Signal modeling and Bayesian estimation for cellular mechanism analysis : study of diffusion at the cell membrane and cell division

Anca-Georgina Caranfil

DOCTORAT BRETAGNE LOIRE / MATHSTIC

UNIVERSITÉ DE RENNES 1

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES 1

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Signal, Image, Vision*

Par

Anca-Georgiana CARANFIL

Signal modeling and Bayesian estimation for cellular mechanism analysis : study of diffusion at the cell membrane and cell division

Thèse présentée et soutenue à Rennes, le 14 novembre 2022
Unité de recherche : INRIA Rennes - IGDR

Rapporteurs avant soutenance :

Philippe Andrey	Directeur de Recherche, INRAE-AgroParisTech, Versailles
Jean-Christophe Thalabard	Professeur, Université Paris Cité, Paris

Composition du Jury :

Président :	Valérie Monbet	Professeure, Université de Rennes 1, Rennes
Examinateurs :	Philippe Andrey	Directeur de Recherche, INRAE-AgroParisTech, Versailles
	Jean-Christophe Thalabard	Professeur, Université Paris Cité, Paris
	Thomas Walter	Professeur, Mines Paris, Université PSL, Paris
Dir. de thèse :	Charles Kervrann	Directeur de Recherche, Inria, Rennes
Co-dir. de thèse :	Yann Le Cunff	Maitre de Conférences, Université de Rennes 1, Rennes

Remerciements

Je tiens à remercier mes deux directeurs de thèse, Charles Kervrann et Yann Le Cunff, pour leur soutien tout au long de cette thèse, de m'avoir guidé, de m'avoir aidé à trouver des solutions lorsque j'étais bloquée, et de m'avoir remonté le moral à plusieurs moments durant ces années. Merci à Jacques Pécréaux, que je considère comme mon troisième co-encadrant, pour la collaboration fructueuse.

Merci à mon jury de thèse, Philippe Andrey, Valérie Monbet, Jean-Christophe Thalabard, Thomas Walter, pour leurs questions et les échanges instructifs durant la soutenance. Merci à mes rapporteurs, Philippe Andrey et Jean-Christophe Thalabard, pour leurs remarques très pertinentes qui m'ont permis d'améliorer ce manuscrit.

Merci également aux membres de mon CSI, Etienne Memin et Mathieu Pinot, pour les réunions d'avancement et pour leur soutien.

Je remercie grandement les équipes Serpico et Cedre pour leur accueil chaleureux. Merci à Thierry, Vincent, Mayela, Emmanuel, Sandeep, Antoine, Anaïs, et à celles et ceux qui sont arrivé.e.s sur la fin de ma thèse, ainsi qu'à l'équipe Fluminance et, en particulier, Valentin et Gilles. Merci pour les coups de main techniques, pour les pauses café détente, et pour les discussions enrichissantes. Merci à Hélène, Laurent, Sylvain, Christophe, Nina, Ostiane, Flora pour leur bienveillance et pour m'avoir aidé à m'intégrer parmi les biologistes. Merci à Laurent de m'avoir montré les expériences sur C.elegans. Merci également aux autres équipes du labo pour les échanges lors des séminaires, en particulier à Marc, Gilles, Florian, Catherine, Sébastien, Grégoire et Anne. Merci d'avoir fait l'effort de comprendre mes formules de matheuse lors de mon premier séminaire.

Enfin, je souhaite remercier mes proches. Merci à mes parents de m'avoir soutenu dans tous mes projets, et de m'avoir toujours encouragé à poursuivre mes rêves. Merci à Mathias de m'avoir supporté pendant toutes ces années, et en particulier de m'avoir aidé à dépasser les moments difficiles, et d'avoir assuré pendant la rédaction de mon manuscrit.

Merci à Pierre, Anne-Lyse, Greg, Joan, et les autres, d'avoir été là.

Merci à Martha de m'avoir aidé à débloquer le processus de rédaction du manuscrit. Merci à Emmanuelle, sans qui la rédaction de ce manuscrit aurait été beaucoup plus difficile. Merci pour les messages motivants et pour les discussions hebdomadaires.

Merci à mes chats : Kali, Diego et Pratek.

Table of Contents

Résumé de thèse

Préambule

Cette thèse est le fruit d'une collaboration entre deux équipes : l'équipe Serpico de l'Inria Rennes et l'équipe CeDRE de l'IGDR (Institut de Génétique et Developpement de Rennes). L'aspect méthodologique ainsi que les approches basées sur la modélisation, l'analyse et le traitement des images sont les points communs qui ont amené les deux équipes à collaborer. Ainsi, un nouveau projet est né, qui avait pour objectif l'élaboration d'une approche basée sur l'assimilation de données pour l'analyse de mécanismes cellulaires.

Cette thèse s'inscrit dans le cadre de cette collaboration, avec le but de proposer une approche de modélisation et d'estimation de paramètres, pour l'étude de la variabilité de mécanismes cellulaires. Ma thèse s'est déroulée à 60% dans l'équipe Serpico, et à 40% dans l'équipe CeDRE, sous la direction de Charles Kervrann, Directeur de Recherche à l'Inria Rennes et responsable de l'équipe Serpico, et Yann Le Cunff, Maître de Conférences à l'Université de Rennes 1, membre de l'équipe DyLiSS à l'IRISA (depuis 2021), et membre de l'équipe CeDRE de l'IGDR pendant l'encadrement de cette thèse, et en étroite collaboration avec Jacques Pécréaux, chercheur CNRS et responsable de l'équipe CeDRE au sein de l'IGDR.

Contexte

Les avancées des méthodes d'imagerie, notamment en microscopie de fluorescence, ont permis l'utilisation de la modélisation mathématique pour l'étude des mécanismes cellulaires. De nouveaux modèles mathématiques ont été introduits pour étudier des effets globaux au sein de la cellule, de façon à révéler les tendances générales des mécanismes étudiés. Notre compréhension des comportements "moyens" est ainsi assez bonne ; en revanche, les comportements individuels et leur influence à l'échelle de la cellule restent à élucider. Bien que des modèles locaux aient également été de plus en plus conçus pour étudier des évènements spécifiques dans la cellule, peu de modèles ou de méthodes existent

qui puissent à la fois décrire un comportement général et fournir des informations sur ce qui se produit localement, à plus petite échelle.

Une des conséquences de ces approches est que les valeurs des paramètres liés à la biophysique ou des paramètres moléculaires sont très partiellement établies. Des connaissances approfondies sont disponibles dans la littérature par rapport à des grandeurs biophysiques moyennes, et des expériences *in vivo* ont permis d'estimer certains paramètres moléculaires ou biophysiques qui sont directement observables en imagerie ; en revanche, d'autres paramètres sont encore très approximatifs. Une approche hybride comme celle que nous proposons a pour objectif à la fois d'estimer des paramètres locaux et d'étudier les interactions entre les différents acteurs impliqués dans le mécanisme d'intérêt.

Contrairement aux modèles préexistants dédiés à l'étude des mécanismes cellulaires, notre approche a l'avantage d'être une méthode prédictive, permettant d'orienter les expériences, et de faciliter une compréhension *a priori* du mécanisme étudié.

Contributions

Cette thèse apporte trois contributions principales : deux nouveaux modèles pour décrire des mécanismes cellulaires, et un cadre d'estimation de paramètres qui permet de prendre en compte la variabilité du mécanisme étudié, ainsi que les interactions entre les différents acteurs qui participent à ce mécanisme. Le cadre d'estimation proposé permet ensuite de tester différents scénarios possibles et d'établir des prédictions pour chacun d'entre eux.

La première contribution porte sur la création d'un modèle de la diffusion locale des protéines transmembranaires en imagerie TIRF. La diffusion est habituellement estimée de façon globale sur une large zone de la cellule : une méthode de corrélation est utilisée pour estimer le coefficient de diffusion apparente, en supposant une diffusion homogène dans la membrane. Cette approche ne permet pas de distinguer les différences de diffusion locales, qui peuvent être induites par la stochasticité des éléments composant la membrane ou participant à la diffusion elle-même, ou par les différents éléments structurels qui peuvent modifier la diffusion des protéines transmembranaires. Nous proposons ici un modèle basé sur une méthode de corrélation, mais sans hypothèse d'homogénéité : notre modèle permet de calculer la diffusion pour chaque évènement de diffusion au sein de fenêtres d'image de petite taille. Cette approche permet alors de construire une carte de diffusion d'une région, afin de détecter des zones de forte variation spatiale de la diffusion.

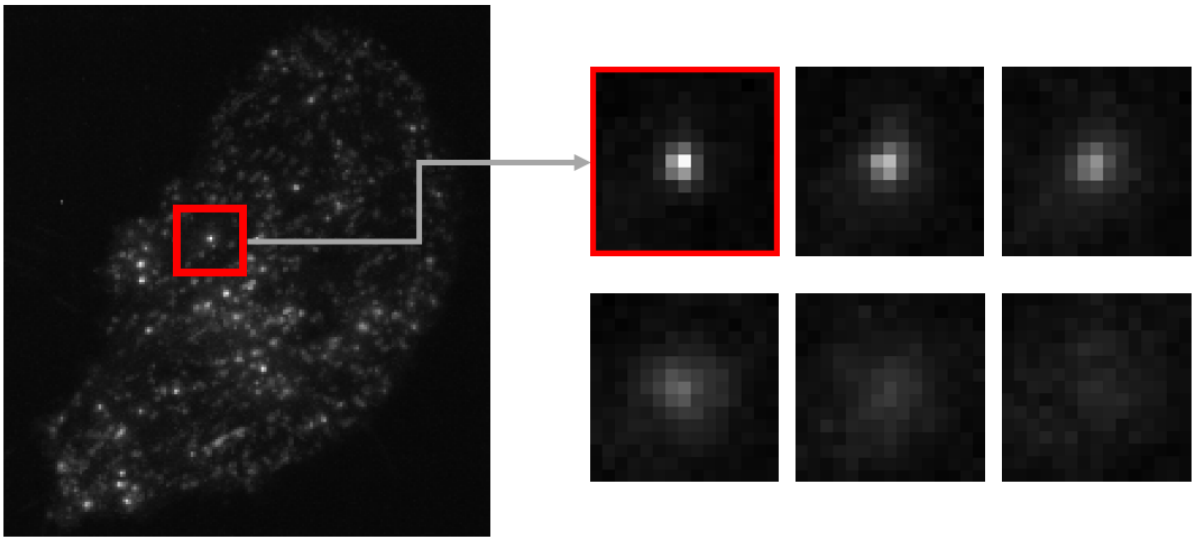


FIGURE 1 – **Illustration d'un spot de diffusion locale en imagerie TIRF.**

La deuxième contribution de cette thèse concerne la modélisation du comportement des centrosomes pendant la division cellulaire, et plus précisément pendant la séparation des chromosomes. Les centrosomes se déplacent vers le côté postérieur de la cellule, et lors de ce déplacement, ils oscillent transversalement à l'axe antérieur-postérieur. Les études préexistantes s'intéressent majoritairement au déplacement postérieur, sans prise en compte des oscillations. Cependant, ces oscillations, fréquemment rencontrées au cours des divisions asymétriques, semblent jouer un rôle-clé dans la division. Par exemple, dans le cas de l'embryon C.elegans, leur absence est le signe d'un dysfonctionnement de la cellule, qui peut conduire à une division erronée, voire à une absence de division et à la mort de la cellule.

Nous nous intéressons ici à ces oscillations transversales et nous proposons un nouveau modèle multi-échelles qui s'appuie sur une approche biophysique du comportement du centrosome. Un bilan des forces, ainsi que des hypothèses biologiques et physiques validées dans la littérature, nous permettent d'établir un modèle des oscillations du centrosome pendant la phase de séparation des chromosomes. Il permet de mieux comprendre quels sont les facteurs déclencheurs des oscillations, quelles sont les conditions nécessaires à leur existence, et quelles interactions se mettent en place pour leur émergence. Il rend notamment possible le test de plusieurs scénarios d'absence d'oscillation chez C.elegans, en combinant les informations données par le modèle et les observations expérimentales. Il permet enfin de représenter les oscillations du centrosome postérieur pour chaque em-

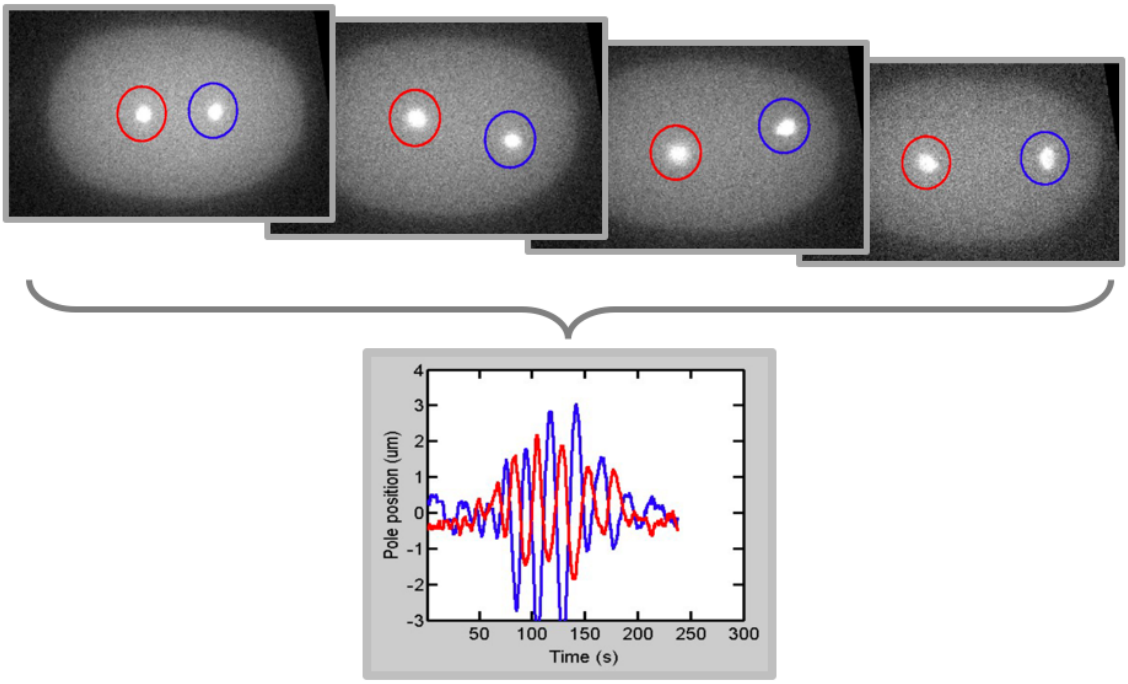


FIGURE 2 – **Illustration des oscillations transversales des centrosomes pendant la division cellulaire.**

bryon, ce qui en fait une base solide pour l'étude de la variabilité d'un embryon à l'autre.

La troisième contribution de cette thèse porte sur la création d'un cadre d'estimation de paramètres en plusieurs étapes, à partir d'un modèle mathématique du mécanisme étudié. La première étape est l'analyse de sensibilité des paramètres, qui permet de regrouper les paramètres en fonction de leurs influences respectives sur le résultat. La deuxième étape est la définition des caractéristiques des paramètres, en fonction des connaissances préalables sur le mécanisme étudié et des observations expérimentales. Enfin, la dernière étape consiste à estimer des paramètres, une tâche qui est effectuée par blocs en fonction des résultats de l'analyse de sensibilité, et est conçue sur le principe bayésien d'estimation des paramètres, l' "Approximate Bayesian Computation" (ABC). Notre approche s'appuie sur très peu d'hypothèses et ne nécessite pas de calculer une vraisemblance, ce qui peut s'avérer complexe. L'estimation des paramètres est faite pour chaque expérience : il est donc possible d'évaluer les paramètres individuels des expériences, mais aussi d'analyser la variabilité globale du mécanisme étudié. Un autre grand intérêt de notre méthode basée sur l'ABC est qu'elle permet de produire des prédictions par rapport aux différents scénarios possibles, et ainsi de guider les expériences nécessaires pour explorer plusieurs

hypothèses. Nous avons appliqué avec succès une version simplifiée de notre méthodologie à la problématique de l'estimation de la diffusion et la version complète à la modélisation des oscillations. Les deux problématiques sont très différentes et se placent à des niveaux de description différents. Le méthodologie proposée est donc polyvalente. Elle s'adapte facilement à des études qui sont en-dehors du cadre de celles présentées dans cette thèse.

Publications et communications

Publications :

— *Soumis au journal Biological Imaging :* A. Caranfil, Y. Le Cunff, C. Kervrann, "BayesTICS : Local temporal image correlation spectroscopy and Bayesian simulation technique for sparse estimation of diffusion in fluorescence imaging" (en révision)

— *En préparation :* A. Caranfil, J. Pécréaux, C. Kervrann, Y. Le Cunff, "Modelling the oscillatory behavior in asymmetric division of C. elegans embryo"

Posters et présentations orales :

— A. Caranfil, C. Kervrann, "Correlation-based method for membrane diffusion estimation in TIRF microscopy". Poster presented at : "France BioImaging : 4th Annual Meeting, Future Challenges in BioImaging", April 2017, Paris

— A. Caranfil, Y. Le Cunff, C. Kervrann, J. Pécréaux, "Modelling oscillatory behavior in asymmetric division of C.elegans embryo". Oral presentation at : "3rd Mathematical Biology Modelling days of Besancon", June 2018, Besancon

— A. Caranfil, Y. Le Cunff, C. Kervrann, J. Pécréaux, "Studying oscillatory behavior in asymmetric division of Caenorhabditis elegans embryo with fluorescence microscopy". Poster presented at : "QBI 2019 - Quantitative BioImaging Conference", January 2019, Rennes

— A. Caranfil, Y. Le Cunff, C. Kervrann, J. Pécréaux, "Modelling oscillatory behavior in asymmetric division of C. elegans embryo". Oral presentation at : "BioHazard - Stochastic Models for Biology", August 2019, Rennes

Aperçu de la thèse

Dans le Chapitre 1, nous expliquons l'importance des études de variabilité pour les mécanismes cellulaires et nous introduisons la méthode ABC (Approximate Bayesian Computation), qui constitue une technique-clé pour l'estimation et la prédiction de paramètres.

Dans le Chapitre 2, nous introduisons les méthodes d'imagerie TIRFM et des méthodes de corrélation associées pour l'estimation de paramètres. La première contribution de cette thèse, un nouveau modèle de diffusion de spots pour l'estimation de la diffusion locale au niveau de la membrane cellulaire, est ensuite présentée. La méthode d'estimation qui en résulte est validée et évaluée sur des données synthétiques, puis sur des données expérimentales.

Le Chapitre 3 explique les mécanismes mis en œuvre lors de l'anaphase de la division cellulaire asymétrique, et présente des travaux liés aux oscillations des centrosomes durant cette phase. Notre deuxième contribution, un modèle mathématique de ces oscillations, est décrite et validée. Nous présentons alors le cadre d'estimation de paramètres qui constitue la troisième contribution de cette thèse. Ce cadre d'estimation est utilisé pour identifier des paramètres de notre modèle d'oscillations, afin de le valider sur des données synthétiques.

SUMMARY

Preamble

This thesis is the result of a collaboration between two teams: the Serpico project-team at Inria Rennes and the CeDRE team at IGDR (Institut of Genetics and Development of Rennes). Methodological aspects, as well as approaches based on modeling, analysis and image processing, are the common threads that led the two teams to collaborate. Thus, a new project was born, with the objective of designing an approach based on data assimilation for cellular mechanisms analysis.

This thesis lies within this collaborative project, with the aim of proposing an approach for modeling and parameter estimation, for the study of the variability of cellular mechanisms. My thesis took place at 60 % in the Serpico team and at 40 % in the CeDRE team, under the supervision of Charles Kervrann, Research Director at Inria Rennes and head of the Serpico team, and Yann Le Cunff, Associate Professor at the University of Rennes 1, member of the DyLiSS team at IRISA, and member of the CeDRE team at IGDR during the supervision of this thesis, and in close collaboration with Jacques Pécréaux, CNRS researcher and head of the team CeDRE at IGDR.

Context

Advances in imaging methods, in particular in fluorescence microscopy, made it possible to use mathematical modelling to study cellular mechanisms. New mathematical models have been introduced in order to study global effects within the cell, in order to reveal general trends of the studied mechanisms. Our understanding of "average" behaviors is thus quite good; in contrast, individual behaviors and their influence at the cell level remain poorly understood. Although local models have also been increasingly designed to study specific events in the cell, few models or methods exist that can both describe general behaviors and provide information about local events at small scales.

As a consequence, values of biophysical or molecular parameters are only partially known. On the one hand, thorough knowledge of average biophysical magnitudes is avail-

able in the literature, and *in vivo* experiments made it possible to estimate some molecular or biophysical parameters that are directly observable from imaging techniques; on the other hand, other parameters are still partially known at best. These issues highlight the need for mixed approaches, that would allow, at the same time, to estimate local parameters and to study the interactions between the different actors involved in the mechanism of interest.

Finally, while predictive studies of cellular mechanisms is not made possible by pre-existing models, the design of a predictive method would allow to direct the experiments, by facilitating an *a priori* understanding of the studied mechanism.

Contributions

This thesis brings three main contributions. The first two contributions are two new models for the description of cellular mechanisms. The third contribution is a parameter estimation framework that allows to take into account the variability of the studied mechanism, but also the interactions between the various actors involved in the studied mechanism. The proposed framework then allows to test several possible scenarios and to establish predictions for each of these scenarios.

The first contribution deals with the design of a model for estimating the local diffusion of transmembrane proteins in TIRF imaging. Diffusion is usually estimated in a global way, in large zones in the cell: a correlation method is used to estimate the apparent diffusion coefficient, under the assumption of a homogeneous diffusion in the membrane. This approach does not allow to distinguish local diffusion, which may be induced by the stochasticity of the elements that compose the membrane and the factors of the diffusion itself, or by the various structural elements that may slow down or even inhibit the diffusion of transmembrane proteins. Here we propose a model based on a correlation method, but without any homogeneity assumption; this model allows to evaluate the diffusion for each diffusion event, within small image windows. This approach then allows to estimate diffusion maps, in order to identify areas of high spatial variation of the diffusion.

The second contribution of this thesis deals with the modeling of centrosome behavior during cell division, more precisely during the separation of chromosomes. Centrosomes move to the posterior side of the cell and, in doing so, they oscillate transversely to the anterior-posterior axis. Existing studies focus on posterior displacement, without taking

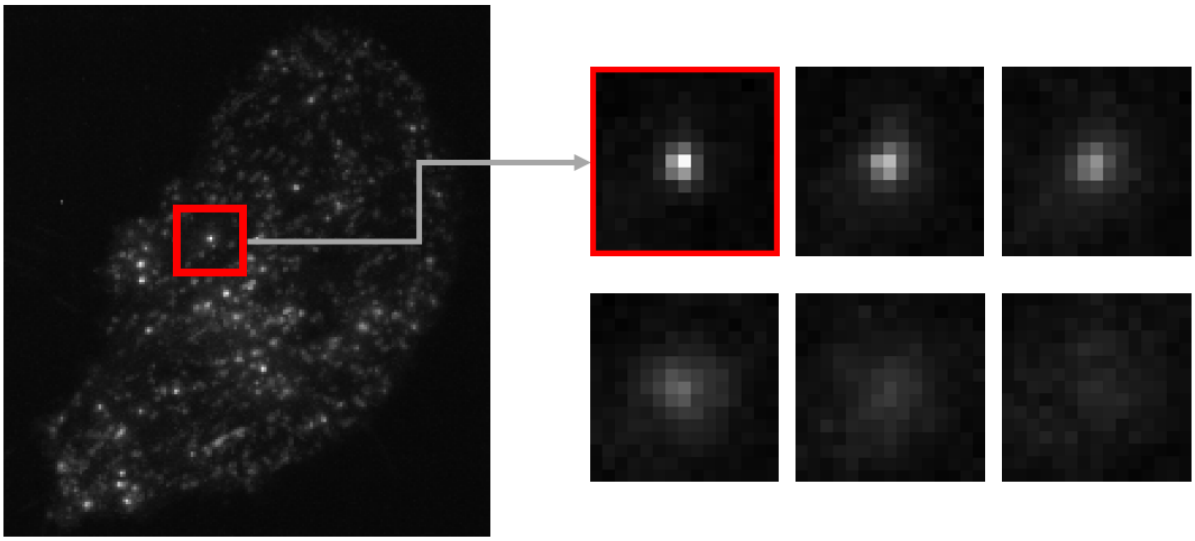


Figure 3 – **Illustration of a local spot of diffusion in TIRF imaging.**

these oscillations into account. However, not only are these oscillations frequently encountered during asymmetric divisions, but they also seem to play a key role in division. For instance, in the case of the C.elegans embryo, their absence is a sign of cell dysfunction, which can lead to erroneous division, or even to a lack of division and, ultimately, the death of the cell.

We focus here on these transverse oscillations and propose a new multi-scale model, which is based on a biophysical approach to the behavior of the centrosome. Force balance, as well as assumptions validated by the literature, allow us to design a model of centrosome oscillations during the chromosome separation phase. This model allows one to better understand the factors that trigger the oscillations, the conditions that are necessary for their existence, and the interactions that are put in place for them to occur. It notably makes it possible to test several scenarios of absence of oscillation in C.elegans, by combining information given by the model and experimental observations. Finally, it enables to represent the oscillations of the posterior centrosome for each embryo, yielding solid grounds for the study of variability from one embryo to another.

The third contribution of this thesis is the design of a multi-step parameter estimation framework, based on a mathematical model of the studied mechanism. The first step is a sensitivity analysis of the model parameters, which allows to group the parameters for estimation, according to their respective influences on the result. The second step is the definition of the common characteristics of the parameters, based on both previous

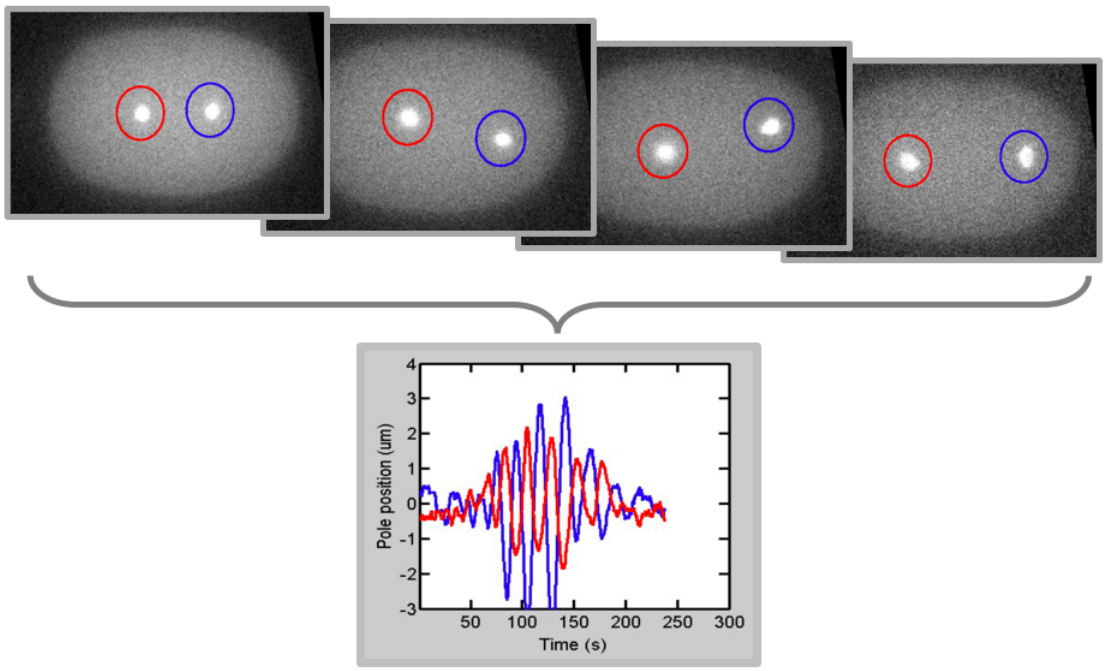


Figure 4 – **Illustration of the transverse oscillations of the spindle poles during cell division.**

knowledge about the studied mechanism and experimental observations. Finally, the last step consists in estimating the parameters, which is performed "blockwise" according to the results of the sensitivity analysis. Approximate Bayesian Computation, a Bayesian method for parameter estimation, is used at this stage. Our method allows to estimate the parameters as soon as a mathematical model is available; it relies on very few assumptions and does not require the computation of a likelihood. Parameter estimation is performed for each experiment: hence, it is possible to evaluate the individual parameters for each experiment, but also to analyze the global variability of the studied mechanism. Another key asset of our method is that it allows to produce predictions with respect to the different possible scenarios and, thus, to guide the experiments so as to efficiently explore various hypotheses. We applied a simplified version of our methodology to the diffusion estimation problematic and the complete version to the oscillations modeling problematic. The two biological issues are widely different, with different description levels. The methodology we propose is therefore versatile, and easily adaptable to studies beyond the scope of those presented in this thesis.

Publications and communications

Publications:

— *Submitted to Biological Imaging:* A. Caranfil, Y. Le Cunff, C. Kervrann, "BayesTICS: Local temporal image correlation spectroscopy and Bayesian simulation technique for sparse estimation of diffusion in fluorescence imaging" (in revision)

— *Under preparation:* A. Caranfil, J. Pécréaux, C. Kervrann, Y. Le Cunff, "Modelling the oscillatory behavior in asymmetric division of C. elegans embryo"

Posters and talks:

— A. Caranfil, C. Kervrann, "Correlation-based method for membrane diffusion estimation in TIRF microscopy". Poster presented at : "France BioImaging : 4th Annual Meeting, Future Challenges in BioImaging", April 2017, Paris

— A. Caranfil, Y. Le Cunff, C. Kervrann, J. Pécréaux, "Modelling oscillatory behavior in asymmetric division of C.elegans embryo". Oral presentation at : "3rd Mathematical Biology Modelling days of Besancon", June 2018, Besancon

— A. Caranfil, Y. Le Cunff, C. Kervrann, J. Pécréaux, "Studying oscillatory behavior in asymmetric division of Caenorhabditis elegans embryo with fluorescence microscopy". Poster presented at : "QBI 2019 - Quantitative BioImaging Conference", January 2019, Rennes

— A. Caranfil, Y. Le Cunff, C. Kervrann, J. Pécréaux, "Modelling oscillatory behavior in asymmetric division of C. elegans embryo". Oral presentation at : "BioHazard - Stochastic Models for Biology", August 2019, Rennes

Thesis outline

In Chapter 1, we explain the importance of variability studies for cellular mechanisms and we introduce the Approximate Bayesian Computation method as a key technique for parameter estimation and prediction.

Chapter 2 introduces TIRFM acquisition methods and associated correlation methods for parameter estimation. The first contribution of this thesis, namely, a new spot diffusion model for the estimation of local diffusion at the cell membrane, is then presented, and the resulting estimation method is validated and assessed on both synthetic and experimental data.

Chapter 3 explains the mechanisms at work during anaphase in asymmetric cell divi-

sions, and presents works related to the oscillations of the mitotic spindle poles during this phase. Our second contribution, a mathematical model of these oscillations, is described and validated. We then present the parameter estimation framework that constitutes the third contribution of this thesis. This framework is used for estimating the parameters of our oscillation model, thus validating it on synthetic data.

Chapter 1

Variability at cell scale, how to handle it ?

1.1 Why variability matters

Variability is visible at all biological scales, it is what makes life so diverse, and what enables cells, organisms, and species to survive in an always changing environment [7]. At the cell scale, variability yields robustness to internal and external fluctuations, and it provides cells with the means to adapt to rapid changes in the environment [2] [82]. At the population level, a diverse spectrum of responses by individual specialization is achieved through variability, and this specialization can either enable a few cells to survive in case of exceptional conditions, or have them trigger a signal resulting in a collective response of the population and thus enable the population to survive [3].

Cellular variability is more and more considered as an evolutionary mechanism. J. J. Kupiec [41] suggested that a Darwinian mechanism for developmental decisions could include variability, as variation generates different cell fate scenarios and regulation mechanisms act as a selection mechanism. In the same vein, H. Dueck et al. [22] proposed that single cell variation is required, at least in part, for system and population-level function, as seen in ecological communities where individuals with different roles enable group functioning. In this context, how does a system manage variability to obtain a coherent behavior that enables survival [52]? Should variability always be reduced or regulated as much as possible or is variability a mechanism that systems can use in their advantage [27]? These ongoing questions have been studied for a long time, and partial answers were given, but much is still to be uncovered.

