frac{N}{N-1} \right) - \left(1 - \frac{\log(S/s_m)}{2(N-1)R_m} \right) \log(\hat{s}_N/s_m) \right) \end{aligned} \quad (3.93)$$

Similarly as in lemma 1.1, we can derive a lower bound for the empirical flow.

$$\begin{aligned} \log(\hat{s}_N(t, s, x, S, \gamma)/s_m) & \geq -\frac{2R_M \log(S/s_m)}{2(N-1)R_M - \log(S/s_m)} \\ & + \left(\log(s/s_m) + \frac{2R_M \log(S/s_m)}{2(N-1)R_M - \log(S/s_m)} \right) \exp \left(-\gamma \left(1 - \frac{\log(S/s_m)}{2(N-1)R_m} \right) t \right) \end{aligned} \quad (3.94)$$

We can derive a global lower bound by considering the inequality

$$-\frac{2R_M \log(S/s_m)}{2(N-1)R_M - \log(S/s_m)} \geq -\frac{2R_M}{2N-3} \quad (3.95)$$

As for the upper-bound, we have for the derivative of the empirical flow

$$\begin{aligned} & \frac{\partial}{\partial t} \log(\hat{s}_N(t, s, x, S, \gamma)/s_m) \\ & \leq \gamma \left(\log(S/s_m) - \left(1 - \frac{\log(S/s_m)}{2(N-1)R_M} \right) \log(\hat{s}_N(t, s, x, S, \gamma)/s_m) \right) \end{aligned} \quad (3.96)$$

Thus, by the Grönwall lemma,

$$\begin{aligned} \log(\hat{s}_N(t, s, x, S, \gamma)/s_m) & \leq \frac{\log(S/s_m)}{1 - \frac{\log(S/s_m)}{2(N-1)R_M}} \\ & - \left(\frac{\log(S/s_m)}{1 - \frac{\log(S/s_m)}{2(N-1)R_M}} - \log(s/s_m) \right) \exp \left(-\gamma \left(1 - \frac{\log(S/s_m)}{2(N-1)R_M} \right) t \right) \end{aligned}$$

□

Let us proceed to the proof of the propagation of chaos for the Schneider system.

Proposition 3.3. Propagation of chaos for the Schneider system

We use the same assumption as in corollary 3.1 for the initial distribution $\mu_0 \in \mathcal{P}(\mathbb{R}_+^* \times \Theta)$. Moreover, we assume that the moment $\int_{\mathbb{R}^2} |x|^2 \mu_0^x(dx)$ is finite. For

any initial configuration $z_{1:N}^0 \in ([s_0^{\min}; s_0^{\max}] \times \Theta)^N$, $\hat{\mu}_N(t, z_{1:N}^0)$ is the empirical population measure associated with this initial configuration. Then we have the following almost sure convergence

$$\mathbb{P} \left\{ (z_n^0)_{n \in \mathbb{N}^*} \sim \mu_0^{\otimes \infty} \mid \forall t \geq 0, \lim_{N \rightarrow \infty} \mathcal{W}_1(\hat{\mu}_N(t, z_{1:N}^0), \mu(t)) = 0 \right\} = 1 \quad (3.97)$$

Proof. Let $z_{1:N}^0 \in ([s_0^{\min}; s_0^{\max}] \times \Theta)^N$ be an initial configuration. Let $\pi_s \in \Pi(\hat{\mu}_N^s(0, z_{1:N}^0), \mu_0^s)$, $\pi_x \in \Pi(\hat{\mu}_N^x(0, z_{1:N}^0), \mu_0^x)$, $\pi_S \in \Pi(\hat{\mu}_N^S(0, z_{1:N}^0), \mu_0^S)$ and $\pi_\gamma \in \Pi(\hat{\mu}_N^\gamma(0, z_{1:N}^0), \mu_0^\gamma)$ be a set of couplings associated with the marginal distributions of the empirical distribution and the initial distribution. From these couplings, we build the initial coupling $\pi_0 = \pi_s \otimes \pi_x \otimes \pi_S \otimes \pi_\gamma$, which is in $\Pi(\hat{\mu}_N(0, z_{1:N}^0), \mu_0)$.

We consider the empirical flow \hat{s}_N associated with the empirical population measure $t \mapsto \hat{\mu}_N(t, z_{1:N}^0)$, whose properties are studied in lemma 3.9. The mean-field flow associated with $t \mapsto \mu(t)$ is denoted by s_∞ . For any $t \geq 0$, we define the coupling $\pi_t = (\hat{s}_N(t), \text{Id}_\Theta, s_\infty(t), \text{Id}_\Theta) \# \pi_0 \in \Pi(\hat{\mu}_N(t, z_{1:N}^0), \mu(t))$. The space $\mathbb{R}_+^* \times \Theta$ is endowed with the metric $m_{\mathcal{Z}}$ defined by

$$\begin{aligned} m_{\mathcal{Z}}(s_1, \theta_1, s_2, \theta_2) &= \frac{|s_1 - s_2| + |S_1 - S_2|}{s_m} + \frac{|x_1 - x_2|}{\ell} + \tau_r |\gamma_1 - \gamma_2| \\ &= \frac{|s_1 - s_2|}{s_m} + m_\Theta(\theta_1, \theta_2) \end{aligned} \quad (3.98)$$

where $\ell, \tau_r > 0$ are arbitrary constants. For this metric and for any time $t \geq 0$, the Wasserstein distance between distributions $\hat{\mu}_N(t, z_{1:N}^0)$ and $\mu(t)$ is expressed as follows

$$\begin{aligned} \mathcal{W}_1(\hat{\mu}_N(t, z_{1:N}^0), \mu(t)) &= \inf \left\{ \iint_{(\mathbb{R}_+^* \times \Theta)^2} m_{\mathcal{Z}}(s_1, \theta_1, s_2, \theta_2) \pi(ds_1, d\theta_1, ds_2, d\theta_2), \right. \\ &\quad \left. \pi \in \Pi(\hat{\mu}_N(t, z_{1:N}^0), \mu(t)) \right\} \end{aligned} \quad (3.99)$$

The coupling π_t provides an upper-bound of the Wasserstein distance at time t .

$$\begin{aligned} \mathcal{W}_1(\hat{\mu}_N(t, z_{1:N}^0), \mu(t)) &\leq \iint_{(\mathbb{R}_+^* \times \Theta)^2} \left(\frac{|s_1 - s_2|}{s_m} + m_\Theta(\theta_1, \theta_2) \right) \pi_t(ds_1, d\theta_1, ds_2, d\theta_2) \\ &\leq \frac{1}{s_m} \iint_{(\mathbb{R}_+^* \times \Theta)^2} |\hat{s}_N(t, s_1, \theta_1) - s_\infty(t, s_2, \theta_2)| \pi_0(ds_1, d\theta_1, ds_2, d\theta_2) \\ &\quad + \iint_{\Theta^2} m_\Theta(\theta_1, \theta_2) \pi_0^\theta(d\theta_1, d\theta_2) \end{aligned} \quad (3.100)$$

Let us focus on the first term of the upper-bound.

$$D_N^{\pi_0}(t) = \iint_{(\mathbb{R}_+^* \times \Theta)^2} |\hat{s}_N(t, s_1, \theta_1) - s_\infty(t, s_2, \theta_2)| \pi_0(ds_1, d\theta_1, ds_2, d\theta_2) \quad (3.101)$$

As detailed in the appendix page 157, We can prove that

$$\begin{aligned} D_N^{\pi_0}(t) &\leq \frac{s_0^{\max} e^{R_M} R_M}{N-1} + \frac{tA(\hat{\mu}_N(0, z_{1:N}^0))}{N-1} + e^{R_M} \iint_{(\mathbb{R}_+^*)^2} |s_1 - s_2| \pi_0^s(ds_1, ds_2) \\ &+ tB(\hat{\mu}_N^x(0, z_{1:N}^0), \mu_0^x) \sqrt{\iint_{\mathbb{R}^4} |x_1 - x_2|^2 \pi_0^x(dx_1, dx_2)} \\ &+ \alpha_S \iint_{[S_m; s_m e^{R_M}]^2} |S_1 - S_2| \pi_0^S(dS_1, dS_2) \\ &+ \alpha_\gamma t \iint_{[0; \gamma_M]^2} |\gamma_1 - \gamma_2| \pi_0^\gamma(d\gamma_1, d\gamma_2) + \beta_N \int_0^t D_N^{\pi_0}(\tau) d\tau \end{aligned} \quad (3.102)$$

where the functionals $A(\mu)$, $B(\mu_1, \mu_2)$ have the following expressions:

$$\begin{aligned} A(\mu) &= \frac{1}{2R_M} \int_{\mathbb{R}_+^* \times [S_m; s_m e^{R_M}] \times [0; \gamma_M]} \left(\frac{sS}{s_m} \gamma \log(s/s_m) + s_0^{\max} e^{R_M} R_M \gamma \log(S/s_m) \right) \\ &\times \mu^{s, S, \gamma}(ds, dS, d\gamma) \\ B(\mu_1, \mu_2) &= \frac{s_0^{\max} e^{R_M} R_M \gamma_M}{\sigma_x^2} \left(2 \int_{\mathbb{R}^2} |x| \mu_2(dx) + \sqrt{2 \int_{\mathbb{R}^2} |x|^2 \mu_2(dx) + 2 \int_{\mathbb{R}^2} |x|^2 \mu_1(dx)} \right. \\ &\left. + \int_{\mathbb{R}^2} \sqrt{2 \int_{\mathbb{R}^2} |x|^2 \mu_1(dx) + 2 \int_{\mathbb{R}^2} |x|^2 \mu_1(dx) - 4x' \cdot \left(\int_{\mathbb{R}^2} x \mu_1(dx) + \int_{\mathbb{R}^2} x \mu_2(dx) \right)} \mu_1(dx') \right) \end{aligned} \quad (3.103)$$

and the coefficients $\alpha_S, \alpha_\gamma, \beta_N$ are

$$\begin{aligned} \alpha_S &= (s_{\max}^0 / S_m) \\ \alpha_\gamma &= s_{\max}^0 \log\left(\frac{s_0^{\max}}{s_m}\right) e^{R_M} + s_{\max}^0 e^{R_M} R_M \\ \beta_N &= \frac{s_0^{\max} e^{R_M} R_M \gamma_M}{s_m^N \sigma_r} (1 + \sigma_r / R_M + 1/2) \end{aligned} \quad (3.104)$$

We can find in the appendix, page 157, the detailed derivation of the upper-bound

of $D_N^{\pi_0}(t)$ in inequation (3.102). By Grönwall lemma, we obtain that

$$\begin{aligned}
 D_N^{\pi_0}(t) &\leq \frac{1}{\beta_N} \left(\frac{A(\hat{\mu}_N(0, z_{1:N}^0))}{N-1} + B(\hat{\mu}_N^x(0, z_{1:N}^0), \mu_0^x) \sqrt{\iint_{\mathbb{R}^4} |x_1 - x_2|^2 \pi_0^x(dx_1, dx_2)} \right. \\
 &+ \alpha_\gamma \iint_{[0; \gamma_M]^2} |\gamma_1 - \gamma_2| \pi_0^\gamma(d\gamma_1, d\gamma_2) \left. (e^{\beta_N t} - 1) + \left(\frac{s_0^{\max} e^{R_M} R_M}{N-1} \right. \right. \\
 &+ \left. \left. e^{R_M} \iint_{[s_0^{\min}; s_0^{\max}]^2} |s_1 - s_2| \pi_0^s(ds_1, ds_2) + \alpha_S \iint_{[s_m; s_m e^{R_M}]^2} |S_1 - S_2| \pi_0^S(dS_1, dS_2) \right) e^{\beta_N t} \right)
 \end{aligned} \tag{3.105}$$

By gathering the terms from inequalities (3.105) and (3.100), we obtain the following upper-bound on the Wasserstein distance between the empirical distribution and the mean-field distribution at time t

$$\begin{aligned}
 \mathcal{W}_1(\hat{\mu}_N(t, z_{1:N}^0), \mu(t)) &\leq \frac{e^{R_M}}{s_m} \mathcal{W}_1(\hat{\mu}_N^s(0, z_{1:N}^0), \mu_0^s) e^{\beta_N t} \\
 &+ \left(\frac{B(\hat{\mu}_N^x(0, z_{1:N}^0), \mu_0^x)}{s_m \beta_N} (e^{\beta_N t} - 1) + \frac{1}{\ell} \right) \mathcal{W}_2(\hat{\mu}_N^x(0, z_{1:N}^0), \mu_0^x) \\
 &+ \frac{(\alpha_S e^{\beta_N t} + 1)}{s_m} \mathcal{W}_1(\hat{\mu}_N^S(0, z_{1:N}^0), \mu_0^S) + \left(\frac{\alpha_\gamma}{s_m} (e^{\beta_N t} - 1) + \tau_r \right) \mathcal{W}_1(\hat{\mu}_N^\gamma(0, z_{1:N}^0), \mu_0^\gamma) \\
 &+ \frac{1}{s_m(N-1)} \left(\frac{A_N(\hat{\mu}_N(0, z_{1:N}^0), \mu_0)}{\beta_N} (e^{\beta_N t} - 1) + s_0^{\max} e^{R_M + \beta_N t} R_M \right)
 \end{aligned} \tag{3.106}$$

From the law of large numbers, we have the following almost sure convergences

$$\begin{aligned}
 \mathbb{P}(\Omega_A) &= \mathbb{P} \left\{ z_{1:N}^0 \sim \mu_0^{\otimes \infty} \mid \lim_{N \rightarrow \infty} A(\hat{\mu}_N(0, z_{1:N}^0)) = A(\mu_0) \right\} = 1 \\
 \mathbb{P}(\Omega_B) &= \mathbb{P} \left\{ z_{1:N}^0 \sim \mu_0^{\otimes \infty} \mid \lim_{N \rightarrow \infty} B(\hat{\mu}_N^x(0, z_{1:N}^0), \mu_0^x) = B(\mu_0^x, \mu_0^x) \right\} = 1
 \end{aligned} \tag{3.107}$$

From Varadarajan's theorem (theorem 3.3), we have also

$$\begin{aligned}
 \mathbb{P}(\Omega_s) &= \mathbb{P} \left\{ z_{1:N}^0 \sim \mu_0^{\otimes \infty} \mid \lim_{N \rightarrow \infty} \mathcal{W}_1(\hat{\mu}_N^s(0, z_{1:N}^0), \mu_0^s) = 0 \right\} = 1 \\
 \mathbb{P}(\Omega_x) &= \mathbb{P} \left\{ z_{1:N}^0 \sim \mu_0^{\otimes \infty} \mid \lim_{N \rightarrow \infty} \mathcal{W}_2(\hat{\mu}_N^x(0, z_{1:N}^0), \mu_0^x) = 0 \right\} = 1 \\
 \mathbb{P}(\Omega_S) &= \mathbb{P} \left\{ z_{1:N}^0 \sim \mu_0^{\otimes \infty} \mid \lim_{N \rightarrow \infty} \mathcal{W}_1(\hat{\mu}_N^S(0, z_{1:N}^0), \mu_0^S) = 0 \right\} = 1 \\
 \mathbb{P}(\Omega_\gamma) &= \mathbb{P} \left\{ z_{1:N}^0 \sim \mu_0^{\otimes \infty} \mid \lim_{N \rightarrow \infty} \mathcal{W}_1(\hat{\mu}_N^\gamma(0, z_{1:N}^0), \mu_0^\gamma) = 0 \right\} = 1
 \end{aligned} \tag{3.108}$$

As the intersection of all these events is of probability one, and as $(\beta_N)_{N \geq 1}$ is convergent, we obtain the result we want to prove. \square

In the previous proof, we have obtained estimations of the Wasserstein distance having exponential growth with respect to time. In practice however, the discrepancy between the two distribution remains bounded in time, since all the dynamics of the Schneider system take place in a compact region of the space. Uniform bounds in time can be obtained, by adopting methodologies developed in Salhi et al. (2018).²⁴ Arguments specific to Maxwell kinetic model are used in Cortez and Fontbona (2018)⁷ to obtain estimates contracting with times. In the case of the Schneider system especially, these bounds can be refined by considering the differential system satisfied by the gradients of the empirical flow and the mean-field flow. As for the dependency with respect to N , sharp estimates are derived for the expectation $\mathbb{E}\mathcal{W}_p$ in Kac model by Cortez and Fontbona (2016).⁸ In the situation where the propagation of chaos is proved by means of Dobrushin's stability, the dependency on N of the estimate is mainly monitored by the quantification of the speed of convergence of the initial distributions, ensured by Varadarajan's theorem. For an estimation of the convergence's speed, we can refer to Lei (2020).¹⁹

It is worth noting that the proof of the propagation of chaos in the absence of smoothness assumption can be done by resorting to Bogolioubov-Born-Green-Kirkwood-Yvon (BBGKY) hierarchy, which studies the dynamics of the population distributions $\mu_{1:N}(t)$, and their convergence to some factorized distribution. This is one of the ingredient of the proof of propagation of chaos for a system with topological interactions in Degond and Pulvirenti (2019).⁹ One can also adopt the approach used in Lagoutière and Vauchelet (2017),¹⁸ consisting in proving the convergence of the empirical population measure by a compactness argument, and thus prove the existence and uniqueness of the mean-field distribution.

Some properties of the macroscopic distribution $t \mapsto \mu(t)$ have a statistical interest. First of all, it enables to describe individual trajectories that do not depend on the size N of the population, which constitutes a latent variable that may be difficult to estimate, as mentioned in section 2.6. Moreover, all the other latent variables associated with the unobserved part of the system are somehow summarized by the distribution $\mu(t)$, and they do not need to be inferred in order to access parameters of actual interest. Finally, the independence of the individuals under the macroscopic regime opens up possibilities to distribute the inference calculations more efficiently. Therefore, we are going to use $\mu(t)$ to build a consistent approximation of an inference problem.

3.4.2 Mean-field approximated inference in the case of the Schneider system

In the Schneider system, let us assume that initially all plants have the size s_0 , which is an unknown value in the interval $[s_0^{\min}; s_0^{\max}]$. The values of the competition parameters σ_x, σ_r are also unknown. We would like to infer $\eta = (s_0, \sigma_x, \sigma_r)$ from the observations of N_0 plants at times t_1, \dots, t_M , with standard error σ . These plants are assumed to be within a population of size N , which is also unknown. The prior on the parameters is uniform, and the prior on the size of the population is a shifted Poisson distribution $N_0 + \text{Poisson}(\lambda)$. Conditionally to the parameter η , the likelihood of the observation $s \in \mathbb{R}^{N_0 M}$ is

$$p(s|\eta) = \frac{1}{(2\pi\sigma^2)^{\frac{N_0 M}{2}}} \sum_{N=N_0}^{+\infty} p_N(N) \int_{\Theta^N} \prod_{i=1}^{N_0} \prod_{j=1}^M \exp\left(-\frac{(s_{ij} - s_i^N(t_j, \eta, \theta_{1:N}))^2}{2\sigma^2}\right) \times (\mu_0^\theta)^{\otimes N} (d\theta_{1:N}) \quad (3.109)$$

$$p(s|\eta) = \sum_{N=N_0}^{+\infty} p_N(N) p_{s|\eta, N}(s|\eta, N) \quad (3.110)$$

where $p_N(N) = \frac{\lambda^{N-N_0}}{(N-N_0)!} e^{-\lambda}$ is the density of the Poisson distribution with respect to the counting measure, and where s_i^N is the empirical flow.

$$s_i^N(t, \eta, \theta_{1:N}) = s_m \left(\frac{s_0}{s_m}\right)^{e^{-\gamma_i t}} \exp\left(\gamma_i e^{-\gamma_i t} \log(S_i/s_m) \int_0^t (1 - \hat{C}_i^N(\tau, \eta, \theta_{1:N})) e^{\gamma_i \tau} d\tau\right) \\ \hat{C}_i^N(t, \eta, \theta_{1:N}) = \frac{1}{N-1} \sum_{j \neq i} C(s_i^N(t, \eta, \theta_{1:N}), s_j^N(t, \eta, \theta_{1:N}), |x_i - x_j|) \quad (3.111)$$

For N large enough, we can consider an approximation of the likelihood density $p_{s|\eta, N}$ by replacing the empirical flow by the mean-field flow s_∞ .

$$s_\infty(t, \eta, \theta_i) = s_m \left(\frac{s_0}{s_m}\right)^{e^{-\gamma_i t}} \exp\left(\gamma_i e^{-\gamma_i t} \log(S_i/s_m) \int_0^t (1 - C_\infty(\tau, \eta, \theta_i)) e^{\gamma_i \tau} d\tau\right) \\ C_\infty(t, \eta, \theta_i) = \int_{\Theta} C(s_\infty(t, \eta, \theta_i), s_\infty(t, \eta, \theta'), |x_i - x'|) \mu_0^\theta(d\theta') \quad (3.112)$$

The approximation of $p_{s|\eta, N}$ by use of the mean-field flow is denoted by $p_{s|\eta}^\infty$.

$$p_{s|\eta}^\infty(s|\eta) = \frac{1}{(2\pi\sigma^2)^{\frac{N_0 M}{2}}} \prod_{i=1}^{N_0} \int_{\Theta} \prod_{j=1}^M \exp\left(-\frac{(s_{ij} - s_\infty(t_j, \eta, \theta_i))^2}{2\sigma^2}\right) \mu_0^\theta(d\theta_i) \quad (3.113)$$

We can notice that the independence between the observed individuals is translated by the fact that the integral and the product signs have permuted, and that there is no latent variable associated with unobserved individual. The other level of simplification concerns the latent variable N . If we consider that the mean-field approximation is relevant above some threshold of the population size N_∞ , we can consider the following approximation of the joint likelihood density $\hat{p}_{s|\eta}^{N_\infty}$.

$$\hat{p}_{s|\eta}^{N_\infty}(s|\eta) = \sum_{N=N_0}^{N_\infty} p_N(N) p_{s|\eta, N}(s|\eta, N) + \left(\sum_{N=N_\infty+1}^{+\infty} p_N(N) \right) p_{s|\eta}^\infty(s|\eta) \quad (3.114)$$

In the original problem (equation 3.110), the support of the latent variable N is infinite, whereas in the approximated problem (equation 3.114) it has a finite support. Let us prove the consistency of this approximation. For convenience, we quantify the convergence of the the mean-field approximated density to the exact density in terms of total variation distance.

Definition 3.6. Total variation distance (from definition 2.4 and lemma 2.1 in Tsybakov (2008)²⁷)

Let $(\mathcal{Z}, \mathcal{B}_\mathcal{Z})$ be a measurable space. The total variation distance is the metric defined over the set of probability measures $\mathcal{P}(\mathcal{Z})$ by

$$\forall \mu, \nu \in \mathcal{P}(\mathcal{Z}), \quad d_{TV}(\mu, \nu) = \sup\{|\mu(A) - \nu(A)|, A \in \mathcal{B}_\mathcal{Z}\} \quad (3.115)$$

Let $\mu, \nu \in \mathcal{P}(\mathcal{Z})$ be two probability measures being absolutely continuous with respect to some σ -finite measure λ . Then we have

$$d_{TV}(\mu, \nu) = \frac{1}{2} \int_{\mathcal{Z}} \left| \frac{d\mu}{d\lambda}(z) - \frac{d\nu}{d\lambda}(z) \right| \lambda(dz) \quad (3.116)$$

Lemma 3.10. Mean convergence of Wasserstein distance

Let $(\mathcal{Z}, m_\mathcal{Z})$ be a Polish metric space and $\mu \in \mathcal{P}_p(\mathcal{Z})$ for some $p \geq 1$. Then:

- if $p = 1$ and if $\text{diam}(\mathcal{Z}) = \sup\{m_\mathcal{Z}(z_1, z_2), (z_1, z_2) \in (\mathcal{Z})^2\} < \infty$, we have

$$\lim_{N \rightarrow \infty} \mathbb{E} \left\{ \mathcal{W}_1 \left(\frac{1}{N} \sum_{n=1}^N \delta_{z_n}, \mu \right), z_{1:N} \sim \mu^{\otimes N} \right\} = 0 \quad (3.117)$$

- if $p > 1$, we have

$$\lim_{N \rightarrow \infty} \mathbb{E} \left\{ \mathcal{W}_p \left(\frac{1}{N} \sum_{n=1}^N \delta_{z_n}, \mu \right), z_{1:N} \sim \mu^{\otimes N} \right\} = 0 \quad (3.118)$$

Proof. Let $\varepsilon > 0$. Let $N \in \mathbb{N}^*$.

$$\begin{aligned} \mathbb{E} \left\{ \mathcal{W}_p(\hat{\mu}(z_{1:N}), \mu), z_{1:N} \sim \mu^{\otimes N} \right\} &\leq \frac{\varepsilon}{2} \mathbb{P} \left\{ z_{1:N} \sim \mu^{\otimes N} \mid \mathcal{W}_p(\hat{\mu}_N(z_{1:N}), \mu) \leq \frac{\varepsilon}{2} \right\} \\ &+ \mathbb{E} \left\{ \mathcal{W}_p(\hat{\mu}_N(z_{1:N}), \mu) \mathbb{I} \left\{ \mathcal{W}_p(\hat{\mu}_N(z_{1:N}), \mu) \geq \frac{\varepsilon}{2} \right\}, z_{1:N} \sim \mu^{\otimes N} \right\} \end{aligned} \quad (3.119)$$

By theorem 3.3, the Wasserstein distance between the empirical measure and the sampling distribution converges in probability to 0. If $p = 1$ and if $\text{diam}(\mathcal{Z}) < \infty$, there exists $N_0 \in \mathbb{N}^*$ such that for all $N \geq N_0$, we have

$$\mathbb{P} \left\{ z_{1:N} \sim \mu^{\otimes N} \mid \mathcal{W}_1(\hat{\mu}_N(z_{1:N}), \mu) \geq \frac{\varepsilon}{2} \right\} \leq \frac{\varepsilon}{2 \text{diam}(\mathcal{Z})} \quad (3.120)$$

Then for any $N \geq N_0$

$$\mathbb{E} \left\{ \mathcal{W}_1(\hat{\mu}_N(z_{1:N}), \mu), z_{1:N} \sim \mu^{\otimes N} \right\} \leq \varepsilon \quad (3.121)$$

If $p > 1$, there exists $N_1 \in \mathbb{N}^*$ such that for all $N \geq N_1$

$$\mathbb{P} \left\{ z_{1:N} \sim \mu^{\otimes N} \mid \mathcal{W}_p(\hat{\mu}_N(z_{1:N}), \mu) \geq \frac{\varepsilon}{2} \right\} \leq \frac{\varepsilon}{2} \left(\iint_{\mathcal{Z}^2} m_{\mathcal{Z}}(z_1, z_2)^p \mu^{\otimes 2}(dz_1, dz_2) \right)^{-\frac{1}{p}} \quad (3.122)$$

By Hölder's inequality, we have that

$$\begin{aligned} &\mathbb{E} \left\{ \mathcal{W}_p(\hat{\mu}_N(z_{1:N}), \mu) \mathbb{I} \left\{ \mathcal{W}_p(\hat{\mu}_N(z_{1:N}), \mu) \geq \varepsilon \right\}, z_{1:N} \sim \mu^{\otimes N} \right\} \\ &\leq \mathbb{E} \left\{ \mathcal{W}_p(\hat{\mu}_N(z_{1:N}), \mu)^p, z_{1:N} \sim \mu^{\otimes N} \right\}^{1/p} \mathbb{P} \left\{ \mathcal{W}_p(\hat{\mu}_N(z_{1:N}), \mu) \geq \frac{\varepsilon}{2} \right\} \end{aligned} \quad (3.123)$$

Besides, we can obtain an upper-bound of the Wasserstein distance by considering the independent coupling $\hat{\pi} = \hat{\mu}_N \otimes \mu$.

$$\begin{aligned} \mathcal{W}_p(\hat{\mu}_N, \mu)^p &\leq \frac{1}{N} \sum_{n=1}^N \int_{\mathcal{Z}} m_{\mathcal{Z}}(z_n, z')^p \mu(dz') \\ \mathbb{E} \left\{ \mathcal{W}_p(\hat{\mu}_N, \mu)^p, z_{1:N} \sim \mu^{\otimes N} \right\} &\leq \iint_{\mathcal{Z}^2} m_{\mathcal{Z}}(z_1, z_2)^p \mu^{\otimes 2}(dz_1, dz_2) \end{aligned} \quad (3.124)$$

We conclude that for all $N \geq N_1$, we have that

$$\mathbb{E} \left\{ \mathcal{W}_p(\hat{\mu}_N(z_{1:N}), \mu), z_{1:N} \sim \mu^{\otimes N} \right\} \leq \varepsilon \quad (3.125)$$

□

Proposition 3.4. Consistency of the mean-field approximation

Let $\mathcal{H} = [s_0^{\min}; s_0^{\max}] \times [\sigma_x^{\min}; \sigma_x^{\max}] \times [\sigma_r^{\min}; \sigma_r^{\max}]$ be the parameter space, endowed with the prior distribution $\mu_\eta \in \mathcal{P}(\mathcal{H})$. Let $\mu_0(s_0) = \delta_{s_0} \otimes \mu_\theta$ be the initial distribution of the system, parameterized by the initial size of the plants $s_0 \in [s_0^{\min}; s_0^{\max}]$. $\mu_\theta \in \mathcal{P}(\Theta)$ is such that $\forall s_0 \in [s_0^{\min}; s_0^{\max}]$ the initial distribution $\mu_0(s_0)$ satisfies the assumptions of corollary 3.1. Let p_N be the prior density of the population size N , with support included within $\llbracket N_0; +\infty \llbracket$. We consider the marginal distribution of the observations s conditionally to the population size N , having a density with respect to the Lebesgue measure of $\mathbb{R}^{N_0 M}$.

$$\forall s \in \mathbb{R}^{N_0 M}, \quad p_{s|N}(s) = \int_{\mathcal{H}} p_{s|\eta, N}(s|\eta, N) \mu_\eta(d\eta) \quad (3.126)$$

where the density $p_{s|\eta, N}$ is defined in equations (3.109) and (3.110). We also introduce the mean-field marginal density of the observation.

$$\forall s \in \mathbb{R}^{N_0 M}, \quad p_s^\infty(s) = \int_{\mathcal{H}} p_{s|\eta}^\infty(s|\eta) \mu_\eta(d\eta) \quad (3.127)$$

where $p_{s|\eta}^\infty$ is defined in equation (3.113). Then we have

$$d_{TV}(p_{s|N}, p_s^\infty) \xrightarrow{N \rightarrow \infty} 0 \quad (3.128)$$

Proof. Let $N \geq N_0$. We have

$$\begin{aligned} d_{TV}(p_{s|N}, p_s^\infty) &= \frac{1}{2} \int_{\mathbb{R}^{N_0 M}} |p_{s|N}(s) - p_s^\infty(s)| \lambda(ds) \\ &\leq \frac{1}{2(2\pi\sigma^2)^{\frac{N_0 M}{2}}} \int_{\mathbb{R}^{N_0 M}} \int_{\mathcal{H}} \int_{\Theta^N} \left| \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N_0} \sum_{j=1}^M (s_{ij} - s_i^N(t_j, \eta, \theta_{1:N}))^2\right) \right. \\ &\quad \left. - \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N_0} \sum_{j=1}^M (s_{ij} - s_\infty(t_j, \eta, \theta_i))^2\right) \right| \mu_\theta^{\otimes N}(d\theta_{1:N}) \mu_\eta(d\eta) \lambda(ds) \end{aligned} \quad (3.129)$$

We consider the term between the absolute values.

$$\begin{aligned}
 & \left| \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N_0} \sum_{j=1}^M (s_{ij} - s_i^N(t_j, \eta, \theta_{1:N}))^2 \right) - \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N_0} \sum_{j=1}^M (s_{ij} - s_\infty(t_j, \eta, \theta_i))^2 \right) \right| \\
 &= \exp \left(-\frac{\|s\|^2}{2\sigma^2} \right) \left| \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N_0} \sum_{j=1}^M (s_i^N(t_j, \eta, \theta_{1:N})^2 - 2s_{ij}s_i^N(t_j, \eta, \theta_{1:N})) \right) \right. \\
 & \quad \left. - \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N_0} \sum_{j=1}^M (s_\infty(t_j, \eta, \theta_i)^2 - 2s_{ij}s_\infty(t_j, \eta, \theta_i)) \right) \right| \\
 &\leq \frac{1}{2\sigma^2} \exp \left(-\frac{\|s\|^2}{2\sigma^2} \right) \left| \sum_{i=1}^{N_0} \sum_{j=1}^M (s_i^N(t_j, \eta, \theta_{1:N}) + s_\infty(t_j, \eta, \theta_i) - 2s_{ij}) \right. \\
 & \quad \left. \times (s_i^N(t_j, \eta, \theta_{1:N}) - s_\infty(t_j, \eta, \theta_i)) \right| \tag{3.130}
 \end{aligned}$$

Besides, we know from corollary 3.1 and proposition 1.3 that $|s_i^N(t_j, \eta, \theta_{1:N})| \leq S_i$ and $|s_\infty(t_j, \eta, \theta_i)| \leq S_i$.

$$\begin{aligned}
 & \left| \sum_{i=1}^{N_0} \sum_{j=1}^M (s_i^N(t_j, \eta, \theta_{1:N}) + s_\infty(t_j, \eta, \theta_i) - 2s_{ij})(s_i^N(t_j, \eta, \theta_{1:N}) - s_\infty(t_j, \eta, \theta_i)) \right| \\
 &\leq 2 \sum_{i=1}^{N_0} \sum_{j=1}^M (S_i + |s_{ij}|) |s_i^N(t_j, \eta, \theta_{1:N}) - s_\infty(t_j, \eta, \theta_i)| \tag{3.131}
 \end{aligned}$$

By integrating with respect to $\theta_{1:N}$, we obtain

$$\begin{aligned}
 & \sum_{i=1}^{N_0} \sum_{j=1}^M \int_{\Theta^N} (S_i + |s_{ij}|) |s_i^N(t_j, \eta, \theta_{1:N}) - s_\infty(t, \eta, \theta_i)| (\mu_0^\theta)^{\otimes N} (d\theta_{1:N}) \\
 &= \sum_{j=1}^M \left(N_0 \int_{\Theta^N} S_1 |s_1^N(t_j, \eta, \theta_{1:N}) - s_\infty(t_j, \eta, \theta_1)| (\mu_0^\theta)^{\otimes N} (d\theta_{1:N}) \right. \\
 & \quad \left. + \sum_{i=1}^{N_0} |s_{ij}| \int_{\Theta^N} |s_1^N(t_j, \eta, \theta_{1:N}) - s_\infty(t_j, \eta, \theta_1)| (\mu_0^\theta)^{\otimes N} (d\theta_{1:N}) \right) \tag{3.132}
 \end{aligned}$$

Let $\theta_{1:N} \in \Theta^N$ and $\eta \in \mathcal{H}$. We can estimate the discrepancy between the empirical and the macroscopic flow by introducing the competition potentials.

$$\begin{aligned}
 & |s_1^N(t, \eta, \theta_{1:N}) - s_\infty(t, \eta, \theta_1)| \\
 &\leq s_0 \left(\frac{S_1}{s_m} \right) \log \left(\frac{S_1}{s_m} \right) \left| \gamma_1 \int_0^t \left(\hat{C}_1^N(\tau, \eta, \theta_{1:N}) - C_\infty(\tau, \eta, \theta_1) \right) e^{\gamma_1(\tau-t)} d\tau \right| \tag{3.133}
 \end{aligned}$$

where the competition potentials \hat{C}_1^N and C_∞ are defined in equations (3.111) and (3.112). By a similar approach than the one used in the proof of proposition 3.3, we obtain the following inequality

$$\begin{aligned}
 & |s_1^N(t, \eta, \theta_{1:N}) - s_\infty(t, \eta, \theta_1)| \\
 & \leq \frac{1}{N-1} (\hat{s}_1(\eta, \theta_1, \hat{\mu}_N(0))e^{\beta(\eta)t} + \hat{s}_2(\eta, \theta_1, \hat{\mu}_N(0))e^{t/\tau(\eta, \theta_1) - \gamma_1 t} + \hat{s}_3(\eta, \theta_1, \hat{\mu}_N(0))) \\
 & + \mathcal{W}_2(\hat{\mu}_N^x(0), \mu_0^x) (a_1^x(\eta, \theta_1, \hat{\mu}_N(0))e^{\beta(\eta)t} + a_2^x(\eta, \theta_1, \hat{\mu}_N(0))e^{t/\tau(\eta, \theta_1) - \gamma_1 t} + a_3^x(\eta, \theta_1, \hat{\mu}_N(0))) \\
 & + \mathcal{W}_1(\hat{\mu}_N^\gamma(0), \mu_0^\gamma) (a_1^\gamma(\eta, \theta_1)e^{\beta(\eta)t} + a_2^\gamma(\eta, \theta_1)e^{t/\tau(\eta, \theta_1) - \gamma_1 t} + a_3^\gamma(\eta, \theta_1)) \\
 & + \mathcal{W}_1(\hat{\mu}_N^S(0), \mu_0^S) (a_1^S(\eta, \theta_1)e^{\beta(\eta)t} + a_2^S(\eta, \theta_1)e^{t/\tau(\eta, \theta_1) - \gamma_1 t})
 \end{aligned} \tag{3.134}$$

The expressions of the coefficients \hat{s}_i, a_i^θ , along with their derivations, are given in the appendix, page 162. It is also proved there that these coefficients are bounded. We can therefore obtain by applying lemma 3.10 that for all $\eta \in \mathcal{H}$ and for all $s \in \mathbb{R}^{N_0 M}$

$$\lim_{N \rightarrow \infty} \sum_{i=1}^{N_0} \sum_{j=1}^M \int_{\Theta^N} (S_i + |s_{ij}|) |s_i^N(t_j, \eta, \theta_{1:N}) - s_\infty(t, \eta, \theta_i)| (\mu_0^\theta)^{\otimes N} (d\theta_{1:N}) = 0 \tag{3.135}$$

By dominated convergence, we can conclude that

$$\lim_{N \rightarrow \infty} \int_{\mathbb{R}^{N_0 M}} \int_{\mathcal{H}} |p_{s|\eta, N}(s|\eta, N) - p_{s|\eta}^\infty(s|\eta)| \mu_\eta(d\eta) \lambda(ds) = 0 \tag{3.136}$$

□

In the previous proof, we can notice that the critical size of the population, the threshold above which we can consider that the mean-field limit is a relevant approximation (for a given tolerance), depends not only on the dynamics of the model, but also on the experimental conditions, on the way data were collected. To clarify this, let us consider the upper-bound of the total variation of the marginal distribution of the observations at time t_j .

$$\begin{aligned}
 d_{TV}(p_{s_j|N}, p_{s_j}^\infty) & \leq \frac{1}{2(2\pi\sigma^2)^{N_0/2}} \int_{\mathbb{R}^{N_0}} \int_{\mathcal{H}} \int_{\Theta^N} \frac{1}{\sigma^2} \exp\left(-\frac{|s|^2}{2\sigma^2}\right) \left(N_0 S_1 + \sum_{i=1}^{N_0} |s_i|\right) \frac{s_0 S_1}{s_m} \\
 & \times \log\left(\frac{S_1}{s_m}\right) \left| \gamma_1 \int_0^{t_j} \left(\hat{C}_1^N(\tau, \eta, \theta_{1:N}) - C_\infty(\tau, \eta, \theta_{1:N})\right) e^{\gamma_1(\tau - t_j)} d\tau \right| \mu_\theta^{\otimes N}(d\theta_{1:N}) \mu_\eta(d\eta) \\
 & \times \lambda(ds)
 \end{aligned} \tag{3.137}$$

$$\leq \frac{N_0}{\sigma^2} \left(s_m e^{R_M} + \sigma \sqrt{\frac{2}{\pi}} \right) s_0^{\max} R_M e^{R_M} \mathbb{E} \left\{ \text{comp}_N(t_j, \eta, \theta_{1:N}), \eta \sim \mu_\eta, \theta_{1:N} \sim \mu_\theta^{\otimes N} \right\}$$

$$\text{where } \text{comp}_N(t_j, \eta, \theta_{1:N}) = \left| \gamma_1 \int_0^{t_j} \left(\hat{C}_1^N(\tau, \eta, \theta_{1:N}) - C_\infty(\tau, \eta, \theta_1) \right) e^{\gamma_1(\tau - t_j)} d\tau \right| \quad (3.138)$$

As it was proved in equation (3.135), the expectation term tends to zero. We can therefore have a qualitative idea of the relation between the critical size N_ε , above which we have $d_{TV}(p_{s_j|N}, p_{s_j}^\infty) \leq \varepsilon$, and the parameters related to the experimental conditions N_0, σ . We can see in particular that for all $\varepsilon > 0$, we can expect N_ε to be increasing with N_0 and decreasing with σ . In a nutshell, the more accurate the experimental conditions, the larger the size of the population that can be approximated by the mean-field dynamics.