Types of cellular variability At the cell scale, variability can be genetic, environmental or due to other factors, this last type being referred to as phenotypic variability

[3]. Although works on genetic and environmental variability have been carried out for quite a long time, the interest in phenotypic variability is relatively new. Here, we focus mainly on phenotypic variability, whether this type of variability should be investigated and when, and what studying phenotypic variability can tell us about a biological system such as a cell or a population of cells.

Studying phenotypic variability The origins of phenotypic variability are not well understood. Studying this type of variability is often complex and difficult to set up in practice, as it has to be measured over several individuals of a population, or over long periods of time for a single individual. This can sometimes be technically difficult, especially for in vivo experiments where some kind of invasive technique is used, and long exposure can result in a variety of effects that can change the phenotype of interest. The impact of genetic and environmental factors of variation has to be diminished as much as possible and, in practice, it is almost impossible to get both genetically identical individuals and homogenous environmental conditions for each observed individual. Nevertheless, several studies demonstrated that variability can emerge from molecular mechanisms that are environment-independent [79] [36]. In addition, it has been established that phenotypic variation can arise at rates that are far greater than the rates of mutational mechanisms, or to be unaffected by the inactivation of known mutational mechanisms [28]. Finally, as variability is often an indirect effect of several complex mechanisms working together in a complex manner, it is mathematically difficult to take every factor and every correlation into account, especially because there is not enough prior information about these factors and their interactions [23] [22]. This last part will probably become less of a problem as more and more phenotypic studies are conducted, and more hypotheses and different scenarios are being tested and validated.

Perhaps the most thoroughly studied case of phenotypic variability concerns microorganisms. Microorganisms, such as bacteria, have been studied for a long time, especially in culture setups. The first studies of variability came from Delbruck, who, in the 1940s, observed that virus-infected E.coli bacteria display a large variability in the number of phages produced per cell. He suggested that this variability could be due to the stochasticity of chemical interactions in the cell, and also that predetermined factors, such as cell size, could have an influence on the observed variability. Since then, studies on microorganisms have helped us better understand how phenotypic variability arises and what its

consequences are.

Stochastic factors In microorganisms, several molecular causes were found to be at the origin of cellular variability in genetic and environmental identical conditions. One of them is the stochasticity of some molecular processes, such as gene expression [24] and distribution of molecules during cell division [36]. One example of phenotypic variability due to stochastic gene expression is observed in genetically modified large cells of Bacilius subtilis that have less variability in the initiation of sporulation than normal cells [79]. Larger cells result in reduced variation in molecular composition and, in turn, reduced variability in initial sporulation. Other mechanisms that could be responsible for phenotypic variability are oscillations and periodic cycles, such as the cell cycle. Cell-cell interaction, through physical contact or diffusable molecules [76] [70] can also result in such variability. The phenotype of one cell can influence the phenotype of another cell [3], as gene expression can be regulated by signaling from nearby cells [91] [13], and variation in such signaling mechanisms can result in variation in gene expression, and thus, in phenotypic variability.

Deterministic factors Interestingly, although some of the phenotypic variability is thought to be due to stochastic mechanisms [67] [45], an increasing number of deterministic mechanisms were found to regulate these stochastic mechanisms. Indeed, gene expression can be regulated through gene-regulatory networks, in a way that can decrease or increase variation, and this can be achieved through negative or positive feedback loops. A regulatory mechanism such as a positive feedback loop [75] can give rise to different groups in a population exhibiting different phenotypes. On the other hand, negative feedback loops can lead to decreased variability in gene expression. In E.coli for example, negative feedback loops are responsible for low levels of variation in the expression noise of essential genes, which increases the fitness of the cells and, thus, their growth rates [43] [90]. Switch-like behaviors, that were previously thought to be greatly due to stochastic processes, were found to be predetermined by cell's physiology, growth rate or population density [71]. B. Snijder and L. Pelkmans proposed in 2011 that phenotypic variability arises largely from the spectrum of microenvironmental differences in a cell population and from cell's history (memory of protein levels or phenotypic state for example), and that studies of cell-to-cell variability need to distinguish between stochastic origins and regulated, deterministic origins in order to better understand phenotypic variability.

Phenotypic variability and population survival Phenotypic variability is a complex trait, that is itself variable between individuals, the underlying mechanisms that cause such variability being diverse in all kinds of ways: different molecular bases, different dynamics and different local and global impact. Phenotypic variability can influence several crucial aspects of the life of an individual or a population, from development to reproduction and up to survival [67], [7]. If phenotypic variability is an evolutionary trait, what are its advantages for a cell, an organism or a population?

Studies from microorganisms gave the first answers to this question. In a fluctuating environment, microorganisms have several strategies to ensure survival. The most common theories state that individuals sense their environment and respond to changes through signal transduction and regulation of gene expression that generate the appropriate phenotype to fit the given conditions. In low-fluctuation environment, or if fluctuations occur over a relatively long period of time, this strategy allows most individuals to survive as they can all adapt to new conditions through this common mechanism. However, this might fail when changes in the environment happen too fast for individuals to have the time to generate optimal phenotypes through this mechanism, or when conditions are too diverse for all individuals to develop a signal transduction pathway. One possible solution is to increase phenotypic variability regardless of the environment, so that a variety of responses is already available in case of rapid fluctuations in the environment: this strategy is known as bet hedging. For example, studies on E.coli show that a small part of the population has a slow growth rate (these individuals are known as persisters), which ensures a better tolerance to antibiotics in case of sudden exposure [9]. Moreover, E.coli cells can switch between normal and persistence states in a stochastic manner [45]. Phenotypic switching is also present in B.subtilis that can switch between competence and sporulation [49] [17], suggesting that phenotypic variability is used as a survival strategy by this organism. Studies on cancer cells suggest that this mechanism could be used to switch between noninvasive and invasive states, which enables survival to chemotherapies that target proliferating cells [34] [93]. Other examples of the use of stochastic phenotypic variability includes the metabolic functions, that enable organisms to adapt to new sources of food [2], or the immune system, that can rapidly respond to changing conditions thanks to individual phenotypic variability [15] [39].

Generating a large variety of phenotypes can play other crucial roles for group survival

and functioning. One such role is the production and secretion of molecules by a small percentage of the population, that can benefit the whole population in case of crisis. In clonal populations of B.subtilis, some individuals produce a protease called subtilisin E in food-poor environments. This molecule is costly to produce, so only a few individuals engage in the production and secretion of subtilisin E. The molecule diffuses freely in the medium of growth, and it allows the degradation of proteins outside of the cell, its degradation products serving the whole population. This phenotype has its origins in the expression of the gene aprE that encodes subtilisin E, and only a minority of individuals were found to express this gene [87]. A second example is the separation of mutually exclusive tasks between cells. This is the case for nitrogen fixation and photosynthesis in cyanobacteria, as the process of photosynthesis damages the enzyme responsible for nitrogen fixation. These two cellular processes are thus accomplished simultaneously by different cells, ensuring the proper functioning at the organism level for the cyanobacteria [4]. A third example is the coordination of a population by a small percentage that responds rapidly to changes in the environment and emits a signal that triggers a group response. Patil et al [61] studied a population of human dendritic cells infected with the Newcastle disease virus; they showed that a small fraction of the cells, not only rapidly activated Ifnb1 in order to fight the disease, but, most importantly, emitted paracrine signals that further activated Ifnb1 expression in the rest of the population. This mechanism is crucial for populations of cells that need to balance between several behaviors, such as the immune system that needs a balance between rapid response to attacks and avoiding self-toxicity [73] [61]. Finally, phenotypic variability can play a major role in contexts that demand a fractional response from the population. In the case of binary decision at the cell scale, phenotypic variability transforms the binary behavior into a continuous behavior at the population level, allowing for fractional or dose-dependent responses, instead of a switch-like behavior that would be triggered from binary decisions in a uniform population. Several studies suggest that this kind of mechanism is responsible for fractional death of cancer cells in response to chemotherapy [40] [60], and fractional killing in drug treatment using apoptosis-inducing receptor ligands such as TRAIL [77] [12] [72].

The consequences of phenotypic variability can be a combination of the previously mentioned functional roles. For example, the variability in virulence genes expression in Salmonella leads to both segregation of tasks and generation of persisters during antibiotic exposure [8]. An increasing amount of evidence shows that phenotypic variability has a

major functional impact from the cell level to the population level, and that its dynamics is more complex than it was previously assumed [3].

1.2 Measuring variability : model parameter estimation with Approximate Bayesian Computation method

Cellular variability can be studied in a variety of ways, depending on the prior knowledge, the dynamics of the observed process, and the type of available data. In this thesis, we propose a direct approach based on the signal modeling of the molecular process and the estimation of model parameters using a Bayesian framework. This approach yields both point estimates and complete distributions of the estimated parameters. Moreover, the model parameters are linked to molecular characteristics. Thus, the variability of these parameters provides direct information about the molecular variability. In what follows, we describe the parameter estimation framework we used to study local variability in Chapter 2 and population variability in Chapter 3.

The Bayesian approach enables model checking and validation as well as parameter inference and prediction. In this framework, both the data and the model parameters are considered as random variables. Then, the conditional probability distribution of the parameters given some observed data contains all the information needed to analyze the model; this distribution is called the posterior distribution.

Let us consider a model with a p-dimensional vector of parameters $\theta \in \mathbb{K}^p$. In order to compute the posterior distribution of θ, a prior distribution for θ, and the so-called likelihood of the data must be given. The prior distribution is chosen by the expert, and can provide information about what is already expected about θ, in which case it is informative; but it can also show that nothing is known about the parameters, and in this case the prior distribution is noninformative. The latter case can take the form of a uniform distribution in the range of allowed possible values for θ. The likelihood gives the probability for each observation, given the model and parameters θ.

The posterior distribution for θ is generated by updating the prior distribution, through

the likelihood, given the observations y_{obs}. Mathematically, Bayes' theorem gives the posterior distribution:

$$\pi(\theta|y_{obs}) = \frac{p(y_{obs}|\theta)\pi(\theta)}{\int p(y_{obs}|\theta)\pi(\theta)\,\mathrm{d}\theta} \tag{1.1}$$

Although this approach is quite simple in its formal description, the complexity of the model, or that of the prior, makes it impossible to obtain an exact solution for the posterior distribution. Most of the times, numerical methods are used to compute the posterior. However, even in this case, the posterior might be impossible to compute because the numerical evaluation of the likelihood can be too computationally costly, or even impossible. The prior distribution is always available as it is chosen, but the likelihood can be difficult to obtain in a mathematical form, and for some models even impossible; this is the case for models that directly simulate the data-generating mechanism (models employing neuron networks for example). Thus, the direct application of Bayesian methods is restricted to fairly simple models.

To solve this problem, several techniques have emerged that approximate the posterior distribution. Among them, methods that are "likelihood-free" have proved effective and easy to implement. In this group, the Approximate Bayesian Computation – ABC - has been of particular interest, as an intuitive and accessible method for model analysis and inference. The idea behind ABC is to approximate the posterior in an indirect way, without computing or numerically evaluating the likelihood.

This method was first developed in [81] and [66], and it was since extended in a number of ways, and for a wide range of applications. A few examples of biology related applications can be found in [69, 68] (protein networks), [44] (single cell gene expression), [89, 88] (cell biology), [80] (pathogen transmission), [47, 1] (infectous disease dynamics).

1.2.1 Standard ABC method

The classical posterior distribution is given by formula (1.1). The integral in the denominator of the right-hand side acts as a normalization constant. The posterior is thus proportional to the product of the likelihood and the prior, i.e.:

$$\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta). \tag{1.2}$$

As the likelihood is usually not available, the posterior cannot be computed using this equation. Instead, the basic ABC method [66, 11] provides a way of approximating the posterior distribution by only using the model itself. A data set $\tilde{y}$ can be directly obtained by simulating the model for a given parameter vector $\tilde{\theta}$. This vector $\tilde{\theta}$ is then considered a valid approximation of θ if the distance between the simulated data $\tilde{y}$ and the observation y_{obs} is small enough, i.e. $\|\tilde{y} - y_{obs}\| \leq h$, with $\|\cdot\|$ a chosen norm, and h close to 0. In this case, $\tilde{\theta}$ is a sample from the desired posterior distribution. Mathematically, this means drawing samples $(\tilde{\theta}, \tilde{y})$ from the joint distribution defined by:

$$\pi_{ABC}(\theta, y|y_{obs}) \propto \mathrm{I}(\|y - y_{obs}\| \leq h)p(y|\theta)\pi(\theta), \tag{1.3}$$

where $\mathrm{I}(\cdot)$ is the indicator function defined by $\mathrm{I}(z) = 1$ if z is true, and $\mathrm{I}(z) = 0$ otherwise. Furthermore, the marginal distribution is given by

$$\pi_{ABC}(\theta|y_{obs}) = \int \pi_{ABC}(\theta, y|y_{obs})\,\mathrm{d}y. \tag{1.4}$$

Then, we have

$$\begin{aligned}
\lim_{h\to 0} \pi_{ABC}(\theta|y_{obs}) &= \lim_{h\to 0} \int \pi_{ABC}(\theta, y|y_{obs})\,\mathrm{d}y \\
&\propto \lim_{h\to 0} \int \mathrm{I}(\|y - y_{obs}\| \leq h)p(y|\theta)\pi(\theta)\,\mathrm{d}y \\
&= \int \Big(\lim_{h\to 0} \mathrm{I}(\|y - y_{obs}\| \leq h)\Big)p(y|\theta)\pi(\theta)\,\mathrm{d}y
\end{aligned}$$

$$\begin{aligned}
\lim_{h\to 0} \pi_{ABC}(\theta|y_{obs}) &\propto \int \delta_{y_{obs}}(y)p(y|\theta)\pi(\theta)\,\mathrm{d}y \\
&= p(y_{obs}|\theta)\pi(\theta) \\
&\propto \pi(\theta|y_{obs}).
\end{aligned}$$

Thus, for $h \to 0$, the $\tilde{\theta}$ from the samples $(\tilde{\theta}, \tilde{y})$ are drawn from the true posterior distribution. This is the basis of the ABC method.

Numerical implementation From a practical point of view, the basic ABC algorithm proceeds as follows:

1. Sample a candidate $\tilde{\theta}$ from the prior.
2. Use $\tilde{\theta}$ to simulate data $\tilde{y}$ from the model ($\tilde{y}$ having the same dimensions as y_{obs}).
3. Compute the distance $\|\tilde{y} - y_{obs}\|$. If the distance is small enough i.e. $\|y - y_{obs}\| \leq h$, then $\tilde{\theta}$ is accepted as a sample from the posterior. If not, discard $\tilde{\theta}$.

The final step of the estimation is to compute the posterior expectation known as the minimum mean square error (MMSE) estimator and defined as $\tilde{\theta}_{\text{MMSE}} = \mathbb{E}(\tilde{\theta}|y) = \int p(\tilde{\theta}|y)\tilde{\theta}\,\mathrm{d}\tilde{\theta}$. The simplest way to compute the MMSE estimator is to draw samples $\{\theta_i\}_{i=1,\ldots,N}$ using the procedure detailed above, then approximate $\tilde{\theta}_{\text{MMSE}}$ by $\tilde{\theta}_{\text{MMSE}} = N^{-1}\sum \theta_i$. Then, the samples can be used to compute the posterior distribution for which the maximum mode equals the maximum a posteriori (MAP) estimator $\tilde{\theta}_{\text{MAP}}$.

This algorithm requires selection of a suitable norm $\|\cdot\|$, as well as a choice of h. As $h \to \infty$, accepted $\tilde{\theta}$ come from the prior, and as $h \to 0$, accepted $\tilde{\theta}$ is drawn from the posterior. The choice of h reflects the balance between computability and accuracy. When $h = 0$, the algorithm is exact and gives samples from the true posterior.

1.2.2 Generalized ABC method

A generalization of this method is possible, in order to improve the approximation [74]. The first point is to replace the indicator function $\mathrm{I}(\cdot)$ with a standard smoothing kernel $\mathrm{K}_h(u)$ defined by:

$$\mathrm{K}_h(u) = \frac{1}{h}\mathrm{K}\left(\frac{u}{h}\right). \tag{1.5}$$

Kernels are symmetric functions with the following characteristics :

- $K(u) \geq 0$ for all u,
- $\int K(u)\mathrm{d}u = 1$,
- $\int uK(u)\mathrm{d}u = 0$,
- $\int u^2K(u)\mathrm{d}u < \infty$.

In particular, this implies $\lim_{h\to 0} K_h(u) = \delta_0(u)$. The kernel $K_h(u)$ introduces weights for the accepted $\tilde{\theta}$, such that the $\tilde{\theta}$ for which h is close to zero will have more weight than those that make h further from 0.

The second point is to generate $\tilde{\theta}$ from an arbitrarily simpler distribution $g(\theta)$ instead of the prior distribution, and accept $\tilde{\theta}$ with probability proportional to $\mathrm{K}_h(u)\frac{\pi(\tilde{\theta})}{g(\tilde{\theta})}$. The reason for this probability of acceptance is that the expression of π_{ABC} as a function of g and p is as follows:

$$\pi_{ABC}(\theta, y|y_{obs}) \propto \mathrm{K}_h(\|y - y_{obs}\|)\frac{\pi(\theta)}{g(\theta)}p(y|\theta)g(\theta). \tag{1.6}$$

Hence,

$$\pi_{ABC}(\theta, y|y_{obs}) \propto \mathrm{K}_h(\|y - y_{obs}\|)p(y|\theta)\pi(\theta), \tag{1.7}$$

which is very similar to (1.3). Indeed, the limit behavior of the marginal distribution is the same:

$$\begin{aligned}
\lim_{h\to 0}\pi_{ABC}(\theta|y_{obs}) &= \lim_{h\to 0}\int \pi_{ABC}(\theta, y|y_{obs})\,\mathrm{d}y \\
&\propto \lim_{h\to 0}\int \mathrm{K}_h(\|y - y_{obs}\|)p(y|\theta)\pi(\theta)\,\mathrm{d}y \\
\lim_{h\to 0}\pi_{ABC}(\theta|y_{obs}) &\propto \int\Big(\lim_{h\to 0}\mathrm{K}_h(\|y - y_{obs}\|)\Big)p(y|\theta)\pi(\theta)\,\mathrm{d}y \\
&= \int \delta_{y_{obs}}(y)p(y|\theta)\pi(\theta)\,\mathrm{d}y \\
&= p(y_{obs}|\theta)\pi(\theta) \\
&\propto \pi(\theta|y_{obs})
\end{aligned}$$

As a result, by generating samples from the joint distribution, $\tilde{\theta}$ is drawn from the true posterior as $h \to 0$.

Numerical implementation In practice, the ABC algorithm is modified as follows:

1. Sample a candidate $\tilde{\theta}$ from g.
2. Use $\tilde{\theta}$ to simulate data $\tilde{y}$ from the model ($\tilde{y}$ having the same dimensions as y_{obs}).

3. Compute the probability of acceptance

$$\frac{K_h\left(\|\tilde{y} - y_{obs}\|\right)\pi(\tilde{\theta})}{Cg(\tilde{\theta})},$$

where the constant C ensures that this expression properly defines a probability. Accept $\tilde{\theta}$ with this probability.

Then, similar to the basic ABC method, one can compute the MMSE estimator $\tilde{\theta}_{\text{MMSE}}$ and the MAP estimator $\tilde{\theta}_{\text{MAP}}$.

The choice of h is paramount in order to balance computability and accuracy: as $h \to \infty$, the accepted $\tilde{\theta}$ is drawn from g, while, as $h \to 0$, the accepted $\tilde{\theta}$ is drawn from the posterior.

The ABC method requires several thousands of samples in order to compute the posterior distribution. Thus, this method becomes appropriate to use when the data $\tilde{\theta}$ is cheap and fast to simulate.

In this thesis, the ABC method was used for parameter inference, in both Chapter 2 and Chapter 3. The complexity of the developed models has guided us to the ABC method, along with the other advantages previously mentioned in this section. The adjustments and improvements made to this method are detailed in the respective sections for each model.

Chapter 2

Local variability: diffusion estimation in fluorescence imaging

Variability at the level of a single cell is essential for cell adaptability to rapidly changing environments. Studying the local variability of cellular mechanisms can reveal internal changes, such as changes in dynamics or in local structure. It can also attest of intrinsic variability due to stochastic factors that the cell could use to its advantage. In this chapter, we study the local variability of the diffusion of transmembrane proteins at the plasma membrane.

Diffusion is the most common way of transport at the microscopic scale, as it requires no energy to generate movement. It accounts for 80% of trafficking in the cell, and there are several crucial processes that rely on it for transport.

One of the key processes relying on diffusion is exocytosis, which consists in the transport of cargo to the cell membrane or to the extracellular medium. The role of exocytosis is to regenerate the membrane, to eliminate waste from the cell and to transport molecules synthesized in the cell that will play a role outside of it, such as neurotransmitters, hormones, or enzymes.

The most used technique to image the exocytosis process is Total Internal Reflection Fluorescence Microscopy, or TIRFM. The principle of this technique, illustrated by Fig. 2.1, is to use the total reflection of light in order to generate a weak electromagnetic field that will excite the fluorescently tagged molecules. This technique yields a series of images showing the evolution in time of the fluorescent intensity during exocytosis. In particular, the different steps of exocytosis have been accurately described using TIRFM.

In order to estimate the diffusion during exocytosis, a standard approach consists in

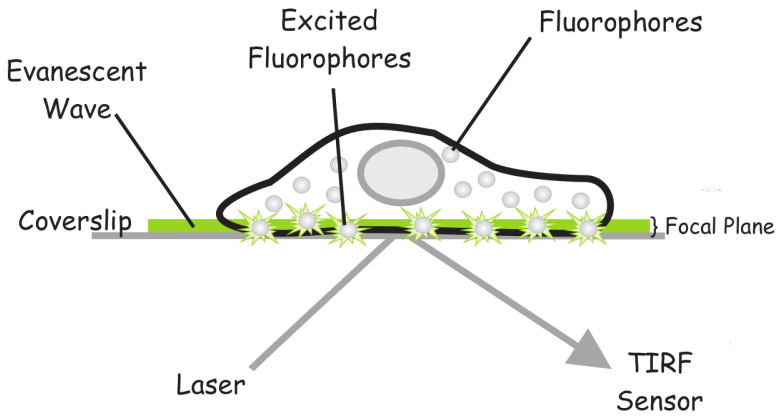


Figure 2.1 – **Illustration of TIRF microscopy principle.**

combining the TIRF acquisition method with a correlation method. This approach yields mean characteristic values for diffusion, but does not provide any information about the local variability of the diffusion between different regions of interest, and in particular, it does not allow for the estimation of diffusion for isolated diffusion events.

In what follows, we propose a new method that allows for the evaluation of local diffusion and its variability in TIRFM images, using a single sequence from classical acquisition. Our approach, based on a standard correlation method, uses initial local conditions in order to better describe local diffusion events during the exocytosis process. This method is well suited for both isolated diffusion events and cases where two or several diffusion events occur in the region of interest.

Sections 2.1 and 2.2 present the biological context, the acquisition method TIRFM and describe the general principle of the correlation methods. Section 2.3 presents the mathematical background of how the method for local diffusion estimation adapted for TIRFM was designed. We describe extensive simulations and demonstrate the effectiveness of our method before applying it to real sequences of TIRFM images depicting the diffusion of transferrin receptors during exocytosis.

2.1 Context : exocytosis mechanisms

In this section we will study the diffusion of transmembrane proteins during the fusion phase of exocytosis. The proper diffusion of the transmembrane proteins is important, not

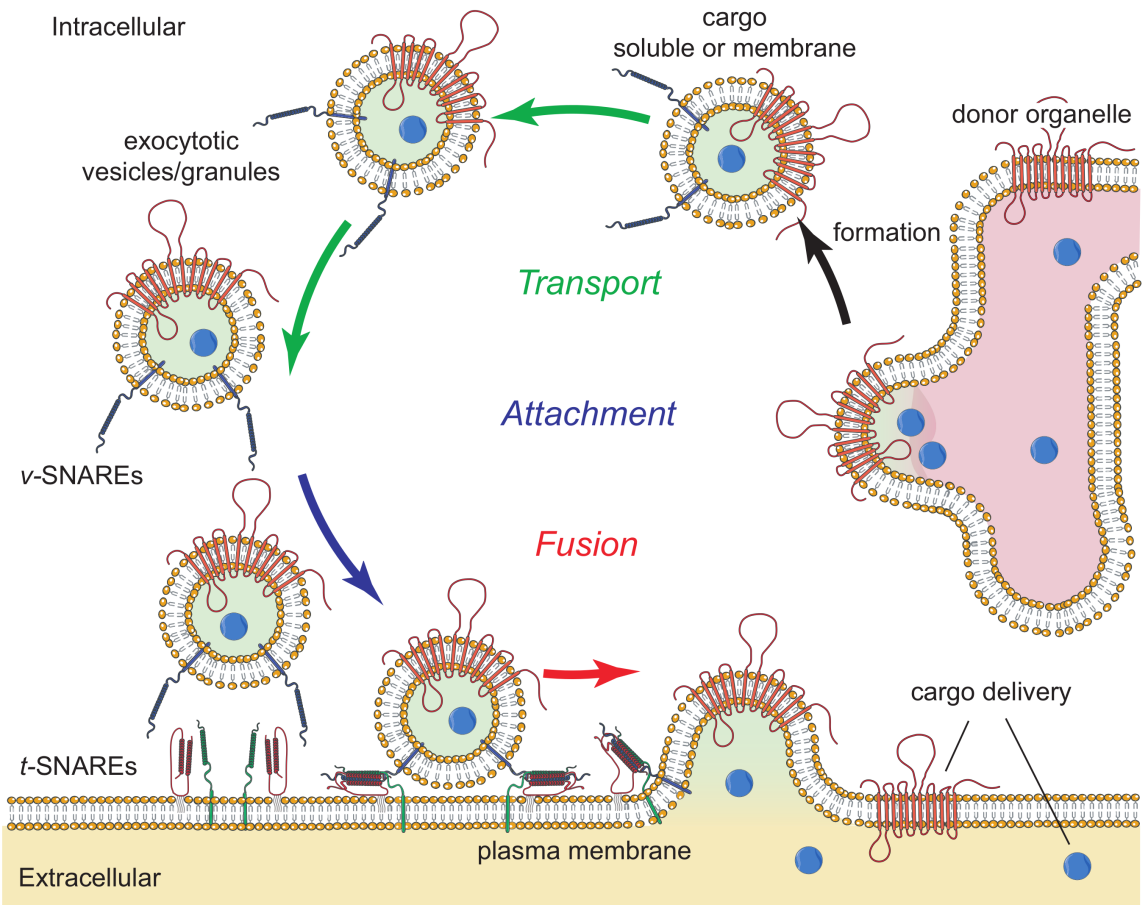


Figure 2.2 – **Illustration of the three steps of exocytosis.** Adapted from [16].

only for its role in the fusion process, but also later in the cycle of the transmembrane proteins, because they serve as receptors for outside molecules that are needed in the cell and help with their transport inside the cell, via the endocytosis process.

The are two types of exocytosis: constitutive exocytosis and regulated exocytosis. On the one hand, constitutive exocytosis is used to regenerate the cell membrane and to eliminate waste from the cell, and is thus a process that takes place in every cell. On the other hand, regulated exocytosis is triggered by an external factor (such as Calcium ions) and happens only in specialized cells; its role is to transport specific molecules to the cell membrane and to deliver molecules outside the cell.

The factors influencing exocytosis, as well as the molecules involved in the mechanism

and their regulation, can be different between cells, or even between parts of the cells, and can lead to different types of exocytosis. Nevertheless, there are three major steps that govern exocytosis, as illustrated by Fig. 2.2:

- transport of the cargo by a vesicle from the donor organelle to the cell membrane;
- attachment of the vesicle to the membrane;
- fusion to the membrane.

The first step of exocytosis is the generation of a vesicle and its transport to the cell membrane, also called plasma membrane. Vesicles form in a donor organelle, and they are composed of a membrane that encapsulates the cargo molecules. The vesicle membrane is made of lipids and associated proteins, just as the plasma membrane, but its specific composition depends on the type of exocytosis, and its role. Once a vesicle is formed, it undergoes active transport along the cytoskeleton with the help of molecular motors. In this stage, microtubules (that are tubular structures of the cytoskeleton) have been shown to play a crucial role. Indeed, when disrupting microtubule growth with nocodazole in adipocytes (fat cells), the transport of the storage vesicles of GLUT4 (glucose transporter 4), from the intracellular sites to the plasma membrane, is inhibited [18].

When the vesicle arrives near the plasma membrane, the cortical actin network moves the vesicle to a fusion site, where the second phase of exocytosis takes place: the attachment of the vesicle to the plasma membrane via specialized docking proteins. These docking proteins are called SNAREs (soluble N-ethylmaleimide-sensitive factor attachment protein receptors), and are present both in the vesicle (v-SNARE) and in the membrane (t-SNARE). These proteins interact in a zipper manner and form a SNARE complex that will attach the vesicle to the membrane. The formation of the SNARE complex is crucial for attachment, as shown in 2002 by a study on pancreatic beta cells [57]. Cells were treated so as to reduce the interactions between SNARE proteins, which resulted in a reduction of proper attachment of secretory vesicles.

During the attachment phase, the actin network and various proteins regulate the attachment of the vesicle. Disruption of actin has been shown to inhibit attachment of secretory vesicles in neuroendocrine PC-12 cells [42] and adrenal chromaffin cells [58]. Myosins, that are molecular motors associated with actin filaments, were shown to play an important role in the transport along the actin network, and in the attachment of

vesicles. Several studies show that reduced expression or defective Myosin Va results in reduced transport and attachment of secretory vesicles [86, 38]. Through its dynamic behavior, the actin network can either enhance the movement of the vesicles, or diminish their movements and even act as a barrier to the plasma membrane. This makes actin and the associated molecular motors an important regulatory system in the attachment phase of exocytosis.

Once a vesicle is correctly attached, it can fuse to the plasma membrane. The membrane of the vesicle becomes a part of the plasma membrane, thus regenerating the cell membrane, and, depending on the type of exocytosis, molecules can be delivered outside the cell. In this phase, several molecular actors are thought to play an important role. The dynamics of the actin network was shown to play a role in the stability of the fusion pore. Indeed, the overexpression of B-actin, actin-related protein 3 or mAbp1 in PC-12 cells, as well as high concentrations of cytochalasin D (that inhibits actin filament polymerization), slow the delivery of big cargo molecules in Ca^{2+} regulated exocytosis [26]. Another molecular actor that was shown to play an important role in vesicle fusion in Ca^{2+} regulated exocytosis is the Synaptotagmin proteins family. In Syt-1 shRNA PC-12 cells (that express partial loss of function), there was a reduction of fusion pore opening and expansion that resulted in a reduction in vesicle fusion events in Ca^{2+} regulated exocytosis [48]. Other studies showed that Synaptotagmins are involved in the regulation of vesicle fusion, either by inducing positive membrane curvature [53], or by modulating fusion pore opening [84]. The Synaptotagmin proteins are transmembrane proteins: they are permanently attached to the membrane, and span it entirely. Transmembrane proteins are present in the plasma membrane as well as in the vesicle membrane, and they are thought to have several roles in exocytosis: they help breaking through the cell membrane, they keep the cargo molecules in place during transport, and once the fusion begins, they diffuse in the plasma membrane, taking the cargo molecules alongside and helping them get outside of the cell.

2.2 Introduction to image correlation spectroscopy and TIRFM methods

2.2.1 Principle of Image Correlation Spectroscopy

Fluorescence Correlation Spectroscopy (FCS), illustrated by Fig. 2.3, is an acquisition and analysis method introduced in 1974 [50, 25, 51], when techniques for tagging molecules with fluorescent particles first started to emerge. The aim of this method is to capture the fluorescent signal from the tagged molecules with an optical system and then use this signal to measure the concentration of particles in the region of focus. The acquired fluorescent signal is a one-dimensional signal, so the analysis is naturally performed with a correlation method, a fairly easy method to implement, and popular in signal processing.

The acquisition setup in FCS is relatively simple. A laser beam, previously focused to a point (using a confocal microscope, usually), is sent to the region of interest. This laser beam excites the tagged molecules that further emit photons. The photon count gives the fluorescent signal, that is captured by an optical sensor and transformed into a digital signal afterwards. This 1D signal depicts the evolution over time of the intensity of the fluorescence.

It is generally assumed that the concentration of particles in the studied region of interest (ROI) is low and only one molecule at a time resides in the ROI, so that the photon count is unambiguously coming from a single molecule. The ROI also needs to be free of background so there is no possible fluorescent signal interfering with the emitted signal. Because the excitation is obtained directly with a laser beam, the probability of photobleaching of the fluorophores is high. A low intensity laser beam is generally recommended. Another consequence of the possible photobleaching is that the time of acquisition must be short before molecules start photobleaching, which means that only fast dynamics can be studied. Lastly, the signal-to-noise ratio increases with time, so FCS is limited to short acquisition times.