3.4.3 Second order approximation in the case of the homogeneous Spring Cloud system

What can be said about a system that has only a mean-field limit after normalization? This is the case, for instance, of the Spring Cloud and the Cucker-Smale models, which require a change of the time scale and a renormalization of the velocity variables to converge towards a dynamical system independent of the size of the population. We can see in a simple example that the convergence in total variation cannot be generalized in all cases.

Let us come back to the example of the homogeneous Spring Cloud system. We assume that the initial conditions are normally distributed and that the parameters κ, m are the same for all the N particles. It is shown in equation (2.7) that the trajectories of the particles are elliptic around the barycenter, itself following a rectilinear uniform motion.

$$\hat{x}_1^N(t, x_{1:N}^0, v_{1:N}^0) = \frac{1}{N} \sum_{i=1}^N x_i^0 + \frac{t\sqrt{N}}{N} \sum_{i=1}^N \frac{v_i^0}{\sqrt{N}} + \left(x_1^0 - \frac{1}{N} \sum_{i=1}^N x_i^0 \right) \cos(\omega t \sqrt{N}) \quad (3.139)$$

$$+ \left(\frac{v_1^0}{\sqrt{N}} - \frac{1}{N} \sum_{i=1}^N \frac{v_i^0}{\sqrt{N}} \right) \frac{\sin(\omega t \sqrt{N})}{\omega} \text{ where } \omega = \frac{\kappa}{\sqrt{m}},$$

$$v_1^N(t, x_{1:N}^0, v_{1:N}^0) = \frac{1}{N} \sum_{i=1}^N v_i^0 - \omega \sqrt{N} \left(x_1^0 - \frac{1}{N} \sum_{i=1}^N x_i^0 \right) \sin(\omega t \sqrt{N}) \quad (3.140)$$

$$+ \left(v_1^0 - \frac{1}{N} \sum_{i=1}^N v_i^0 \right) \cos(\omega t \sqrt{N}).$$

The velocity of a particle is all the more high as the population is large, with an order of magnitude of \sqrt{N} if the cloud is not initially concentrated at a single point. It seems therefore reasonable to assume that the initial distribution of the variable v depends on N , with for instance a normal distribution like $\mathcal{N}(0, N\Sigma_v)$. By the law of large numbers, we obtain an almost sure convergence towards the mean-field flow after normalization of the time-scale and of the velocity variables.

$$\begin{aligned} & \mathbb{P} \left\{ (x_n^0)_{n \in \mathbb{N}^*} \sim \mathcal{N}(\mu_x, \Sigma_x)^{\otimes \infty}, (v_n^0)_{n \in \mathbb{N}^*} \sim \mathcal{N}(0, \Sigma_v)^{\otimes \infty} \mid \right. \\ & \forall t \geq 0, \lim_{N \rightarrow \infty} x_1^N(t/\sqrt{N}, x_{1:N}^0, v_{1:N}^0 \sqrt{N}) = x_\infty(t, x_1^0, v_1^0), \\ & \left. \lim_{N \rightarrow \infty} \frac{v_1^N(t/\sqrt{N}, x_{1:N}^0, v_{1:N}^0 \sqrt{N})}{\sqrt{N}} = v_\infty(t, x_1^0, v_1^0) \right\} = 1, \quad (3.141) \\ & \text{with } x_\infty(t, x_1^0, v_1^0) = \mu_x + (x_1^0 - \mu_x) \cos(\omega t) + v_1^0 \frac{\sin(\omega t)}{\omega}, \\ & v_\infty(t, x_1^0, v_1^0) = -\omega(x_1^0 - \mu_x) \sin(\omega t) + v_1^0 \cos(\omega t). \end{aligned}$$

In the inference problems we consider, the observation times are fixed, and they therefore cannot change with the size of the population, which may be unknown. By applying the reverse normalization, we project the mean-field dynamics onto the temporality of the observer. Let us study the gap between the two flows within the non-normalized temporality.

$$\begin{aligned} x_1^N(t, x_{1:N}^0, v_{1:N}^0) - x_\infty(t\sqrt{N}, x_1^0, v_1^0/\sqrt{N}) &= \left(\frac{1}{N} \sum_{i=1}^N x_i^0 - \mu_x \right) (1 - \cos(\omega t \sqrt{N})) \\ &+ \frac{1}{N} \sum_{i=1}^N \frac{v_i^0}{\sqrt{N}} \left(t\sqrt{N} - \frac{\sin(\omega t \sqrt{N})}{\omega} \right) \end{aligned} \quad (3.142)$$

If the initial conditions are normally distributed, the distribution of this gap is also Gaussian, with the following mean and variance

$$\begin{aligned} & \mathbb{E} \left\{ x_1^N(t, x_{1:N}^0, v_{1:N}^0) - x_\infty(t\sqrt{N}, x_1^0, v_1^0/\sqrt{N}), x_{1:N}^0 \sim \mathcal{N}(\mu_x, \Sigma_x)^{\otimes N}, \right. \\ & \left. v_{1:N}^0 \sim \mathcal{N}(0, N\Sigma_v)^{\otimes N} \right\} = 0, \\ & \text{Cov} \left\{ x_1^N(t, x_{1:N}^0, v_{1:N}^0) - x_\infty(t\sqrt{N}, x_1^0, v_1^0/\sqrt{N}), x_{1:N}^0 \sim \mathcal{N}(\mu_x, \Sigma_x)^{\otimes N}, \right. \\ & \left. v_{1:N}^0 \sim \mathcal{N}(0, N\Sigma_v)^{\otimes N} \right\} = (1 - \cos(\omega t \sqrt{N}))^2 \frac{\Sigma_x}{N} + \left(t\sqrt{N} - \frac{\sin(\omega t \sqrt{N})}{\omega} \right)^2 \frac{\Sigma_v}{N}. \end{aligned} \quad (3.143)$$

If $\Sigma_v \neq 0$, we can notice that the approximation by x_∞ is not consistent in L^2 . This is due to the fact that the sequence of the empirical distributions of the velocities is not tight. We need to have a higher order expansion to obtain a consistent approximation. We have the following convergence in distribution

$$\begin{aligned} & \forall \varphi \in \mathcal{C}_b(\mathbb{R}^2), \mathbb{E} \left\{ \varphi \left(\sqrt{N} \left(x_1^N(t/\sqrt{N}, x_{1:N}^0, v_{1:N}^0) - x_\infty(t, x_1^0, v_1^0/\sqrt{N}) \right) \right), \right. \\ & \left. x_{1:N}^0 \sim \mathcal{N}(\mu_x, \Sigma_x)^{\otimes N}, v_{1:N}^0 \sim \mathcal{N}(0, N\Sigma_v)^{\otimes N} \right\} \\ & \xrightarrow{N \rightarrow \infty} \mathbb{E} \left\{ \varphi \left(u_x(1 - \cos(\omega t)) + u_v \left(t - \frac{\sin(\omega t)}{\omega} \right) \right), u_x \sim \mathcal{N}(0, \Sigma_x), u_v \sim \mathcal{N}(0, \Sigma_v) \right\} \end{aligned} \quad (3.144)$$

We can deduce from the above convergence a second order approximation of the microscopic dynamics, where auxiliary variables u_x and u_v are introduced to represent the variance of the gap between the empirical flow and the mean-field flow.

$$\begin{aligned} & u_x, u_v \sim \mathcal{N}(0, \mathbf{I}_2) \\ & x_1^0 \sim \mathcal{N}(\mu_x, \Sigma_x), v_1^0 \sim \mathcal{N}(0, N\Sigma_v) \\ & \hat{x}_\infty^N(t, x_1^0, v_1^0, u_x, u_v) = x_\infty(t\sqrt{N}, x_1^0, v_1^0/\sqrt{N}) + \Sigma_x^{1/2} u_x \frac{1 - \cos(\omega t\sqrt{N})}{\sqrt{N}} \\ & + \Sigma_v^{1/2} u_v \left(t - \frac{\sin(\omega t\sqrt{N})}{\omega\sqrt{N}} \right) \end{aligned} \quad (3.145)$$

To simulate a group of N_0 individuals within a population of size N , we need first to sample variables u_x and u_v and then to apply the equation (3.145) for N_0 initial conditions x_i^0, v_i^0 sampled independently. Under this approximated simulation, the number of latent variables does not depend on the size of the population.

Such an approach cannot be applied straightforwardly in the case of a non-linear system, such as the Cucker-Smale model, but the methodology seems to be roughly the same.

1. First, we derive the normalizations of the time-scale and the state variables to obtain a mean-field limit.
2. We then study the asymptotic distribution of the gap between the two flows within the current temporality, after performing the reverse normalizations.
3. If the mean-field flow is not sufficient to provide a consistent approximation for the inference problem, then we consider a higher order expansion by deriving the speed of convergence of the microscopic flow to the mean-field flow (in the previous example, the speed of convergence is \sqrt{N}).

Thus, the derivation of the mean-field limit appears as an unavoidable step in the simplification of microscopic dynamics. This simplification is effective in the case of linear or analytical systems, but in the general case, it is strongly conditioned by the numerical accessibility of the mean-field flow, the simulation of which is an active subject of research.

3.5 Lagrangian simulations of the mean-field characteristic flow: two case studies

In this section, we detail two methodologies to obtain numerical approximations of the mean-field limit distribution in two specific cases: the Spring Cloud system and the Schneider system. These preliminary results remain relatively dependent on the properties of the underlying system, but they might open up the prospect of being extended to more generalized situations. In both cases, an approximation of the mean-field characteristic flow is built by estimating the population statistics, which is the barycenter in the case of Spring Cloud, and the competition potential in the case of the Schneider system. The time-evolutions of these statistics are approached by piecewise polynomial functions, exactly as in the microscopic case, but in the macroscopic case, the coefficients might be spatially dependent. The conditions of consistency of the numerical scheme are only considered qualitatively. Consistent numerical methods to solve non-local transport equations in low dimensions can be found in Carrillo et al. (2008),⁶ Lafitte et al. (2016),¹⁷ Lagoutière and Vauchelet (2017),¹⁸ but these methods cannot be used directly in our case, due to the high dimension of the phase space.

3.5.1 The Spring Cloud system

Let us come back to the expression of the mean-field characteristic flow of the Spring Cloud system. Let μ_0 be a probability measure in $\mathcal{P}_1(\mathbb{R}^4 \times \Theta)$, with $\Theta = [\kappa_{\min}; \kappa_{\max}] \times [m_{\min}; m_{\max}]$ the space of masses and attractivenesses κ . Then the mean-field characteristic defined in equation (3.84) can be expressed using the

barycenter of the attractivenesses.

Let $\bar{\kappa} = \mathbb{E}\{\kappa, \kappa \sim \mu_0^\kappa\}$, and $\omega_{\kappa,m} = \sqrt{\frac{\kappa\bar{\kappa}}{m}}$.

We have:

$$\begin{aligned}
 x_\infty(t, x_0, v_0, \kappa, m) &= x_0 \cos(\omega_{\kappa,m}t) + v_0 \frac{\sin(\omega_{\kappa,m}t)}{\omega_{\kappa,m}} + \int_0^t \omega_{\kappa,m} x_\kappa^\infty(\tau) \sin(\omega_{\kappa,m}(t-\tau)) d\tau \\
 v_\infty(t, x_0, v_0, \kappa, m) &= -\omega_{\kappa,m} x_0 \sin(\omega_{\kappa,m}t) + v_0 \cos(\omega_{\kappa,m}t) \\
 &+ \int_0^t \omega_{\kappa,m}^2 x_\kappa^\infty(\tau) \cos(\omega_{\kappa,m}(t-\tau)) d\tau, \\
 \text{where } x_\kappa^\infty(t) &= \frac{1}{\bar{\kappa}} \mathbb{E} \{ \kappa x_\infty(t, x_0, v_0, \kappa, m), (x_0, v_0, \kappa, m) \sim \mu_0 \}.
 \end{aligned} \tag{3.146}$$

We consider a piecewise linear approximation of the barycenter.

$$\hat{x}_\kappa^\infty(t) = \sum_{k=0}^{M-1} (a_k + b_k(t-t_k)) \mathbb{I}\{t_k \leq t < t_{k+1}\} \tag{3.147}$$

where the coefficients $a_k, b_k \in \mathbb{R}^2$ are to be determined and where $\{t_0 = 0, \dots, t_M = T\}$ is a subdivision of the interval $[0; T]$. In the microscopic case, such an approximation leads to a consistent numerical scheme of second order. Under this approximation, the mean-field characteristic is a linear combination of sinusoidal functions, with time-varying coefficients.

$$\begin{aligned}
 \hat{x}_\infty(t, x_0, v_0, \kappa, m) &= x_0 \cos(\omega_{\kappa,m}t) + v_0 \frac{\sin(\omega_{\kappa,m}t)}{\omega_{\kappa,m}} + \delta\hat{x}(t, \kappa, m) \\
 \delta\hat{x}(t, \kappa, m) &= \sum_{k=0}^{M-1} \int_0^t \omega_{\kappa,m} (a_k + b_k(t'-t_k)) \mathbb{I}\{t_k \leq t' < t_{k+1}\} \sin(\omega_{\kappa,m}(t-t')) dt' \\
 \hat{v}_\infty(t, x_0, v_0, \kappa, m) &= -\omega_{\kappa,m} x_0 \sin(\omega_{\kappa,m}t) + v_0 \cos(\omega_{\kappa,m}t) + \delta\hat{v}(t, \kappa, m) \\
 \delta\hat{v}(t, \kappa, m) &= \sum_{k=0}^{M-1} \int_0^t \omega_{\kappa,m}^2 (a_k + b_k(t'-t_k)) \mathbb{I}\{t_k \leq t' < t_{k+1}\} \cos(\omega_{\kappa,m}(t-t')) dt'
 \end{aligned} \tag{3.148}$$

The coefficients a_k, b_k correspond to the position and the velocity of the barycenter at the level of the time of the subdivision.

$$\begin{aligned}
 \forall k \in \llbracket 0; M-1 \rrbracket, a_k &= \frac{1}{\bar{\kappa}} \mathbb{E} \{ \kappa \hat{x}_\infty(t_k, x_0, v_0, \kappa, m), (x_0, v_0, \kappa, m) \sim \mu_0^\kappa \} \\
 b_k &= \frac{1}{\bar{\kappa}} \mathbb{E} \left\{ \kappa \frac{\partial \hat{x}_\infty}{\partial t}(t_k, x_0, v_0, \kappa, m), (x_0, v_0, \kappa, m) \sim \mu_0 \right\}
 \end{aligned} \tag{3.149}$$

We can therefore obtain the coefficients by induction. Initially, the position and the speed of the barycenter are a_0 and b_0 .

$$a_0 = \frac{1}{\bar{\kappa}} \mathbb{E} \{ \kappa x_0, (x_0, \kappa) \sim \mu_0^{x, \kappa} \}, \quad b_0 = \frac{1}{\bar{\kappa}} \mathbb{E} \{ \kappa v_0, (v_0, \kappa) \sim \mu_0^{v, \kappa} \} \quad (3.150)$$

We have then an inductive equation to compute the position of the barycenter. For all $k \in \llbracket 0; M - 2 \rrbracket$

$$\begin{aligned} a_{k+1} &= \mu_{x,x}^\kappa(t_{k+1}) + \mu_{x,v}^\kappa(t_{k+1}) + \sum_{\ell=0}^k a_\ell [c(t_{k+1} - t_{\ell+1}) - c(t_{k+1} - t_\ell)] \\ &+ \sum_{\ell=0}^k b_\ell [(t_{\ell+1} - t_\ell)c(t_{k+1} - t_{\ell+1}) + s(t_{k+1} - t_{\ell+1}) - s(t_{k+1} - t_\ell)] \\ \mu_{x,x}^\kappa(t) &= \frac{1}{\bar{\kappa}} \int_{\mathbb{R}^2 \times \Theta} \kappa x_0 \cos(\omega_{\kappa,m} t) \mu_0^{x,\kappa,m}(\mathrm{d}x_0, \mathrm{d}\kappa, \mathrm{d}m) \\ \mu_{x,v}^\kappa(t) &= \frac{1}{\bar{\kappa}} \int_{\mathbb{R}^2 \times \Theta} \kappa v_0 \frac{\sin(\omega_{\kappa,m} t)}{\omega_{\kappa,m}} \mu_0^{v,\kappa,m}(\mathrm{d}v_0, \mathrm{d}\kappa, \mathrm{d}m) \\ c(t) &= \frac{1}{\bar{\kappa}} \int_{\Theta} \kappa \cos(\omega_{\kappa,m} t) \mu_0^{\kappa,m}(\mathrm{d}\kappa, \mathrm{d}m) \\ s(t) &= \frac{1}{\bar{\kappa}} \int_{\Theta} \kappa \frac{\sin(\omega_{\kappa,m} t)}{\omega_{\kappa,m}} \mu_0^{\kappa,m}(\mathrm{d}\kappa, \mathrm{d}m) \end{aligned} \quad (3.151)$$

We have a similar inductive equation for the speed of the barycenter.

$$\begin{aligned} b_{k+1} &= \frac{d\mu_{x,x}^\kappa}{dt}(t_{k+1}) + \frac{d\mu_{x,v}^\kappa}{dt}(t_{k+1}) + \sum_{\ell=0}^k a_\ell \left(\frac{dc}{dt}(t_{k+1} - t_{\ell+1}) - \frac{dc}{dt}(t_{k+1} - t_\ell) \right) \\ &+ \sum_{\ell=0}^k b_\ell \left((t_{\ell+1} - t_\ell) \frac{dc}{dt}(t_{k+1} - t_{\ell+1}) + c(t_{k+1} - t_{\ell+1}) - c(t_{k+1} - t_\ell) \right) \\ \frac{d\mu_{x,x}^\kappa}{dt}(t) &= -\frac{1}{\bar{\kappa}} \mathbb{E} \{ \kappa \omega_{\kappa,m} x_0 \sin(\omega_{\kappa,m} t), (x_0, \kappa, m) \sim \mu_0^{x,\kappa,m} \} \\ \frac{d\mu_{x,v}^\kappa}{dt}(t) &= \frac{1}{\bar{\kappa}} \mathbb{E} \{ \kappa v_0 \cos(\omega_{\kappa,m} t), (v_0, \kappa, m) \sim \mu_0^{v,\kappa,m} \} \\ \frac{dc}{dt}(t) &= -\frac{1}{\bar{\kappa}} \int_{\Theta} \kappa \omega_{\kappa,m} \sin(\omega_{\kappa,m} t) \mu_0^{\kappa,m}(\mathrm{d}\kappa, \mathrm{d}m) \end{aligned} \quad (3.152)$$

Once the coefficients are computed, we can derive the time-evolution of the terms $\delta \hat{x}(t, \kappa, m)$ and $\delta \hat{v}(t, \kappa, m)$, the components of the particle trajectory that is related

to the motion of the barycenter. For the position, we have

$$\begin{aligned}
 \delta\hat{x}(t, \kappa, m) &= \omega_{\kappa, m} \sum_{k=0}^{M-1} \mathbb{I}\{t_k \leq t < t_{k+1}\} \int_{t_k}^t (a_k + b_k(t' - t_k)) \sin(\omega_{\kappa, m}(t - t')) dt' \\
 &+ \omega_{\kappa, m} \sum_{k=0}^{M-1} \mathbb{I}\{t_{k+1} \leq t\} \int_{t_k}^{t_{k+1}} (a_k + b_k(t' - t_k)) \sin(\omega_{\kappa, m}(t - t')) dt' \\
 &= \sum_{k=0}^{M-1} \mathbb{I}\{t_k \leq t < t_{k+1}\} \left(a_k(1 - \cos(\omega_{\kappa, m}(t - t_k))) - b_k \frac{\sin(\omega_{\kappa, m}(t - t_k))}{\omega_{\kappa, m}} + b_k(t - t_k) \right) \\
 &+ \sum_{k=0}^{M-1} \mathbb{I}\{t_{k+1} \leq t\} (a_k (\cos(\omega_{\kappa, m}(t - t_{k+1})) - \cos(\omega_{\kappa, m}(t - t_k))) \\
 &+ b_k \left((t_{k+1} - t_k) \cos(\omega_{\kappa, m}(t - t_{k+1})) + \frac{\sin(\omega_{\kappa, m}(t - t_{k+1})) - \sin(\omega_{\kappa, m}(t - t_k))}{\omega_{\kappa, m}} \right))
 \end{aligned} \tag{3.153}$$

Similarly, for the velocity component, we have

$$\begin{aligned}
 \delta\hat{v}(t, \kappa, m) &= \sum_{k=0}^{M-1} \mathbb{I}\{t_k \leq t < t_{k+1}\} (b_k(1 - \cos(\omega_{\kappa, m}(t - t_k))) + a_k \omega_{\kappa, m} \sin(\omega_{\kappa, m}(t - t_k))) \\
 &+ \sum_{k=0}^{M-1} \mathbb{I}\{t_{k+1} \leq t\} (a_k \omega_{\kappa, m} (\sin(\omega_{\kappa, m}(t - t_k)) - \sin(\omega_{\kappa, m}(t - t_{k+1}))) + b_k (\cos(\omega_{\kappa, m}(t - t_{k+1})) \\
 &- \cos(\omega_{\kappa, m}(t - t_k)) - (t_{k+1} - t_k) \omega_{\kappa, m} \sin(\omega_{\kappa, m}(t - t_{k+1}))))
 \end{aligned} \tag{3.154}$$

The mean-field characteristic flows can be written as an linear transformation of the initial conditions (x_0, v_0) .

$$\begin{aligned}
 \begin{pmatrix} \hat{x}_\infty(t, x_0, v_0, \kappa, m) \\ \hat{v}_\infty(t, x_0, v_0, \kappa, m) \end{pmatrix} &= \Omega_{\kappa, m}(t) \begin{pmatrix} x_0 \\ v_0 \end{pmatrix} + \begin{pmatrix} \delta\hat{x}(t, \kappa, m) \\ \delta\hat{v}(t, \kappa, m) \end{pmatrix} \\
 \Omega_{\kappa, m}(t) &= \begin{pmatrix} \cos(\omega_{\kappa, m}t) \mathbf{I}_2 & \frac{\sin(\omega_{\kappa, m}t)}{\omega_{\kappa, m}} \mathbf{I}_2 \\ -\omega_{\kappa, m} \sin(\omega_{\kappa, m}t) \mathbf{I}_2 & \cos(\omega_{\kappa, m}t) \mathbf{I}_2 \end{pmatrix}
 \end{aligned} \tag{3.155}$$

The expression of the inverse of the characteristic flow can be derived analytically as the solution of a linear system. $x_0(t, x, v, \kappa, m)$ and $v_0(t, x, v, \kappa, m)$ gives the initial condition of a particle located at (x, v) at time t , with the parameters (κ, m) .

$$\begin{aligned}
 \hat{x}_0(t, x, v, \kappa, m) &= (x - \delta x(t, \kappa, m)) \cos(\omega_{\kappa, m}t) - (v - \delta v(t, \kappa, m)) \frac{\sin(\omega_{\kappa, m}t)}{\omega_{\kappa, m}} \\
 \hat{v}_0(t, x, v, \kappa, m) &= (x - \delta x(t, \kappa, m)) \omega_{\kappa, m} \sin(\omega_{\kappa, m}t) + (v - \delta v(t, \kappa, m)) \cos(\omega_{\kappa, m}t)
 \end{aligned} \tag{3.156}$$

If the initial distribution μ_0 is absolutely continuous with respect to the Lebesgue measure, with density f_0 , it follows that the mean-field distributions is also absolutely continuous for all time $t \geq 0$, and is expressed using the reverse characteristic flows.

$$\hat{f}(t, x, v, \kappa, m) = f_0(x_0(t, x, v, \kappa, m), v_0(t, x, v, \kappa, m), \kappa, m) \quad (3.157)$$

Notice that the Jacobian of the reverse characteristic flow is constant and unitary for all time, so it does not appear in the above equation. Equation (3.157) provides a consistent approximation, when the time-step of the subdivision tends to zero, of the strong solution of the non-local transport equation associated with Spring-Cloud system.

$$\begin{aligned} & \frac{\partial f}{\partial t}(t, x, v, \kappa, m) + v \cdot \frac{\partial f}{\partial x}(t, x, v, \kappa, m) \\ & + \frac{\kappa}{m} \frac{\partial f}{\partial v}(t, x, v, \kappa, m) \cdot \int_{\mathbb{R}^2 \times [\kappa_1; \kappa_2]} \kappa'(x' - x) f^{x, \kappa}(t, x', \kappa') \lambda(dx', d\kappa') = 0 \quad (3.158) \\ & f(0, x, v, \kappa, m) = f_0(x, v, \kappa, m) \end{aligned}$$

In practice however, none of the expectations used to expressed the coefficients $a_k, b_k, \mu_{x,x}, \mu_{x,v}$ are analytical, and one has to resort to Monte-Carlo integration to compute them. Typically, for the coefficient $c(t)$ for instance, we need a sample $(\kappa_j, m_j)_{1 \leq j \leq K}$ from the distribution $\mu_0^{\kappa, m}$.

$$c_K(t) = \frac{1}{K\bar{\kappa}} \sum_{j=1}^K \kappa_j \cos(\omega(\kappa_j, m_j)t) \quad (3.159)$$

If we use the same sample size K for all the Monte-Carlo integrations, we need to choose K jointly with the time-step Δt of the subdivision. Indeed, if the time-step is too small and if the Monte-Carlo approximation is too rough, the numerical errors accumulate through time and this may result in a significant bias in the simulation. The relationship between K and Δt is still to be investigated.

Let us now illustrate the suggested numerical scheme in an example of Spring-Cloud system with a bimodal initial distribution, with initial conditions normally distributed.

$$\mu_0 = \frac{1}{2} \sum_{i=1}^2 \mathcal{N}(x_i, \sigma_x^2 \mathbf{I}_2) \otimes \mathcal{N}(v_i, \sigma_v^2 \mathbf{I}_2) \otimes \mathcal{U}([\kappa_i - \delta\kappa; \kappa_i + \delta\kappa]) \otimes \mathcal{U}([m_i - \delta m; m_i + \delta m]) \quad (3.160)$$

From equation (3.155), we can derive the conditional distribution of the state

variables at time t knowing the parameters (κ, m) .

$$(x_t, v_t) | \kappa, m \sim \sum_{i=1}^2 \mathbb{I}\{(\kappa, m) \in [\kappa_i \pm \delta\kappa] \times [m_i \pm \delta m]\} \mathcal{N} \left(\begin{pmatrix} x_\infty(t, x_i, v_i, \kappa, m) \\ v_\infty(t, x_i, v_i, \kappa, m) \end{pmatrix}, \Sigma_{\kappa, m}(t) \right) \quad (3.161)$$

$$\begin{aligned} \Sigma_{\kappa, m}(t) &= \Omega_{\kappa, m}(t) \begin{pmatrix} \sigma_x^2 \mathbf{I}_2 & 0 \\ 0 & \sigma_v^2 \mathbf{I}_2 \end{pmatrix} \Omega_{\kappa, m}(t)^\top \\ &= \begin{pmatrix} \left(\sigma_x^2 \cos(\omega_{\kappa, m} t)^2 + \sigma_v^2 \left(\frac{\sin(\omega_{\kappa, m} t)}{\omega_{\kappa, m}} \right)^2 \right) \mathbf{I}_2 & \frac{(\sigma_v^2 - \sigma_x^2 \omega_{\kappa, m}^2) \sin(2\omega_{\kappa, m} t)}{2\omega_{\kappa, m}} \mathbf{I}_2 \\ \frac{(\sigma_v^2 - \sigma_x^2 \omega_{\kappa, m}^2) \sin(2\omega_{\kappa, m} t)}{2\omega_{\kappa, m}} \mathbf{I}_2 & (\sigma_x^2 \omega_{\kappa, m}^2 \sin(\omega_{\kappa, m} t)^2 + \sigma_v^2 \cos(\omega_{\kappa, m} t)^2) \mathbf{I}_2 \end{pmatrix} \end{aligned} \quad (3.162)$$

So we can derive the marginal distribution of the x variable at time t :

$$f^x(t, x) = \frac{1}{4\pi} \sum_{i=1}^2 \mathbb{E} \left\{ \frac{\exp \left(-\frac{|x - x_\infty(t, x_i, v_i, \kappa, m)|^2}{2 \left(\sigma_x^2 \cos(\omega_{\kappa, m} t)^2 + \sigma_v^2 \frac{\sin(\omega_{\kappa, m} t)^2}{\omega_{\kappa, m}^2} \right)} \right)}{\sigma_x^2 \cos(\omega_{\kappa, m} t)^2 + \sigma_v^2 \frac{\sin(\omega_{\kappa, m} t)^2}{\omega_{\kappa, m}^2}}, (\kappa, m) \sim \mu_i^{\kappa, m} \right\} \quad (3.163)$$

$$\mu_i^{\kappa, m} = \mathcal{U}([\kappa_i \pm \delta\kappa]) \otimes \mathcal{U}([m_i \pm \delta m])$$

So the trajectories of the particles remain around their respective barycenters, with a time varying variance. Figure (3.1) shows the trajectories of the mean position vectors (for the conditional distribution). The two clouds are chosen to be non-symmetric, the cloud initially on the left has lower κ, m in average, whereas the cloud initially on the right has higher κ, m . The initial speeds are parallel to the y -axis, with opposite directions. We can notice on the figure the motions of the two centers do not correspond to the trajectories of two particles in interaction, in other words the macroscopic systems cannot be summarized by two particles.

Simulating the marginal distribution $f^x(t, x)$ can be done efficiently by first simulating a sample (κ, m) from $\mu_i^{\kappa, m}$ ($i = 1, 2$) and by applying the following formula:

$$x_t = x_\infty(t, x_i, v_i, \kappa, m) + u_x \sqrt{\sigma_x^2 \cos(\omega_{\kappa, m} t)^2 + \sigma_v^2 \frac{\sin(\omega_{\kappa, m} t)^2}{\omega_{\kappa, m}^2}}, \quad u_x \sim \mathcal{N}(0, \mathbf{I}_2). \quad (3.164)$$

Figure 3.2 shows the time-evolution of the marginal density $f^x(t, x)$, reconstructed by kernel density estimation. We can notice that the two clouds are initially well separated, and that at time $t = 10$ for instance, they are almost merged together.

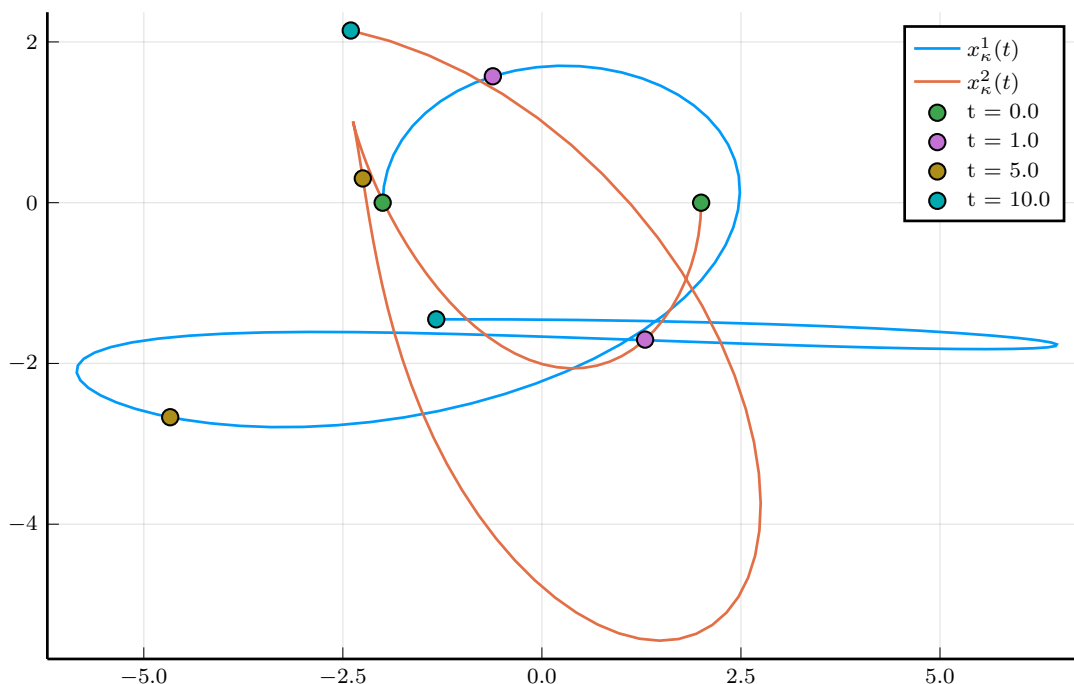


Figure 3.1: Trajectories of the barycenter of the two initial clouds: $x_\kappa^1(t) = x_\infty(t, x_1, v_1, \kappa_1, m_1)$ and $x_\kappa^2(t) = x_\infty(t, x_2, v_2, \kappa_2, m_2)$. These trajectories give an indication on the location of the centers of the two clouds. The detailed configuration of this simulation is presented in the appendix, page 166.

3.5.2 The Schneider system

We now consider the case of the Schneider system, which is non-linear with respect to the state variables, as opposed to the Spring Cloud system. The principle of the methodology is roughly the same as in the previous section, i.e., it is based on an approximation of the dynamics of the population statistics driving the motion of the individuals. For the Schneider system, this population statistic is the competition potential. It is therefore no longer a question of approaching the motion of a point in phase space (as for the Spring Cloud barycenter), it is a question of modelling the evolution of a function depending on time and state variables. As previously done and in keeping with the numerical scheme developed in the microscopic case, the time evolution of the competition potential is approximated by a piecewise polynomial function, specifically piecewise constant in the following example. In the macroscopic case, we additionally need to approximate the spatial dependency of the interaction function, i.e. the competition potential in the Schneider case. To this end, we consider a parametric family that is dense in the

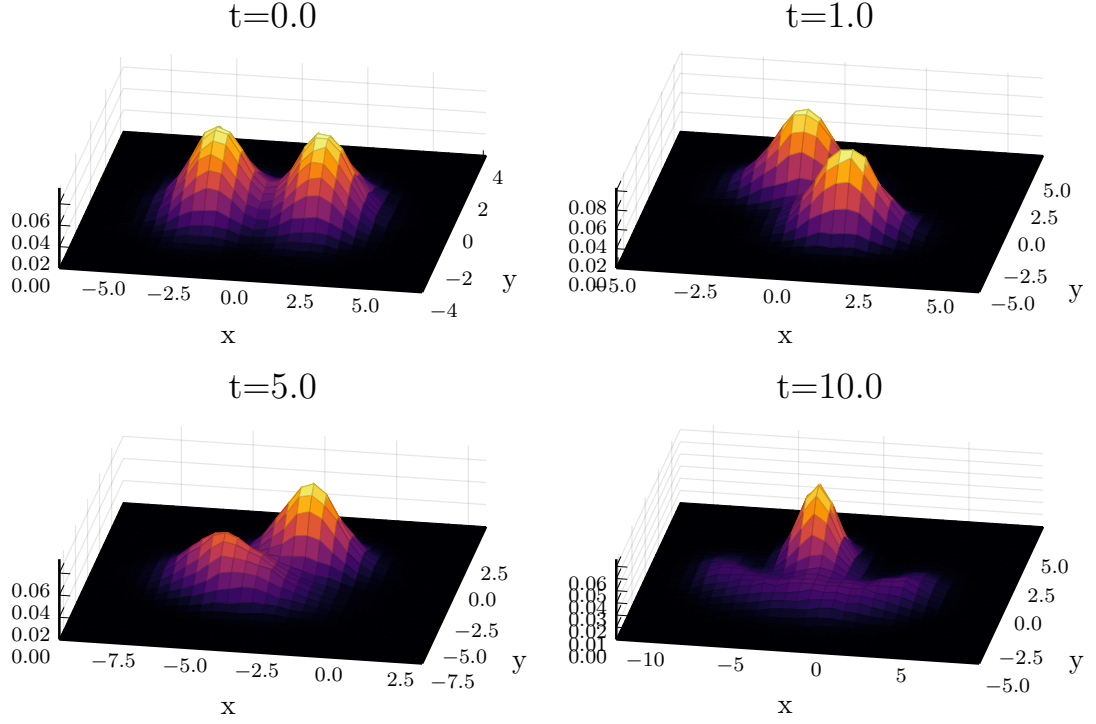


Figure 3.2: Evolution of the marginal density of the Spring Cloud system $f^x(t, x)$ with time. The detailed configuration of this simulation can be found in the appendix, page 166.

space of continuous and bounded functions defined over the phase space, namely polynomial functions of nonlinear and bounded transformations of the variables (s, x, S, γ) .