In classical FCS, one assumes that the intensity of the emitted fluorescent light depends on the local concentration of the fluorophores and the laser intensity profile. Once the intensity of emitted light is acquired, the autocorrelation of the fluctuations of the

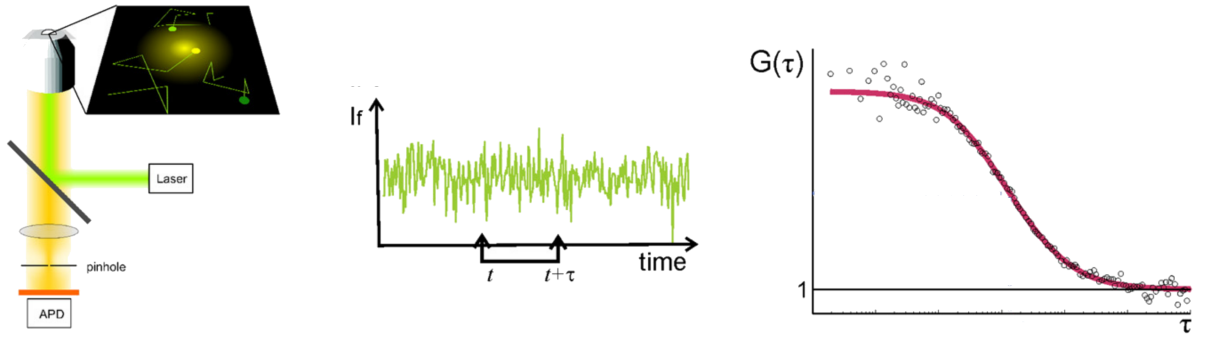


Figure 2.3 – **Principle of the FCS method.** Left: schematic experimental setup. Center: acquired fluorescent signal. Right: autocorrelation of the signal. Adapted from [29].

signal is computed. Depending on the assumptions on the observed molecule population, the theoretical form of the autocorrelation can differ.

Two methods derived from FCS have rapidly emerged, to overcome its limitations in some sense:

ICS - Image Correlation Spectroscopy [64]: ICS uses the same principle as FCS but the laser moves in space, such that a larger zone is covered. With ICS, an image is acquired showing different fluorescent points corresponding to fluorescent molecules. The spatial correlation of this image is then used to access information about the observed population. Using ICS, one can compute the average number of molecules in the image domain as well as the average concentration of the population. The acquisition in ICS is fast and has less problems with photobleaching and bad signal-to-noise ratio. However, as for FCS, the number of molecules in the ROI must be low and the image must be free of background.

TICS - Temporal Image Correlation Spectroscopy [92]: Based on the same principle as FCS and ICS, TICS combines their advantages; it gives access to the concentration of the population in the ROI, as well as the dynamical properties such as the characteristic time of diffusion. It also gives the percentage of immobile particles in the region of observation. TICS can be applied to both low and fast dynamical populations, and it can detect diffusion and flow at the same time. TICS uses laser scanning as ICS to cover a whole region, and it goes further by acquiring a set

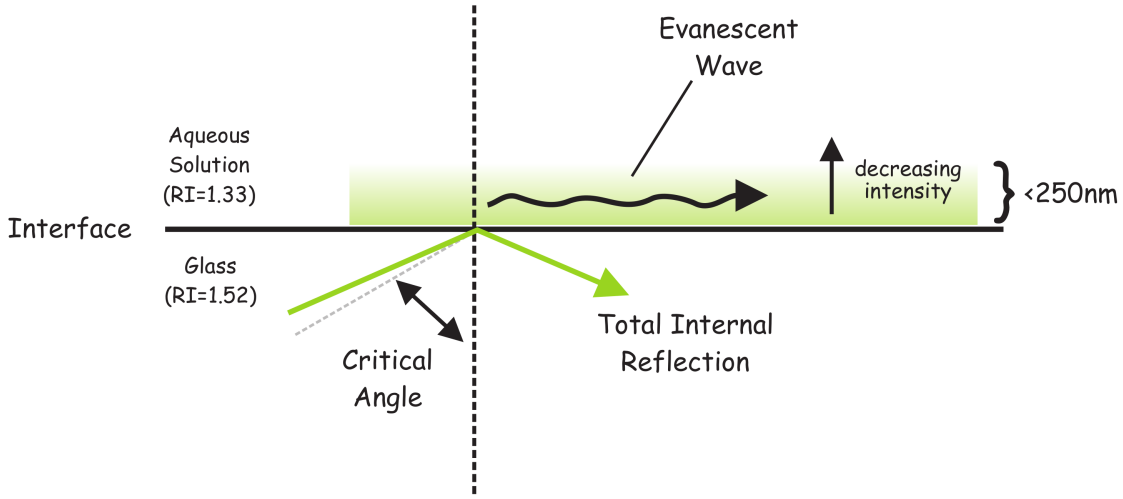


Figure 2.4 – **Principle of total reflection of light and evanescent wave.**

of images that show the evolution of the studied population in time, in the ROI. TICS uses the same temporal correlation as FCS, but for a 2D signal, that is, an image. As for FCS, the concentration of the observed population must be low.

2.2.2 Total Internal Reflection Fluorescence Microscopy

Total Internal Reflection Fluorescence Microscopy (TIRF), is used to capture images of fluorescently tagged molecules within a thin layer, with a depth of 70-250 nm, at the surface of the illuminated plane. This method can be used to illuminate large areas, such as an entire cell or even several cells, as it does not use a focused laser. TIRF imaging has very little background and minimum impact on the live cell. As the exposure of the cell(s) to light is minimal, it is possible to acquire long 2D movies. Therefore, it is an excellent method for getting images of near-surface single molecules and their displacements, populations of molecules and their dynamics, as well as distinct structures such as vesicles, granules, or contact surfaces between domains.

TIRFM uses the total reflection of light and the associated electromagnetic radiation, called evanescent wave, as the bases of the imaging method (see Fig. 2.4). When a light beam goes from a dense medium (high refractive index) to a less dense medium (low refractive index), and the angle between the beam and the contact surface (the incidence

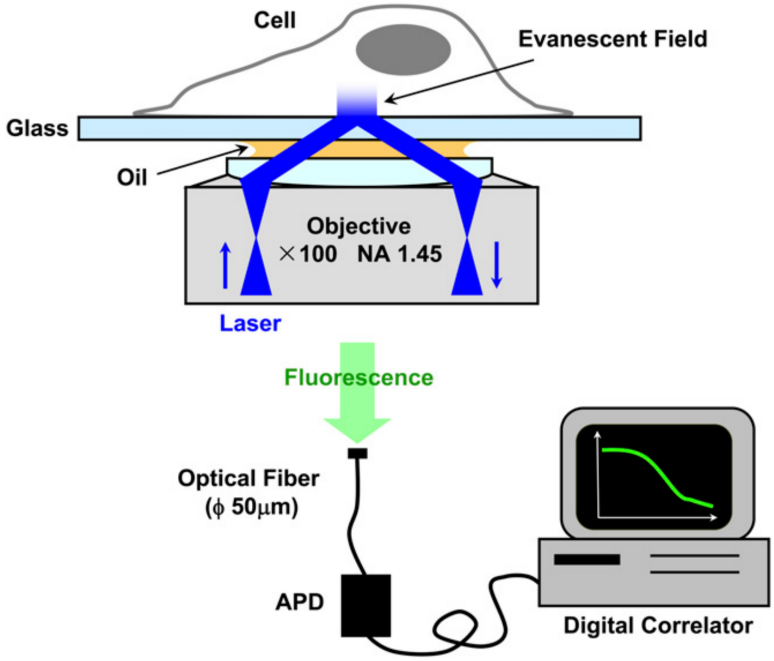


Figure 2.5 – **TIRF microscopy setup.** Adapted from [59].

angle) is greater than the so-called critical angle (angle at which the light reflects completely, and does not refract through the interface and into the low density medium), the light is totally reflected, so the light returns to the medium it came from, but a part of the energy of the light still goes into the second medium. This energy is a weak electromagnetic field at the interface between the two mediums that propagates into the second one with exponential decay in intensity with distance from the interface, and with the same frequency as the light beam: this field is called the evanescent wave. The depth of the evanescent wave is short, but sufficient to excite the fluorescent molecules that are near the surface; on the contrary, it is not sufficient to excite fluorescent molecules that are inside the cell, in the background of the illuminated field.

One possible setup of the TIRFM method, illustrated by Fig. 2.5, is as follows. The cell is placed on a glass, that is a medium denser than the cell, and an intermediate layer such as oil can be used to reduce even further the depth of the evanescent field. A laser beam points toward the glass, and the fluorescent signal emitted by the tagged molecules is focused and captured by an optical fiber, to be transformed further down the line into an electrical signal by an Avalanche photodiode (APD); this signal is recorded and pro-

cessed by a digital correlator.

In the case of the study of the late stage of exocytosis, TIRFM presents all the assets needed. The fact that the tagged proteins are excited by an evanescent wave, instead of a direct light beam, makes this method less prone to noise and background signal, and it does not affect the physiology or movement of the proteins as the excitation from the evanescent wave is weak. This enables long acquisition times, which makes it possible to study slow processes such as diffusion. Last but not least, the method is efficient with a high acquisition rate (50 – 100 ms/image), and its costs are low compared to other imaging methods.

Through TIRFM, it is possible to identify all important stages of exocytosis. Fluorescently tagged vesicles appear in TIRF as spots of light, more or less bright depending on their location/depth from the surface (i.e., the contact surface between the cell and the glass). The first phase, the transport phase, corresponds to a roughly linear movement of a spot in time, while the vesicle is being transported to the plasma membrane via cytoskeletal elements of the cell. When they arrive near the plasma membrane, vesicles can undergo attachment. This second phase of exocytosis is visible in TIRFM images as a restricted movement of the spot, that keeps a constant intensity throughout this phase. Several studies, though, show that stationary or restricted movement is not enough for the vesicle to undergo fusion after this phase. A regulatory step is probably involved, between the attachment and the fusion, such as the calcium concentration of the specific fusion site at the membrane. The last phase of exocytosis is the fusion to the plasma membrane. This phase is easily identified in TIRFM images as a sudden increase or burst in fluorescent intensity, thought to be due to higher excitation of the fluorophores at the exterior part of the cell than in the inside, or to the change in pH between the two mediums; the fluorescence of the tagged proteins is increased almost twofold, and this burst in intensity is followed by a spreading of the fluorescent signal as the proteins diffuse in a lateral manner in the plasma membrane. The study of this last phase in different cell types and under different conditions help better understand the fusion phase and its regulation. Thanks to TIRFM, several types of fusion were identified: from a “kiss and run” type where some of the cargo of the vesicle is released but the vesicle does not fuse to the membrane, to “mixed kiss and run” type where some of the cargo is released and a part of the vesicle fuses to the membrane, to a “full kiss and run” type where all cargo is released and

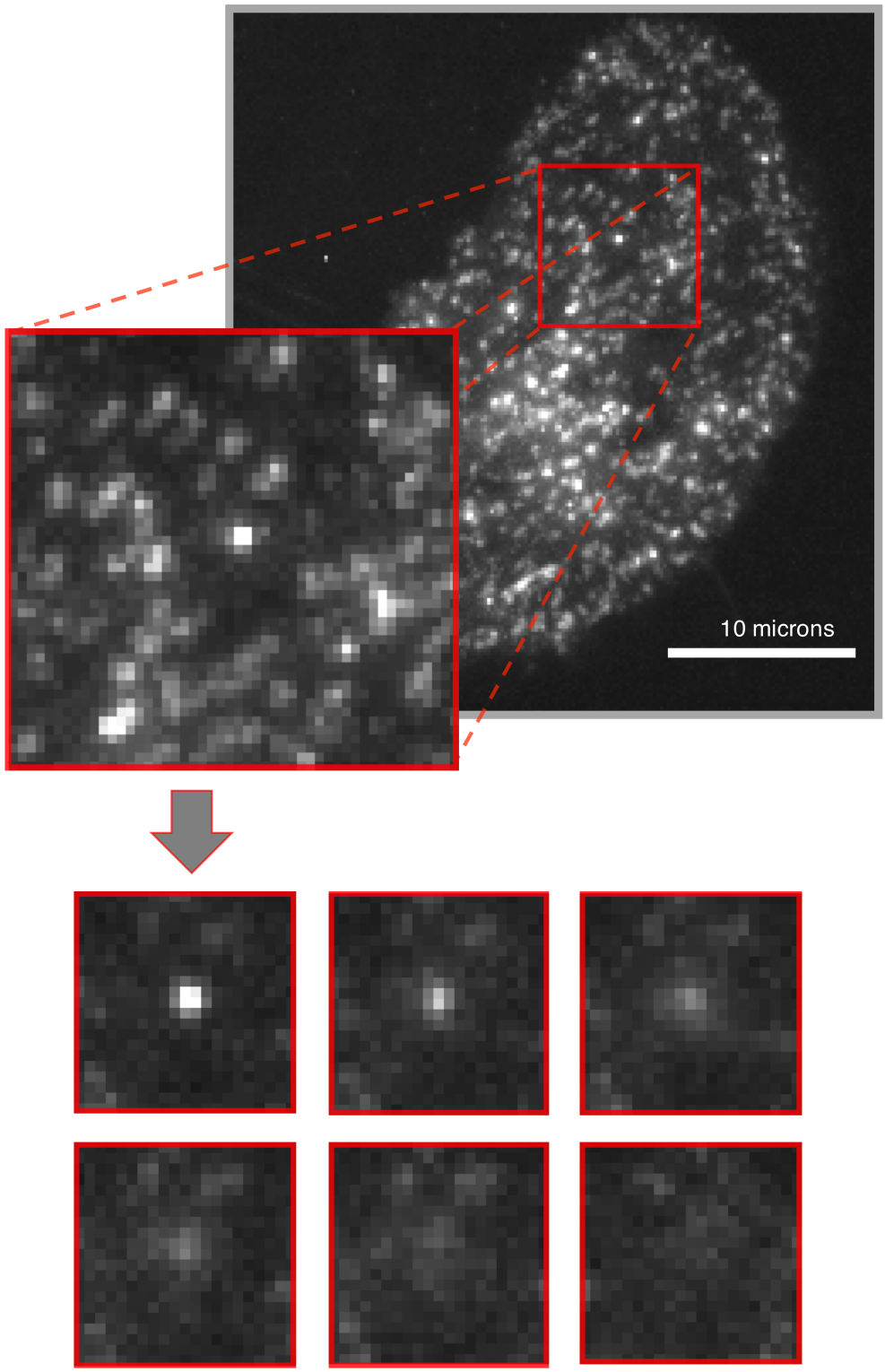


Figure 2.6 – **Illustration of a TIRFM sequence depicting TfR protein diffusion, as well as a selected diffusion event to be analyzed.**

the entire vesicle fuses to the membrane. Finally, for some cell types, when several vesicles arrive at the same fusion spot, they may fuse with each other and with the membrane.

The different types of fusion were qualitatively studied thanks to TIRF microscopy and the three main types were identified. However, quantitative methods for the local characterization of the different types of fusion from TIRFM imaging data are still lacking. Moreover, measures of local changes in the membrane can provide significant insights into the regulatory mechanisms involved in exocytosis. Here, we propose a quantitative method for the local characterization of diffusion during fusion in exocytosis, from standard TIRFM images.

2.3 Local diffusion estimation at the cell membrane

In this section, we first describe the mathematical model of local diffusion and propose a correlation-based method for diffusion estimation. Then, we analyze the behavior of the model in the space of possible parameter values. We validate the estimation method on data obtained by directly simulating the model. Finally, we show extensive results on synthetic images of local diffusion, before applying our method to real TIRFM images depicting Transferrin Receptor proteins (TfR) diffusion at the plasma membrane after vesicle fusion (see Fig. 2.6). The estimation method was developed in Matlab 9.4.

2.3.1 ABC-like method for local diffusion estimation in TIRF images

2.3.1.1 Model for spot diffusion

In TIRFM images, flourescently-tagged vesicles appear as spots of light, and fusion is identified by a burst in fluorescent intensity followed by a spreading of the fluorescent signal as proteins diffuse in the plasma membrane. In what follows, we describe the model for the fluorescent intensity of diffusing spots as proposed in [10], which will later be used to estimate local diffusion.

Let $\begin{cases} f : \Omega \times \mathbb{N} \to \mathbb{R} \\ \quad (p,t) \mapsto f(p,t) \end{cases}$ be the fluorescent intensity, with $\Omega \subset \mathbb{R}^2$ the ROI in the

image domain. The intensity at point $p = (x, y) \in \Omega$ and time t is noted $f(p, t)$.

The intensity f is assumed proportional (with factor B) to the convolution (denoted $*$) of the microscopic number density (concentration of molecules) C, and the instrumental point spread function (PSF) h:

$$f = B\,(C * h)\,, \tag{2.1}$$

where $B = \rho\epsilon Q$, ρ is the efficiency of the instrument to collect photons, ϵ is the molecular absorption coefficient and Q is the quantum yield of the fluorophore.

Diffusion of transmembrane proteins is mainly modeled by lateral diffusion in the membrane [6], and it can be described by Fick's second law [65]:

$$\frac{\partial C(p,t)}{\partial t} = D\Delta_p C(p,t)$$

where $C(p, t)$ denotes the concentration of molecules at time t and location $p \in \Omega \subset \mathbb{R}^2$, and Δ is the Laplacian operator. Here, the diffusion coefficient D is assumed to be constant over Ω.

In TIRFM images, a vesicle appears to be smaller than the pixel size before fusion. We thus assume that all tagged proteins are concentrated in $p_0 \in \Omega$, at initial time t_0, that is:

$$C(p, t_0) = C_0\delta(p - p_0) \tag{2.2}$$

where C_0 is the initial concentration of molecules and δ is the Kronecker symbol.

The closed-form solution to this partial derivative equation is [21]:

$$C(p,t) = \frac{C_0}{4\pi(t-t_0)D}\,\exp\left(\frac{\|p-p_0\|^2}{4(t-t_0)D}\right).$$

We obtain:

$$C(x,y,t) = \frac{C_0}{4\pi(t-t_0)D}\exp\left(-\frac{x^2+y^2}{4(t-t_0)D}\right),\ \forall\ \ t > t_0.$$

We consider that the PSF h is approximated by a Gaussian function with isotropic bandwidth σ_{PSF}, that is:

$$h(p) = \frac{1}{2\pi\sigma^2_{\text{PSF}}}\exp\left(-\frac{\|p\|^2}{2\sigma^2_{\text{PSF}}}\right).$$

This writes:

$$h(x,y) = \frac{1}{2\pi\sigma^2_{\text{PSF}}} \exp\left(-\frac{x^2+y^2}{2\sigma^2_{\text{PSF}}}\right).$$

One can then compute the fluorescent intensity using (2.1), that yields :

$$f(x,y,t) = B \times \frac{C_0}{2\pi\left(2(t-t_0)D + \sigma^2_{\text{PSF}}\right)} \times \exp\left(-\frac{x^2+y^2}{4(t-t_0)D + 2\sigma^2_{\text{PSF}}}\right). \tag{2.3}$$

Finally, the expression of the fluorescent intensity at spatial position p given an initial spot position located at p_0 is given by :

$$\boxed{f(p,t) = \frac{A_0}{2(t-t_0)D + \sigma^2_{\text{PSF}}} \times \exp\left(-\frac{\|p-p_0\|^2}{4(t-t_0)D + 2\sigma^2_{\text{PSF}}}\right)}. \tag{2.4}$$

with $A_0 = \dfrac{C_0 B}{2\pi}$.

2.3.1.2 Autocorrelation of the fluorescent intensity in TIRF images

In the TICS method, one calculates the temporal autocorrelation of the fluorescent intensity, which is further used to compute the concentration of the population and the characteristic time of diffusion. In this method, one assumes that the process associated with the image sequence is stationary. Or, considering (2.2), this assumption does not hold in our modeling framework, and we can not use the classical TICS formula. Here, we propose two new models for the temporal autocorrelation of the fluorescent intensity: G_1 for images with uniform background, and G_2 for non-uniform background. We use the model for the fluorescent intensity described above in conjunction with the TICS approach to establish our models. Our approach has two main advantages : it uses classical TIRF sequences so there is no need for new data acquisition, and the models allow for both local and global estimation of the apparent diffusion.

Model G_1 for uniform background

Here, the following expectation formula for the autocorrelation function is used:

$$G_1(t,\tau) = \frac{\langle\, f(t+\tau)f(t)\,\rangle}{\langle\, f\,\rangle_t^2}, \tag{2.5}$$

where $\langle \cdot \rangle$ and $\langle \cdot \rangle_t$ denotes the spatial and spatio-temporal averages respectively, and τ is the temporal lag.

Using the expression of the fluorescent intensity adapted to TIRF sequences, given by (2.4), with $t_0 = 0$ (without loss of generality), let us compute $G_1(t, \tau)$:

$$G_1(t,\tau) = \frac{\langle f(t+\tau)f(t) \rangle}{\langle f \rangle_t^2}$$

$$= \frac{\left\langle \frac{A_0^2}{(2Dt+\sigma_{\text{PSF}}^2)(2D(t+\tau)+\sigma_{\text{PSF}}^2)} \exp\left(-\frac{\|p-p_0\|_2^2}{4Dt+2\sigma_{\text{PSF}}^2} - \frac{\|p-p_0\|_2^2}{4D(t+\tau)+2\sigma_{\text{PSF}}^2}\right) \right\rangle}{\left\langle \frac{A_0}{2Dt+\sigma_{\text{PSF}}^2} \exp\left(-\frac{\|p-p_0\|_2^2}{4Dt+2\sigma_{\text{PSF}}^2}\right) \right\rangle^2}$$

Let $\sigma^2(t) = 2Dt + \sigma_{\text{PSF}}^2$. Then, we get

$$G_1(t,\tau) = \frac{\left\langle \frac{A_0^2}{\sigma^2(t)\sigma^2(t+\tau)} \exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(t)} - \frac{\|p-p_0\|_2^2}{2\sigma^2(t+\tau)}\right) \right\rangle}{\left\langle \frac{A_0}{\sigma^2(t)} \exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(t)}\right) \right\rangle^2}$$

$$= \frac{A_0^2}{\sigma^2(t)\sigma^2(t+\tau)} \frac{\sigma^4(t)}{A_0^2} \frac{\left\langle \exp\left(-\frac{\|p-p_0\|_2^2 (\sigma^2(t)+\sigma^2(t+\tau))}{2\sigma^2(t)\sigma^2(t+\tau)}\right) \right\rangle}{\left\langle \exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(t)}\right) \right\rangle^2}$$

$$= \frac{\sigma^2(t)}{\sigma^2(t+\tau)} \frac{\left\langle \exp\left(-\frac{\|p-p_0\|_2^2 (\sigma^2(t)+\sigma^2(t+\tau))}{2\sigma^2(t)\sigma^2(t+\tau)}\right) \right\rangle}{\left\langle \exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(t)}\right) \right\rangle^2}$$

$$G_1(t,\tau) = \frac{\sigma^2(t)}{\sigma^2(t+\tau)} \frac{\dfrac{1}{|\Omega|}\displaystyle\int_{p\in\Omega} \exp\left(-\frac{\|p-p_0\|_2^2\,(\sigma^2(t)+\sigma^2(t+\tau))}{2\sigma^2(t)\sigma^2(t+\tau)}\right) \mathrm{d}p}{\left[\dfrac{1}{|\Omega|}\displaystyle\int_{p\in\Omega} \exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(t)}\right) \mathrm{d}p\right]^2}$$

$$= |\Omega|\,\frac{\sigma^2(t)}{\sigma^2(t+\tau)} \frac{\displaystyle\int_{p\in\Omega} \exp\left(-\frac{\|p-p_0\|_2^2\,(\sigma^2(t)+\sigma^2(t+\tau))}{2\sigma^2(t)\sigma^2(t+\tau)}\right) \mathrm{d}p}{\left[\displaystyle\int_{p\in\Omega} \exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(t)}\right) \mathrm{d}p\right]^2}.$$

Moreover, the integral of the exponential on the domain Ω is approximately equal to the integral on $\mathbb{R}^2$ when p is far from p_0, which results in a Gauss integral. It follows that

$$\boxed{\begin{aligned} G_1(t,\tau) &\approx |\Omega|\,\frac{\sigma^2(t)}{\sigma^2(t+\tau)}\,\frac{2\pi\,\sigma^2(t)\sigma^2(t+\tau)}{\sigma^2(t)+\sigma^2(t+\tau)}\,\frac{1}{[\,2\pi\,\sigma^2(t)]^2}, \\ &\approx \frac{|\Omega|}{2\pi\,(\sigma^2(t)+\sigma^2(t+\tau))}, \\ G_1(t,\tau) &\approx \frac{|\Omega|}{4\pi\,(D\tau+2Dt+\sigma_{\mathrm{PSF}}^2)}. \end{aligned}} \tag{2.6}$$

Thus, for the process we are modeling, the autocorrelation function depends on both the time lag τ and time t.

In model G_1 (2.6), there are two parameters to estimate: the diffusion coefficient D and the variance σ_{PSF}^2 of the PSF. The parameter $|\Omega|$ depends on the data and its default value will be discussed in Section 2.3.2.

Model G_2 for non-uniform background

In the case of non-uniform background, the following expectation formula for the autocorrelation function is used:

$$G_2(t,\tau) = \frac{\langle\, \delta f(t)\delta f(t+\tau)\,\rangle}{\langle\, f\,\rangle_t^2}, \tag{2.7}$$

where $\delta f(t) = f(t) - \bar{f}_t$ is the fluorescent intensity fluctuation around the temporal average value, and $\bar{f}_t = \frac{1}{T-t}\int_t^T f(\theta)\,\mathrm{d}\theta$.

Let us now calculate the explicit formula for $G_2(\tau)$:

$$G_2(t,\tau) = \frac{\langle\, \delta f(t)\delta f(t+\tau)\,\rangle}{\langle\, f\,\rangle_t^2}$$

$$= \frac{\langle\, \left(f(t)-\bar{f}_t\right)\left(f(t+\tau)-\bar{f}_t\right)\,\rangle}{\langle\, f\,\rangle_t^2}$$

$$= \frac{\langle\, f(t)f(t+\tau)\,\rangle}{\langle\, f\,\rangle_t^2} - \frac{\langle\, f(t+\tau)\bar{f}_t\,\rangle}{\langle\, f\,\rangle_t^2} - \frac{\langle\, f(t)\bar{f}_t\,\rangle}{\langle\, f\,\rangle_t^2} + \frac{\langle\, \bar{f}_t^2\,\rangle}{\langle\, f\,\rangle_t^2}\,.$$

Let us now calculate each term of the right hand side of the previous equation. The first term was already explicitly calculated for the first model, so :

$$\frac{\langle\, f(t)f(t+\tau)\,\rangle}{\langle\, f\,\rangle_t^2} = G_1(t,\tau) \approx \frac{|\Omega|}{4\pi\left(D\tau + 2Dt + \sigma_{\text{PSF}}^2\right)}\,. \tag{2.8}$$

For the second term :

$$\langle\, f(t+\tau)\bar{f}_t\,\rangle = \left\langle\, \frac{A_0}{2D(t+\tau)+\sigma_{\text{PSF}}^2} \exp\left(-\frac{\|p-p_0\|_2^2}{4D(t+\tau)+2\sigma_{\text{PSF}}^2}\right) \frac{1}{T-t}\int_t^T f(\theta)\mathrm{d}\theta \,\right\rangle$$

$$\langle\, f(t+\tau)\bar{f}_t\,\rangle = \frac{A_0}{(2D(t+\tau)+\sigma_{\text{PSF}}^2)(T-t)}\frac{1}{|\Omega|}\int_{p\in\Omega}\exp\left(-\frac{\|p-p_0\|_2^2}{4D(t+\tau)+2\sigma_{\text{PSF}}^2}\right)\int_t^T f(\theta)\mathrm{d}\theta\,\mathrm{d}p$$

$$= \frac{A_0}{\sigma^2(t+\tau)(T-t)|\Omega|}\int_{p\in\Omega}\int_t^T \exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(t+\tau)}\right)\frac{A_0}{\sigma^2(\theta)}\exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(\theta)}\right)\mathrm{d}\theta\,\mathrm{d}p,$$

where $\sigma^2(t) = 2Dt + \sigma_{\text{PSF}}^2$. By inverting the two integrals, and using the Gauss integral, as done previously for the first model, as well as a change of variable at the end ($x = D(\theta + t + \sigma) + \sigma_{\text{PSF}}^2$, $\mathrm{d}x = D\mathrm{d}\theta$) one obtains :

$$\langle\, f(t+\tau)\bar{f}_t\,\rangle = \frac{\pi A_0^2}{D|\Omega|(T-t)}\ \ln\frac{D(t+T+\tau)+\sigma_{\text{PSF}}^2}{D(2t+\tau)+\sigma_{\text{PSF}}^2} \tag{2.9}$$

In the same way, one obtains :

$$\langle\, f(t)\bar{f}_t \,\rangle = \frac{\pi A_0^2}{D|\Omega|(T-t)}\ \ln\frac{D(t+T)+\sigma^2_{\text{PSF}}}{2Dt+\sigma^2_{\text{PSF}}}\,. \tag{2.10}$$

For the third term, we have

$$\begin{aligned}
\langle\, \bar{f}_t\ \bar{f}_t \,\rangle &= \left\langle\ \frac{1}{(T-t)^2}\int_t^T f(\theta_1)\mathrm{d}\theta_1 \int_t^T f(\theta_2)\mathrm{d}\theta_2\ \right\rangle \\
&= \frac{1}{|\Omega|(T-t)^2}\int_{p\in\Omega}\int_t^T\int_t^T f(\theta_1)f(\theta_2)\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}p \\
&= \frac{1}{|\Omega|(T-t)^2}\int_{p\in\Omega}\int_t^T\int_t^T \frac{A_0}{2D\theta_1+\sigma^2_{\text{PSF}}}\exp\left(-\frac{\|p-p_0\|_2^2}{4D\theta_1+2\sigma^2_{\text{PSF}}}\right)\frac{A_0}{2D\theta_2+\sigma^2_{\text{PSF}}}\exp\left(-\frac{\|p-p_0\|_2^2}{4D\theta_2+2\sigma^2_{\text{PSF}}}\right)\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}p \\
&= \frac{A_0^2}{|\Omega|(T-t)^2}\int_{p\in\Omega}\int_t^T\int_t^T \frac{1}{\sigma^2(\theta_1)}\exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(\theta_1)}\right)\frac{1}{\sigma^2(\theta_2)}\exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(\theta_2)}\right)\mathrm{d}\theta_1\mathrm{d}\theta_2\mathrm{d}p,
\end{aligned}$$

where $\sigma^2(\theta) = 2D\theta + \sigma^2_{\text{PSF}}$. As previously, by inverting the integrals, and using the Gauss integral, as well as a triple change of variable (first $x = D(\theta_1+\theta_2)+\sigma^2_{\text{PSF}}$, $\mathrm{d}x = D\mathrm{d}\theta_1$, then $y = D(t+\theta_2)+\sigma^2_{\text{PSF}}$, $\mathrm{d}y = D\mathrm{d}\theta_2$ and then $z = y + D(T-t)$, $\mathrm{d}z = \mathrm{d}y$) one obtains :

$$\langle\, \bar{f}_t\ \bar{f}_t \,\rangle = \frac{\pi A_0^2}{D^2|\Omega|(T-t)^2}\left[\left(2DT+\sigma^2_{\text{PSF}}\right)\ln\left(\frac{2DT+\sigma^2_{\text{PSF}}}{D(T+t)+\sigma^2_{\text{PSF}}}\right)+\left(2Dt+\sigma^2_{\text{PSF}}\right)\ln\left(\frac{2Dt+\sigma^2_{\text{PSF}}}{D(T+t)+\sigma^2_{\text{PSF}}}\right)\right]. \tag{2.11}$$

The last term is obtained as follows:

$$\begin{aligned}
\langle\, f\,\rangle_t^2 &= \left\langle\ \frac{A_0}{2Dt+\sigma^2_{\text{PSF}}}\exp\left(-\frac{\|p-p_0\|_2^2}{4Dt+2\sigma^2_{\text{PSF}}}\right)\ \right\rangle^2 \\
&= \left[\int_{p\in\Omega}\frac{A_0}{2Dt+\sigma^2_{\text{PSF}}}\exp\left(-\frac{\|p-p_0\|_2^2}{4Dt+2\sigma^2_{\text{PSF}}}\right)\mathrm{d}p\right]^2 \\
&= \left[\int_{p\in\Omega}\frac{A_0}{\sigma^2(t)}\exp\left(-\frac{\|p-p_0\|_2^2}{2\sigma^2(t)}\right)\mathrm{d}p\right]^2.
\end{aligned}$$

Again, the integral of the exponential is the Gauss integral, which yields :

$$\langle\, f\,\rangle_t^2 = \frac{4\pi^2 A_0^2}{|\Omega|^2}. \tag{2.12}$$

Finally, the autocorrelation model for non-uniform background is defined as:

$$G_2(t,\tau) = \frac{\langle f(t)f(t+\tau)\rangle}{\langle f\rangle_t^2} - \frac{\langle f(t+\tau)\bar{f}_t\rangle}{\langle f\rangle_t^2} - \frac{\langle f(t)\bar{f}_t\rangle}{\langle f\rangle_t^2} + \frac{\langle \bar{f}_t^2\rangle}{\langle f\rangle_t^2}$$

$$= G_1(\tau) - \frac{\frac{\pi A_0^2}{D|\Omega|(T-t)}\ \ln\frac{D(t+T+\tau)+\sigma_{\text{PSF}}^2}{D(2t+\tau)+\sigma_{\text{PSF}}^2}}{\frac{4\pi^2A_0^2}{|\Omega|^2}} - \frac{\frac{\pi A_0^2}{D|\Omega|(T-t)}\ \ln\frac{D(t+T)+\sigma_{\text{PSF}}^2}{2Dt+\sigma_{\text{PSF}}^2}}{\frac{4\pi^2A_0^2}{|\Omega|^2}}$$

$$+ \frac{\frac{\pi A_0^2}{D^2|\Omega|(T-t)^2}\left[(2DT+\sigma_{\text{PSF}}^2)\ln\left(\frac{2DT+\sigma_{\text{PSF}}^2}{D(T+t)+\sigma_{\text{PSF}}^2}\right)+(2Dt+\sigma_{\text{PSF}}^2)\ln\left(\frac{2Dt+\sigma_{\text{PSF}}^2}{D(T+t)+\sigma_{\text{PSF}}^2}\right)\right]}{\frac{4\pi^2A_0^2}{|\Omega|^2}}.$$

After all simplifications, one gets:

$$\boxed{\begin{aligned} G_2(t,\tau) = G_1(t,\tau) + |\Omega|\Bigg[&\frac{1}{4\pi D(T-t)}\ \ln\left(\frac{D\tau+2Dt+\sigma_{\text{PSF}}^2}{D\tau+D(T+t)+\sigma_{\text{PSF}}^2}\right)\\ &+\frac{2DT+\sigma_{\text{PSF}}^2}{4\pi D^2(T-t)^2}\ \ln\left(\frac{2DT+\sigma_{\text{PSF}}^2}{D(T+t)+\sigma_{\text{PSF}}^2}\right)+\frac{D(T+t)+\sigma_{\text{PSF}}^2}{4\pi D^2(T-t)^2}\ \ln\left(\frac{2Dt+\sigma_{\text{PSF}}^2}{D(T+t)+\sigma_{\text{PSF}}^2}\right)\Bigg]. \end{aligned}} \tag{2.13}$$

Let us define :

$$K_1(t,\tau) = \frac{1}{4\pi D(T-t)}\ \ln\left(\frac{D\tau+2Dt+\sigma_{\text{PSF}}^2}{D\tau+D(T+t)+\sigma_{\text{PSF}}^2}\right), \tag{2.14}$$

and

$$K_2(\tau) = \frac{2DT+\sigma_{\text{PSF}}^2}{4\pi D^2(T-t)^2}\ \ln\left(\frac{2DT+\sigma_{\text{PSF}}^2}{D(T+t)+\sigma_{\text{PSF}}^2}\right)+\frac{D(T+t)+\sigma_{\text{PSF}}^2}{4\pi D^2(T-t)^2}\ \ln\left(\frac{2Dt+\sigma_{\text{PSF}}^2}{D(T+t)+\sigma_{\text{PSF}}^2}\right). \tag{2.15}$$

Finally, we get

$$\boxed{G_2(t,\tau) = G_1(t,\tau) + |\Omega|\left(K_1(t,\tau)+K_2(t)\right).} \tag{2.16}$$

As in the case of uniform background, the autocorrelation function depends on both the

time lag τ and time t.
In model G_2 (2.13), the parameters to be estimated are: the diffusion coefficient D and the variance σ^2_{PSF} of the PSF. The parameters $|\Omega|$ and T depend on the data and their default values will be discussed in Section 2.3.2.

2.3.1.3 Local diffusion estimation

In our approach, a simple to implement version of the ABC method was used. Here, a simulation of the model is directly and quickly computed with formulas (2.6) or (2.13). For a typical diffusion event, the algorithm requires 100 000 evaluations of the model and takes about 1 minute for analyzing a single spot. The different steps of the method are the following ones :

Step 0 : Compute the autocorrelation of the fluorescent intensity in the ROI using the discrete form of (2.5) for model G_1, or (2.7) for model G_2, that is :

$$z_1(t,\tau)|_{t=0} = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \frac{f_{i,j}(t+\tau) f_{i,j}(t)}{\bar{f}_t^2}$$
$$z_2(t,\tau)|_{t=0} = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \frac{(f_{i,j}(t+\tau) - \bar{f})(f_{i,j}(t) - \bar{f})}{\bar{f}_t^2}$$

respectively, where (i,j) denotes a pixel in Ω, W and H are the width and height of the ROI ($|\Omega| = W \times H$), $\bar{f} = \frac{1}{T|\Omega|} \sum_{i=1}^{W} \sum_{j=1}^{H} \sum_{k=0}^{T-1} f_{i,j}(k)$, and $\bar{f}_t = \frac{1}{|\Omega|} \sum_{i=1}^{W} \sum_{j=1}^{H} f_{i,j}(t)$, $t \in \{0, ..., T-1\}$.
The following default values were used :

- ROI size : $|\Omega| = 20 \times 20$ pixels.
- Image length T : 50 images (or more) for both G_1 and G_2. Ideally, 300 images for G_2.

Step 1 : Generate $\theta = (D, \sigma^2_{\text{PSF}})$ from prior, where the parameters D and σ^2_{PSF} are generated independently.

- Prior distribution for both D and σ^2_{PSF}: Uniform $D \sim \mathcal{U}[0.1; 2]$ and $\sigma^2_{\text{PSF}} \sim \mathcal{U}[0.1; 2]$.
- Number of samples $\theta = (D, \sigma^2_{\text{PSF}})$ is 100000 samples.

Step 2 : Compute simulations and compare with the autocorrelation computed at Step 0.

- Compute simulations G_1 and G_2 for each of the 100000 samples.
- Compute the error (in terms of $L2$ distance) between G_1 and G_2 and the autocorrelation from the observed sequence.

Step 3 : Accept "the best" (in terms of $L2$ error) simulations and compute the estimates for parameters D and σ^2_{PSF}.

- Accept the best 1% of the simulations.
- Keep the corresponding parameters of the best 1% simulations, and compute the estimates $\tilde{\theta}_{\text{MAP}}$ and $\tilde{\theta}_{\text{MMSE}}$.

In practice, the estimates are computed as follows :

- $\tilde{\theta}_{\text{MMSE}} = \frac{1}{N_{samples}} \sum \tilde{\theta}$ where $N_{samples} = 100000$ and $\tilde{\theta}$ represents the accepted samples corresponding to the best 1% of the simulations;
- $\tilde{\theta}_{\text{MAP}} = \text{argmin}_\theta \ \text{error}L2(Autocorr_{obs}, G_{1/2})$.

2.3.2 Model Validation

The final objective of this part was the implementation and first validation of a proof of concept of the proposed method in Matlab. The results shown in the remainder of this Chapter were obtained in this context. A complete statistical analysis on large samples was out of the scope of this thesis, but would be required for the full validation of the method.

2.3.2.1 Analysis of the expected autocorrelation function

The aim of this section is to analyze the influence of the parameters on the expected autocorrelation function, for the two models derived in the previous section : Model G_1 is suitable for isolated, single diffusing spots, without background, while Model G_2 is suitable for diffusing spots with background and possible other diffusion spots in the region of interest.

We performed experiments with the following default parameters:

$$\begin{cases} D = 0.5\ pixels^2/frame \\ \sigma^2_{\text{PSF}} = 1\ pixels^2 \\ |\Omega| = 400\ pixels^2 \\ T = 50\ frames \end{cases} \tag{2.17}$$

where D is the diffusion coefficient, σ^2_{PSF} is the standard deviation for the PSF, $|\Omega|$ is the image size, and T is the length of the sequence.

The parameters of interest are D and σ^2_{PSF} for both models. Other parameters of interest are $|\Omega|$ and T that will be discussed when appropriate.

Expected autocorrelation function with model G_1 The influence of the parameters on the expected autocorrelaton function with model G_1 (2.6) is studied here.

The influence of the parameters D and σ^2_{PSF} on the autocorrelation function is first analyzed. Figure 2.7 shows the autocorrelation function G_1 when parameters D and σ^2_{PSF} are varying in the same time. 100 pairs of values $(D, \sigma^2_{\text{PSF}})$ were randomly chosen, with $D \in [0.1; 1.5]$ and $\sigma^2_{\text{PSF}} \in [0.5; 2]$. The starting value of G_1 is significantly influenced by the variation of the parameters, and it reaches higher values than in the case of a single varying parameter (see Figs. A.1a and A.1b). The slope of the the autocorrelation curve is also influenced in a significant way, and can be steeper than in the case of single varying parameter, meaning that the combined impact of the two parameters can lead to faster dynamics. Meanwhile, the end value of G_1 is barely influenced by the variation in these parameters.

Another parameter of interest in G_1 is the size $|\Omega|$ of each image. We tested the influence of the ROI size as follows: 100 values were randomly chosen between 50 and 2500, the other parameters being fixed to default values. Figure 2.8 shows the autocorrelation function when parameter $|\Omega|$ varies. The impact of $|\Omega|$ on G_1 is global, as expected from (2.6), and it acts as a stretching parameter. This makes sense, as it is a multiplicative factor in the right-hand side of (2.6): it amplifies the effects of D and σ^2_{PSF} in the same way. The starting values of G_1 are thus higher, and the dynamics is faster, for larger values of $|\Omega|$.

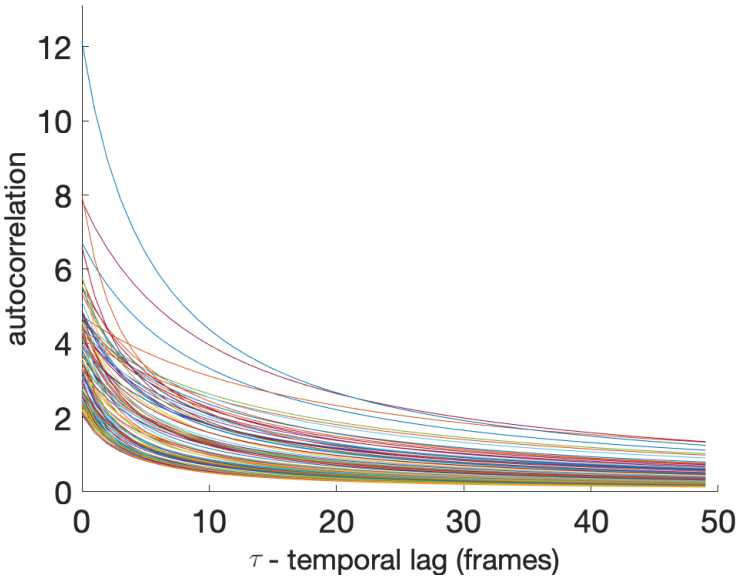


Figure 2.7 – **Influence of model parameters on the autocorrelation function for model G_1.** Each curve illustrates the simulation of model G_1 for a pair of values (D,σ^2_{PSF}), all other parameters being fixed to the default values.

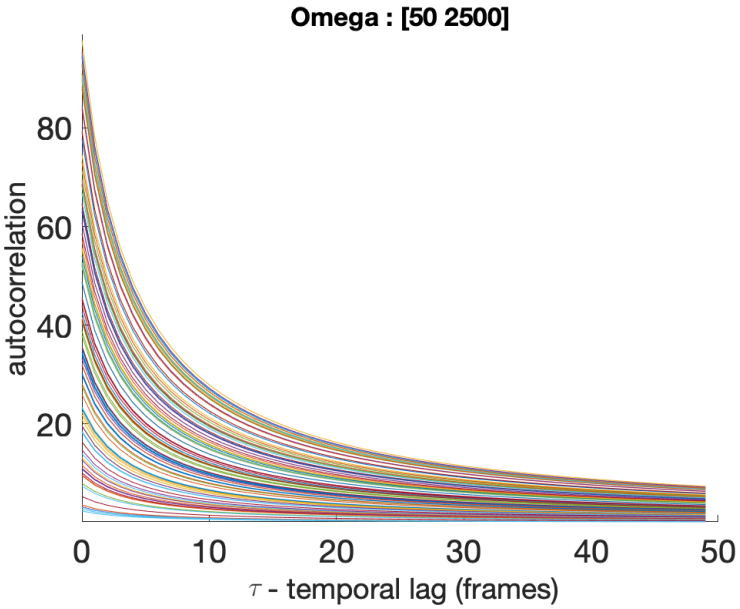


Figure 2.8 – **Influence of the parameter $|\Omega|$ on the autocorrelation function for model G_1.** Each curve illustrates the simulation of the model for a given value of $|\Omega|$, all other parameters being fixed to the default values.

Expected autocorrelation function with model G_2 We then analyzed the influence of the parameters on the expected autocorrelaton function with model G_2 (2.13).

As for model G_1, the joint influence of the parameters D and σ^2_{PSF} is first studied. Figure 2.9 shows the autocorrelation function when D and σ^2_{PSF} are jointly simulated. As previously, 100 pairs of values $(D, \sigma^2_{\text{PSF}})$ were randomly chosen, with $D \in [0.1; 1.5]$ and $\sigma^2_{\text{PSF}} \in [0.5; 2]$. The effects on model G_2 are different than on model G_1. While the starting value is still influenced by the joint variation of the parameters, the range of variation is smaller than for model G_1. Meanwhile, the influence on the ending value of G_2 is slightly higher. The slope of the autocorrelation curve varies mostly in the same way as for model G_1.
The influence of the ROI size $|\Omega|$ is shown in Fig. 2.10a. $|\Omega|$ has a similar influence on G_1 and G_2, it multiplies the effects of the variation of D and σ^2_{PSF}. The influence of the length T of the sequence, is shown in Fig. 2.10b. The value of T has an impact on the dynamics of G_2. This could affect the estimation of the parameters, in particular when T is small (less than 100). When T is large, i.e., close to 300, the impact is much weaker. In conclusion, T should be chosen large enough when using the model G_2 in order to reduce its influence on the estimation.

2.3.2.2 Validation of the parameter estimation method

In this section, we examine the performance of models G_1 and G_2. The proposed estimation method is applied to simulated data obtained directly from (2.6) and (2.13). We assess the parameter estimation error for each model. We first show the estimation results for model G_1, then for model G_2. The parameters of interest D and σ^2_{PSF} are jointly estimated while the other parameters are fixed to default values.

The default parameters are set as follows:

$$\begin{cases} |\Omega| = 400\ pixels^2 \\ T = 50\ frames \\ N_{samples} = 20000 \\ R = 0.01 \end{cases} \tag{2.18}$$

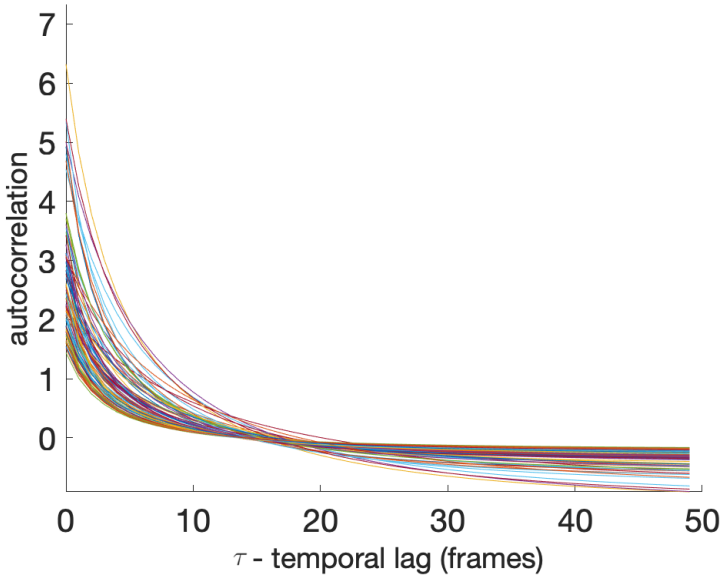
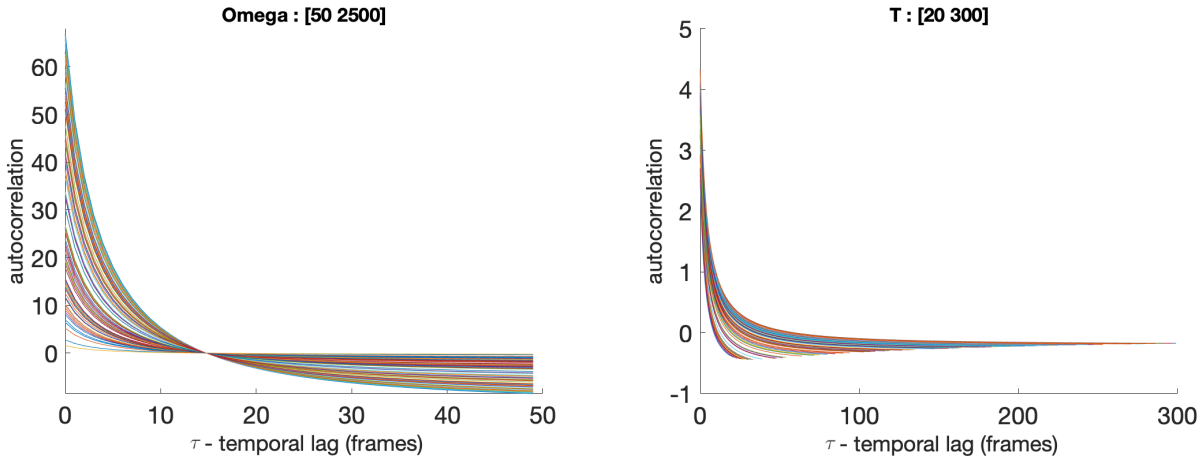


Figure 2.9 – **Influence of model parameters on the autocorrelation function for model G_2.** Each curve illustrates the simulation of model G_2 for a pair of values (D,σ^2_{PSF}), all other parameters being fixed to the default values.



(a) $|\Omega|$ (b) T

Figure 2.10 – **Influence of the parameters $|\Omega|$ and T on the autocorrelation function for model G_2.** Each curve illustrates the simulation of model G_2 for a given value of $|\Omega|$ a) or T (b), all other parameters being fixed to the default values.

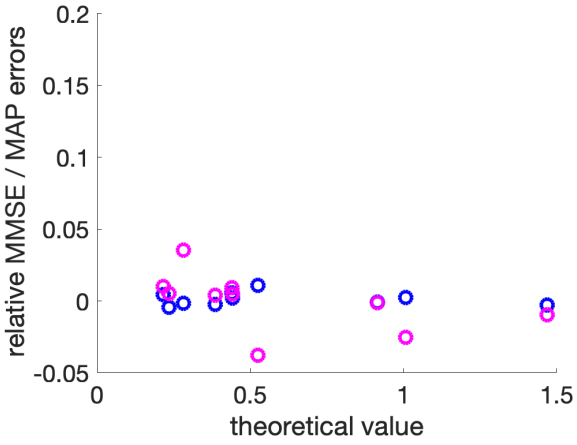
The parameter $N_{samples}$ corresponds to the number of samples generated by the method, from which the final solution is chosen. The parameter R corresponds to the fraction of samples that are accepted by the estimation method (here the best 1% of the samples are accepted).

Parameter estimation with model G_1 The parameter estimation is applied here to simulated data obtained from model G_1 (2.6).

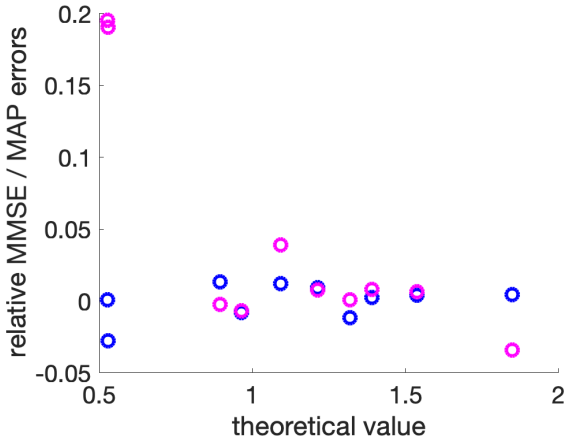
The joint estimation of the parameters was performed, and the results are shown in Fig. 2.11. The a priori distribution interval for D is set to $[0.1, 1.5]$ and the one for σ^2_{PSF} is set to $[0.5, 2]$.The estimation of the parameter D is accurate as shown in Fig. 2.11a, with relative errors on $\tilde{\theta}_{\text{MAP}}$ and $\tilde{\theta}_{\text{MMSE}}$ lower than 5%. The standard deviation is also small, as seen in Fig. 2.11c. The estimation of D is thus very accurate when both parameters are estimated together from model G_1. The results of the estimation of parameter σ^2_{PSF} are given in the right-hand side pictures in Figure 2.11. The estimation of σ^2_{PSF} is less accurate than the estimation of D for two values out of the ten tested values. This can be observed in Fig. 2.11b, where the MMSE relative error for the first two values is around 20%. The theoretical values for these two cases are on the edge of the a priori distribution interval, which seems to influence the quality of the estimation of this parameter. Nevertheless, the relative error on $\tilde{\theta}_{\text{MAP}}$ is in the same range as the other 8 estimations, that is, lower than 5%. The boxplots in Fig. 2.11d show that the standard deviations for the estimation of σ^2_{PSF} are higher than those of the estimation of D. The range is reasonable though, and the MMSE and MAP estimates are accurate, so the estimation of σ^2_{PSF} is overall reliable when both parameters are estimated from model G_1.

Parameter estimation with model G_2 The parameter estimation is applied here to simulated data obtained from model G_2 (2.13).

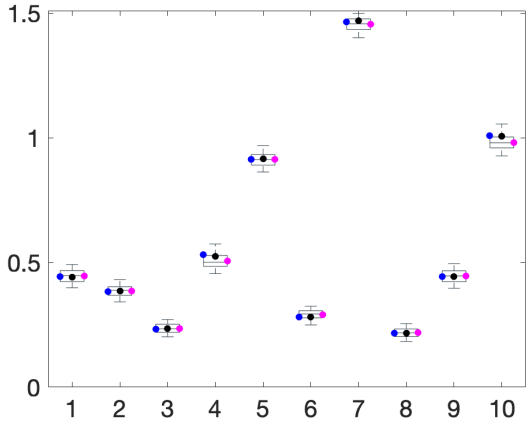
The estimation is performed in the same way as for model 1: the parameters D and σ^2_{PSF} are studied jointly. In what follows, the diffusion coefficient D ranges between 0.1 and 1.5 and the parameter characterizing the PSF, σ^2_{PSF}, ranges between 0.5 and 2. The results, shown in Fig. 2.12d, are similar to those obtained with model G_1. The estimation of D is slightly more accurate than the estimation of σ^2_{PSF}, with a maximum relative error of 5%



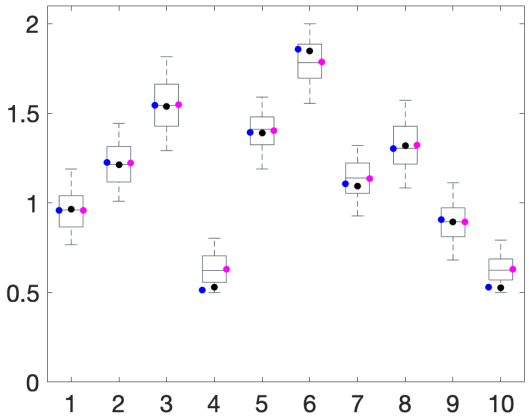
(a) D relative error estimation



(b) σ^2_{PSF} relative error estimation



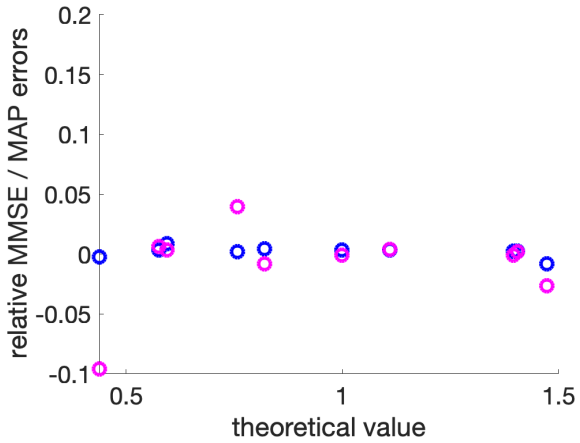
(c) D posterior distribution



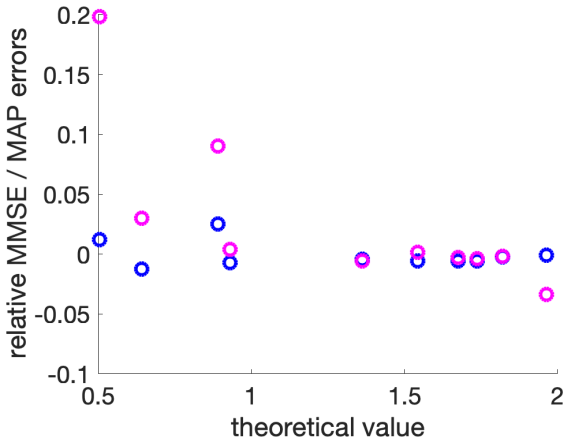
(d) σ^2_{PSF} posterior distribution

Figure 2.11 – **Estimation of both D and σ^2_{PSF} with model G_1.** (a,b) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the theoretical value of the parameter. (c,d) Estimated distribution of the parameters, for the ten tested cases. The true value of the parameters is represented by black circles.

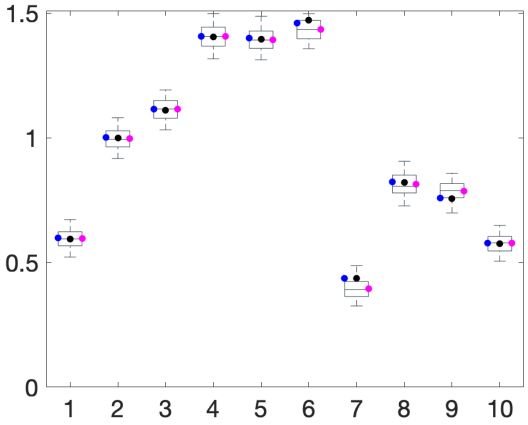
for D while, for σ^2_{PSF}, the relative error is about 5% for 8 tested cases and two cases yield MMSE relative errors of $10\% - 20\%$ (see Figs. 2.12a and 2.12b). The standard deviation for the estimation of D is slightly higher for model G_2 than for model G_1, whereas the standard deviation for the estimation of σ^2_{PSF} is slightly smaller for model G_2. Globally, the joint estimation of the parameters D and σ^2_{PSF} is accurate for all tested cases.



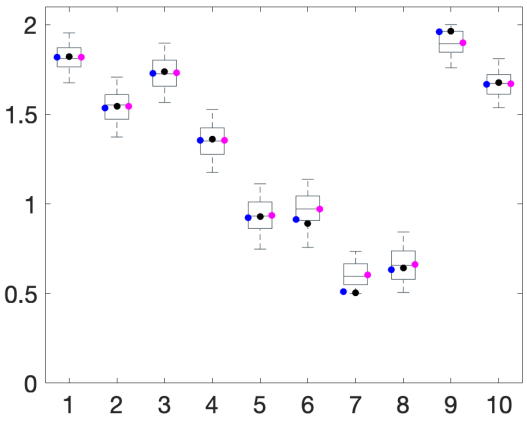
(a) D relative error estimation



(b) σ^2_{PSF} relative error estimation



(c) D posterior distribution



(d) σ^2_{PSF} posterior distribution

Figure 2.12 – **Estimation of both D and σ^2_{PSF} with model G_2.** (a,b) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the theoretical value of the parameter. (c,d) Estimated distribution of the parameters, for the ten tested cases. The true value of the parameters is represented by black circles.

2.3.2.3 Influence of parameters $|\Omega|$ and T

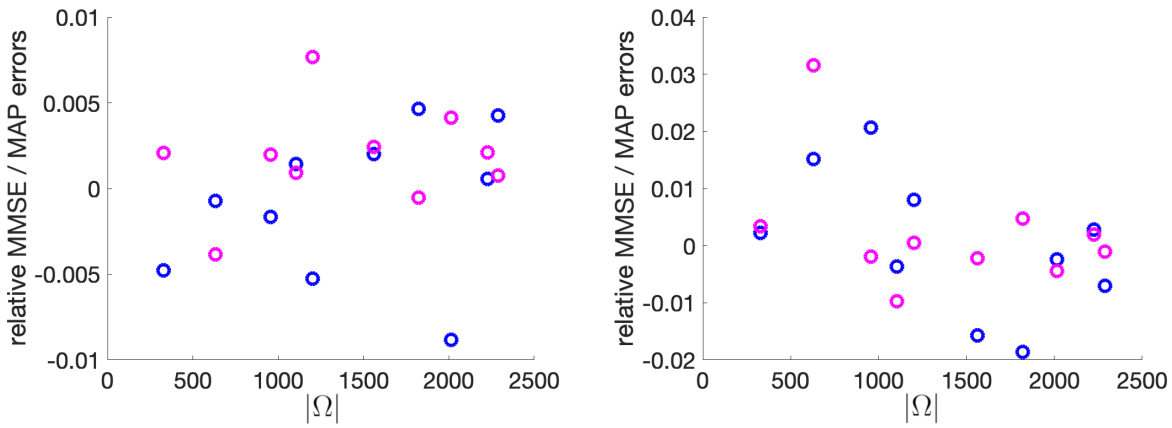
In this section, the influence of the parameters $|\Omega|$ and T is studied. For this purpose, joint estimations of parameters D and σ^2_{PSF} are performed for models G_1 and G_2, while the parameters $|\Omega|$ and T vary in a given range. For the parameters $N_{samples}$ and R, that could influence the estimation indirectly, see Appendices Figs. (C.3), (C.4), (C.7) and (C.8). Values for parameter $|\Omega|$ are chosen randomly in the range $[50; 2500]$, and the values for parameter T are chosen in the range $[20; 300]$.

Influence of the parameter $|\Omega|$ on the estimation with model G_1 The parameter estimation is applied here to simulated data obtained from model G_1 (2.6), while varying the parameter $|\Omega|$. The other parameters are set to default values.

The results of the joint estimation of the parameters D and σ^2_{PSF}, with varying parameter $|\Omega|$, are shown in Fig. 2.13. The MMSE and MAP estimates of both parameters are accurate, with a maximum relative error of 1% for D and 4% for σ^2_{PSF} as shown in Figs. 2.13a and 2.13b depincting plots of the relative errors of the MMSE and MAP estimates with respect to the value of $|\Omega|$. The standard deviation for the estimation of D is very small, and the standard deviation for the estimation of σ^2_{PSF} is also quite small, as shown in Fig. 2.14. From these results, variations of the parameter $|\Omega|$ do not seem to have a specific influence on parameter estimation. Thus, the ROI size can be chosen as convenient, as long as there is no information added or subtracted from the image by varying the ROI size, i.e., the background stays the same.

Influence of the parameters $|\Omega|$ and T on the estimation with model G_2 The parameter estimation is applied here to simulated data obtained from model G_2 (2.13), while varying the parameter $|\Omega|$, then T. The other parameters are set to default values.

The estimation of the parameters D and σ^2_{PSF} with varying parameter $|\Omega|$ was performed in the same way as for model G_1. Figure 2.15 illustrates the results. The MMSE and MAP estimates are accurate for most of the tested cases. In Figs. 2.15a and 2.15b, the relative errors for D are lower than 4%, while relative errors for σ^2_{PSF} are lower than 2% in 9 out the the 10 tested cases; only one case has a relative error of 10% for the MMSE estimate of σ^2_{PSF}. The standard deviation is relatively low for both D and σ^2_{PSF}, and they are relatively constant for all tested cases. The estimation of both parameters for this model is thus accurate even when varying the size of the image. As with model G_1, the size of the image can be chosen as convenient.