Expression of the approximated flow with piecewise competition potential: Let us recall the expression of the mean-field flow as a function of the competition potential. For any $\theta = (x, S, \gamma) \in \Theta$:

$$s_\infty(t, s, \theta) = s_m \left(\frac{s}{s_m} \right)^{e^{-\gamma t}} \exp \left(\gamma \log \left(\frac{S}{s_m} \right) \int_0^t (1 - C_\infty(\tau, s, \theta)) e^{\gamma(\tau-t)} d\tau \right)$$

where $C_\infty(t, s, \theta) = \mathbb{E} \{ C(s_\infty(t, s, \theta), s_\infty(t, s', \theta'), |x - x'|), (s', \theta') \sim \mu_0 \}$

(3.165)

Let $\{t_0 = 0, t_1, \dots, t_M = T\}$ be a subdivision of the observation interval $[0; T]$ with regular time-step Δt . We consider a piecewise constant approximation of the

competition potential.

$$\begin{aligned}\hat{s}_\infty(t, s, \theta) &= s_m \left(\frac{s}{s_m}\right)^{e^{-\gamma t}} \left(\frac{S}{s_m}\right)^{1-e^{-\gamma t}-\hat{C}_\infty(t, s, \theta)} \\ \hat{C}_\infty(t, s, \theta) &= \gamma \sum_{k=0}^{M-1} \int_0^t C_k(s, \theta) \mathbb{I}\{t_k \leq t < t_{k+1}\} e^{\gamma(\tau-t)} d\tau\end{aligned}\tag{3.166}$$

The $C_k(s, \theta)$ are approximations of the competition potential exerted on a plant of initial size s and of parameter θ at time t_k .

Approximation of the initial competition potential: Let us build these approximations step by step. We start by the initial competition potential, which has the following expression.

$$C_\infty(0, s, \theta) = \mathbb{E} \left\{ \frac{\log(s'/s_m)}{2R_M \left(1 + \frac{|x-x'|^2}{\sigma_x^2}\right)} \left(1 + \tanh\left(\frac{1}{\sigma_r} \log(s'/s)\right)\right), (s', x') \sim \mu_0^{s, x} \right\}\tag{3.167}$$

In particular, we can notice that initially this potential does not depend on parameters S and γ . This expectation is not analytical in general, and a natural way to obtain a consistent approximation of it is by resorting to Monte-Carlo integration.

$$\tilde{C}_\infty(0, s, x; s'_{1:N}, x'_{1:N}) = \frac{1}{N} \sum_{i=1}^N C(s, s'_i, |x - x'_i|)\tag{3.168}$$

where $s'_{1:N}, x'_{1:N}$ are sampled from $\mu_0^{s, x}$. This approximation is more or less equivalent to simulating the microscopic dynamics that we know to tend towards the mean-field dynamics. By doing so, there is not really any computational advantage in the use of the mean-field flow, as it appears only as an individual trajectory within a large enough population. One way to get rid of the dependency with respect to the sample is to construct a parametric approximation of the map in equation (3.168). We can consider the parametric family consisting of polynomial functions of some bounded transformations of the variables s, x .

$$\begin{aligned}\mathfrak{C}_0 &= \{(s, x, y, \beta) \in [s_0^{\min}; s_0^{\max}] \times \mathbb{R}^2 \times \mathbb{R}^{n(3, d)} \mapsto \beta \cdot f_d^3(s, x, y), d \in \mathbb{N}\} \\ \text{where } \forall d \in \mathbb{N}, f_d^3(s, x, y) &= \frac{q_d^3 \left(\log(s/s_m), \arctan\left(\frac{x-\mu_x}{L_x}\right), \arctan\left(\frac{y-\mu_y}{L_y}\right) \right)}{1 + \frac{(x-\mu_x)^2 + (y-\mu_y)^2}{\sigma_x^2}}\end{aligned}\tag{3.169}$$

In the above equation q_d^3 is the polynomial feature function of three variables and of degree d . For instance, the polynomial feature function of degree 2 with two variables is

$$q_2^2(x_1, x_2) = (1, x_1, x_1^2, x_2, x_1x_2, x_2^2) \in \mathbb{R}^6 \quad (3.170)$$

More generally, we denote by q_d^k the polynomial feature function with k variables and of degree d .

$$\begin{aligned} q_d^k(x_{1:k}) &= (x_1^{\alpha_1} \dots x_k^{\alpha_k})_{\alpha_{1:k} \in A_d^k} \\ A_d^k &= \{\alpha_{1:k} \in \mathbb{N}^k \mid \alpha_1 + \dots + \alpha_k \leq d\} \end{aligned} \quad (3.171)$$

The cardinality of A_d^k is a classical result of combinatorics.

$$\text{Card}(A_d^k) = \sum_{\ell=0}^d \binom{k + \ell - 1}{k - 1} = n(k, d) \quad (3.172)$$

In equation (3.169), we have represented the position variable (x, y) by the bijective transformation $\left(\arctan\left(\frac{x - \mu_x}{L_x}\right), \arctan\left(\frac{y - \mu_y}{L_y}\right) \right)$ where (μ_x, μ_y) can be chosen as the mean position, and L_x, L_y are typical lengths, such as the standard deviation of x and y , or the competition parameter σ_x . With this parametrization of the variable, we can express the competition potential as a continuous function defined over a compact domain, that can be uniformly approximated by a polynomial function thanks to the Stone-Weierstrass theorem. We also divide the polynomial function by a factor $\frac{1}{1 + \frac{(x - \mu_x)^2 + (y - \mu_y)^2}{\sigma_x^2}}$ to ensure that the approximation has roughly the same behaviour as the target function when $|x| \rightarrow \infty$.

β is the vector of coefficients of the polynomial function at the numerator. We can choose β so that it minimizes the quadratic risk between the target function $\tilde{C}_\infty(0, \cdot)$ and the class \mathfrak{C}_0 for a fixed degree d .

$$\beta_0^* = \underset{\beta \in \mathbb{R}^{n(3,d)}}{\text{argmin}} \mathbb{E}\{(\tilde{C}_\infty(0, s, x; s_{1:N}, x_{1:N}) - \beta \cdot f_d^3(s, x))^2, (s, x) \sim \mu_0^{s,x}\},$$

which is equivalent to find β_0^* such that:

$$\mathbb{E}\{f_d^3(s, x)f_d^3(s, x)^\top, (s, x) \sim \mu_0^{s,x}\} \beta_0^* = \mathbb{E}\{\tilde{C}_\infty(0, s, x)f_d^3(s, x), (s, x) \sim \mu_0^{s,x}\} \quad (3.173)$$

The above linear system is not necessarily invertible, according to the distribution μ_0 : for instance, if μ_0 is a Dirac distribution, the system is of rank 1. However, the system always admits at least one solution². In practice, we consider a solution of

²One can decompose the right hand side in the orthogonal sum formed by the image of the matrix and the orthogonal of the image.

the linear system

$$\left(\sum_{k=1}^K f_d^3(s_k, x_k) f_d^3(s_k, x_k)^\top \right) \beta_0^* = \sum_{k=1}^K \tilde{C}_\infty(0, s_k, x_k) f_d^3(s_k, x_k) \quad (3.174)$$

where $s_{1:K}, x_{1:K}$ is a training set consisting in a sample of $\mu_0^{s,x}$ independent of $s'_{1:N}, x'_{1:N}$. The approximation of the initial competition potential is then

$$\begin{aligned} \hat{C}_0(s, x) &= p_{[0;1]}(\beta \cdot f_d^3(s, x)) \\ \text{where } p_{[0;1]}(x) &= \max(\min(1, x), 0) \end{aligned} \quad (3.175)$$

We have incorporated in this reconstruction our knowledge on the boundedness of the potential by simply projecting the values of the linear combination into $[0; 1]$. We can assess the relevance of this approximation by computing the coefficient of determination over a testing set $(s_{1:K}^t, x_{1:K}^t, \tilde{C}_{1:K}^t)$.

$$\begin{aligned} R^2 &= 1 - \frac{\sum_{k=1}^K (\tilde{C}_k^t - \hat{C}_0(s_k^t, x_k^t))^2}{\sum_{k=1}^K (\tilde{C}_k^t - m_{\tilde{C}^t})^2} \\ m_{\tilde{C}^t} &= \frac{1}{K} \sum_{k=1}^K \tilde{C}_k^t \end{aligned} \quad (3.176)$$

The coefficient of determination is used especially to calibrate the degree d of the polynomial approximation, that need to be precise but also light in terms of computation, as the dimension of the coefficient space $n(k, d)$ increases like the factorial function with d .

Approximation of the subsequent competition potentials: If the reconstruction is accurate enough, and if the sample size is large enough, equation (3.175) enables to sample $\hat{s}_\infty(t, s, \theta)$ for any t in the subinterval $[0; \Delta t]$, that is close to the actual mean-field flow $s_\infty(t, s, \theta)$. We then use the same methodology to build the next approximation of the competition potential, but we need to integrate in the arguments of the approximation the parameters S, γ , that have an influence on the competition potential at time Δt .

$$\begin{aligned} \hat{s}_\infty(\Delta t, s, \theta) &= s_m \left(\frac{s}{s_m} \right)^{e^{-\gamma \Delta t}} \left(\frac{S}{s_m} \right)^{(1-e^{-\gamma \Delta t})(1-\hat{C}_0(s, x))} \\ \tilde{C}_\infty(\Delta t, s, \theta; s'_{1:N}, \theta'_{1:N}) &= \frac{1}{N} \sum_{i=1}^N C(\hat{s}_\infty(\Delta t, s, \theta), \hat{s}_\infty(\Delta t, s'_i, \theta'_i), |x - x'_i|) \end{aligned} \quad (3.177)$$

The class of functions used to approximate the above competition potential has the same structure as \mathfrak{C}_0 , but with additional arguments.

$$\begin{aligned} \mathfrak{C}_1 &= \{(s, x, y, S, \gamma, \beta) \in [s_0^{\min}; s_0^{\max}] \times \Theta \times \mathbb{R}^{n(5,d)} \mapsto \beta \cdot f_d^5(s, x, y, S, \gamma), d \in \mathbb{N}\} \\ f_d^5(s, x, y, S, \gamma) &= \frac{q_d^5 \left(\log(s/s_m), \arctan\left(\frac{x-\mu_x}{L_x}\right), \arctan\left(\frac{y-\mu_y}{L_y}\right), \log(S/s_m), e^{-\gamma\Delta t} \right)}{1 + \frac{(x-\mu_x)^2 + (y-\mu_y)^2}{\sigma_x^2}} \end{aligned} \quad (3.178)$$

Other choices of transformations for the variables S and γ are possible. These specific transformations $\log(S/s_m)$ and $e^{-\gamma\Delta t}$ are used here because they better describe the relationship between the competition potential and the parameters S, γ .

The identification of the linear combination coefficient β is done exactly as previously by minimization of the square loss between the empirical potential $\hat{C}_\infty(\Delta t, \dots; s'_{1:N}, \theta'_{1:N})$ and the class \mathfrak{C}_1 over a training set. The degree of the polynomial approximation is calibrated by computing the coefficient of determination R^2 over a testing set. We similarly learn by recurrence the functions $\hat{C}_1(s, \theta), \dots, \hat{C}_{M-1}(s, \theta)$. To simplify the procedure, we choose the same value of degree d for all the f_d^5 at each time step $\Delta t, \dots, T - \Delta t$, that is potentially different from the degree chosen for the function f_d^3 , used to approximate the initial potential. The final expression of the approximated mean-field characteristic flows is obtained by computing the integral over the time defining the reconstructed potential $\hat{C}_\infty(t, s, \theta)$ in equation (3.166).

$$\begin{aligned} \hat{C}_\infty(t, s, \theta) &= \sum_{k=0}^{M-1} C_k(s, \theta) [\mathbb{I}\{t_k \leq t < t_{k+1}\} (1 - e^{\gamma(t_k - t)}) \\ &\quad + \mathbb{I}\{t_{k+1} \leq t\} (e^{\gamma(t_{k+1} - t)} - e^{\gamma(t_k - t)})] \end{aligned} \quad (3.179)$$

The consistency of this numerical scheme is yet to be proved, although we have good reason to believe it holds. When the sample size $N \rightarrow \infty$, the empirical potential converges to the mean-field potential uniformly and almost surely. This empirical potential is well approximated by the functions of the families \mathfrak{C}_0 or \mathfrak{C}_1 if the degree d is chosen large enough. As it was noticed in the case of the Spring-Cloud system, the sample size and the time-step must be chosen jointly, to avoid an accumulation of numerical errors. In the Schneider case, we also have to choose at the same time the degree d of the polynomial approximation that monitors the accuracy of the reconstruction of the empirical potential.

Example of a simulation: We applied this numerical scheme on the following initial distribution

$$\begin{aligned} s &\sim \mathcal{U}([s_0^{\min}; s_0^{\max}]) \\ x &\sim \mathcal{N}(0, L^2 \mathbf{I}_2) \\ S|x &\sim \mathcal{N}_{[s_m; s_m e^{R_M}]}(\bar{S}(x), \delta S^2), \quad \gamma|x \sim \mathcal{N}_{[0; \gamma_M]}(\bar{\gamma}(x), \delta \gamma^2) \end{aligned} \tag{3.180}$$

where the functions $\bar{S}(x)$ and $\bar{\gamma}(x)$ are defined in equation (1.58) and $\mathcal{N}_{[a;b]}(\mu, \sigma^2)$ denotes a truncated Gaussian distribution over the segment $[a; b]$. The spatial variations of functions $\bar{S}(x)$ and $\bar{\gamma}(x)$ are represented on figure 3.3. The configuration of the initial distribution is chosen in order to have four spatial regions distinguishing the parameters: regions with high and low S , and regions with high and low γ . In the absence of competition, we expect the mean size of plants at a given position to converge to $\bar{S}(x)$. We can notice on figure 3.3 that, due to the competition, the surface $x \mapsto \hat{s}_\infty(T, \bar{s}, x, \bar{S}(x), \bar{\gamma}(x))$ is quite different from $x \mapsto \bar{S}(x)$, and that the region where the plants remain small in average is wider than in the case without competition.

3.6 Conclusion

The mean-field limit can be used to construct simplifications of inference problems in the cases where the system variables do not need to be normalized to obtain a tight sequence of empirical measure $(\hat{\mu}_N)_{n \geq 2}$. When the sequence is not tight, it seems necessary to develop the analysis one order further by studying the asymptotic behavior of $\epsilon_i^N(t, x_{1:N}, \theta_{1:N}) = x_i^N(t, x_{1:N}^0, \theta_{1:N}) - x_\infty(\rho_t^{-1}(t, N), \rho_x^{-1}(x_i^0, N), \theta_i)$ where ρ_t, ρ_x are the normalization functions of the variables t, x respectively, which are such that

$$x_i^N(\rho_t(t, N), \rho_x(x_{1:N}^0, N), \theta_{1:N}) \xrightarrow{N \rightarrow \infty} x_\infty(t, x_i^0, \theta_i) \tag{3.181}$$

in distribution. It is then a matter of estimating the speed of convergence of the sequence ϵ_i^N , i.e., finding a sequence $\alpha_N \rightarrow \infty$ such that the limit distribution of the sequence $\alpha_N \epsilon_i^N$ is not degenerated (equal to δ_0). For the Spring Cloud system, this speed of convergence is derived from the Central Limit Theorem, but this argument does not seem to be generalizable directly to other systems. There are several questions to be investigated about this second order development. How to prove the existence of this non-degenerate distribution in a general setting? How to simulate this distribution? What are the characteristics of the systems for which the second order development is still not enough?

Concerning the simulation of the mean-field limit, this chapter presents only a preliminary work that needs to be supported by consistency proofs in a more

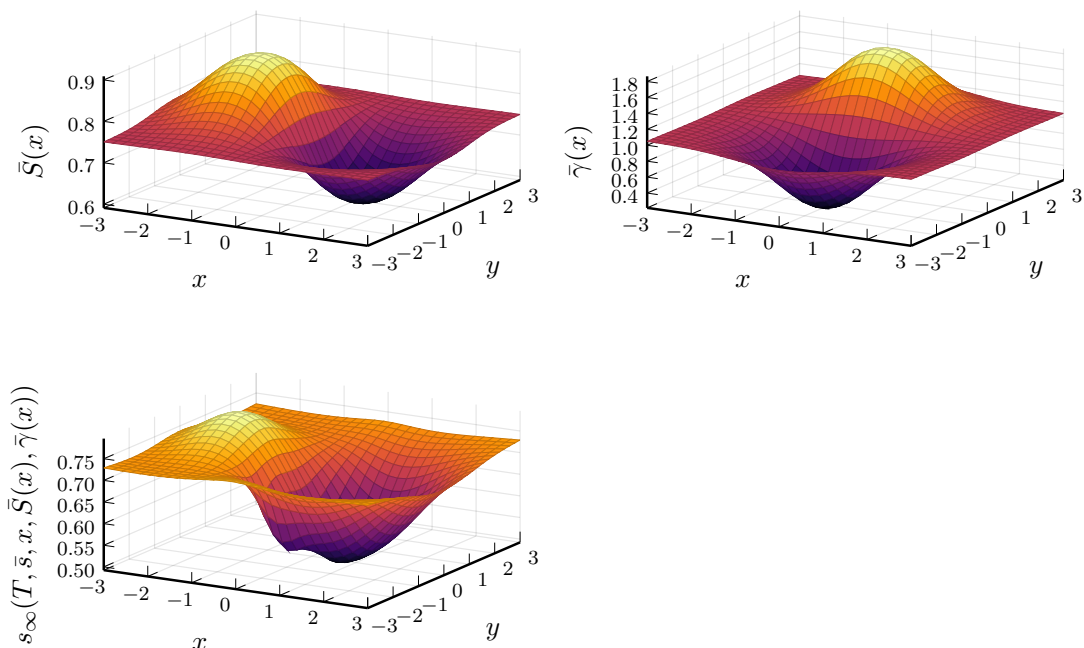


Figure 3.3: Top: Mean values of the parameters S, γ according to the position of the plant $\bar{S}(x)$ and $\bar{\gamma}(x)$. Bottom: Comparison with the values of the mean-field flow at the final time $\hat{s}_\infty(T, \bar{s}, x, \bar{S}(x), \bar{\gamma}(x))$ where $\bar{s} = \frac{s_0^{\min} + s_0^{\max}}{2}$. The detailed configuration of the simulation used to generate this graph can be found on page [167](#)

general setting than in the case of the Schneider and the Spring Cloud systems. Several qualitative arguments lead us to believe that the simulations displayed in this chapter are consistent approximations, but prudence must be kept regarding the mode of convergence of the numerical flow with respect to the different meta-parameters of the scheme, especially to the relationship between the Monte-Carlo sample size K and the time step Δt , that may lead to a significant accumulation of errors if K and Δt are chosen independently. Finally, this numerical scheme needs to be compared with other methods for non-local transport equations, especially those that do not use a mesh over the state space.