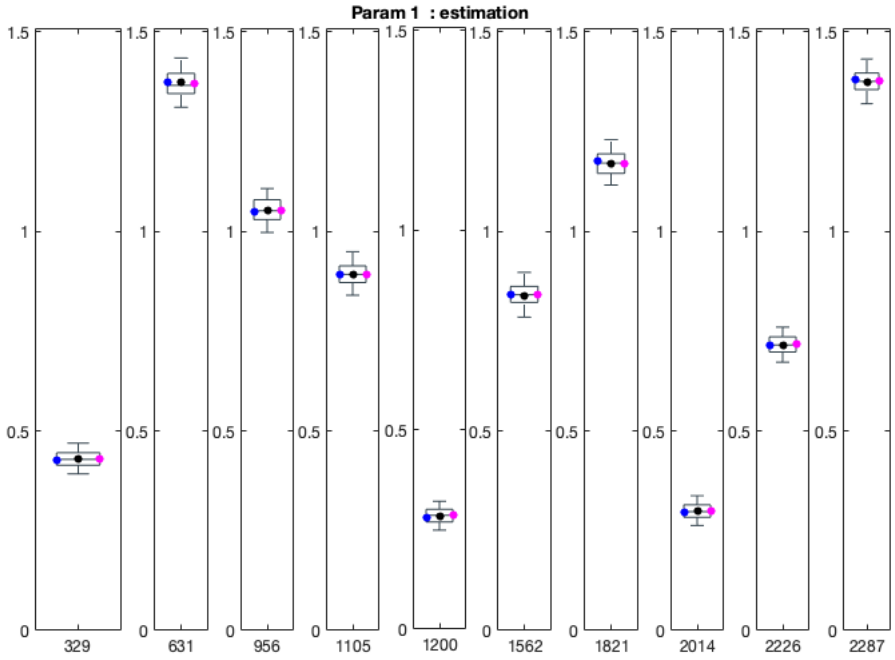


(a) D relative error estimation

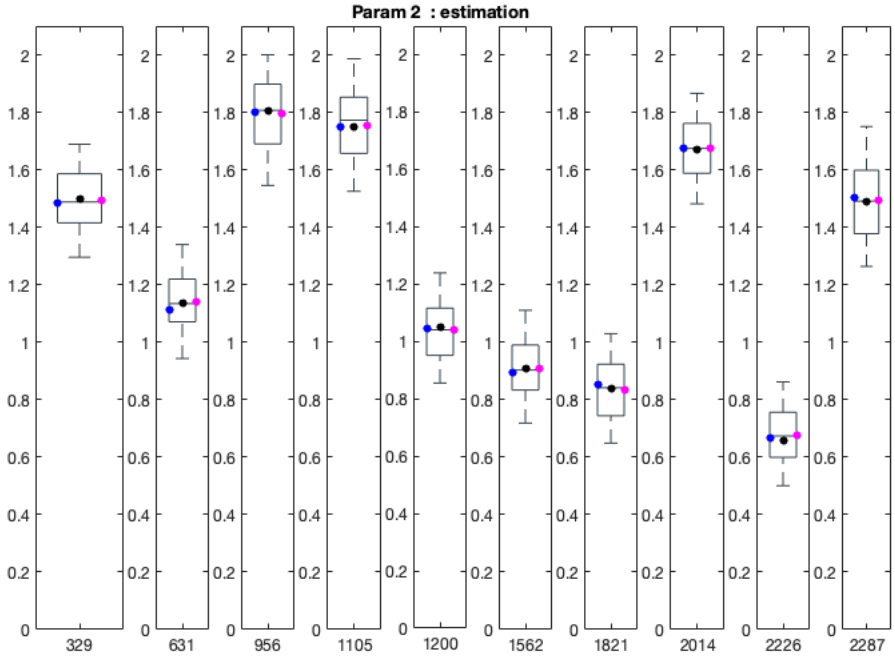
(b) σ^2_{PSF} relative error estimation

Figure 2.13 – **Influence of $|\Omega|$ on parameter estimation with model G_1.** (a,b) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the value of $|\Omega|$.

The estimation of the parameters D and σ^2_{PSF} with varying parameter T was then performed. In Fig. 2.18, the MMSE and MAP estimates are accurate for most of the tested cases (8 out of 10) for both parameters. The MMSE relative error for D is lower than 5% for 8 cases out of 10 and around 10% for the 2 remaining cases, as shown in Fig. 2.16a. For parameter σ^2_{PSF} the MMSE relative error is also lower than 5% for 8 out of 10 cases, and close to 10% and 20%, respectively, for the two remaining cases, as illustrated in Fig. 2.16b. Figure 2.18a shows the estimator for the parameter D for each tested value of T. The standard deviation of the estimator is negatively correlated to T, with a Pearson's linear correlation coefficient of about -0.74 and a corresponding p-value around 0.015. However, the standard deviation is relatively small, even when T is small. Figure 2.18b shows the estimator for σ^2_{PSF} for the 10 tested values of T. Here, there does not seem to be any correlation between the value of T and the standard deviation. As with the previous results, the standard deviation for σ^2_{PSF} is slightly higher than the one for D. Some caution is thus needed when choosing the value of T in order to have the best estimation for D. Globally, the results are accurate for all tested cases.

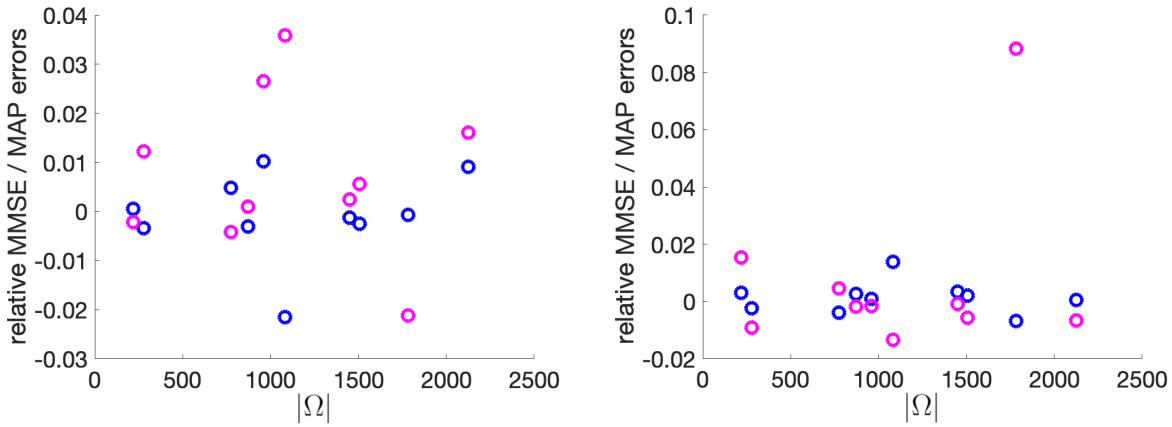


(a) D posterior distribution



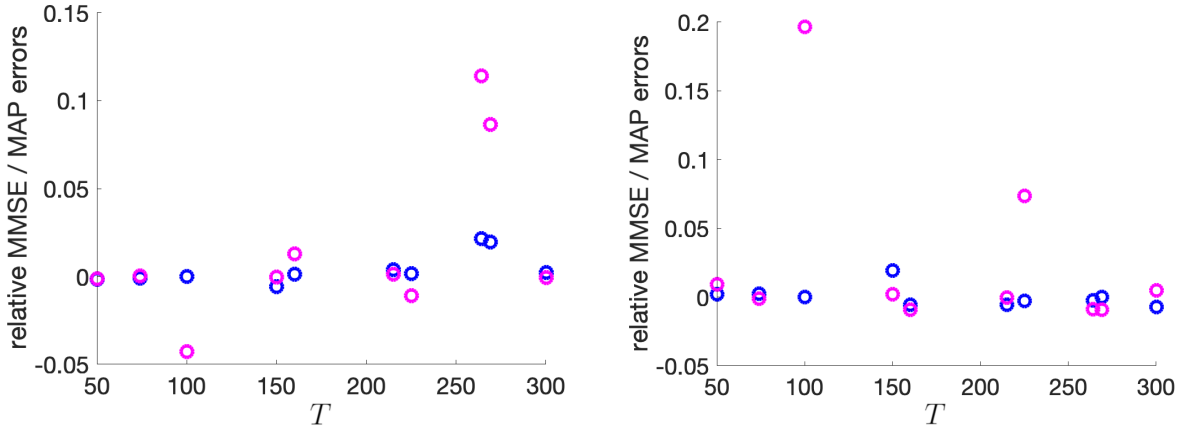
(b) σ^2_{PSF} posterior distribution

Figure 2.14 – **Influence of $|\Omega|$ on parameter estimation with model G_1.** (a,b) Estimated distribution of parameters D and σ^2_{PSF}, for the ten tested cases corresponding to ten values of $|\Omega|$. The true values of D and σ^2_{PSF} are represented by black circles.



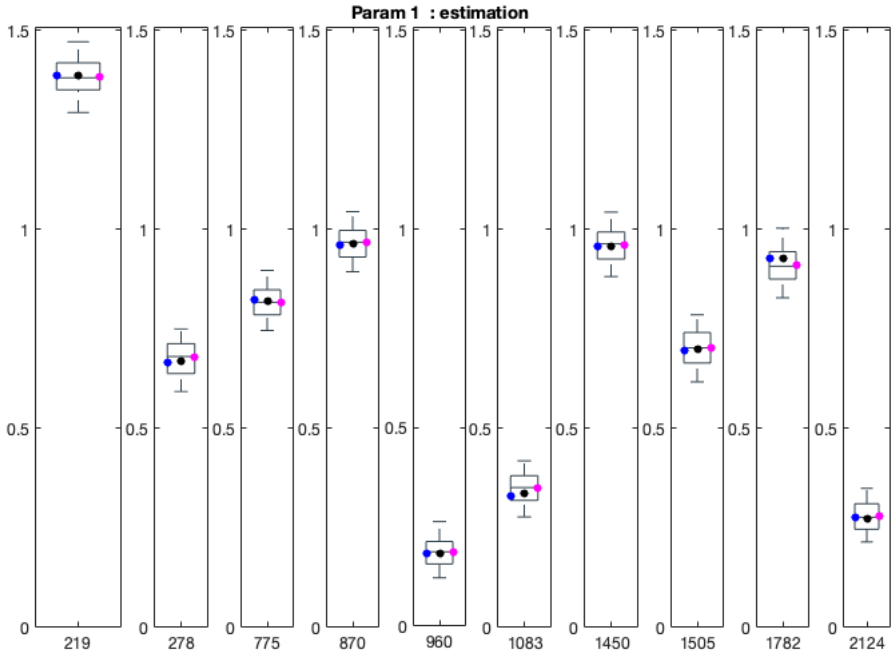
(a) D relative error estimation (b) σ^2_{PSF} relative error estimation

Figure 2.15 – **Influence of $|\Omega|$ on parameter estimation with model G_2.** (a,b) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the value of $|\Omega|$.

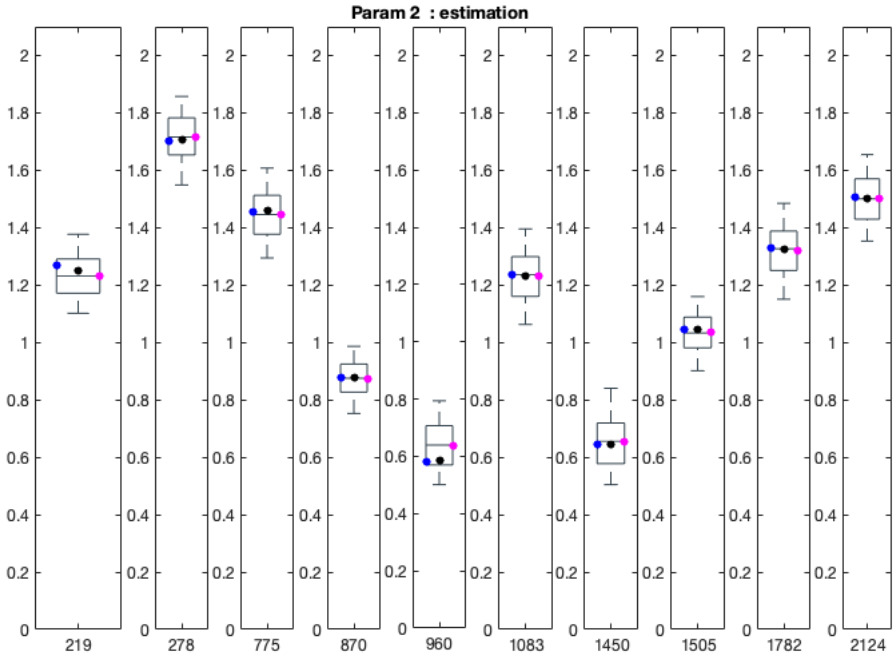


(a) D relative error estimation (b) σ^2_{PSF} relative error estimation

Figure 2.16 – **Influence of T on parameter estimation with model G_2.** (a,b) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the value of T.

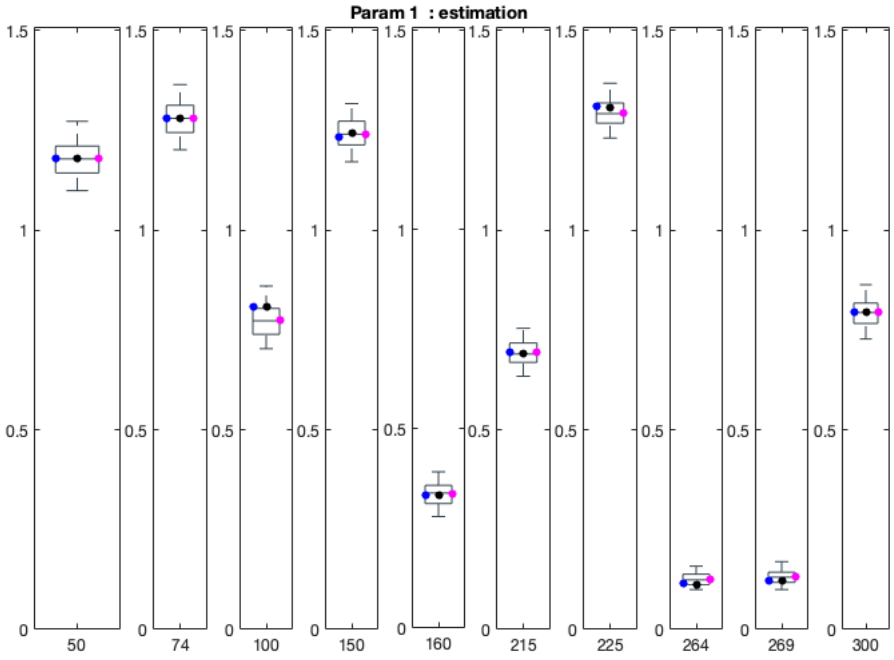


(a) D posterior distribution

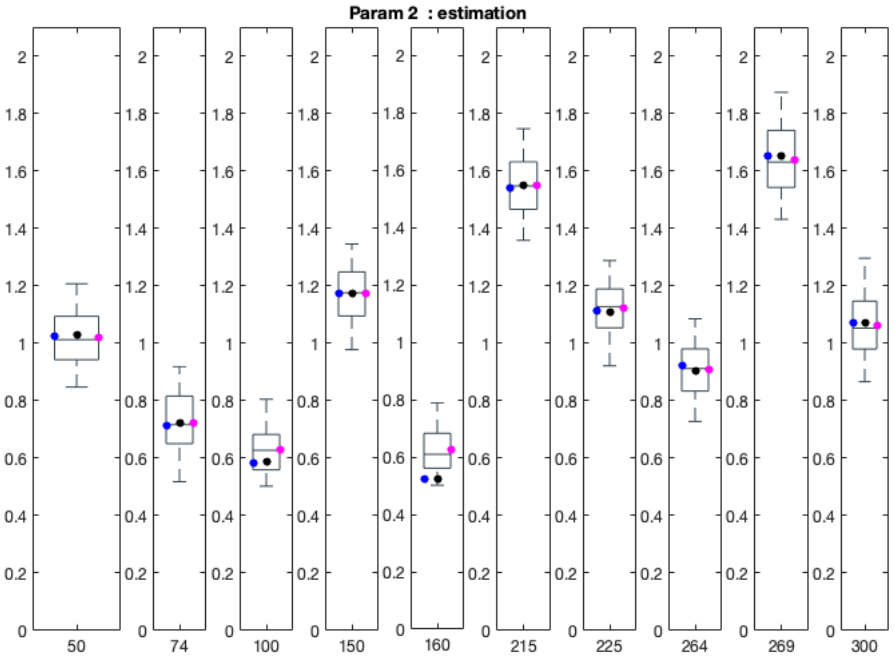


(b) σ^2_{PSF} posterior distribution

Figure 2.17 – **Influence of $|\Omega|$ on parameter estimation with model G_2.** (a,b) Estimated distribution of parameters D and σ^2_{PSF}, for the ten tested cases corresponding to ten values of $|\Omega|$. The true values of D and σ^2_{PSF} are represented by black circles.



(a) D posterior distribution



(b) σ^2_{PSF} posterior distribution

Figure 2.18 – **Influence of T on parameter estimation with model G_2.** (a,b) Estimated distribution of parameters D and σ^2_{PSF}, for the ten tested cases corresponding to ten values of T. The true values of D and σ^2_{PSF} are represented by black circles.

2.3.3 Results on synthetic data

In this section, we assess our method on synthetic images, that is, simulated images of real data. In particular, we estimate the model parameters, i.e., the diffusion coefficient and the standard deviation for the PSF, on synthetic images depicting diffusing spots, and in different scenarios that could challenge our method. The previous section gave us insights on how the model parameters influence the model, and how varying these parameters impacts the estimation. From this knowledge, a step further can be taken by testing the method on synthetic images that contain spots simulating a vesicle that diffuses according to the hypothesis of pure diffusion from an initial point of concentration.

2.3.3.1 Presentation of the synthetic image dataset

The synthetic images were generated by using the method from [10]. A sequence of images containing several diffusing spots was obtained, then corrupted with Gaussian noise with 6 different levels (yielding 6 *general sequences* with different signal-to-noise ratios). Each sequence contains 300 frames, and each image in the sequence is of size 256×256 pixels. Diffusing spots appear and disappear at different times of the sequence. Each diffusing spot is concentrated in one pixel at the first frame, and the values of the parameters are $D = 0.10$ and $\sigma^2_{\text{PSF}} = 1$.

Four scenarios were considered:

Scenario #1: A set of sequences, each containing a diffusing spot, centered in the image, with the same image size. All tested spots are isolated.

Scenario #2: A set of sequences, starting from the same diffusing spot, with different image sizes. Additional spots might be present in the image for large ROI sizes.

Scenario #3: A set of sequences, starting from the same diffusing spot. The spot is located in different places in the image. Additional spots might also be present in the image.

Scenario #4: A set of sequences, starting from the same diffusing spot, centered in the image, with noise levels that vary between sequences. In all sequences, the spot is isolated.

Across these scenarios, both models G_1 and G_2 are tested and the joint estimation of the model parameters is performed. Assessing our method on synthetic images gives us

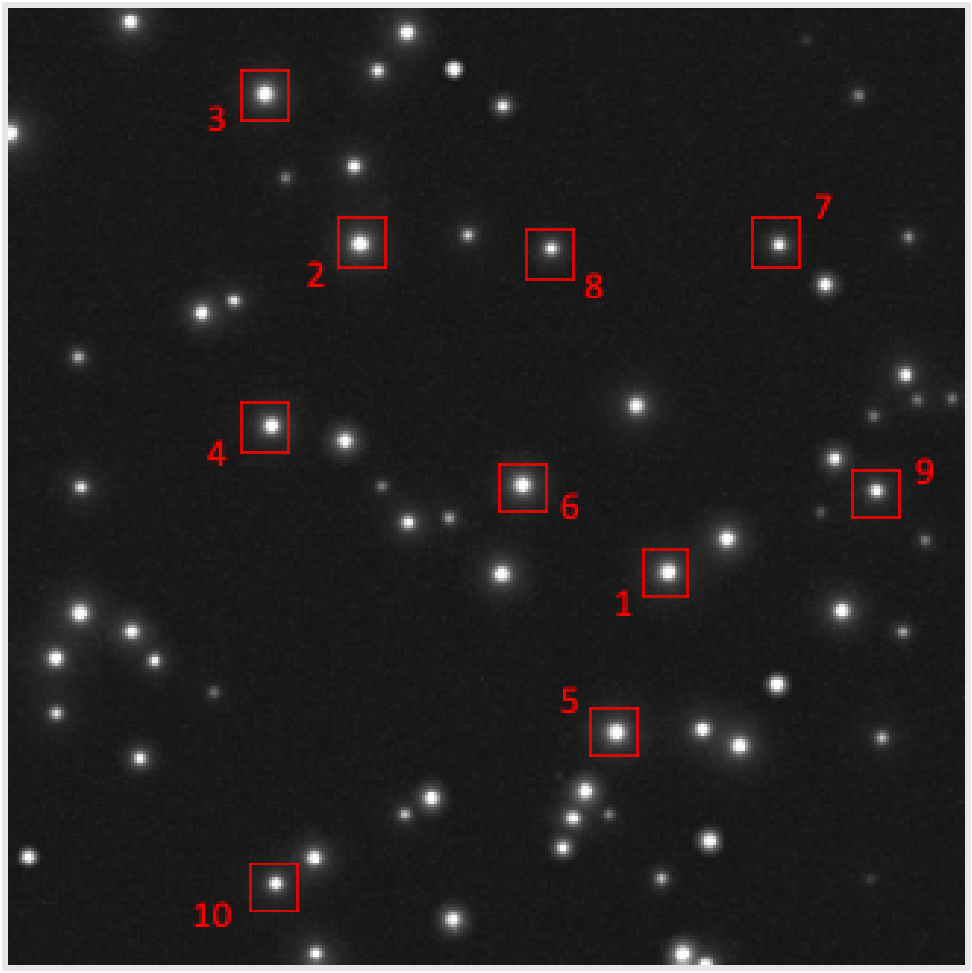


Figure 2.19 – **Synthetic images depicting selected spots for sparse estimation.**

more information on the accuracy of the method in conditions close to those observed with real microscopy acquisition images, on its range of applicability, and on its possible limitations.

2.3.3.2 Parameter estimation on synthetic images

For each tested scenario, the different diffusing spots were selected from the 6 general sequences generated. The model parameters are known: $D = 0.10$ and $\sigma^2_{\text{PSF}} = 1$. The parameters $|\Omega|$ and T, related to the image sequence, vary among the different scenarios. The parameters related to the estimation are fixed to the following values: $N_{samples} = 20000$, $R = 0.01$.

Scenario #1: Sparse estimation on synthetic images In scenario #1, 10 diffusing spots were selected and extracted from a general synthetic sequence with a low noise level. A typical image sequence depicting spots is shown in Fig. 2.19. Each spot is close to the center of the image, and the patch size is 24x22 pixels. For model G_1, the length of the sequence is $T = 100$ frames; for model G_2, the length T is the maximum length possible, from the frame the spot appears up to frame 300. For models G_1 and G_2, the joint estimation of the parameters D and σ^2_{PSF} was performed.

The results of the estimation for all 10 sequences are reported in Fig. 2.21 for model G_1, and Fig. 2.23 for model G_2. Each figure shows four results for a given diffusing spot (from left to right) :

- Autocorrelation computed from the image sequence (in black) together with the kept simulations that form the proposed solution (in green);
- Autocorrelation computed from the image sequence (in black) together with the simulations corresponding to the MMSE and MAP estimates (in magenta and green, respectively);
- Histogram of the estimator for parameter D;
- Histogram of the estimator for parameter σ^2_{PSF}.

The histograms are a visual representation of the estimated distributions for each parameter, which are extracted from the joint estimated distribution of $(D, \sigma^2_{\text{PSF}})$. They provide a general idea of the estimated distribution. The following description of the results also relies on the punctual estimates, which are computed using the algorithm and the formulas given in section 2.3.1.3.

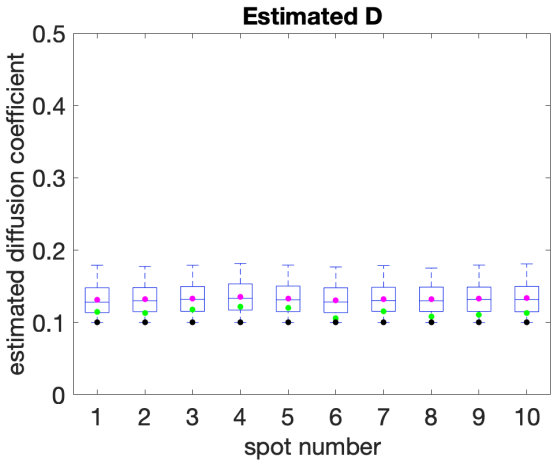
For model G_1, the fitting is very good for all ten tested spots. In particular, both simulations corresponding to the MMSE and MAP estimates fit almost perfectly the curve of the computed autocorrelation. The theoretical value for the parameter D is 0.1. Here, the range of D is $[0.1; 1.5]$. Note that this scenario can be considered an edge case, as the theoretical value is an endpoint of the interval. As such a situation is expectedly hard to handle, we assess the quality of the estimation with respect to both the relative estimation error and the range of the a priori distribution of the parameter.

The distributions for the parameter D for the ten spots is asymmetric, with a MMSE around $[0.12; 0.13]$, a minimum value of 0.1 and a maximum value of 0.18. The theoretical

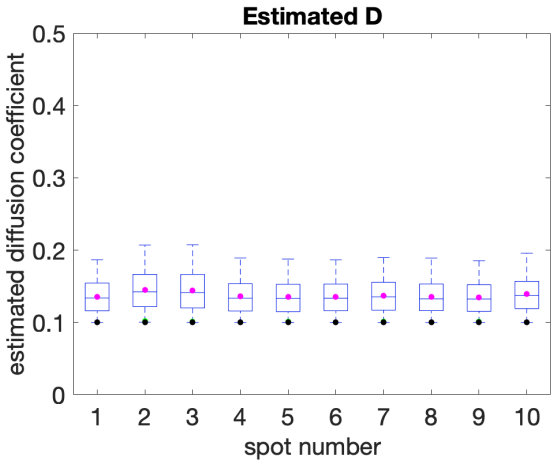
value of σ^2_{PSF} is 1. The histograms show a distribution with a MMSE value between 0.9 and 1, with a minimum value of 0.7 and a maximum value of 1.4. The distribution of σ^2_{PSF} is also asymmetric, which might mean that the distribution of D influences the global form of the distribution of σ^2_{PSF}.

For model G_2, the quality of the fitting is not as good as with model G_1. As explained in the previous section, the parameter corresponding to the length of the sequence can hinder the estimation for this model. While the quality of the fitting is very good at the beginning of the curve, it then deteriorates. Nevertheless, the quality seems to improve when getting from 100 frames to 300 frames. The estimation of the parameter D is similar to the one for model G_1, with a mean around $[0.12; 0.13]$, a minimum value of 0.1 and a maximum value of 0.2. As for the estimation of the parameter σ^2_{PSF}, it slightly differs from that of model G_1. The mean value is in the range $[0.7; 1]$, with a minimum value of 0.6 and a maximum value of 1.2. Thus, σ^2_{PSF} seems to be slightly underestimated for some spots.

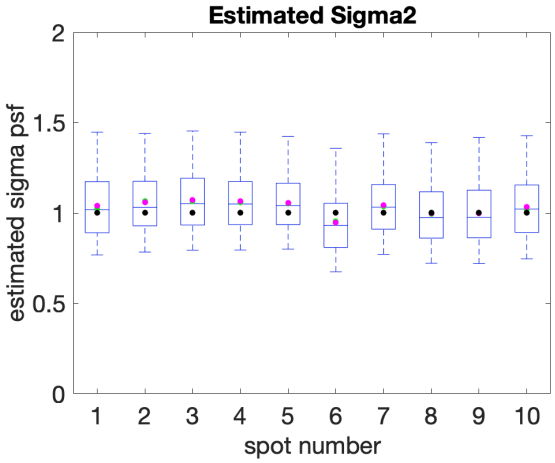
For a better comparison, Fig. 2.20 shows the boxplots of the estimators of both parameters, for model G_1 and model G_2. Each figure displays the ten results for the ten tested spots. On the one hand, the estimation of D with model G_1 is very consistent, the mean values and standard deviations being almost identical for all spots (see Fig. 2.20a), whereas there are some discrepancies between the estimators with model G_2 (see Fig. 2.20b); in particular, the mean value and the standard deviation vary more between spots with model G_2. On the other hand, the MAP estimate with model G_2 is always equal to 0.1, the real value of D, which is not the case for model G_1. A similar observation can be made for the estimation of parameter σ^2_{PSF} : the estimation seems more consistent with model G_1, especially in terms of standard deviation, whereas, with model G_2, the estimation slightly varies between spots (see Fig. 2.20c and 2.20d). In addition, model G_2 seems to consistently underestimate the value of σ^2_{PSF}. All in all, the estimation of both D and σ^2_{PSF} for this edge case is satisfactory with both models. These results are promising in terms of the robustness of the estimation method.



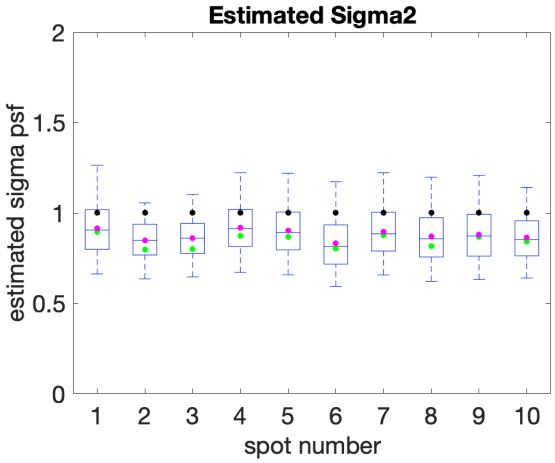
(a) Model G_1 - D estimation



(b) Model G_2 - D estimation



(c) Model G_1 - σ^2_{PSF} estimation



(d) Model G_2 - σ^2_{PSF} estimation

Figure 2.20 – **Sparse estimation of the model parameters with both models on synthetic images.** (a,b) Estimated distribution of parameter D with model G_1 (a) and model G_2 (b) for the ten selected spots. (c,d) Estimated distribution of parameter σ^2_{PSF} with model G_1 (c) and model G_2 (d) for the ten selected spots. The theoretical value is represented by a black dot, the MAP estimate by a green dot and the MMSE estimate by a magenta dot.

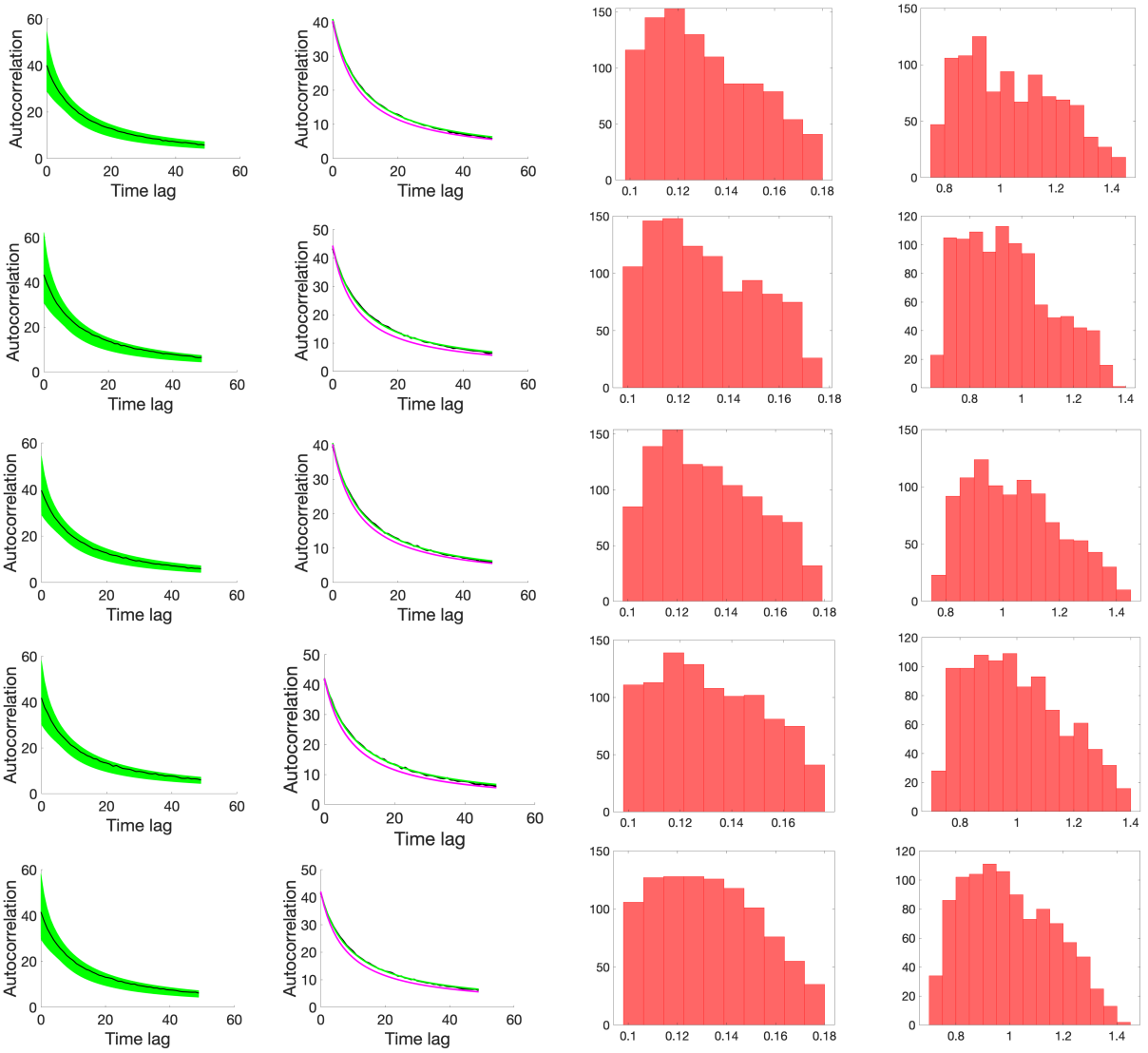


Figure 2.21 – **Sparse estimation for model G_1 on synthetic images, for the first five tested cased.** The first column gives the computed autocorrelation (in black), and the simulation results for the solution proposed by our method (in green). The second column gives the computed autocorrelation (in black), the simulation corresponding to the MAP estimate (in green) and the simulation corresponding to the MMSE estimate (in magenta). The third column gives the histogram of the estimated distribution of D. The fourth column gives the histogram of the estimated distribution of σ^2_{PSF}.

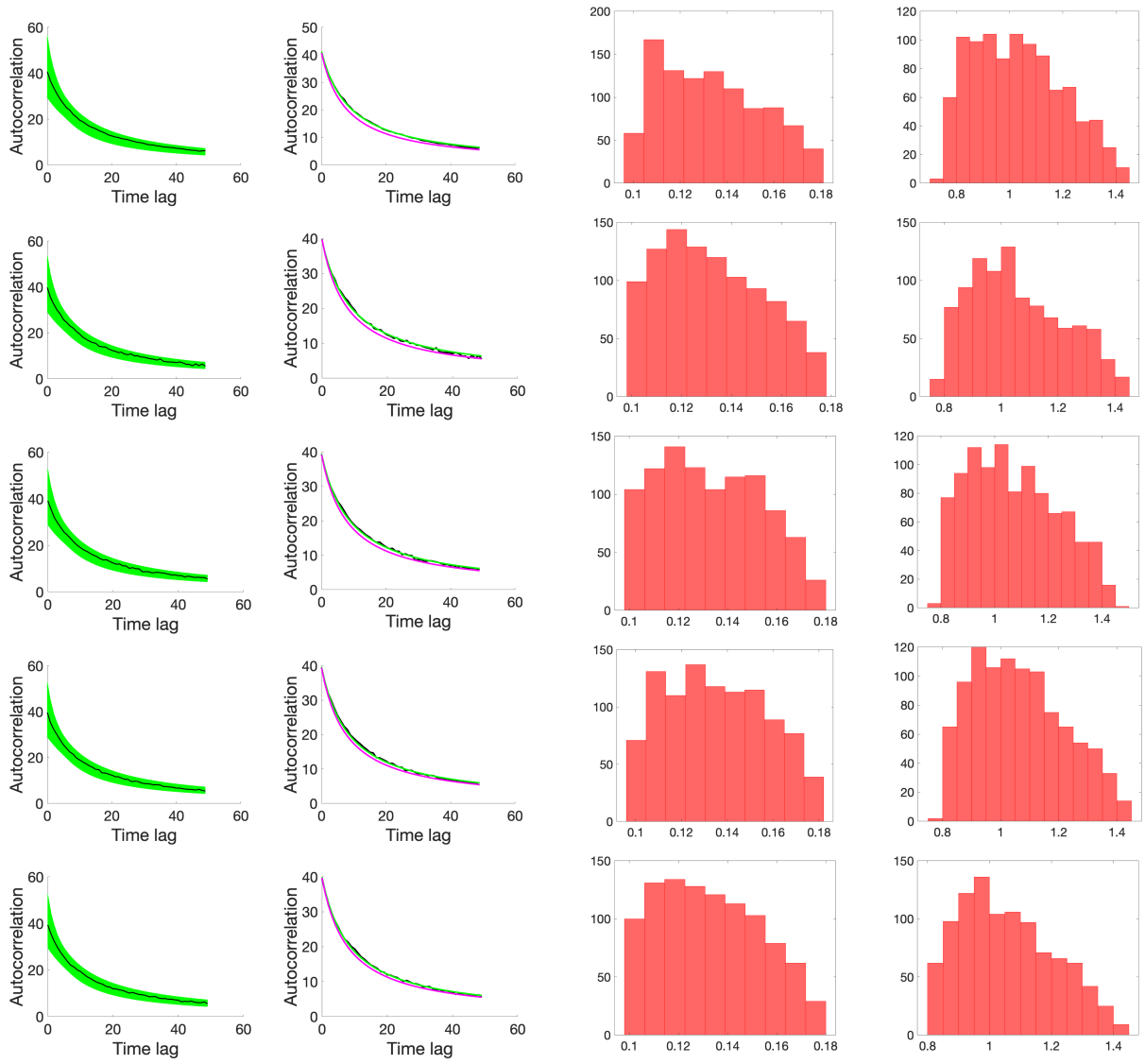


Figure 2.22 – **Sparse estimation for model G_1 on synthetic images,for the last five tested cased.** The first column gives the computed autocorrelation (in black), and the simulation results for the solution proposed by our method (in green). The second column gives the computed autocorrelation (in black), the simulation corresponding to the MAP estimate (in green) and the simulation corresponding to the MMSE estimate (in magenta). The third column gives the histogram of the estimated distribution of D. The fourth column gives the histogram of the estimated distribution of σ^2_{PSF}.

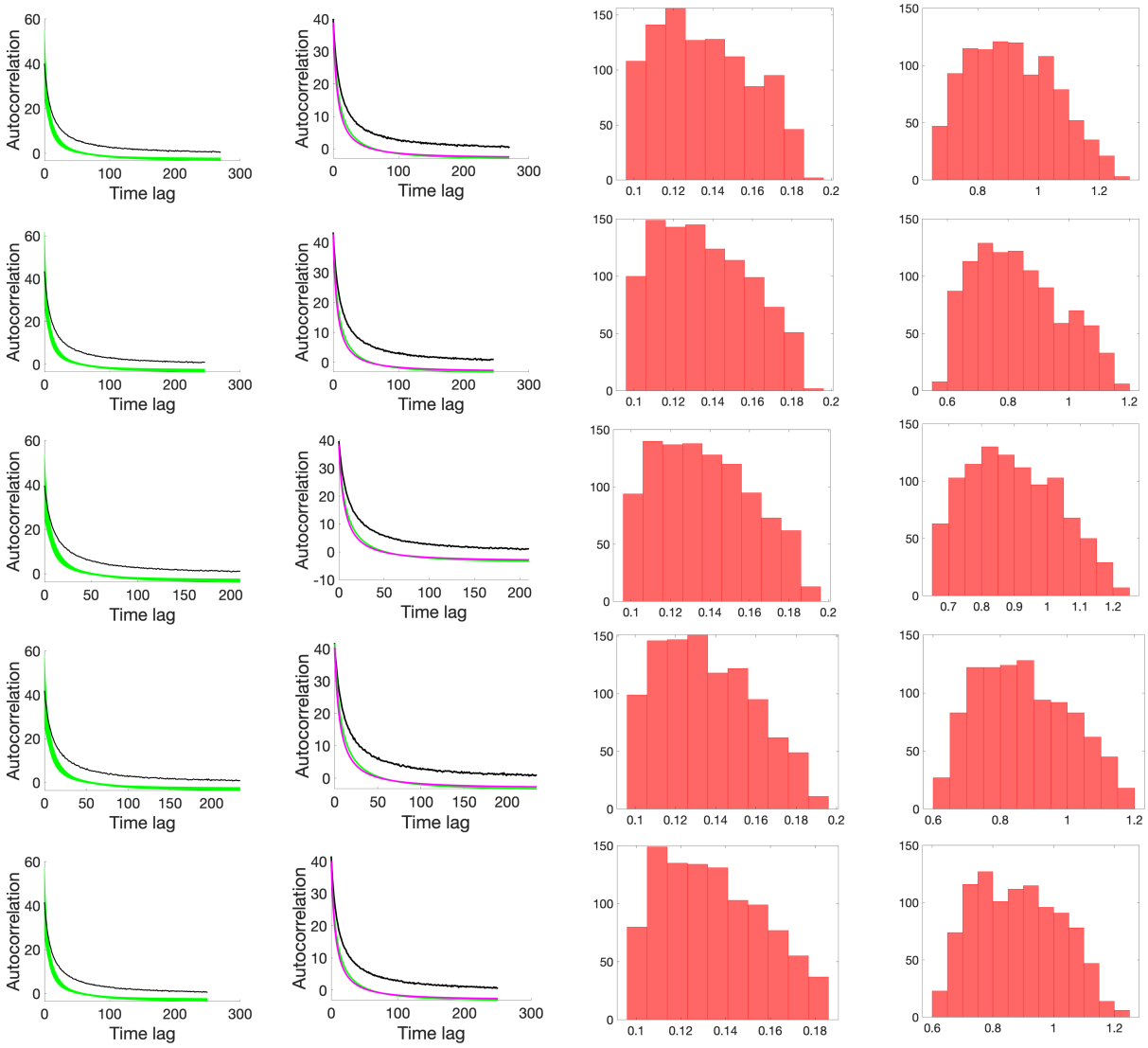


Figure 2.23 – **Sparse estimation for model G_2 on synthetic images, for the first five tested cased.** The first column gives the computed autocorrelation (in black), and the simulation results for the solution proposed by our method (in green). The second column gives the computed autocorrelation (in black), the simulation corresponding to the MAP estimate (in green) and the simulation corresponding to the MMSE estimate (in magenta). The third column gives the histogram of the estimated distribution of D. The fourth column gives the histogram of the estimated distribution of σ^2_{PSF}.

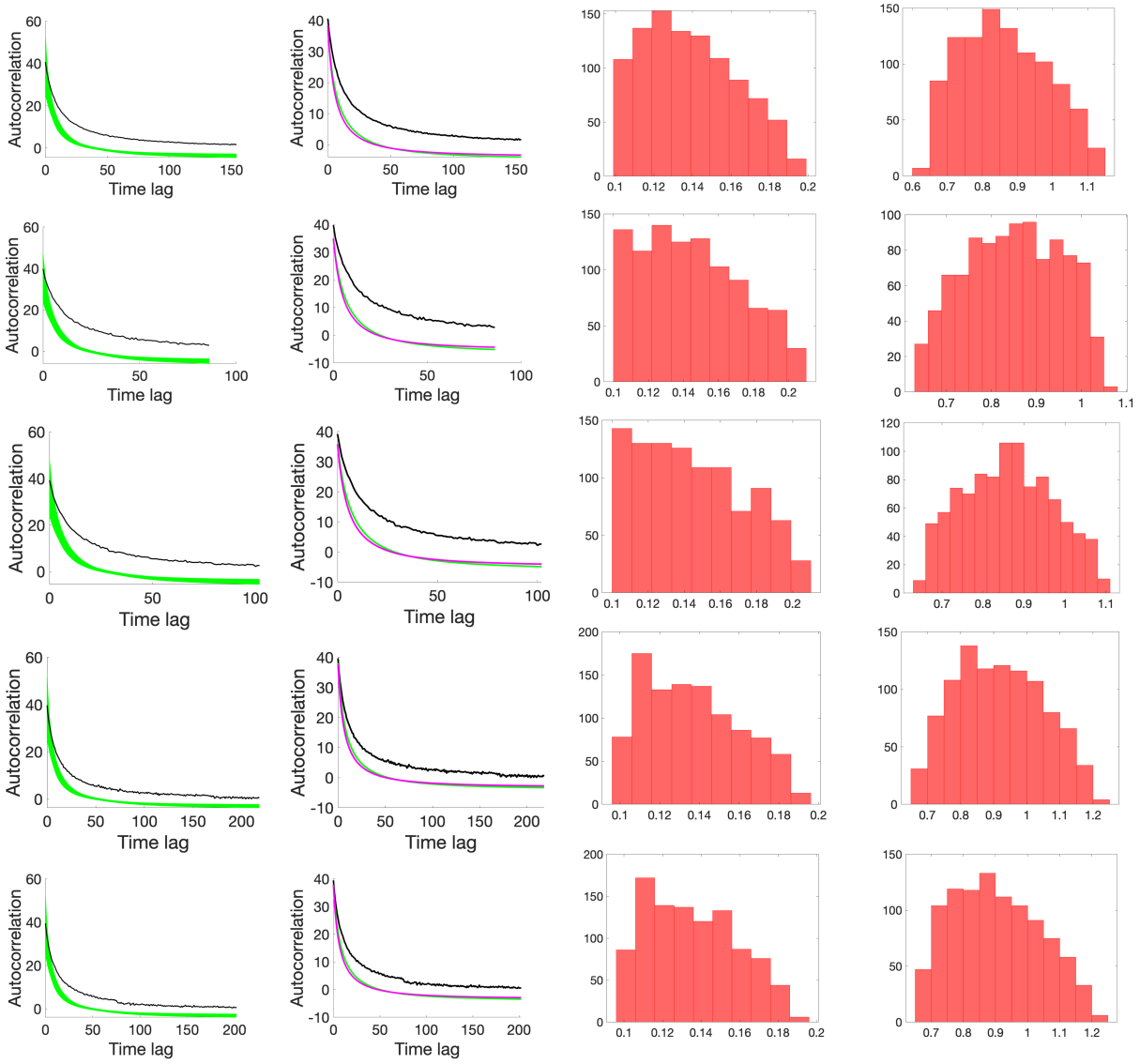


Figure 2.24 – **Sparse estimation for model G_2 on synthetic images, for the last five tested cased.** The first column gives the computed autocorrelation (in black), and the simulation results for the solution proposed by our method (in green). The second column gives the computed autocorrelation (in black), the simulation corresponding to the MAP estimate (in green) and the simulation corresponding to the MMSE estimate (in magenta). The third column gives the histogram of the estimated distribution of D. The fourth column gives the histogram of the estimated distribution of σ^2_{PSF}.

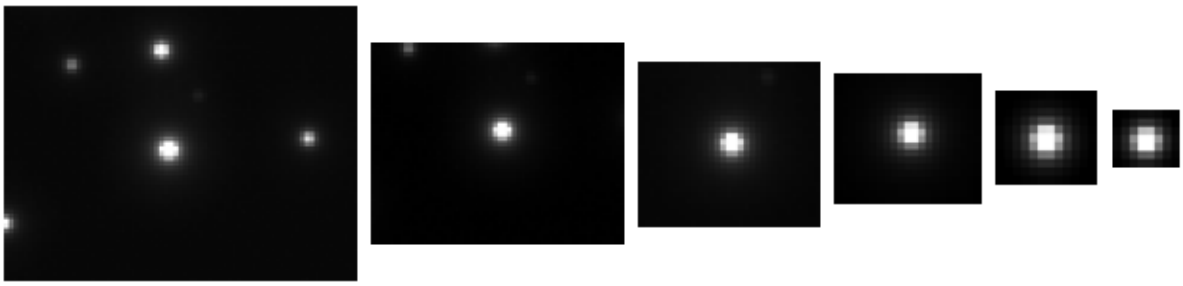


Figure 2.25 – **Synthetic images: tested image sizes (ROI).**

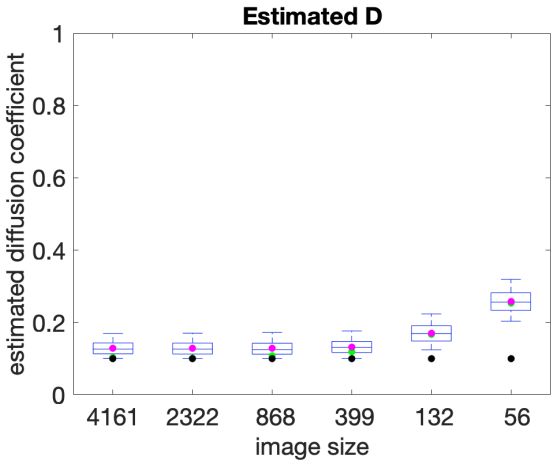
Scenario #2: Varying image size In scenario #2, one spot was selected from a general synthetic sequence with a low noise level, and ten different sequences centered on this spot were extracted from this data, with different image sizes. Image sizes are the following: 73×57 pixels, 54×43 pixels, 31×28 pixels, 21×19 pixels, 12×11 pixels and 8×7 pixels .The length T of each sequence is 100 frames for model G_1, and 247 frames for model G_2. In the large image sizes, additional spots appear and diffuse. Figure 2.25 illustrates the six image sequences that were used in this scenario.

The results of the joint estimation of D and σ^2_{PSF}, for both model G_1 and model G_2, for the 6 tested sequences, are shown in Figure 2.26.

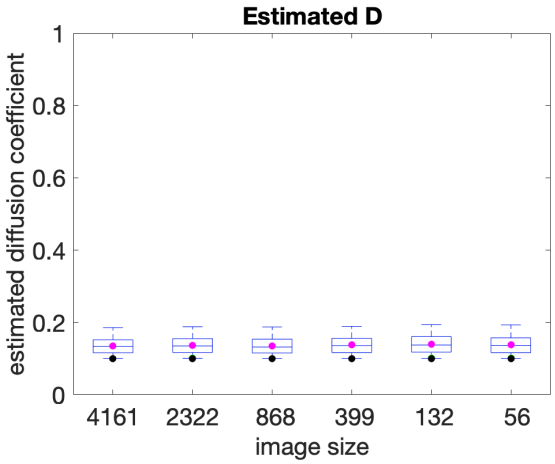
For both models, the estimation of D is very good for all image sizes larger than 400 $\approx (21 \times 19)$ pixels. However, while the estimation with model G_2 is not influenced by the image size, the estimation with model G_1 seems to be less accurate when the image size is lower than 400 pixels. We can notice a correlation with image size: the lower the image size, the lower the quality of the estimation (see Figs. 2.26a and 2.26b). The estimation of the parameter σ^2_{PSF} is shown in Figs. 2.26c and 2.26d. One can see that the results are different for this parameter than for D. For model G_1, the MMSE and MAP estimates get better when the image size is larger, although the standard deviation increases with image size. For model G_2 however, the estimation gets better for lower image sizes.

The choice of the size of the image needs to balance the effects on the estimation of the two parameters, and in particular, for model G_1. In the case tested here, an image size of more or less 400 pixels seems to be at equilibrium.

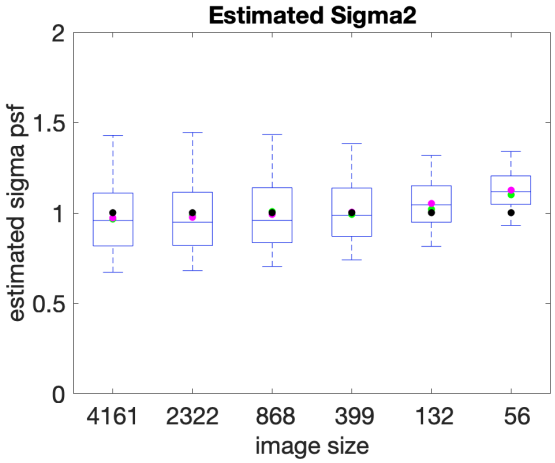
Scenario #3: Varying spot position For scenario #3, a diffusing spot was picked from a general synthetic sequence with low noise, and 6 different sequences of the same



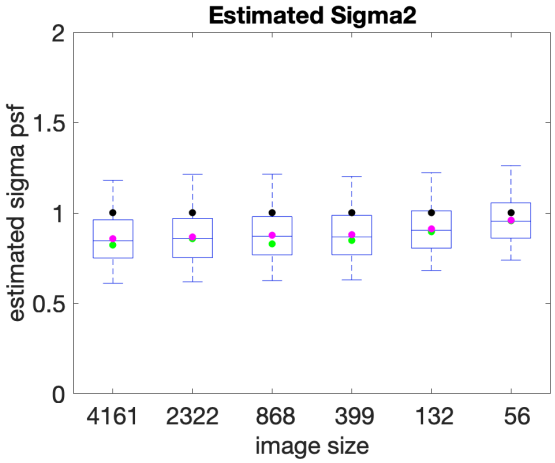
(a) Model G_1 - D estimation



(b) Model G_2 - D estimation



(c) Model G_1 - σ^2_{PSF} estimation



(d) Model G_2 - σ^2_{PSF} estimation

Figure 2.26 – **Estimation of the model parameters for varying ROI size, with both models, on synthetic images.** (a,b) Estimated distribution of parameter D with model G_1 (a) and model G_2 (b) for the six tested cases. (c,d) Estimated distribution of parameter σ^2_{PSF} with model G_1 (c) and model G_2 (d) for the six tested cases. The theoretical value is represented by a black dot, the MAP estimate by a green dot and the MMSE estimate by a magenta dot.

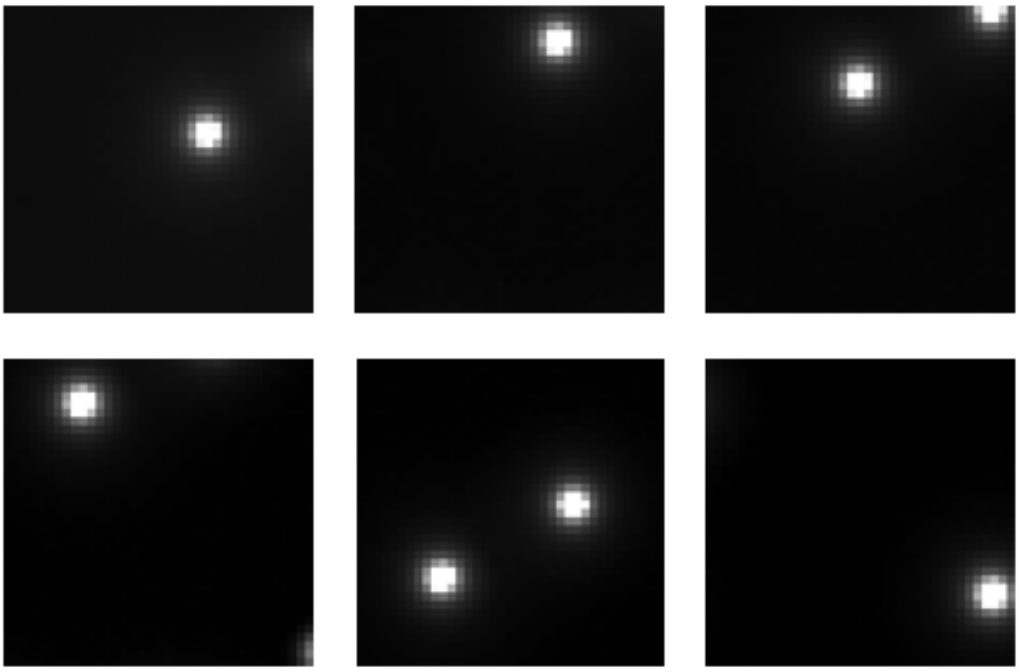
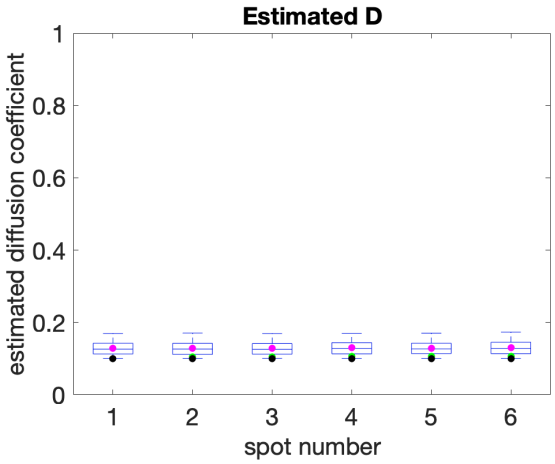


Figure 2.27 – **Synthetic images: tested diffusing spot positions.**

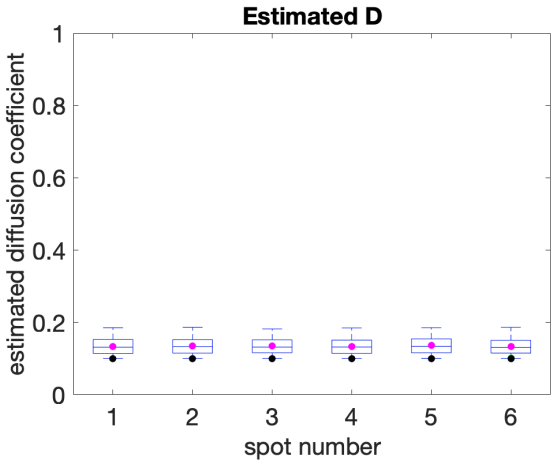
size were extracted by moving the ROI around this spot between sequences. As a result, the spot is not always isolated: a second spot can appear and diffuse in the image. In each sequence, the region of interest is of size 37×37 pixels; the length T of the sequence is 100 frames for model G_1, and 271 frames for model G_2. Figure 2.27 shows frames from all six sequences used for this scenario.

The results of the estimation of the parameters are reported in Fig. 2.28. The estimation is consistent between all tested sequences, for both parameters and both models: there does not seem to be any influence of the spot position on the estimation. The estimation of D is very accurate for both models (Figs. 2.28a and 2.28b). The estimation of σ^2_{PSF} is better in terms of both MMSE and MAP estimates with model G_1, but the standard deviation is lower for model G_2. Overall, the quality of estimation is good for both parameters. The region on interest for the estimation can thus be chosen as convenient.

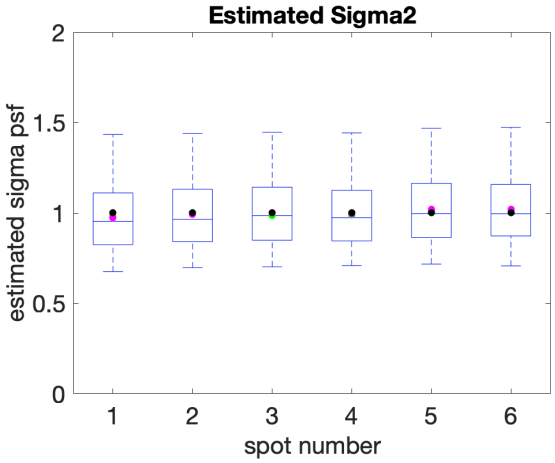
Scenario #4: Varying noise level In scenario #4, six noise levels, ranging from low to medium, were tested for the same isolated diffusing spot. For each sequence, the spot is located in the center of the image, the image size is 28×28 pixels, and the length



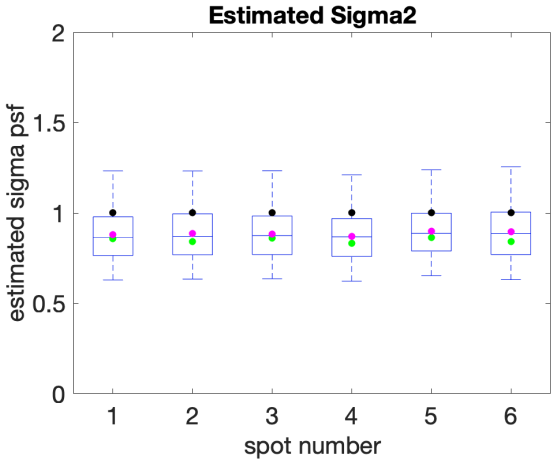
(a) Model G_1 - D estimation



(b) Model G_2 - D estimation



(c) Model G_1 - σ^2_{PSF} estimation



(d) Model G_2 - σ^2_{PSF} estimation

Figure 2.28 – **Estimation of the model parameters for varying spot position in the ROI, with both models, on synthetic images.** (a,b) Estimated distribution of parameter D with model G_1 (a) and model G_2 (b) for the six tested cases. (c,d) Estimated distribution of parameter σ^2_{PSF} with model G_1 (c) and model G_2 (d) for the six tested cases. The theoretical value is represented by a black dot, the MAP estimate by a green dot and the MMSE estimate by a magenta dot.

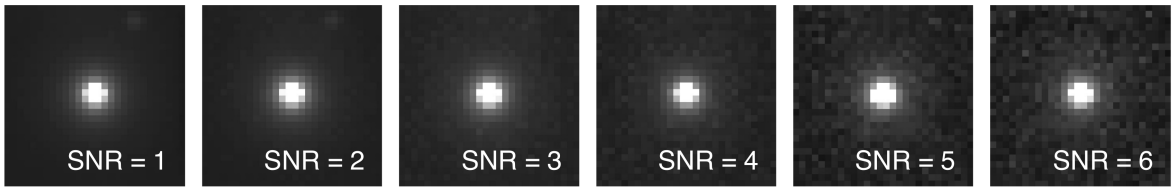


Figure 2.29 – **Synthetic images: tested noise levels.**

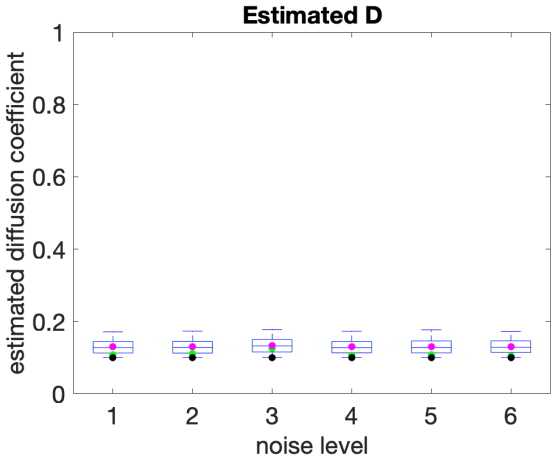
T is 100 frames for model G_1 and 247 frames for model G_2. Figure 2.29 displays the 6 image sequences used for this scenario. The results of the estimation of the parameters are given in Fig. 2.30. The estimation of D is pretty accurate with both models (see Figs. 2.30a and 2.30b), with MMSE and MAP estimates close to the real value and a small standard deviation. The estimation of σ^2_{PSF} is less consistent. With model G_1, the MMSE and MAP estimates are more precise than with model G_2, but the standard deviation of the estimation is higher with model G_1. Thus, some caution might be needed for the estimation of the parameter σ^2_{PSF} when noise is present, but not for the estimation of the parameter D, which is consistent, and very accurate no matter the noise level.

2.3.4 Results on experimental data

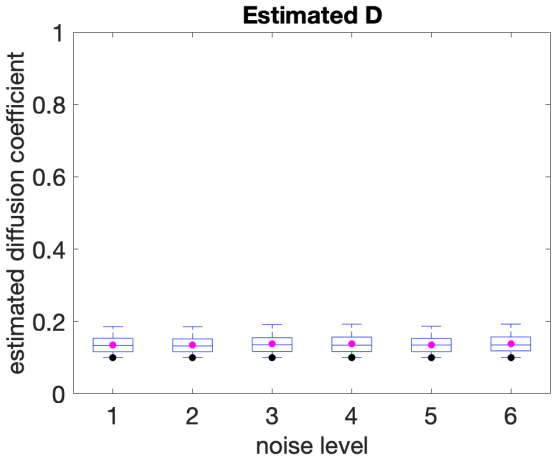
In this section, diffusion is estimated on experimental real data using the same model and estimation method as in the previous section. The data is a TIRFM 2D image sequence (acquired with a frame rate of 100 ms, and pixel size of 160 nm), shown in Fig. 2.31, depicting Transferrin Receptor molecules (TfR) tagged with pHluorin in M10 cells, undergoing exocytosis. Five spots were manually selected, and local diffusion for these spots was estimated.