3.7 Appendix of chapter 3

Proof of inequality (3.102): Let us express the empirical flow and the mean-field flow.

$$\begin{aligned} \hat{s}_N(t, s_1, \theta_1) &= s_m (s_1/s_m)^{e^{-\gamma_N(\theta_1)t}} \exp(\gamma_N(\theta_1) \log(S_1/s_m)) \\ &\times \int_0^t \left(1 - \frac{N}{N-1} \int_{\mathbb{R}_+^* \times \mathbb{R}^2} C(\hat{s}_N(\tau, s_1, \theta_1), s'_1, |x_1 - x'_1|) \hat{\mu}_N^{s,x}(\tau, ds'_1, dx'_1) \right) e^{\gamma_N(\theta_1)(\tau-t)} d\tau \end{aligned}$$

where $\gamma_N(\theta) = \gamma_N(x, S, \gamma) = \gamma \left(1 - \frac{\log(S/s_m)}{2(N-1)R_M} \right)$

$$\begin{aligned} s_\infty(t, s_2, \theta_2) &= s_m (s_2/s_m)^{e^{-\gamma_2 t}} \exp(\gamma_2 \log(S_2/s_m)) \\ &\times \int_0^t \left(1 - \int_{\mathbb{R}_+^* \times \mathbb{R}^2} C(s_\infty(\tau, s_2, \theta_2), s'_2, |x_2 - x'_2|) \mu^{s,x}(t, ds'_2, dx'_2) \right) e^{\gamma_2(\tau-t)} d\tau \end{aligned} \quad (3.182)$$

Let us consider the function

$$\begin{aligned} \mathcal{S} : (t, s, \gamma, S, C) &\in \mathbb{R}_+ \times [s_0^{\min}; s_0^{\max}] \times [0; \gamma_M] \times [S_m; s_m \exp(R_M)] \times [-1; 1] \\ &\mapsto s_m \left(\frac{s}{s_m} \right)^{e^{-\gamma t}} \left(\frac{S}{s_m} \right)^C \end{aligned} \quad (3.183)$$

We introduce additional notations for the competition terms. Let $s_1, s_2 \in [s_0^{\min}; s_0^{\max}]$ and $\theta_1 = (x_1, S_1, \gamma_1), \theta_2 = (x_2, S_2, \gamma_2) \in \Theta$.

$$\begin{aligned} C_N(t, s_1, \theta_1) &= \gamma_N(\theta_1) \int_0^t \left(1 - \frac{N}{N-1} \int_{\mathbb{R}_+^* \times \mathbb{R}^2} C(\hat{s}_N(\tau, s_1, \theta_1), s'_1, |x_1 - x'_1|) \hat{\mu}_N^{s,x}(\tau, ds'_1, dx'_1) \right) e^{\gamma_N(\theta_1)(\tau-t)} d\tau \\ C_\infty(t, s_2, \theta_2) &= \gamma_2 \int_0^t \left(1 - \int_{\mathbb{R}_+^* \times \mathbb{R}^2} C(s_\infty(\tau, s_2, \theta_2), s'_2, |x_2 - x'_2|) \mu^{s,x}(t, ds'_2, dx'_2) \right) e^{\gamma_2(\tau-t)} d\tau \end{aligned} \quad (3.184)$$

We decompose the difference between the two flows into three terms.

$$\begin{aligned} \hat{s}_N(t, s_1, \theta_1) - s_\infty(t, s_2, \theta_2) &= \mathcal{S}(t, s_1, \gamma_N(\theta_1), S_1, C_N(t, s_1, \theta_1)) - \mathcal{S}(t, s_1, \gamma_1, S_1, C_N(t, s_1, \theta_1)) \\ &+ \mathcal{S}(t, s_1, \gamma_1, S_1, C_N(t, s_1, \theta_1)) - \mathcal{S}(t, s_2, \gamma_2, S_2, C_N(t, s_1, \theta_1)) \\ &+ \mathcal{S}(t, s_2, \gamma_2, S_2, C_N(t, s_1, \theta_1)) - \mathcal{S}(t, s_2, \gamma_2, S_2, C_\infty(t, s_2, \theta_2)) \end{aligned} \quad (3.185)$$

Let us consider the first term.

$$\begin{aligned}
 \mathcal{S}_1 &= |\mathcal{S}(t, s_1, \gamma_N(\theta_1), S_1, C_N(t, s_1, \theta_1)) - \mathcal{S}(t, s_1, \gamma_1, S_1, C_N(t, s_1, \theta_1))| \\
 &\leq \int_0^1 \left| \frac{\partial \mathcal{S}}{\partial \gamma}(t, s_1, \gamma_1 + \alpha(\gamma_N(\theta_1) - \gamma_1), S_1, C_N(t, s_1, \theta_1)) d\alpha \right| \cdot |\gamma_N(\theta_1) - \gamma_1| \\
 \frac{\partial \mathcal{S}}{\partial \gamma}(t, s, \gamma, S, C) &= -te^{-\gamma t} s_m \log(s/s_m) \left(\frac{s}{s_m}\right)^{e^{-\gamma t}} \left(\frac{S}{s_m}\right)^C \\
 \mathcal{S}_1 &\leq \gamma_1 t s_1 \log(s_1/s_m) (S_1/S_m) \frac{\log(S_1/s_m)}{2(N-1)R_M}
 \end{aligned} \tag{3.186}$$

Let us consider the second term.

$$\begin{aligned}
 \mathcal{S}_2 &= |\mathcal{S}(t, s_1, \gamma_1, S_1, C_N(t, s_1, \theta_1)) - \mathcal{S}(t, s_2, \gamma_2, S_2, C_N(t, s_1, \theta_1))| \\
 \frac{\partial \mathcal{S}}{\partial s}(t, s, \gamma, S, C) &= e^{-\gamma t} (S/s_m)^C (s/s_m)^{e^{-\gamma t}-1} \\
 \left| \frac{\partial \mathcal{S}}{\partial s}(t, s, \gamma, S, C) \right| &\leq e^{R_M} \\
 \frac{\partial \mathcal{S}}{\partial S}(t, s, \gamma, S, C) &= C \left(\frac{s}{s_m}\right)^{e^{-\gamma t}} \left(\frac{S}{s_m}\right)^{C-1} \\
 \left| \frac{\partial \mathcal{S}}{\partial S}(t, s, \gamma, S, C) \right| &\leq \frac{s_0^{\max}}{s_m} \\
 \mathcal{S}_2 &\leq e^{R_M} |s_1 - s_2| + t s_{\max}^0 \log(s_0^{\max}/s_m) e^{R_M} |\gamma_1 - \gamma_2| + (s_0^{\max}/s_m) |S_1 - S_2|
 \end{aligned} \tag{3.187}$$

Let us consider the third term.

$$\begin{aligned}
 \frac{\partial \mathcal{S}}{\partial C}(t, s, \gamma, S, C) &= s_m \left(\frac{S}{s_m}\right)^C \log\left(\frac{S}{s_m}\right) \left(\frac{s}{s_m}\right)^{e^{-\gamma t}} \\
 \left| \frac{\partial \mathcal{S}}{\partial C}(t, s, \gamma, S, C) \right| &\leq s_0^{\max} R_M e^{R_M} \\
 |\mathcal{S}(t, s_2, \gamma_2, S_2, C_N(t, s_1, \theta_1)) - \mathcal{S}(t, s_2, \gamma_2, S_2, C_\infty(t, s_2, \theta_2))| &\leq \\
 s_0^{\max} e^{R_M} R_M |C_N(t, s_1, \theta_1) - C_\infty(t, s_2, \theta_2)|
 \end{aligned} \tag{3.188}$$

Let us expand the difference of the competition terms:

$$\begin{aligned}
 c_N(t, s_1, \theta_1) &= \int_{\mathbb{R}_+^* \times \mathbb{R}^2} C(\hat{s}_N(t, s_1, \theta_1), s'_1, |x_1 - x'_1|) \hat{\mu}_N^{s,x}(t, ds'_1, dx'_1) \\
 c_\infty(t, s_2, \theta_2) &= \int_{\mathbb{R}_+^* \times \mathbb{R}^2} C(s_\infty(t, s_2, \theta_2), s'_1, |x_2 - x'_2|) \mu(t, ds'_2, dx'_2)
 \end{aligned} \tag{3.189}$$

$$\begin{aligned}
 & |C_N(t, s_1, \theta_1) - C_\infty(t, s_2, \theta_2)| \\
 & \leq \left| \int_0^t \left(1 - \frac{N}{N-1} c_N(\tau, s_1, \theta_1) \right) (\gamma_N(\theta_1) e^{\gamma_N(\theta_1)(\tau-t)} - \gamma_1 e^{\gamma_1(\tau-t)}) d\tau \right| \\
 & + \left| \int_0^t \left(1 - \frac{N}{N-1} c_N(\tau, s_1, \theta_1) \right) (\gamma_1 e^{\gamma_1(\tau-t)} - \gamma_2 e^{\gamma_2(\tau-t)}) d\tau \right| \\
 & + \frac{\gamma_2}{N-1} \left| \int_0^t c_N(\tau, s_1, \theta_1) e^{\gamma_2(\tau-t)} d\tau \right| + \gamma_2 \int_0^t |c_N(\tau, s_1, \theta_1) - c_\infty(\tau, s_2, \theta_2)| e^{\gamma_2(\tau-t)} d\tau
 \end{aligned} \tag{3.190}$$

Let us consider the first term.

$$\begin{aligned}
 & \left| 1 - \frac{N}{N-1} c_N(\tau, s_1, \theta_1) \right| \leq 1 \\
 & \frac{\partial}{\partial \gamma} \left(\gamma \int_0^t \left(1 - \frac{N}{N-1} c_N(\tau, s_1, \theta_1) \right) e^{\gamma(\tau-t)} d\tau \right) \\
 & = \int_0^t \left(1 - \frac{N}{N-1} c_N(\tau, s_1, \theta_1) \right) (1 + \gamma(\tau-t)) e^{\gamma(\tau-t)} d\tau \\
 & \left| \frac{\partial}{\partial \gamma} \left(\gamma \int_0^t \left(1 - \frac{N}{N-1} c_N(\tau, s_1, \theta_1) \right) e^{\gamma(\tau-t)} d\tau \right) \right| \leq \int_0^t (1 + \gamma(t-\tau)) e^{\gamma(\tau-t)} d\tau \\
 & = \frac{2 - e^{-\gamma t} (2 + \gamma t)}{\gamma} \leq t
 \end{aligned} \tag{3.191}$$

We can deduce from the above inequality the following upper-bound on the competition terms.

$$\begin{aligned}
 & |C_N(t, s_1, \theta_1) - C_\infty(t, s_2, \theta_2)| \leq \gamma_1 t \frac{\log(S_1/s_m)}{2R_M(N-1)} + t|\gamma_1 - \gamma_2| + \frac{1}{N-1} \\
 & + \gamma_2 \int_0^t \left| \int_{\mathbb{R}_+^* \times \Theta} C(\hat{s}_N(\tau, s_1, \theta_1), s'_1, |x_1 - x'_1|) \hat{\mu}_N^{s,x}(\tau, ds'_1, dx'_1) \right. \\
 & \left. - \int_{\mathbb{R}_+^* \times \Theta} C(s_\infty(\tau, s_2, \theta_2), s'_2, |x_2 - x'_2|) \mu(\tau, ds'_2, dx'_2) \right| e^{\gamma_2(\tau-t)} d\tau
 \end{aligned} \tag{3.192}$$

We use the coupling π_0 introduced at the beginning of the proof to express the last term.

$$\begin{aligned}
 & \left| \int_{\mathbb{R}_+^* \times \Theta} C(\hat{s}_N(t, s_1, \theta_1), s'_1, |x_1 - x'_1|) \hat{\mu}_N^{s,x}(t, ds'_1, dx'_1) \right. \\
 & \left. - \int_{\mathbb{R}_+^* \times \Theta} C(s_\infty(t, s_2, \theta_2), s'_2, |x_2 - x'_2|) \mu(t, ds'_2, dx'_2) \right|
 \end{aligned} \tag{3.193}$$

$$\begin{aligned}
 &\leq \iint_{(\mathbb{R}_+^* \times \Theta)^2} |C(\hat{s}_N(t, s_1, \theta_1), \hat{s}_N(t, s'_1, \theta'_1), |x_1 - x'_1|) - C(s_\infty(t, s_2, \theta_2), s_\infty(t, s_2, \theta'_2), |x_2 - x'_2|)| \\
 &\quad \times \pi_0(ds'_1, d\theta'_1, ds'_2, d\theta'_2)
 \end{aligned} \tag{3.194}$$

We compute the derivatives of the competition potential to obtain upper-bounds of the variations. For any $\tilde{s}_1 \in [s_m^N, \hat{S}_N(S_1)]$, any $\tilde{s}_2 \in [s_m; S_2]$, and $\delta \geq 0$, we have

$$\begin{aligned}
 \frac{\partial \bar{C}}{\partial s}(\tilde{s}_1, \tilde{s}_2, \delta) &= -\frac{\log(\tilde{s}_2/s_m)(1 - \tanh^2(\frac{1}{\sigma_r} \log(\tilde{s}_2/\tilde{s}_1)))}{2R_M \tilde{s}_1 \sigma_r (1 + \delta^2/\sigma_x^2)} \\
 \left| \frac{\partial \bar{C}}{\partial s}(\tilde{s}_1, \tilde{s}_2, \delta) \right| &\leq \frac{1}{2s_m^N \sigma_r} \\
 \frac{\partial \bar{C}}{\partial s'}(\tilde{s}_1, \tilde{s}_2, \delta) &= \frac{\sigma_x^2(\sigma_r(1 + \tanh(\frac{1}{\sigma_r} \log(\tilde{s}_2/\tilde{s}_1)) + \log(\tilde{s}_2/\tilde{s}_1)(1 - \tanh^2(\frac{1}{\sigma_r} \log(\tilde{s}_2/\tilde{s}_1))))}{2R_M \tilde{s}_2 \sigma_r (\sigma_x^2 + \delta^2)} \\
 \left| \frac{\partial \bar{C}}{\partial s'}(\tilde{s}_1, \tilde{s}_2, \delta) \right| &\leq \frac{2\sigma_r + R_M}{2R_M s_m \sigma_r} \\
 \frac{\partial \bar{C}}{\partial \delta^2}(\tilde{s}_1, \tilde{s}_2, \delta) &= -\frac{\log(\tilde{s}_2/s_m)(1 + \tanh(\frac{1}{\sigma_r} \log(\tilde{s}_2/\tilde{s}_1)))}{2R_M \sigma_x^2 (1 + \delta^2/\sigma_x^2)^2} \\
 \left| \frac{\partial \bar{C}}{\partial \delta^2}(\tilde{s}_1, \tilde{s}_2, \delta) \right| &\leq \frac{1}{\sigma_x^2}
 \end{aligned} \tag{3.195}$$

It follows that

$$\begin{aligned}
 &\iint_{(\mathbb{R}_+^* \times \Theta)^2} |C(\hat{s}_N(t, s_1, \theta_1), \hat{s}_N(t, s'_1, \theta'_1), |x_1 - x'_1|) \\
 &\quad - C(s_\infty(t, s_2, \theta_2), s_\infty(t, s_2, \theta'_2), |x_2 - x'_2|)| \pi_0(ds'_1, d\theta'_1, ds'_2, d\theta'_2) \\
 &\leq \frac{|\hat{s}_N(t, s_1, \theta_1) - s_\infty(t, s_2, \theta_2)|}{2s_m^N \sigma_r} \\
 &\quad + \frac{2\sigma_r + R_M}{2R_M s_m \sigma_r} \iint_{(\mathbb{R}_+^* \times \Theta)^2} |\hat{s}_N(t, s'_1, \theta'_1) - s_\infty(t, s'_2, \theta'_2)| \pi_0(ds'_1, d\theta'_1, ds'_2, d\theta'_2) \\
 &\quad + \frac{|x_1 - x_2|}{\sigma_x^2} \int_{\mathbb{R}^2} |x_1 + x_2 - 2x'_2| \mu_0^x(dx'_2) \\
 &\quad + \frac{1}{\sigma_x^2} \iint_{\mathbb{R}^4} |x'_1 + x'_2 - 2x_1| \cdot |x'_1 - x'_2| \pi_0^x(dx'_1, dx'_2)
 \end{aligned} \tag{3.196}$$

For the last terms relative to x , we have used the relation

$$|x_1 - x'_1|^2 - |x_2 - x'_2|^2 = (x_1 + x_2 - 2x'_2) \cdot (x_1 - x_2) + (x'_1 + x'_2 - 2x_1) \cdot (x'_1 - x'_2) \tag{3.197}$$

Let us further consider the x terms.

$$\begin{aligned}
 & \int_{\mathbb{R}^2} |x_1 + x_2 - 2x'_2| \mu_0^x(dx'_2) \leq |x_1 + x_2| + 2 \int_{\mathbb{R}^2} |x'_2| \mu_0^x(dx'_2) \\
 & \iint_{\mathbb{R}^4} |x'_1 + x'_2 - 2x_1| \cdot |x'_1 - x'_2| \pi_0^x(dx'_1, dx'_2) \\
 & \leq \sqrt{\iint_{\mathbb{R}^4} |x'_1 + x'_2 - 2x_1|^2 \pi_0^x(dx'_1, dx'_2)} \cdot \sqrt{\iint_{\mathbb{R}^4} |x'_1 - x'_2|^2 \pi_0^x(dx'_1, dx'_2)} \\
 & \iint_{\mathbb{R}^4} |x'_1 + x'_2 - 2x_1|^2 \pi_0^x(dx'_1, dx'_2) \\
 & \leq 2 \int_{\mathbb{R}^2} |x'_1|^2 \hat{\mu}_N^x(0, dx'_1) + 2 \int_{\mathbb{R}^2} |x'_2|^2 \mu_0^x(dx'_2) - 4x_1 \cdot \left(\int_{\mathbb{R}^2} x'_1 \hat{\mu}_N^x(0, dx'_1) + \int_{\mathbb{R}^2} x'_2 \mu_0^x(dx'_2) \right) \\
 & = k^x(x_1, \hat{\mu}_N^x(0), \mu_0^x)^2
 \end{aligned} \tag{3.198}$$

We gather all the previous inequalities to obtain an upper-bound on the gap between the microscopic and the mean-field flows:

$$\begin{aligned}
 |\hat{s}_N(t, s_1, \theta_1) - s_\infty(t, s_2, \theta_2)| & \leq \frac{\gamma_1 t s_1 \log(s_1/s_m)(S_1/s_m) \log(S_1/s_m)}{2R_M(N-1)} + e^{R_M} |s_1 - s_2| \\
 & + t s_{\max}^0 \log(s_0^{\max}/s_m) e^{R_M} |\gamma_1 - \gamma_2| + (s_0^{\max}/S_m) |S_1 - S_2| \\
 & + s_0^{\max} e^{R_M} R_M \left(\gamma_1 t \frac{\log(S_1/s_m)}{2R_M(N-1)} + t |\gamma_1 - \gamma_2| + \frac{1}{N-1} \right) \\
 & + s_0^{\max} e^{R_M} R_M \gamma_M \int_0^t \left(\frac{|\hat{s}_N(\tau, s_1, \theta_1) - s_\infty(\tau, s_2, \theta_2)|}{s_m^N \sigma_r} \right. \\
 & + \frac{2\sigma_r + R_M}{2R_M s_m \sigma_r} \iint_{(\mathbb{R}_+^* \times \Theta)^2} |\hat{s}_N(\tau, s'_1, \theta'_1) - s_\infty(\tau, s'_2, \theta'_2)| \pi_0(ds'_1, d\theta'_1, ds'_2, d\theta'_2) \\
 & + \frac{1}{\sigma_x^2} \left(|x_1 + x_2| + \int_{\mathbb{R}^2} |x| \mu_0^x(dx) \right) |x_1 - x_2| \\
 & \left. + \frac{k^x(x_1, \hat{\mu}_N^x(0), \mu_0^x)}{\sigma_x^2} \iint_{\mathbb{R}^4} |x'_1 - x'_2| \pi_0^x(dx'_1, dx'_2) \right) d\tau
 \end{aligned} \tag{3.199}$$

Finally, we obtain the required inequality by integrating over π_0 .

Proof of the inequality (3.134): Let us expand the expressions of the competition potentials.

$$\begin{aligned}
 \hat{C}_1^N(t, \eta, \theta_{1:N}) - C_\infty(t, \eta, \theta_1) &= \int_{\mathbb{R}_+^* \times \mathbb{R}^2} C(s_1^N(t, \eta, \theta_{1:N}), s'_1, |x_1 - x'_1|) \hat{\mu}_N^{s,x}(t, ds'_1, dx'_1) \\
 &- \int_{\mathbb{R}_+^* \times \mathbb{R}^2} C(s_\infty(t, \eta, \theta_1), s'_2, |x_1 - x'_2|) \mu(t, ds'_2, dx'_2) \\
 &+ \frac{1}{N-1} \int_{\mathbb{R}_+^* \times \mathbb{R}^2} C(s_1^N(t, \eta, \theta_{1:N}), s'_1, |x_1 - x'_1|) \hat{\mu}_N^{s,x}(t, ds'_1, dx'_1) \\
 &- \frac{C(s_1^N(t, \eta, \theta_{1:N}), s_1^N(t, \eta, \theta_{1:N}), 0)}{N-1}
 \end{aligned} \tag{3.200}$$

Let π_t be a coupling between the distributions $\hat{\mu}_N(t, \theta_{1:N})$ and $\mu(t)$. Then we have the following upper-bound on the variations of the competition term.

$$\begin{aligned}
 |\hat{C}_1^N(t, \eta, \theta_{1:N}) - C_\infty(t, \eta, \theta_1)| &\leq \frac{3}{2(N-1)} + \iint_{(\mathbb{R}_+^* \times \mathbb{R}^2)^2} |C(s_1^N(t, \eta, \theta_{1:N}), s'_1, |x_1 - x'_1|) \\
 &- C(s_\infty(t, \eta, \theta_{1:N}), s'_2, |x_1 - x'_2|)| \pi_t^{s,x}(ds'_1, dx'_1, ds'_2, dx'_2)
 \end{aligned} \tag{3.201}$$

From equation (3.195), we have the following inequalities for any $\tilde{s}_1 \in [s_m; S_1]$, $\tilde{s}_2 \in [s_m; S_2]$ and $\delta \geq 0$.

$$\begin{aligned}
 \left| \frac{\partial C}{\partial s}(s_1, s_2, \delta) \right| &\leq \frac{1}{2s_m \sigma_r} \\
 \left| \frac{\partial C}{\partial s'}(s_1, s_2, \delta) \right| &\leq \frac{2\sigma_r + R_M}{2R_M \sigma_r s_m} \\
 \left| \frac{\partial C}{\partial \delta^2}(s_1, s_2, \delta) \right| &\leq \frac{1}{\sigma_x^2}
 \end{aligned} \tag{3.202}$$

We deduce from these inequalities an estimation of the variation of the competition potentials in function of the gap between the empirical and the mean-field flow.

$$\begin{aligned}
 &\int_{(\mathbb{R}_+^* \times \mathbb{R}^2)^2} |C(s_1^N(t, \eta, \theta_{1:N}), s'_1, |x_1 - x'_1|) - C(s_\infty(t, \eta, \theta_{1:N}), s'_2, |x_1 - x'_2|)| \pi_t^{s,x}(ds'_1, dx'_1, ds'_2, dx'_2) \\
 &\leq \frac{|s_1^N(t, \eta, \theta_{1:N}) - s_\infty(t, \eta, \theta_{1:N})|}{2s_m \sigma_r} + \frac{2\sigma_r + R_M}{2R_M s_m \sigma_r} \iint_{(\mathbb{R}_+^*)^2} |s'_1 - s'_2| \pi_t^s(ds'_1, ds'_2) \\
 &+ \frac{1}{\sigma_x^2} \iint_{\mathbb{R}^4} \left| |x_1 - x'_1|^2 - |x_1 - x'_2|^2 \right| \pi_t^x(dx'_1, dx'_2)
 \end{aligned} \tag{3.203}$$

Let us simplify the x terms:

$$\begin{aligned}
 & \left| |x_1 - x'_1|^2 - |x_1 - x'_2|^2 \right| = |(x'_1 + x'_2 - 2x_1) \cdot (x'_1 - x'_2)| \\
 & \iint_{\mathbb{R}^4} \left| |x_1 - x'_1|^2 - |x_1 - x'_2|^2 \right| \pi_t^x(dx'_1, dx'_2) \\
 & \leq \sqrt{\iint_{\mathbb{R}^4} |x'_1 + x'_2 - 2x_1|^2 \pi_t^x(dx'_1, dx'_2)} \times \sqrt{\iint_{\mathbb{R}^4} |x'_1 - x'_2|^2 \pi_t^x(dx'_1, dx'_2)} \quad (3.204) \\
 & \leq k^x(x_1, \hat{\mu}_N^x(0), \mu_0^x) \sqrt{\iint_{\mathbb{R}^4} |x'_1 - x'_2|^2 \pi_t^x(dx'_1, dx'_2)}
 \end{aligned}$$

where $k^x(x_1, \hat{\mu}_N^x(0), \mu_0^x)$ is defined in equation (3.198). We deduce the following inequality on the gap between the empirical flow and the mean-field flow.

$$\begin{aligned}
 & |s_1^N(t, \eta, \theta_{1:N}) - s_\infty(t, \eta, \theta_1)| \leq s_0 \left(\frac{S_1}{s_m} \right) \log \left(\frac{S_1}{s_m} \right) \gamma_1 \\
 & \times \left(\frac{1}{2s_m \sigma_r} \int_0^t |s_1^N(\tau, \eta, \theta_{1:N}) - s_\infty(\tau, \eta, \theta_1)| e^{\gamma_1(\tau-t)} d\tau \right. \\
 & + \frac{2\sigma_r + R_M}{2R_M s_m \sigma_r} \int_0^t \mathcal{W}_1(\hat{\mu}_N^s(\tau, z_{1:N}^0), \mu^s(\tau)) e^{\gamma_1(\tau-t)} d\tau \\
 & \left. + k^x(x_1, \hat{\mu}_N^x(0), \mu_0^x) \mathcal{W}_2(\hat{\mu}_N^x(0), \mu_0^x) (1 - e^{-\gamma_1 t}) \right) \quad (3.205)
 \end{aligned}$$

From inequality (3.105), we have an estimation of $\mathcal{W}_1(\hat{\mu}_N^s(t), \mu^s(t))$:

$$\begin{aligned}
 & \mathcal{W}_1(\hat{\mu}_N^s(t, z_{1:N}^0), \mu^s(t)) \leq E_N(\hat{\mu}_N(0, z_{1:N}^0), \mu_0) (e^{\beta_N t} - 1) + F_N(\hat{\mu}_N(0, z_{1:N}^0), \mu_0) e^{\beta_N t} \\
 & E_N(\hat{\mu}_N(0, z_{1:N}^0), \mu_0) = \frac{1}{\beta_N} \left(\frac{A(\hat{\mu}_N(0, z_{1:N}^0)}{N-1} + B(\hat{\mu}_N^x(0, z_{1:N}^0), \mu_0^x) \mathcal{W}_2(\hat{\mu}_N^x(0, z_{1:N}^0), \mu_0^x) \right. \\
 & \left. + \alpha_\gamma \mathcal{W}_1(\hat{\mu}_N^\gamma(0, z_{1:N}^0), \mu_0^\gamma) \right) \\
 & F_N(\hat{\mu}_N(0, z_{1:N}^0), \mu_0) = \frac{s_0^{\max} e^{R_M} R_M}{N-1} + e^{R_M} \mathcal{W}_1(\hat{\mu}_N^s(0, z_{1:N}^0), \mu_0^s) + \alpha_S \mathcal{W}_1(\hat{\mu}_N^S(0, z_{1:N}^0), \mu_0^S) \quad (3.206)
 \end{aligned}$$

We can notice that in this particular case, since $\mu_0^s = \delta_{s_0}$, then $\mathcal{W}_1(\hat{\mu}_N^s(0, z_{1:N}^0), \mu_0^s) = 0$ for all N .

$$\begin{aligned}
 & |s_1^N(t, \eta, \theta_{1:N}) - s_\infty(t, \eta, \theta_1)| e^{\gamma_1 t} \leq s_0 \left(\frac{S_1}{s_m} \right) \log \left(\frac{S_1}{s_m} \right) \gamma_1 \\
 & \times \left(\frac{1}{2s_m \sigma_r} \int_0^t |s_1^N(\tau, \eta, \theta_{1:N}) - s_\infty(\tau, \eta, \theta_1)| e^{\gamma_1 \tau} d\tau + \frac{2\sigma_r + R_M}{2R_M s_m \sigma_r} \int_0^t (E_N (e^{(\beta_N + \gamma_1)\tau} - e^{\gamma_1 \tau}) \right. \\
 & \left. + F_N e^{(\beta_N + \gamma_1)\tau}) d\tau + k^x(x_1, \hat{\mu}_N^x(0), \mu_0^x) \int_0^t \mathcal{W}_1(\hat{\mu}_N^x(0, \theta_{1:N}), \mu_0^x) e^{\gamma_1 \tau} d\tau \right) \quad (3.207)
 \end{aligned}$$

By Grönwall lemma, we obtain the following estimation.

$$\begin{aligned}
 |s_1^N(t, \eta, \theta_{1:N}) - s_\infty(t, \eta, \theta_1)| &\leq \frac{\tau(\eta, \theta_1)e^{(1/\tau(\eta, \theta_1) - \gamma_1)t}}{(\gamma_1\tau(\eta, \theta_1) - 1)(\tau(\eta, \theta_1)(\beta_N(\eta) + \gamma_1) - 1)} \\
 &\times \left[\frac{2\sigma_r + R_M}{R_M\tau(\eta, \theta_1)} E_N(\hat{\mu}_N(0), \mu_0) ((\gamma_1\tau(\eta, \theta_1) - 1)e^{(\beta_N(\eta) + \gamma_1 - 1/\tau(\eta, \theta_1))t} \right. \\
 &+ (1 - \tau(\eta, \theta_1)(\beta_N(\eta) + \gamma_1))e^{(\gamma_1 - 1/\tau(\eta, \theta_1))t} + \beta_N(\eta)\tau(\eta, \theta_1)) \\
 &+ \frac{2\sigma_r + R_M}{R_M\tau(\eta, \theta_1)} F_N(\hat{\mu}_N(0), \mu_0)(\gamma_1\tau(\eta, \theta_1) - 1) (e^{(\beta_N(\eta) + \gamma_1 - 1/\tau(\eta, \theta_1))t} - 1) \\
 &\left. + \tilde{s}(\eta, \theta_1)\gamma_1 k^x(x_1, \hat{\mu}_N^x(0), \mu_0^x) \mathcal{W}_2(\hat{\mu}_N^x(0), \mu_0^x)(\tau(\eta, \theta_1)(\beta_N(\eta) + \gamma_1) - 1) (e^{(\gamma_1 - 1/\tau(\eta, \theta_1))t} - 1) \right] \quad (3.208)
 \end{aligned}$$

where we have used the following notations

$$\begin{aligned}
 \tilde{s}(\eta, \theta_1) &= s_0 \left(\frac{S_1}{s_m} \right) \log \left(\frac{S_1}{s_m} \right) \\
 \tau(\eta, \theta_1) &= \frac{2s_m\sigma_r}{\tilde{s}(\eta, \theta_1)\gamma_1}
 \end{aligned} \quad (3.209)$$

and $\beta_N(\eta)$ is defined in equation (3.104). By reordering the terms, we obtain the expressions of the coefficients in inequality (3.134). For the $\frac{1}{N-1}$ terms.

$$\begin{aligned}
 \hat{s}_1(\eta, \theta_1, \hat{\mu}_N(0)) &= \frac{2\sigma_r + R_M}{R_M(\beta(\eta)\tau(\eta, \theta_1) + \gamma_1\tau(\eta, \theta_1) - 1)} \left(\frac{A(\hat{\mu}_N(0))}{\beta(\eta)} + s_0^{\max} R_M e^{R_M} \right) \\
 \hat{s}_2(\eta, \theta_1, \hat{\mu}_N(0)) &= \frac{2\sigma_r + R_M}{R_M(\beta(\eta)\tau(\eta, \theta_1) + \gamma_1\tau(\eta, \theta_1) - 1)} \left(\frac{A(\hat{\mu}_N(0))\tau(\eta, \theta_1)}{\gamma_1\tau(\eta, \theta_1) - 1} - s_0^{\max} R_M e^{R_M} \right) \\
 \hat{s}_3(\eta, \theta_1, \hat{\mu}_N(0)) &= -\frac{2\sigma_r + R_M}{R_M(\gamma_1\tau(\eta, \theta_1) - 1)} \cdot \frac{A(\hat{\mu}_N(0))}{\beta(\eta)}
 \end{aligned} \quad (3.210)$$

For the $\mathcal{W}_2(\hat{\mu}_N^x(0), \mu_0^x)$ coefficients, we have

$$\begin{aligned}
 a_1^x(\eta, \theta_1, \hat{\mu}_N(0)) &= \frac{B(\hat{\mu}_N(0), \mu_0^x)(2\sigma_r + R_M)}{R_M\beta(\eta)(\beta(\eta)\tau(\eta, \theta_1) + \gamma_1\tau(\eta, \theta_1) - 1)} \\
 a_2^x(\eta, \theta_1, \hat{\mu}_N(0)) &= \frac{\tau(\eta, \theta_1)\tilde{s}(\eta, \theta_1)\gamma_1}{\gamma_1\tau(\eta, \theta_1) - 1} \left(\frac{B(\hat{\mu}_N^x(0), \mu_0^x)(2\sigma_r + R_M)\tau(\eta, \theta_1)}{2R_M s_m \sigma_r (\beta(\eta)\tau(\eta, \theta_1) + \gamma_1\tau(\eta, \theta_1) - 1)} - k^x(x_1, \hat{\mu}_N^x(0), \mu_0^x) \right) \\
 a_3^x(\eta, \theta_1, \hat{\mu}_N(0)) &= \frac{\tau(\eta, \theta_1)\tilde{s}(\eta, \theta_1)\gamma_1}{\gamma_1\tau(\eta, \theta_1) - 1} \left(k^x(x_1, \hat{\mu}_N^x(0), \mu_0^x) - \frac{B(\hat{\mu}_N^x(0), \mu_0^x)}{2R_M s_m \sigma_r \beta(\eta)(\gamma_1\tau(\eta, \theta_1) - 1)} \right)
 \end{aligned} \quad (3.211)$$

As for the coefficients of $\mathcal{W}_1(\hat{\mu}_N^\gamma(0), \mu_0^\gamma)$,

$$\begin{aligned}
 a_1^\gamma(\eta, \theta_1) &= \frac{(2\sigma_r + R_M)\alpha_\gamma}{R_M\beta(\eta)(\beta(\eta)\tau(\eta, \theta_1) + \gamma_1\tau(\eta, \theta_1) - 1)} \\
 a_2^\gamma(\eta, \theta_1) &= \frac{\tau(\eta, \theta_1)(2\sigma_r + R_M)\alpha_\gamma}{R_M\beta(\eta)\tau(\eta, \theta_1) + \gamma_1\tau(\eta, \theta_1) - 1}(\gamma_1\tau(\eta, \theta_1) - 1) \\
 a_3^\gamma(\eta, \theta_1) &= -\frac{(2\sigma_r + R_M)\alpha_\gamma}{R_M\beta(\eta)(\gamma_1\tau(\eta, \theta_1) - 1)}
 \end{aligned} \tag{3.212}$$

Finally, for the coefficients related to $\mathcal{W}_1(\hat{\mu}_N^S(0), \mu_0^S)$,

$$\begin{aligned}
 a_1^S(\eta, \theta_1) &= \frac{\alpha_S}{R_M} \cdot \frac{2\sigma_r + R_M}{\beta(\eta)\tau(\eta, \theta_1) + \gamma_1\tau(\eta, \theta_1) - 1} \\
 a_2^S(\eta, \theta_1) &= -a_1^S(\eta, \theta_1)
 \end{aligned} \tag{3.213}$$

mean states of the first cloud	$x_1 = (-2, 0), v_1 = (0, 2)$
mean parameters of the first cloud	$\kappa_1 = 0.8, m_1 = 0.8$
mean states of the second cloud	$x_2 = (2, 0), v_2 = (0, -2)$
mean parameter of the second cloud	$\kappa_2 = 1.5, m_2 = 1.5$
variance parameter of the two clouds	$\sigma_x = 1, \sigma_v = 1, \delta\kappa = 0.2, \delta m = 0.2$
sample size for Monte-Carlo integration	$M = 1000$
time-step of the subdivision	$\Delta t = 0.1$
sample size for kernel density estimation	$M' = 10\ 000$

Configuration of the simulation of the Spring Cloud mean-field flow

Configuration of the simulation of the Schneider system

Bounds of the initial size	$s_0^{\min} = 0.1, s_0^{\max} = 0.3$
Extremal sizes	$s_m = 5 \times 10^{-2}, R_M = 3$
Position variance	$L = 1$
Extremal values of $\bar{S}(x)$	$S_m = 0.5, S_M = 1.0, S_0 = 0.75$
Positions of $\bar{S}(x)$ extrema	$x_{\max}^S = (-L, 0), x_{\min}^S = (L, 0)$
Spatial curvatures of $\bar{S}(x)$	$H_{\max}^S = H_{\min}^S = I_2/L^2$
Standard deviation of S	$\delta S = 0.1$
Extremal values of $\bar{\gamma}(x)$	$\gamma_M = 2, \gamma_m = 0.1, \gamma_0 = 1.05$
Positions of $\bar{\gamma}(x)$ extrema	$x_{\min}^\gamma = (0, -L), x_{\max}^\gamma = (0, L)$
Spatial curvature of $\bar{\gamma}(x)$	$H_{\max}^\gamma = H_{\min}^\gamma = I_2/L^2$
Standard deviation of γ	$\delta\gamma = 0.1$
Competition parameters	$\sigma_r = 1.32, \sigma_x = L/2$
Time step and time horizon	$\Delta t = 1, T = 10$
sample size	$N = 1\ 000$
size of training and testing sets	$K = 1\ 000$
random number generator	MersenneTwister(1234)
initial degree of polynomial	$\deg(f_d^3) = 5$
other degree	$\deg(f_d^5) = 3$