Figure 2.32 displays the diffusing spots, and the obtained results for each of them. The diffusing spots display some variability, in terms of maximum intensity and spreading in the region of interest. In addition, background is present in the five images, and other spots are present in the region of interest for cases 2 and 3. Finally, the noise level is medium to high in all images. The estimation is thus challenged by all these factors.

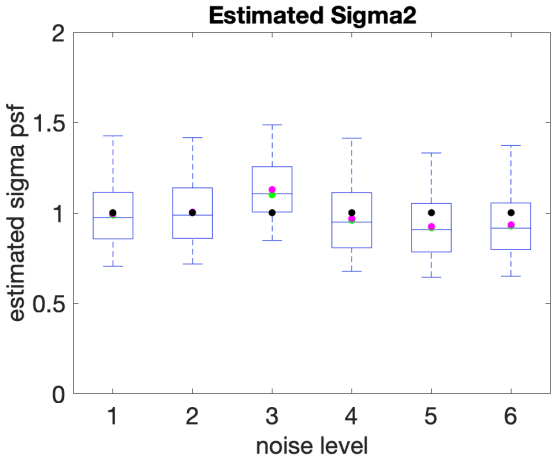
The fitting is good for almost all spots, except for spots #2 and #3 with model G_2, where the beginning and the end are well fitted, but the model does not manage to capture the proper dynamics of the middle part. As the influence of D on the autocorrelation is stronger on this middle part than that of σ_{PSF} (see Figs. A.1 and A.2), this may indicate



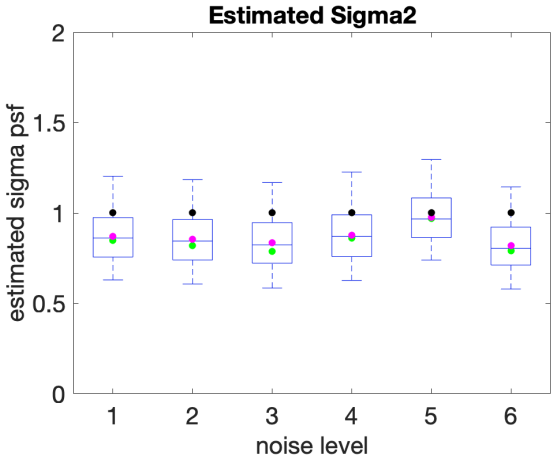
(a) Model G_1 - D estimation



(b) Model G_2 - D estimation



(c) Model G_1 - σ^2_{PSF} estimation



(d) Model G_2 - σ^2_{PSF} estimation

Figure 2.30 – **Estimation of the model parameters for varying noise level, with both models, on synthetic images.** (a,b) Estimated distribution of parameter D with model G_1 (a) and model G_2 (b) for the six tested cases. (c,d) Estimated distribution of parameter σ^2_{PSF} with model G_1 (c) and model G_2 (d) for the six tested cases. The theoretical value is represented by a black dot, the MAP estimate by a green dot and the MMSE estimate by a magenta dot.

that the estimation of D is less accurate. Similar questions are raised by the estimated distributions, that seem further from Gaussian or Gamma distributions than the ones obtained on synthetic data. The first spot is the one that seems to have the best conditions for the estimation: distributions for both D and σ_{PSF} are closer to the ones obtained from synthetic data, and the estimates are similar for both models.

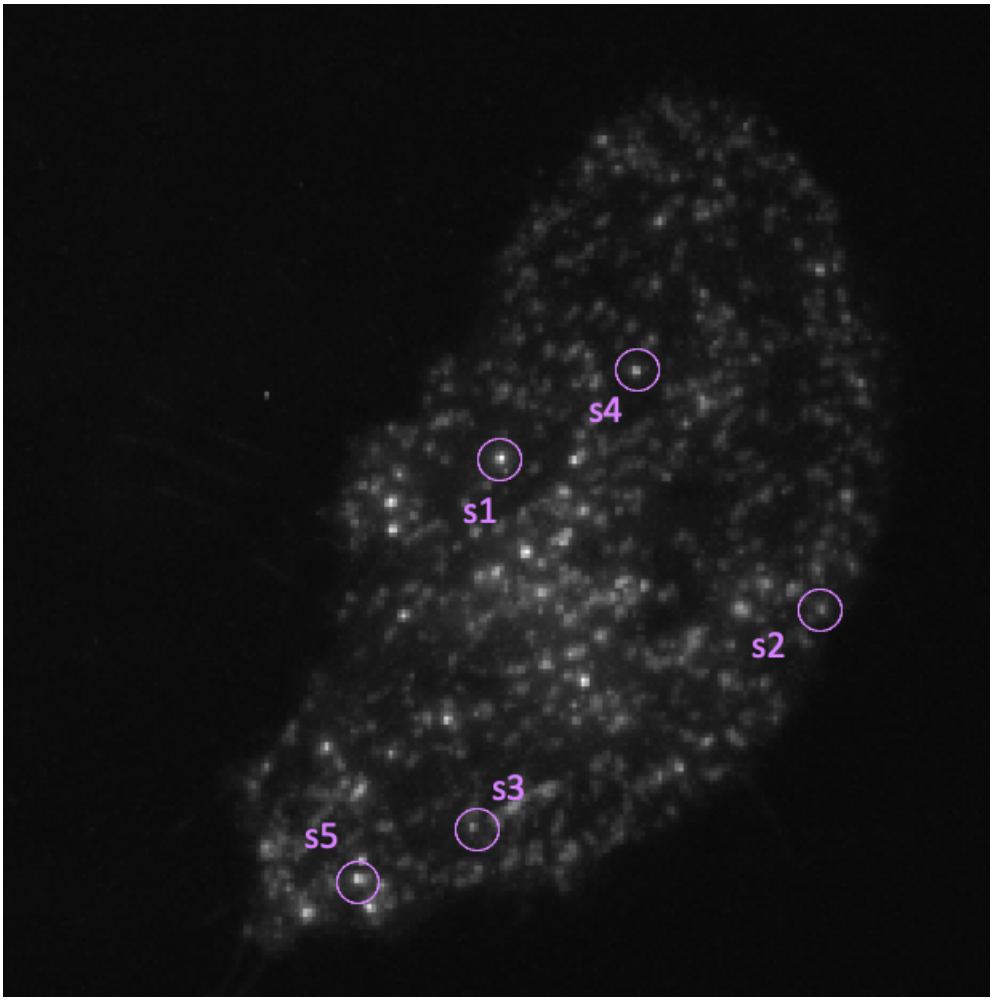


Figure 2.31 – **TIRFM image depicting diffusing pHluorin-tagged spots at the plasma membrane (courtesy of PICT facility, UMR144-CNRS Institut Curie).**

Finally, a surprising result was obtained for spot #4 with model G_1, where the estimation of both parameters strongly differs from the estimation on all the other spots. This might be due to the poor fitting of the beginning of the experimental data, which then influences the estimation of σ_{PSF}. This, in turn, might force the algorithm in a local minimum, where D is not well estimated.

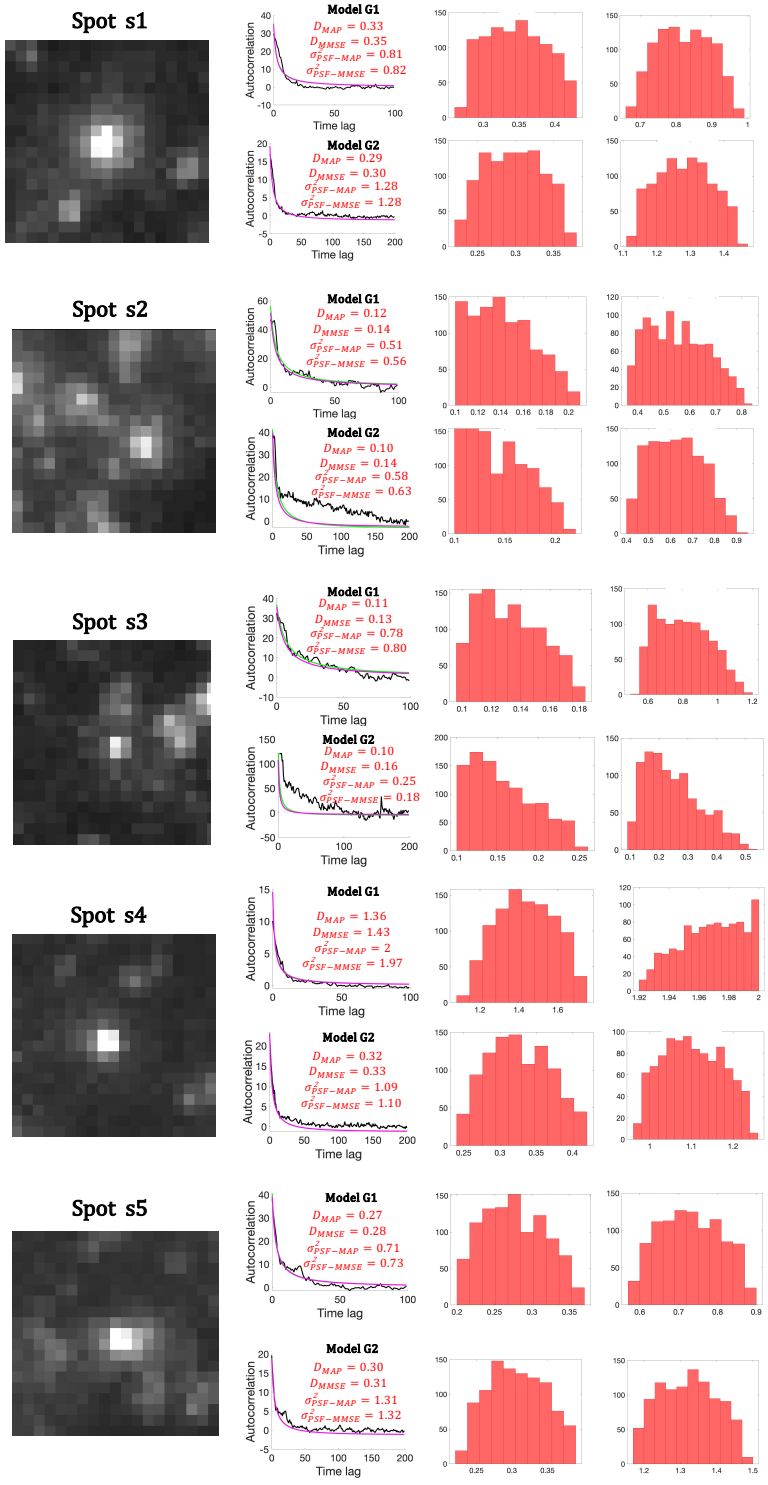


Figure 2.32 – **Estimation results on a real TIRF sequence.** The first column illustrates the selected spot. The second column shows the computed autocorrelation from the real sequence, and the simulations corresponding to the MAP (green) and MMSE (magenta) estimates. The third column gives the histogram of the estimated distribution of D. The fourth column gives the histogram of the estimated distribution of σ^2_{PSF}.

2.4 Discussion and conclusion

In this chapter, we studied on the estimation of diffusion in TIRF imaging and, in particular, on the case of diffusing spots observed during exocytosis. The aim was to build a method that would enable the estimation of local diffusion and the evaluation of its spatial variability from standard sequences obtained from TIRF microscopy.

In Sections 2.1 and 2.2, the context of this study was introduced. Both the Total Internal Reflection Fluorescence Microscopy acquisition technique and the Fluorescence Correlation Spectroscopy analysis method were described together with their advantages and limitations.

In Sections 2.3.1 and 2.3.2, we proposed a proof-of-concept for local diffusion estimation in fluorescence microscopy imaging. We detailed how the method was built and we validated it by analyzing the model with respect to involved parameters and by evaluating the estimation error from data sets built from the autocorrelation model. Preexisting estimation methods use correlation methods that yield ensemble diffusion values; however, these methods do not provide information related to the variability between diffusing spots, as they do not allow to estimate diffusion for isolated diffusion events. Our method was designed to address these issues.

In Sections 2.3.3 and 2.3.4, we evaluated the proposed method on both synthetic and experimental data sets. We assessed its accuracy as well as its robustness to ROI size, SNR, spot position in the ROI and cluttered background, and we showed the added value of our method and discussed its limitations. Our method yielded promising results on synthetic images and proved reliable for all tested scenarios. The application on real TIRF images showed its assets, and in particular, its robustness to medium to high SNR, to other spots diffusing in the ROI, and to cluttered background.

Two limitations need to be adressed: first, the need for long image sequences of at least 100 frames, due to the correlation basis itself; second, the sensitivity of the estimation to interfering dynamics in the ROI, when a process other than pure diffusion is present. This could be improved by introducing a broader model for the expected dynamics in the region of focus and by considering a dynamic background, which should greatly improve

the estimation in such cases.

A short-term application of the work presented above would be the computation of a diffusion map, by computing estimates of the diffusion on a large number of small-sized regions of interest in the frame. The model could also be enhanced by integrating flow components, so that parameters related to flow-diffusion behaviors could also be evaluated. This would enable, among others, the detection of "kiss and run" or "mixed kiss and run" vesicle fusions. Finally, a three-dimensional extension of the method could be envisioned.

Chapter 3

Cell population variability : the mitotic spindle poles behavior in asymmetric division

Asymetric cell division is a key mechanism for the development of living organisms, as it yields daughter cells with different cellular fates. In particular, the series of asymmetric divisions occurring in the early stages of the embryogenesis of the nematode Caenorhabditis elegans (C.elegans) has been thoroughly studied.s Nevertheless, this process is not yet fully understood: while a global understanding and very precise information about some specific proteins are present in the literature, interactions between the actors involved in the division mechanism are still relatively obscure. Light could be shed on these "missing links" between the microscopic and macroscopic scales by developing a biophysical understanding of the process. However, this is a pretty difficult task, due to both the complexity of the mathematical approach and the limited reliability of the image acquisition and processing methods in use.

Anaphase is a crucial stage of the division: as the separation of the chromosomes occurs during this stage, errors can lead to the death of the organism. The behavior of the mitotic spindle, which is the cytoskeletal structure formed for orchestrating the separation of chromosomes, has a strong impact on the whole division. Its poles generally exhibit transversal oscillations in asymmetric division, and the lack of these oscillations is an indicator of an erroneous division.

In this chapter, we propose a new mathematical model of these spindle pole oscillations during anaphase, based on preexisting knowledge about the process and its biophysical factors and molecular actors. In order to be able to put this model into operation on experimental data, we designed a novel framework for model-based parameter estimation, that

is also presented in what follows. The resulting method aims at providing more insight into several aspects of the division process, including its variability and limit behaviors, but also its robustness to parameter perturbations and the underlying interactions between molecular actors.

This chapter begins with a summary of the current knowledge about the first division of the C.elegans embryo and the spindle pole oscillations occurring during anaphase. Our mathematical model for these oscillations is then presented, and it is analyzed in terms of phenotypic variability and parameter influence. The third part of this chapter introduces our novel parameter estimation framework, explaining how it combines sensitivity analysis and Bayesian estimation; this framework is then used with our spindle pole oscillations model.

3.1 Introduction and Related Work

3.1.1 Description of the first division of the C.elegans embryo

3.1.1.1 C.elegans: general considerations

The C.elegans nematode is a simple organism that has been extensively used as a biological model in the last 50 years. It is a small transparent worm, that is 1 mm long and has only 959 cells at adult stage. The C.elegans embryo is approximately 50 microns long and is big enough to be observed by different optic techniques. Despite being a fairly simple organism, C.elegans has complex components such as basic nervous and digestive systems, and an innate immune system that have several mechanisms in common with those in mammalian systems. The C.elegans model is thus a perfect starting point to better understand fundamental biological mechanisms that can be transferred to human biology. C.elegans was first used by Sydney Brenner in 1974 to study the nervous system functioning [14], and since then, this biological model has been used to better understand neurodegenerative diseases such as Parkinson and Alzheimer, bacterial and fungus infections, drug response, and toxicity related mechanisms.

The development of C.elegans is fast, going from a one-cell embryo of 50 microns, to a 1 mm adult stage in roughly 3 days at 20 C [56]. The nematode lives approximately 20 days and in temperatures of 12 to 25 C. Its life cycle is composed of different stages (see

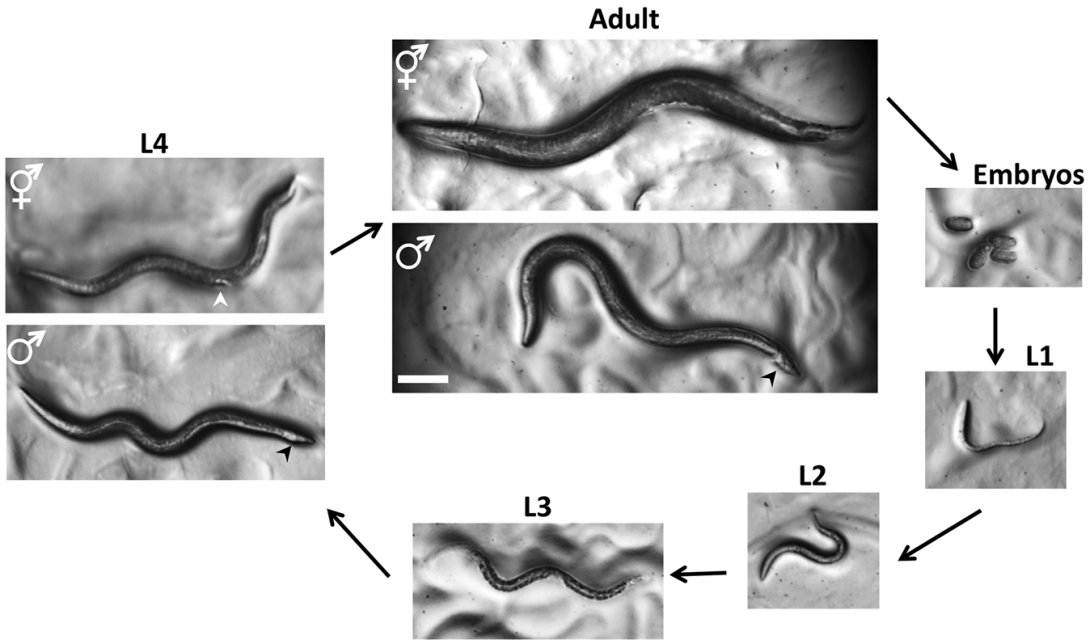


Figure 3.1 – **Developpement of the C.elegans : embryo, L1 stage, L2 stage, L3 stage, L4 stage and adult stage.** Adapted from [20].

Fig. 3.1), starting with the embryonic stage, that lasts for about 14 hours. At the end of the embryonic stage, the eggshell breaks and the embryo gets to a first larva stage, called L1 phase, that is able to feed itself. The larva goes through three more larvae stages L2, L3 and L4 that are easily identified thanks to size and morphology. After phase L4, the larva transforms into an adult worm that is able to reproduce.

One of the most remarkable features of the C.elegans nematode is the reproducibility and fidelity of its first division. This characteristic facilitates the analysis of biological mechanisms occurring during cell division. Another advantage of C.elegans is the facility to deplete proteins using RNAi. Coupled with techniques for expressing fluorescently tagged proteins, it makes it possible to study the effects of malfunctioning proteins and related mechanisms on cell division.

3.1.1.2 C.elegans: first division

Phases of division The first division of the C.elegans nematode is asymmetric, resulting in two daughter cells with different sizes. We focus here on mitosis in the C.elegans

embryo, that takes place after two meioses. This process is illustrated by Fig. 3.2. At the beginning of mitosis, the one-cell embryo has two pronuclei, one male and one female, that are at opposite sides of the cell. When the trigger for the start of mitosis is activated, the two pronuclei move toward the center of the cell, the female pronucleus traveling further than the male one. At this stage, chromosomes are condensed, and the nuclear envelope is intact for both pronuclei. In addition, the male pronucleus has the two centrosomes formed, and they are at opposite sides of the nuclear envelope. The two pronuclei meet at about 70% of the embryo length, forming a nuclear-centrosome complex. This complex then moves toward the center of the cell, while also rotating, thanks to microtubules emanating from the centrosomes. At the end of this phase, nuclear envelopes start to dissipate and thus the first phase of mitosis, called prophase, is achieved. In the next phase, as the nuclear envelopes have disappeared, condensed chromosomes can now be catched by microtubules. The mitotic spindle starts to form as the chromosomes get attached to the corresponding centrosome. This phase is called prometaphase, and it is followed by metaphase, when the mitotic spindle finishes its assembly, and the chromosomes align vertically, on the so-called metaphase plate. When all chromosomes are aligned, and correctly attached by microtubules (as illustrated by Fig. 3.3), there is a trigger to get to the next phase of division, that is anaphase. During anaphase, sister chromatids are separated, and move toward the pole they were attached to, while the mitotic spindle elongates and moves toward the posterior side of the cell. When the movement of the mitotic spindle and the separation of chromosomes end, the cell enters the last phase of mitosis, the so-called telophase. In this last stage, sister chromatids decondense while the nuclear envelope forms around each chromosome. The mitotic spindle is disassembled, and the cell achieves its division by cytokinesis, a process that separate the cytoplasm between the two daughter cells.

Anaphase regulation The transition between metaphase and anaphase is highly regulated by the spindle assembly checkpoint. During metaphase, chromosomes align on the metaphase plate. This is done through a mechanism called search and catch where microtubules nucleating from the centrosomes explores the space until they find a chromosome to attach to. During this time, the two sister chromatids are held together by kinetochores, and more precisely by a complex called cohesin complex. Microtubules attach to kinetochores creating tension between sister chromatids. Any kinetochore that is not under tension produces a signal that blocks the activation of the anaphase-promoting complex.

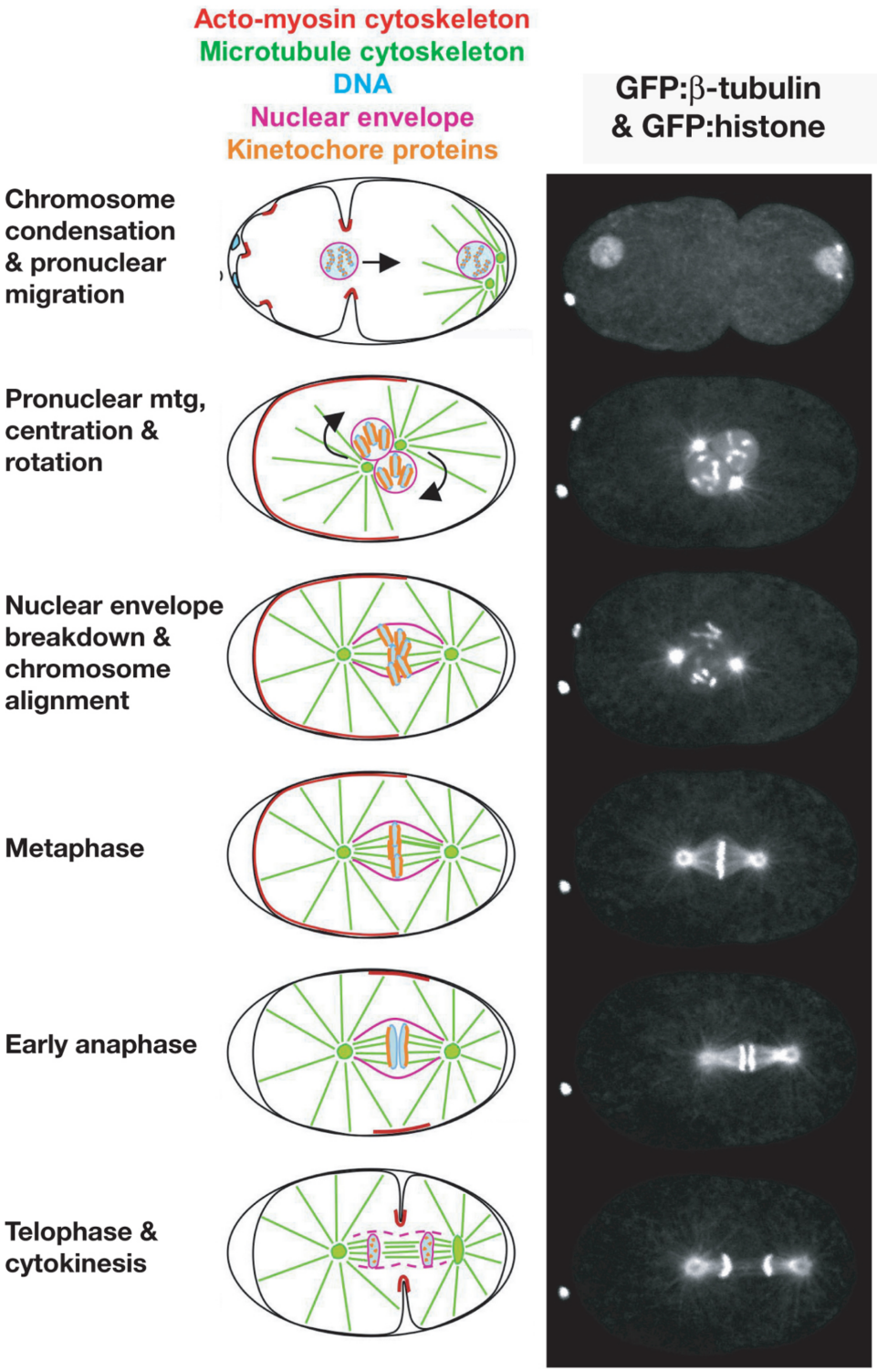


Figure 3.2 – **Phases of the first division of C.elegans.** Left: Schematic illustrations of the major features of the division. Right: Images of each division phase, collected by spinning disk confocal microscopy. Adapted from [56].

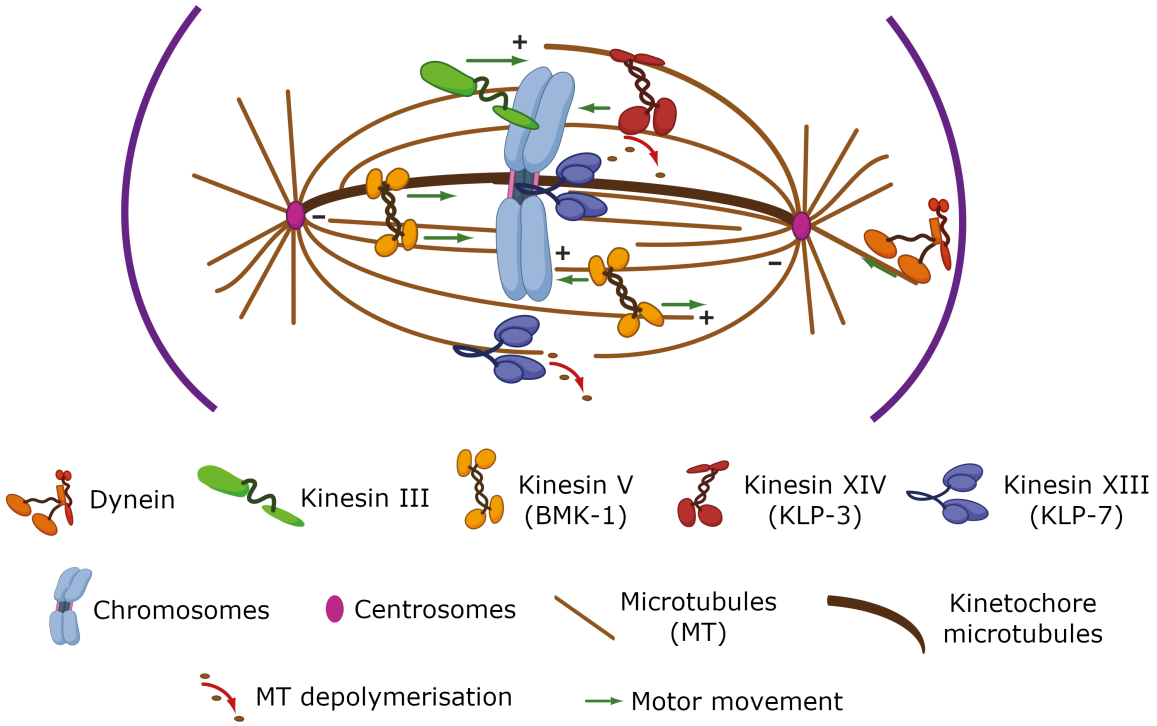


Figure 3.3 – **Schematic representation of the mitotic spindle and associated molecular actors.** Adapted from [46] and [54].

When all chromosomes are attached, all kinetochores are under tension, and thus there is no signal that stops the activation of the anaphase-promoting complex. The activation of the anaphase-promoting complex triggers the activation of separase, an enzyme that breaks the cohesin complex that keeps sister chromatids linked. Sister choromatids are now free to separate, and anaphase can take place.

Spindle poles separation and forces involved in anaphase Anaphase in C.elegans is driven by the separation of the two spindle poles. In other organisms, anaphase is separated in two phases : anaphase A during which only the sister chromatids separate thanks to internal forces of the mitotic spindle, mainly by the microtubules that pull them towards their respective spindle poles; anaphase B when the spindle poles separate thanks to exterior forces pulling on astral microtubules. In C.elegans, anaphase is mainly driven by the spindle poles separation, external forces having a bigger impact on the chromosomes and centrosomes separation than internal forces. Internal forces might still

be needed at the beginning of anaphase when centrosomes are closer to each other. Also, internal forces are needed for spindle elongation as mitotic spindles still elongate when there are no external forces acting on them, but displacement towards the posterior pole, and oscillations are absent in these cases.

How the molecular actors participate to the robustness of anaphase The main components driving anaphase are microtubules and microtubules-associated molecular motors. Microtubules are fiber-like polymers that can grow and shrink in the cytoplasm, and have several roles in the cell. Microtubules can generate and transmit forces at the cell-scale. They can also be used as highways for cargo transporting for molecular motors that can move along the microtubules while carrying/transporting cargos. Microtubules are hollow cylinders made of tubulin subunits. The dense structure of these tubulin subunits makes microtubules difficult to break as a high number of bonds are present. This feature gives rigidity to microtubules, and high resistance to lateral forces. In addition, through the orientation of the tubulin subunits, microtubules have a given polarity with a plus-end and a minus-end. Both ends are changing through association-dissociation (i.e., polymerization-depolymerization) processes, but the minus-end is a lot more dynamic than the plus-end, which is more stable. When a microtubule changes state from growing to shrinking, it is said to undergo a catastrophe event, and when it goes from shrinking to growing, the microtubule undergoes a rescue event. Microtubules are thus in a state of dynamic instability at any given moment. Figure 3.4 summarizes the association-dissociation cycle of microtubules.

Microtubule-associated molecular motors are protein complexes that can apply forces on microtubules, and transport cargos. These complexes go through continuous transformation by binding, hydrolyzing and unbinding of ATP. This continuous movement together with the polarized microtubule fibers makes molecular motors walk on microtubules. Molecular motors transmit the force they generate to the microtubule, and as microtubules are rigid structures, the force is transmitted to centrosomes. One family of molecular motors of particular interest for our study, is the dyneins family. In particular, we are interested in cytoplasmic dynein, which is present in the cytoplasm, and has the role of transporting cargo and generate force on the microtubule. Dynein is a minus-end oriented molecular motors, so it transports cargo toward the minus-end of the microtubule. Cytoplasmic dynein has two heavy chains that are responsible for energy and movement

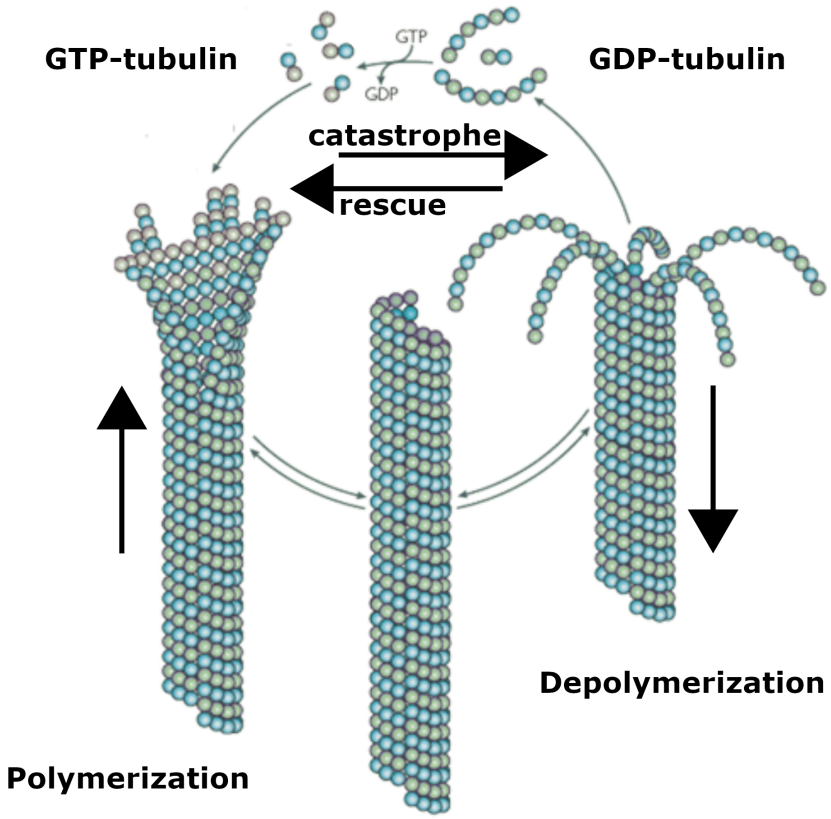


Figure 3.4 – **Polymerization-depolymerization cycle of microtubules.** This cycle is driven by GTP-tubulin and GDP-tubulin hydrolysis. Adapted from [5].

generation, and different intermediate and light chains that help the motor attach to cargo.