Time	R^2 associated with the potential reconstruction
0	0.993
1	0.976
2	0.978
3	0.980
4	0.980
5	0.980
6	0.980
7	0.981
8	0.981
9	0.981

Bibliography

- [1] Bolley, F., Canizo, J. A., and Carrillo, J. A. “Stochastic mean-field limit: non-Lipschitz forces and swarming”. In: *Mathematical Models and Methods in Applied Sciences* vol. 21, no. 11 (2011), pp. 2179–2210.
- [2] Bongini, M., Fornasier, M., Hansen, M., and Maggioni, M. “Inferring interaction rules from observations of evolutive systems I: The variational approach”. In: *Mathematical Models and Methods in Applied Sciences* vol. 27, no. 5 (2017), pp. 909–951. arXiv: [1602.00342](#).
- [3] Boudin, L. and Salvarani, F. “Modelling opinion formation by means of kinetic equations”. In: *Mathematical modeling of collective behavior in socio-economic and life sciences*. Springer, 2010, pp. 245–270.
- [4] Carrillo, J. a., Fornasier, M., Toscani, G., and Vecil, F. “Particle, kinetic, and hydrodynamic models of swarming”. In: *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences SE - 12* (2010), pp. 297–336.
- [5] Carrillo, J. A., James, F., Lagoutiere, F., and Vauchelet, N. “The Filippov characteristic flow for the aggregation equation with mildly singular potentials”. In: *Journal of Differential Equations* vol. 260 (2014), pp. 304–338.
- [6] Carrillo, J. A., Goudon, T., Lafitte, P., and Vecil, F. “Numerical schemes of diffusion asymptotics and moment closures for kinetic equations”. In: *Journal of Scientific Computing* vol. 36, no. 1 (2008), pp. 113–149.
- [7] Cortez, R. and Fontbona, J. “Quantitative uniform propagation of chaos for Maxwell molecules”. In: *Communications in Mathematical Physics* vol. 357, no. 3 (2018), pp. 913–941.
- [8] Cortez, R., Fontbona, J., et al. “Quantitative propagation of chaos for generalized Kac particle systems”. In: *The Annals of Applied Probability* vol. 26, no. 2 (2016), pp. 892–916.
- [9] Degond, P. and Pulvirenti, M. “Propagation of chaos for topological interactions”. In: *The Annals of Applied Probability* vol. 29, no. 4 (2019), pp. 2594–2612. arXiv: [arXiv:1803.01922v1](#).

-
- [10] Dobrushin, R. L. “Vlasov equations”. In: *Functional Analysis and Its Applications* vol. 13, no. 2 (1979), pp. 115–123.
- [11] Golse, F. *Mean Field Kinetic Equations*. 2013.
- [12] Golse, F. “On the Dynamics of Large Particle Systems in the Mean Field Limit”. In: *arXiv preprint arXiv:1301.5494* (2013), pp. 1–144.
- [13] Gorbach, N. S., Bauer, S., and Buhmann, J. M. “Scalable variational inference for dynamical systems”. In: *Advances in Neural Information Processing Systems*. 2017, pp. 4806–4815.
- [14] Kac, M. “Foundations of kinetic theory”. In: *Proceedings of The third Berkeley symposium on mathematical statistics and probability*. Vol. 3. University of California Press Berkeley and Los Angeles, California. 1956, pp. 171–197.
- [15] Kadanoff, L. P. “More is the same; phase transitions and mean field theories”. In: *Journal of Statistical Physics* vol. 137, no. 5-6 (2009), p. 777.
- [16] Kindermann, R. and Snell, L. *Markov random fields and their applications*. Vol. 1. American Mathematical Society, 1980.
- [17] Lafitte, P., Lejon, A., and Samaey, G. “A high-order asymptotic-preserving scheme for kinetic equations using projective integration”. In: *SIAM Journal on Numerical Analysis* vol. 54, no. 1 (2016), pp. 1–33.
- [18] Lagoutière, F. and Vauchelet, N. “Analysis and simulation of nonlinear and nonlocal transport equations”. In: (2017), pp. 1–22.
- [19] Lei, J. “Convergence and concentration of empirical measures under wasserstein distance in unbounded functional spaces”. In: *Bernoulli* vol. 26, no. 1 (2020), pp. 767–798.
- [20] Lu, F., Maggioni, M., and Tang, S. “Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories”. In: *arXiv preprint arXiv:2007.15174* (2020).
- [21] Lu, F., Zhong, M., Tang, S., and Maggioni, M. “Nonparametric inference of interaction laws in systems of agents from trajectory data”. In: *Proceedings of the National Academy of Sciences* vol. 116, no. 29 (2019), pp. 14424–14433.
- [22] Marnissi, Y., Zheng, Y., Chouzenoux, E., and Pesquet, J.-C. “A Variational Bayesian Approach for Image Restoration—Application to Image Deblurring With Poisson–Gaussian Noise”. In: *IEEE Transactions on Computational Imaging* vol. 3, no. 4 (2017), pp. 722–737.
- [23] Perthame, B., Salort, D., and Wainrib, G. “Distributed synaptic weights in a LIF neural network and learning rules”. In: *Physica D: Nonlinear Phenomena* vol. 353 (2017), pp. 20–30.

-
- [24] Salhi, J., MacLaurin, J., and Toumi, S. “On uniform propagation of chaos”. In: *Stochastics* vol. 90, no. 1 (2018), pp. 49–60.
- [25] Šmídl, V. and Quinn, A. *The variational Bayes method in signal processing*. Springer Science & Business Media, 2006.
- [26] Sznitman, A.-S. “Topics in propagation of chaos”. In: *Ecole d’été de probabilités de Saint-Flour XIX—1989*. Springer, 1991, pp. 165–251.
- [27] Tsybakov, A. B. *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008.
- [28] Varadarajan, V. S. “On the convergence of sample probability distributions”. In: *Sankhyā: The Indian Journal of Statistics (1933-1960)* vol. 19, no. 1/2 (1958), pp. 23–26.
- [29] Villani, C. “Optimal transport: old and new”. In: (2009).
- [30] Vlasov, A. A. “The vibrational properties of an electron gas”. In: *Soviet Physics Uspekhi* vol. 10, no. 6 (1968), p. 721.
- [31] Vrettas, M. D., Opper, M., and Cornford, D. “Variational mean-field algorithm for efficient inference in large systems of stochastic differential equations”. In: *Physical Review E* vol. 91, no. 1 (2015), p. 012148.
- [32] Weiss, P. “L’hypothèse du champ moléculaire et la propriété ferromagnétique”. In: (1907).

Conclusion

This is the end
– Jim Morrisson, THE END (1967)

Our initial aim was to build population models derived from individual-based models, with *minimal loss of information*. The three chapters of this thesis have enabled us to clarify the main specifications of a macroscopic approximation of a population formed by interacting individuals with potentially different characteristics.

1. a macroscopic approximation of a population is a consistent approximation of the probability distribution representing all the individuals composing the population when the size $N \rightarrow \infty$.
2. this approximation should have dynamics only depending on the initial distribution μ_0 and the transition function h_N .
3. The numerical method's computational cost to simulate this approximation should be independent of the size N of the population.

Our focus on symmetric and heterogeneous differential equations systems has led us to consider that such an approximation can be expressed using the system's mean-field limit when this limit is well-defined. By construction, the mean-field limit of a population model does not depend on the population's size, and this property makes it meet the two last specifications. If the limit is not well-defined, though, it still seems possible to find the proper normalization of the state and time variables to obtain a mean-field limit when $N \rightarrow \infty$. This statement, on the existence of a proper normalization, needs to be checked on a broader class of symmetric systems than the ones considered in this thesis, especially in the case of systems with stochastic transitions (Bolley et al., 2011³), where the expression of the diffusion coefficient may also influence the normalization. Depending on the behavior of the sequence of population empirical measures $(\hat{\mu}_N(t))_{N \geq 2}$, especially its tightness property, the mean-field limit may or may not provide a consistent approximation of the population, this nuanced result being discussed in sections

3.4.2 and 3.4.3 in the specific cases of the Schneider and the Spring Cloud systems. For the homogeneous Spring Cloud system, the mean-field limit only gives the mean values of the state variables evolving in a large population. To obtain a more comprehensive description of the population dynamics, we need to develop the population's asymptotic behavior at a higher order than the mean-field limit. For now, we do not have any general methodology to derive a consistent approximation of any symmetric population model, but the previous example gives us potential directions to explore.

We hope that the main results of this thesis can be extended to less smooth systems, in particular functional-structural plant models, having the requirement of being symmetric, which is a widespread property in reference plant interaction models, such as Cournède et al. (2008),⁴ Sievänen et al. (2008),¹² Hemmerling et al. (2008).⁸ Our hopes are fueled by the research carried out on propagation of chaos (or asymptotic factorization) in non-smooth systems such as hard spheres, in Gallagher et al. (2014),⁷ or topological interactions, in Degond and Pulvirenti (2019).⁵

It is also necessary to check whether mean-field approximation, when it applies, constitutes a simplification of the original global dynamics in practice. To this end, we need to prove the consistency of the numerical schemes described in section 3.5 and quantify the spatial and temporal errors they induce to use this information in the context of statistical inference. Modeling the uncertainty introduced by the numerical resolution error has to be done as in section 2.7, but this time, with a Gaussian process potentially depending on the variables (t, x, θ) . For now, we do not know whether the simulation of such an error model could be done efficiently.

The application of kinetic theory to agricultural model opens up the possibility to use flocking control methodologies, such as the ones developed by Piccoli et al. (2015),¹¹ to address problems such as yield optimization with limited resources in crops. To study crop optimal management, the crop is generally modeled at a macroscopic level, and the interactions between the plants are neglected in comparison to the influence of the external environment (Della Noce et al., 2019).⁶ Controlling individual-based models allows the user to have much more flexibility, and the monitoring of the system is simplified, as the observation of a single plant can constitute a feedback on the adopted strategy. If optimal control and statistical inference are carried out at the same time, on-line, to deal with the external sources of uncertainty, such as weather, we are dealing with partially observed Markov decision processes (Krishnamurthy, 2016),⁹ which can be naturally expressed within the Bayesian setting developed in chapter 2.

This thesis has always considered a population of constant size. We did not take into account the eventuality for an individual to exit the system (death) or enter the system (birth), which is of particular interest when studying populations from

an ecological perspective. Modeling the mortality is proposed within the Schneider system by Nakagawa et al. (2015),¹⁰ with a probability of death expressed as a non-decreasing function of the competition potential exerted on the individual. One could also have modeled the event of reproduction as a function of the individual state (low competition, closeness to the asymptotic size) and the spatial process associated with the dispersal of the offspring. Formally, when considering these birth-death processes, the transport equations defining the population distribution and the mean-field distribution are no longer conservative. A research direction that may be explored consists in studying the population dynamics at the pace of generations. In this context, we can follow the approach developed in Berestycki et al. (2009)² and in Barton et al. (2010)¹ to derive an evolution process from an individual-based model.

Bibliography

- [1] Barton, N., Etheridge, A., Véber, A., et al. “A new model for evolution in a spatial continuum”. In: *Electronic journal of probability* vol. 15 (2010), pp. 162–216.
- [2] Berestycki, N, Etheridge, A., and Hutzenthaler, M. “Survival, extinction and ergodicity in a spatially continuous population model”. In: *Markov Process. Related Fields* vol. 15, no. 3 (2009), pp. 265–288.
- [3] Bolley, F., Canizo, J. A., and Carrillo, J. A. “Stochastic mean-field limit: non-Lipschitz forces and swarming”. In: *Mathematical Models and Methods in Applied Sciences* vol. 21, no. 11 (2011), pp. 2179–2210.
- [4] Cournède, P.-H., Mathieu, A., Houllier, F., Barthélémy, D., and De Reffye, P. “Computing competition for light in the GREENLAB model of plant growth: a contribution to the study of the effects of density on resource acquisition and architectural development”. In: *Annals of Botany* vol. 101, no. 8 (2008), pp. 1207–1219.
- [5] Degond, P. and Pulvirenti, M. “Propagation of chaos for topological interactions”. In: *The Annals of Applied Probability* vol. 29, no. 4 (2019), pp. 2594–2612. arXiv: [arXiv:1803.01922v1](https://arxiv.org/abs/1803.01922v1).
- [6] Della Noce, A, Carrier, M, and Cournède, P.-H. “Optimal control of non-smooth greenhouse models”. In: *International Symposium on Advanced Technologies and Management for Innovative Greenhouses: GreenSys2019 1296*. 2019, pp. 125–132.
- [7] Gallagher, I., Saint-Raymond, L., and Texier, B. “From Newton to Boltzmann: Hard spheres and short range potentials”. In: *Zurich Lectures in Advanced Mathematics* (2014).
- [8] Hemmerling, R., Kniemeyer, O., Lanwert, D., Kurth, W., and Buck-Sorlin, G. “The rule-based language XL and the modelling environment GroIMP illustrated with simulated tree competition”. In: *Functional plant biology* vol. 35, no. 10 (2008), pp. 739–750.

- [9] Krishnamurthy, V. *Partially observed Markov decision processes*. Cambridge university press, 2016.
- [10] Nakagawa, Y., Yokozawa, M., and Hara, T. “Competition among plants can lead to an increase in aggregation of smaller plants around larger ones”. In: *Ecological Modelling* vol. 301 (2015), pp. 41–53.
- [11] Piccoli, B., Rossi, F., and Trélat, E. “Control to flocking of the kinetic Cucker–Smale model”. In: *SIAM Journal on Mathematical Analysis* vol. 47, no. 6 (2015), pp. 4685–4719.
- [12] Sievänen, R., Perttunen, J., Nikinmaa, E., and Kaitaniemi, P. “Toward extension of a single tree functional–structural model of Scots pine to stand level: effect of the canopy of randomly distributed, identical trees on development of tree structure”. In: *Functional Plant Biology* vol. 35, no. 10 (2008), pp. 964–975.

Title: Symmetric and heterogeneous population models: approximations for simulation and statistical inference

Keywords: Particle systems; Mean-field limit; Bayesian inference; Plant growth model; Collective motion; Gaussian processes

Abstract: Collective motions describe populations in which individuals' interactions are the driving force behind their displacements and their transformation over time. Understanding and controlling collective motions are significant issues in many fields, especially for the study of ecosystems (swarm dynamics), safety in large gatherings and buildings (crowd movement), or agriculture (the study of plant growth). The population models we consider are systems of differential equations with the property of being heterogeneous, i.e., made up of individuals with different characteristics influencing the dynamics. This assumption is motivated by the agricultural application, to study interactions between plants of different varieties or even different species. These systems are also assumed to be symmetric, i.e., having dynamics invariant by permutation of individuals' labels, which is a widespread property within collective motion models, enabling numerous simplifications. However, several challenges remain to be addressed before these models can be used in real-life applications. We focus on the problems related to statistical inference, i.e., matching the model with experimental data and observations made on the system under study.

The first level of difficulty is computational: the simulation of a sizeable interacting population can be too costly in terms of computing time, and, therefore, it is a first impediment to the study of the population at a macroscopic scale. The second level of difficulty relates to the quality of the data: because of the complexity of the modeled system, experimental observations cannot characterize the system's dynamics exactly, in particular because they generally only concern a subset of the population. It is necessary to quantify the uncertainties related to the imperfections in the acquisition of such data. In this thesis, we characterize all the uncertainty sources related to partial observations of symmetric and heterogeneous systems in a Bayesian framework. Some sources of uncertainty, notably the ones arising from inaccurate knowledge of the population size, result in particularly complex inference problems, which we propose to approach using a macroscopic representations of the population. This statistical approximation of the global population motion is based on numerical simulations of the mean-field limit distribution, i.e., a probability distribution expressed as the solution of a non-local transport equation associated with the symmetric system.

Titre : Modèles de population symétriques et hétérogènes: approximations pour la simulation et l'inférence statistique

Mots clés : Système de particules; Limite de champ moyen; Inférence bayésienne; Modèle de croissance de plantes; Mouvement collectif; Processus gaussiens

Résumé : Les mouvements collectifs décrivent des populations dans lesquelles les interactions entre individus sont le moteur de leurs déplacements dans l'espace et de leurs transformations dans le temps. La compréhension et le contrôle des mouvements collectifs constituent des enjeux majeurs dans de nombreux domaines, notamment pour l'étude des écosystèmes (dynamique des essaims d'animaux), la sécurité dans les grands rassemblements et les bâtiments (mouvement de foule), ou encore l'agriculture (étude de la croissance des plantes). Les modèles de population que nous considérons sont des systèmes d'équations différentielles ayant la propriété d'être hétérogènes, i.e., constituées d'individus avec des caractéristiques différentes, et ces caractéristiques ont une influence sur la dynamique. Cette hypothèse est motivée par l'application agricole, où il est question d'étudier les interactions entre plantes de différentes variétés, voire de différentes espèces. Ces systèmes sont également supposés symétriques, i.e., ayant une dynamique invariante par permutation des individus, ce qui est une caractéristique largement répandue au sein des modèles de mouvement collectif, et qui permet de nombreuses simplifications. Un certain nombre de défis restent toutefois à relever pour que ces modèles soient utilisés dans des cas d'application concrets, et nous nous concentrons en particulier sur les problèmes liés à l'inférence statistique, i.e., la confrontation du modèle à des données expérimentales,

des observations réalisées sur le système réel étudié. Un premier niveau de difficulté est d'ordre computationnel: la simulation de grandes populations en interaction peut s'avérer trop coûteuse en temps de calcul, et elle constitue ainsi un premier obstacle à l'étude de la population à une échelle macroscopique. Un second niveau de difficulté a trait à la qualité des données: du fait de la complexité du système modélisé, les observations expérimentales ne peuvent permettre de caractériser exactement la dynamique du système (en particulier car elles ne portent généralement que sur une sous-partie de la population), et il est nécessaire de quantifier les incertitudes liées aux imperfections dans l'acquisition de ces données. Dans cette thèse, nous caractérisons l'ensemble des sources d'incertitude liées aux observations partielles des systèmes symétriques et hétérogènes dans un cadre bayésien. Certaines sources d'incertitude, notamment celle venant d'une connaissance inexacte de la taille de la population et des propriétés des individus non observés, donnent lieu à des problèmes d'inférence particulièrement difficiles, que nous nous proposons d'approcher en utilisant des représentations macroscopiques de la population. Ces approximations statistiques du mouvement global de la population sont basées sur des simulations numériques des distributions limites de champ moyen associées au mouvement collectif, distribution s'exprimant comme la solution d'une équation de transport non-locale.