Although the global mechanism involved in anaphase during the first division of C.elegans is well established, our understanding of how the external forces drive the behavior of the spindle poles, and how specific parameters are linked to this process, is still limited. In the next section, we propose an overview of the main related work and discuss the main assumptions and conclusions we used to build our bio-physical model of the spindle poles oscillations during anaphase.

3.1.2 Spindle modeling during anaphase

Some of the early experiments conducted to better understand the behavior of the spindle poles during anaphase in asymmetric division have given the main axis of expla-

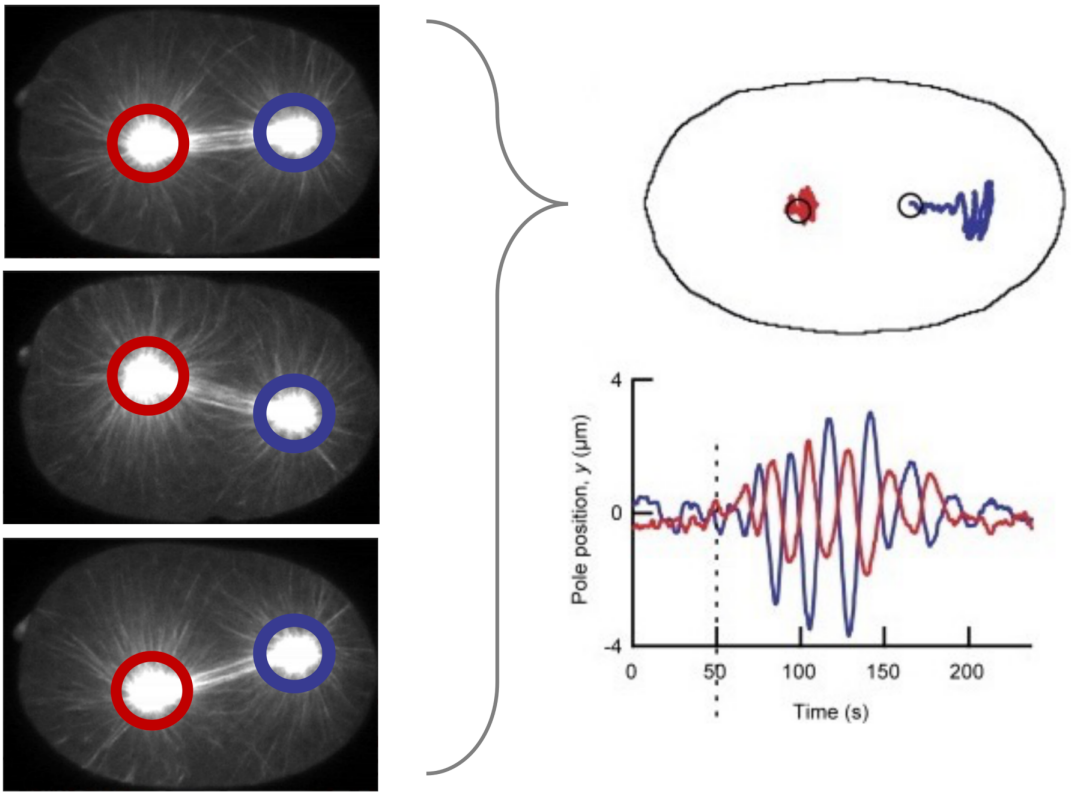


Figure 3.5 – **Anaphase and spindle poles oscillations.** Left: Images of a cell during anaphase, obtained by fluorescence microscopy. The anterior and posterior poles are circled in red and blue, respectively. Top right: Illustration of the movement of the spindle poles. Bottom right: Evolution of the position of the poles (transversal to the AP axis) over time. Adapted from [62].

nation for the observed behavior. During anaphase, the spindle poles move toward the posterior side of the cell, they move apart from each other as the mitotic spindle elongates, and the posterior pole oscillates transversely to the Anterior-Posterior (AP) axis, as shown in Fig. 3.5. To understand what drives this behavior, Grill *et al.* [32] have tested several hypotheses by cutting the mitotic spindle with an ultraviolet laser beam at the beginning of anaphase. The two main hypothesis were that the 3 types of movement undergone by the poles during anaphase (elongation, displacement towards posterior and oscillations) are: either the interior forces of the spindle are responsible for this behavior, either exterior forces are the ones driving the spindle poles behavior. The first hypothesis was fueled by other experiments done in Drosophila for example, that show that overlapping microtubules in the mitotic spindle can create pushing forces on the poles, thus,

moving the poles apart from each other. The second hypothesis was motivated by other experiments showing that astral microtubules can pull on the spindle poles (thus external force being responsible for the behavior in anaphase).

By cutting the mitotic spindle at the beginning of anaphase, Grill *et al.* observed that the poles moved even further apart than in wild-type regular embryos, that the peak velocities of the poles were higher that in regular embryos, and that the posterior pole oscillated when arriving close to the cortex. This means that the forces driving the elongation, displacement, and oscillations of the poles are external to the mitotic spindle. Also, the posterior pole travelled further then the anterior one and its peak velocity was 40% higher. In theory, this could be explained by a change in viscous drag between the anterior and posterior sides of the cell. Grill *et al.* tested this hypothesis through a yolk-granule motion analysis. They found that the diffusion of granules at the anterior side were similar to those at the posterior side, only a 6% difference in average diffusion coefficient was found, which could not account for the 40% difference in peak velocities for the poles. This means that the viscous drag, although it probably has some effect on the peak velocity differences, is not the main cause of this observation. The behavior of the spindle poles during anaphase seems to be mainly driven by forces exterior to the mitotic spindle in the first asymmetric division of the C.elegans embryo.

In their paper, the authors propose a simple model to account for the behavior of the spindle pole after spindle severing. For this, they assume that the pulling forces acting on the spindle pole are generated by astral microtubules. This assumption is motivated by experiments where the astral microtubules are cut, that showing that poles move toward the opposite side of the cell after cutting. Microtubules-driven pulling forces is thus a plausible hypothesis.

Later on, in [33] Grill *et al.* confirmed this latter hypothesis and took a step further in understanding how pulling forces are generated by astral microtubules. In this experiments, they ablated the central region of the centrosome, that cause fragments of the centrosome to move away from the initial position, toward the respective cortex (anterior side for the anterior centrosome, and posterior side for the posterior centrosome). Also, the other centrosome did not move from its original position. This confirms that pulling forces are generated by external (from the mitotic spindle) forces, and that astral micro-

tubules can be the basis of this forces. In this experiment, the authors also observed that the fragments, that were of similar size, expanded radially around the initial position, which means that force generators are not confined to a specific region of the cortex, but rather distributed on the posterior and anterior side. Interestingly, the fragments from the anterior centrosome moved symmetrically around the initial position (still toward the cortex), whereas for the posterior side, when the ablation took place during oscillations, fragments that moved in the initial direction of movement of the centrosome travelled at higher speeds than fragments that moved in the opposite direction of the initial movement of the centrosome. But as the viscous drag is the same on the anterior and posterior side, and the fragments were similar in size and the direction of movement and the size of the fragment were uncorrelated, authors propose that this difference in speed is due to differences in pulling forces, that are higher for the fragments travelling faster, and lower for the fragments travelling slower. This could be explained by either a larger number of force generators per fragments, or by a larger force per individual force generator. By analyzing the variance of the speed of the fragments, authors concluded that the first option is more realistic than the second one. Moreover, as the variance increases and then decreases, the authors suggest that pulling forces are governed by a two-stage process: first the number of force generators increases which makes the variance in speed increase, then this number gets to a saturation number, and the variance in speed decreases.

Finally, the authors conducted the same experiment on GPR-1/2 RNAi embryos. GPR-1/2 proteins are involved in forces regulation, via the Gα-protein activity inhibition [19, 31, 78]. Authors observed that, in these embryos, spindle displacement is disrupted. When ablating the center region of the centrosome, fragments did not expand toward the cortex, and the mean speed was low compared to wild-type embryos. This means that, in GPR-1/2 RNAi embryos, pulling forces were almost absent, which suggests that there are no active force generators at the cortex.

In 2006, Pécréaux *et al.* [62] went one step further in the study of spindle displacement and oscillations. In this paper, the authors investigated how the force generators can give rise to oscillations while also displacing the spindle toward the posterior side of the cortex. The main hypothesis tested here is that a threshold number of active force generators is needed in order to start oscillations. Furthermore, to account for the build up and die down of oscillations together with the spindle displacement, they propose that

force generators stay attached to a microtubule for a longer period of time as the division progresses. This means that the unbinding average rate of a force generator decreases over time, which is explained by the authors in terms of increasing processivity of the force generators.

An antagonistic-motors model is considered, and the hypothesis of a threshold is tested by comparing the behavior given by the model and the observed behavior in RNAi embryos with reduced levels of dynein light intermediate chain and GPR-1/2 that are involved in the regulation of force generators. Model prediction of a threshold number of active force generators was confirmed by the experiments. In the RNAi embryos, oscillations were either greatly reduced, or completely abolished when the RNAi effect was strong. Reducing GPR-1/2 expression will reduce the motor activity, and thus, when the motors activity is too low, the oscillations are either reduced, or completely inexistent if the needed motor activity threshold is not crossed. Thanks to this model, the authors could also test what parameter could account for the build up and die down of oscillations while also preserving the posterior displacement. They found that, the only parameter that could account for both of these behaviors, is the average motor activity, by a monotonic increase in this parameter. Finally, the authors propose that one motor could generate both oscillations and posterior displacement, and the most probable motor that could have this role is cortical dynein as it was previously showed that dynein is crucial for spindle centering in metaphase and for spindle oscillations in anaphase.

These papers shed light on crucial bio-physical aspects of how the spindle poles separate during the first division of C.elegans. Experiments involving cutting the mitotic spindle at the beginning of anaphase and ablating the central region of the centrosomes showed that (i) pulling forces are responsible for the three types of movements of the poles during anaphase, (ii) these pulling forces are external to the mitotic spindle and most probably generated by the astral microtubules, and (iii) they are governed by a two-stage process which involves an increase of the number of force generators until a threshold is reached. An antagonistic-motors model and further experiments including GPR-1/2 RNAi embryos, confirmed that this process is able to produce both posterior displacement and poles oscillations and that it requires a threshold number of active force generators; it was further explained by a monotonic increase in the mean motor activity. These papers set the bases for understating the spindle poles separation during anaphase,

but there are many aspects that need to be further studied, in particular: how the different parameters influence the behavior of the poles; how they are linked to molecular motors participating in anaphase; how they vary between individuals, or between populations, beyond the mean behavior previously studied.

3.2 Model of spindle pole oscillations during anaphase

Our aim is to build a mathematical model that describes the spindle pole oscillations during anaphase in the first division of the C.elegans embryo. An important step in this direction was performed in [62], where a simplified model of the oscillations is proposed by Pécréaux *et al.* We show, in what follows, how a similar approach helped us create a more elaborate model, that is adapted to a wider range of oscillatory behaviors.

3.2.1 Pole oscillations from a biophysical point of view

While the two poles of the embryo influence each other via the mitotic spindle during anaphase, their interaction does not have a significant impact on the oscillations [32]. Thus, we focus our study on single pole behavior. This study can be applied to both anterior and posterior pole independently, as the same assumptions hold for both cases. In what follows, we describe the model for the posterior pole, as it exhibits stronger oscillations, and we refer to it as the "spindle pole". The anterior pole will be studied in Section 3.3.3.3, where we provide a compared analysis between the anterior and the posterior poles.

Tracking of the spindle pole shows that the pole starts to elongate and move toward the posterior side of the cell when anaphase is triggered. Then, when it reaches about 70% of the embryo length, it starts to oscillate vertically for 100 to 200 seconds. The oscillations build up and reach a maximum that is consistent between embryos of the same condition. Then the oscillations die out, the pole reaching an asymmetric equilibrium position close to the posterior cortex, and centered in vertical position. The oscillations are symmetric with respect to the maximum amplitude, the build up and die down periods of time being similar.

To model the spindle pole behavior, a biophysical approach was used (as in [62]), and

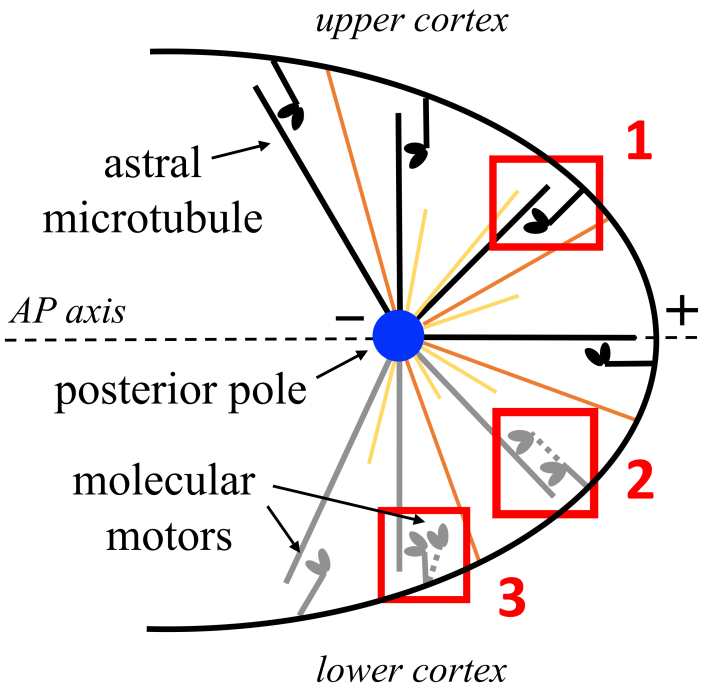


Figure 3.6 – **Schematic representation of the posterior pole, microtubules and molecular motors.** In (1), the motor is bound to a microtubule. In (2), the motor is walking on the microtubule, thus generating a pulling force. In (3), the motor unbinds from the microtubule.

in particular, a force-balance equation gave the basis for our model. The pole is under the action of several forces. First, there are pulling forces generated at the cortex by active motors as seen the previous section ([32, 33]). These motors can bind to a microtubule, pull on the microtubule by "walking" on it, and by consequence, pull on the pole (assuming that microtubules are "rigid enough" structures). These forces pull the spindle pole off-center toward the cortex. Opposing the movement of the pole, there is a viscous force arising from the microtubule network ([35]) and viscosity of the cytoplasm ([30]). The higher the speed of the pole, the higher this force will be, thus slowing down the pole. Finally, there is a force that brings the pole back to a centered vertical position ([35],[63]), that we call a centering force. This force is thought to be generated by the microtubules reaching the cortex that are not actively in a pulling state. This centering mechanism is essential to maintain the spindle centered during metaphase while the chromosomes get aligned and attached to their respective pole, and its action continues through anaphase.

Our model is based on three main modelling assumptions (see Fig. 3.6):

1. Astral microtubules can be in a pulling or non-pulling state. When microtubules (MT) are growing from the centrosome, they are in a non-pulling state until they reach the cell cortex. At the cortex, a MT can either encounter an active force generator and have a certain probability to switch to a pulling state (by having a force generator binding to it), or it does not encounter an active force generator, and it will be in a non-pulling state. When in a non-pulling state, a MT can continue to grow and starts pushing on the cortex (this pushing force contributes to the centering force), or it can bundle when it has been pushing for too long. At any moment, a MT can start to depolymerize and thus start shrinking. Sometimes the depolymerization can be "reversed" when the MT starts growing again and the same possible scenarios as described in this paragraph can take place.
2. A force generator can bind to a MT, pull on the MT, and then it can unbind from the MT. This process has been described via a relation between the force and the unbinding rate of the force generator from the MT. When the force increases, the unbinding rate increases exponentially, thus the higher the force, the higher the chance for the force generator to unbind from the MT, in an exponential manner.
3. Force is generated by having a force generator walking on the MT. This has been described via an inverse linear relation between the speed of the force generator and the force it creates. When the force generator speeds up, the force in the direction of movement gets weaker, and its magnitude decreases linearly with increasing speed.

With these assumptions, the behavior of the spindle pole during anaphase can be completely described. When anaphase is triggered, the mitotic spindle, and thus the poles, move along the AP-axis toward the posterior side of the cell, while also elongating, the two poles getting further apart from each other. When it reaches 70% of the embryo length, the system gets to an unstable state. When in this state, oscillations can be triggered when a threshold of active force generators is passed, by the binding and unbinding of the force generators. Noise or any perturbation (like thermal perturbation) can make the pole move, e.g., vertically toward the upper cortex. In this case the speed increases toward the upper cortex, and as a consequence, the force per motor at the upper cortex decreases, which, in turn, will decrease the unbinding rate. This will increase the binding time of force

generators to microtubules. The binding rate is superior to the unbinding rate, so with larger binding times, the total pulling force from the upper cortex will increase, despite having lower individual force per force generator. This increase in total force is enough to counteract the viscous and the centering forces for a small period of time. Meanwhile, as the pole gets further from the center, the centering force increases, until its magnitude is big enough to compensate the increase of the total force generated at the cortex. The pole will thus reach a maximum amplitude, stop and then reverse movement toward the lower cortex.

The pole is now pulled toward the AP axis by the centering force, and its speed starts to increase again but in the other direction. On the upper cortex this increase in opposite speed creates an increase in the individual force, and therefore an increase of the unbinding rate which ultimately means a decrease of the total pulling force for the upper cortex as microtubules unbind quicker. From the lower cortex perspective, the speed increases, which brings the system in the same position as discussed previously for the upper cortex. The same scenario takes place while the pole moves vertically toward the lower cortex.

At this stage, the pole can continue to oscillate vertically while moving horizontally toward the posterior side of the cell. But the oscillations build up, and then die down after 100 to 200 seconds. To account for this behavior and horizontal displacement, Pécréaux *et al.* [62] propose that the average unbinding rate decreases over time at given force. This would have an effect on the frequency of oscillation, that would decrease over time. Pécréaux *et al.* confirmed this experimentally by measuring the change in frequency over time in wild-type embryo, and found that the frequency is indeed decreasing. We thus keep this assumption in our model. The monotonic decrease in the unbinding rate, accounts for the build up and die down of oscillations. The pole then reaches its final posterior position, vertically centered, but horizontally displaced toward the posterior side of the cell, and this sets the end of anaphase in C.elegans.

3.2.2 Proposed mathematical model and modeling assumptions

The construction of the model is based on a physical approach and relies on a force balance. In the context of the studied problem, four forces act on the centrosome, as illustrated by Fig. 3.7:

- F_+: total force exerted by the motors of the upper cortex;

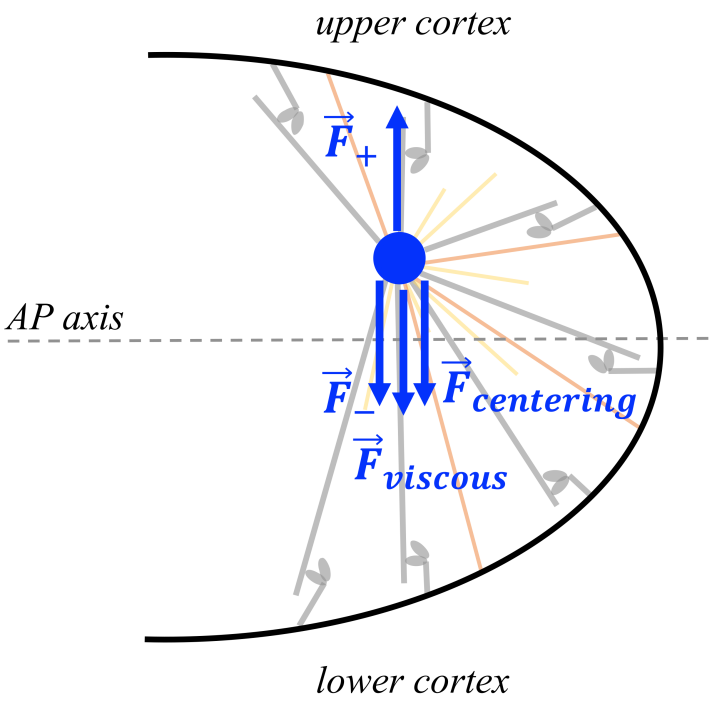


Figure 3.7 – **Schematic representation of the forces acting on the posterior pole,** assuming the pole is moving upwards.

- F_-: total force exerted by the motors of the lower cortex;
- $F_{viscous}$: viscous force, that opposes the movement of the centrosome;
- $F_{centering}$: centering force, that pulls the centrosome back towards the AP axis.

The standard expression for the centering force is $F_{centering} = -K_0\, y$ where K_0 is the elasticity coefficient and y is the vertical position (perpendicular to the AP axis) of the centrosome. In a similar fashion, the viscous force is written as $F_{viscous} = -\Gamma\, \dot{y}$, where Γ is the viscosity coefficient and $\dot{y}$ the velocity of the centrosome.

Three assumptions are used for building our model.

Assumption 1 concerns the "state" of the microtubules: they can be either pulling on the centrosome or not. A microtubule is not pulling on the centrosome, either because there is no molecular motor where it is attached to the cortex, or because it is not long enough to be in contact with the cortex. In what follows, we will be referring to microtubules that are pulling on the centrosome as *attached*, and

to microtubules that are not pulling as *detached*. Let p be the probability that a microtubule is attached (thus, $1-p$ is the probability that it is detached). This probability depends on binding and unbinding rates of the microtubules to the cortex, denoted by k_{on} and k_{off}. The evolution law of p is then given by:

$$\dot{p}(t) = -k_{off}(t)p(t) + k_{on}(1-p(t)). \tag{3.1}$$

Note that, when a microtubule is detached, it may contribute to the viscosity force.

Assumption 2 concerns the relation between the force f exerted on a centrosome by a motor, and the velocity $\dot{y}$ of the motor, taken to be equal to the velocity of the centrosome (i.e., rigid microtubules are considered). A first-order relationship between the force and the velocity is assumed, that is:

$$f \approx \bar{f} - f_p \dot{y}. \tag{3.2}$$

where $\bar{f}$ denotes the "reversal force", above which the direction of the motor is reversed, and f_p is the slope of the force-velocity curve.

Assumption 3 concerns the relation between the force exerted on a centrosome by a motor, and the detachment rate of the corresponding microtubule: the greater the exerted force, the higher the chance for the microtubule to detach. An exponential relation is assumed, under the form:

$$k_{off}(t) = k_0 \exp\left(\frac{|f|}{f_c}\right), \tag{3.3}$$

where f_c represents the sensitivity to the pulling force generated by the motor.

The force balance equation writes as follows:

$$\Gamma\, \dot{y} + K_0\, y = F_+ + F_- \tag{3.4}$$

where

$$\begin{cases} F_+ = Np_+f_+ & \text{(3.5a)} \\ F_- = Np_-f_- & \text{(3.5b)} \\ f_+ \approx \bar{f} - f_p\dot{y} & \text{(3.5c)} \\ f_-(\dot{y}) = -f_+(-\dot{y}) & \text{(3.5d)} \\ p_-(\dot{y},\ddot{y}) = p_+(-\dot{y},-\ddot{y}) & \text{(3.5e)} \end{cases}$$

In this system, N denotes the number of active molecular motors, and p_+ (respectively, p_-) is the probability of attachment of a microtubule at the upper (resp., lower) cortex.

Let us first compute p. We use *Assumption 1*, and write:

$$\dot{p}(t) = -k_{off}(t)p + k_{on}(1-p). \tag{3.6}$$

Hence, for the upper cortex and the lower cortex respectively, one has:

$$\dot{p}_+(t) = -k^+_{off}(t)p + k_{on}(1-p), \tag{3.7}$$
$$\dot{p}_-(t) = -k^-_{off}(t)p + k_{on}(1-p). \tag{3.8}$$

Then, by using *Assumption 3*, then *Assumption 2*:

$$\begin{aligned} k^+_{off}(t) &= k_0 \exp\left(\frac{|f_+|}{f_c}\right) = k_0 \exp\left(\frac{\bar{f} - f_p\dot{y}}{f_c}\right) \\ &= k_0 \exp\left(\frac{\bar{f}}{f_c}\right) \exp\left(-\frac{f_p}{f_c}\dot{y}\right). \end{aligned}$$

Assuming $-\frac{f_p}{f_c}\dot{y}$ small, a second order Taylor expansion of the exponential is used: [1]

$$\exp\left(-\frac{f_p}{f_c}\dot{y}\right) \approx 1 - \frac{f_p}{f_c}\dot{y} + \frac{1}{2}\left(\frac{f_p}{f_c}\right)^2 (\dot{y})^2. \tag{3.9}$$

1. We chose to use a second-order expansion here after we developed and tested a first complete model using a first-order Taylor expansion. We observed that this first model did not capture oscillations with large amplitudes. The second-order expansion makes it possible to model a wider range of behaviors.

Thus :

$$k_{off}^{+}(t) = k_0 \exp\left(\frac{\bar{f}}{f_c}\right)\left(1 - \frac{f_p}{f_c}\dot{y} + \frac{1}{2}\left(\frac{f_p}{f_c}\right)^2(\dot{y})^2\right) \tag{3.10}$$

$$= \overline{k_{off}}\left(1 - \frac{f_p}{f_c}\dot{y} + \frac{1}{2}\left(\frac{f_p}{f_c}\right)^2(\dot{y})^2\right), \tag{3.11}$$

where $\overline{k_{off}}$ is defined as:

$$\overline{k_{off}} = k_0 \exp\left(\frac{\bar{f}}{f_c}\right). \tag{3.12}$$

In the same manner, for the lower cortex, one obtains:

$$k_{off}^{-}(t) = k_0 \exp\left(\frac{|f_-|}{f_c}\right) = k_0 \exp\left(\frac{\bar{f} + f_p\dot{y}}{f_c}\right) \tag{3.13}$$

$$= k_0 \exp\left(\frac{\bar{f}}{f_c}\right)\exp\left(\frac{f_p}{f_c}\dot{y}\right) \tag{3.14}$$

$$= k_0 \exp\left(\frac{\bar{f}}{f_c}\right)\left(1 + \frac{f_p}{f_c}\dot{y} + \frac{1}{2}\left(\frac{f_p}{f_c}\right)^2(\dot{y})^2\right) \tag{3.15}$$

$$= \overline{k_{off}}\left(1 + \frac{f_p}{f_c}\dot{y} + \frac{1}{2}\left(\frac{f_p}{f_c}\right)^2(\dot{y})^2\right) \tag{3.16}$$

Using (3.11) and (3.16) in the (3.7) and (3.8) respectively, it yields:

$$\dot{p}_+(t) = k_{on}(1 - p_+) - p_+\overline{k_{off}}\left(1 - \frac{f_p}{f_c}\dot{y} + \frac{1}{2}\left(\frac{f_p}{f_c}\right)^2(\dot{y})^2\right). \tag{3.17}$$

$$\dot{p}_-(t) = k_{on}(1 - p_-) - p_-\overline{k_{off}}\left(1 + \frac{f_p}{f_c}\dot{y} + \frac{1}{2}\left(\frac{f_p}{f_c}\right)^2(\dot{y})^2\right). \tag{3.18}$$

At this point in their modeling, Pécréaux et al. [62] make use of a supplementary assumption on the dynamics of p_+ and p_-, which are directly related to y via a linear second-order ODE; this results in a constraint on the shape of the solutions. We choose instead to keep the complete expressions of these variables, in order to keep our model closer to the dynamics of the oscillations.

Let us define $p_1 = p_+ - p_-$ and $p_2 = p_+ + p_-$. Then, by adding and subtracting (3.17)

and (3.18), one obtains:

$$\dot{p}_2(t) = 2k_{on} - (k_{on} + \overline{k_{off}})p_2 + \overline{k_{off}}\frac{f_p}{f_c}\dot{y}p_1 - \frac{\overline{k_{off}}}{2}\left(\frac{f_p}{f_c}\right)^2(\dot{y})^2\,p_2, \tag{3.19}$$

$$\dot{p}_1(t) = -(k_{on} + \overline{k_{off}})p_1 + \overline{k_{off}}\frac{f_p}{f_c}\dot{y}p_2 - \frac{\overline{k_{off}}}{2}\left(\frac{f_p}{f_c}\right)^2(\dot{y})^2\,p_1. \tag{3.20}$$

Let us now determine an expression for $\dot{y}(t)$. From the force balance (3.4) and (3.5a), (3.5b), (3.5c), (3.5d), it yields:

$$\begin{aligned}
\Gamma\,\dot{y} + K_0\,y &= F_+ + F_- \\
\Gamma\,\dot{y} + K_0\,y &= Np_+f_+ + Np_-f_- \\
&= Np_+\left(\bar{f} - f_p\dot{y}\right) + Np_-\left(-\bar{f} - f_p\dot{y}\right) \\
&= N\bar{f}\,(p_+ - p_-) + Nf_p\,(-p_+ - p_-)\,\dot{y} \\
&= N\bar{f}\,(p_+ - p_-) - Nf_p\,(p_+ + p_-)\,\dot{y}.
\end{aligned}$$

Then:

$$\Gamma\,\dot{y} + Nf_p\,(p_+ + p_-)\,\dot{y} = N\bar{f}\,(p_+ - p_-) - K_0\,y. \tag{3.21}$$

Hence:

$$\dot{y} = \frac{N\bar{f}\,(p_+ - p_-) - K_0\,y}{\Gamma + Nf_p\,(p_+ + p_-)}. \tag{3.22}$$

By introducing p_1 and p_2, one obtains:

$$\dot{y} = \frac{N\bar{f}p_1 - K_0\,y}{\Gamma + Nf_pp_2}. \tag{3.23}$$

Finally, the proposed mathematical model is summarized as follows:

$$\boxed{\begin{cases}
\dot{y} = \dfrac{N\bar{f}\,p_1 - K_0\,y}{\Gamma + Nf_p\,p_2} \\
\dot{p}_1 = -(k_{on} + \overline{k_{off}})\,p_1 + \overline{k_{off}}\dfrac{f_p}{f_c}\,\dot{y}\,p_2 - \dfrac{\overline{k_{off}}}{2}\left(\dfrac{f_p}{f_c}\right)^2(\dot{y})^2\;p_1 \\
\dot{p}_2 = 2k_{on} - (k_{on} + \overline{k_{off}})\,p_2 + \overline{k_{off}}\dfrac{f_p}{f_c}\,\dot{y}\,p_1 - \dfrac{\overline{k_{off}}}{2}\left(\dfrac{f_p}{f_c}\right)^2(\dot{y})^2\;p_2
\end{cases}} \tag{3.24}$$

In model (3.24), the parameters to be estimated are N, $\bar{f}$, K_0, Γ, k_{on}, $\overline{k_{off}}$, f_p and f_c.

In addition, Pécréaux *et al.* [62] showed that a monotonic decrease of $\overline{k_{off}}$ accounts for the build-up and die down of the oscillations. In what follows, we model $\overline{k_{off}}$ in the same way, that is:

$$\overline{k_{off}}(t) = \frac{a - \tilde{k}_{end}}{1 + \exp(dt - b)} + \tilde{k}_{end}. \tag{3.25}$$

Parameters a, b, d and $\tilde{k}_{end}$ will thus be estimated alongside the other parameters of (3.24).

3.2.3 Simulations and model analysis

As discussed above, there is a lack of quantitative evaluation of the parameters involved in the oscillations of the spindle poles during anaphase in C.elegans. Our aim is to better understand the relation between these parameters and the observed phenotype, both for individual embryos and between populations of embryos. To this purpose, we start by analyzing the simulations given by the proposed model with respect to the characteristics of the oscillations, then we connect these characteristics to the parameter values. Matlab (version 9.4) was used to simulate the oscillations model.

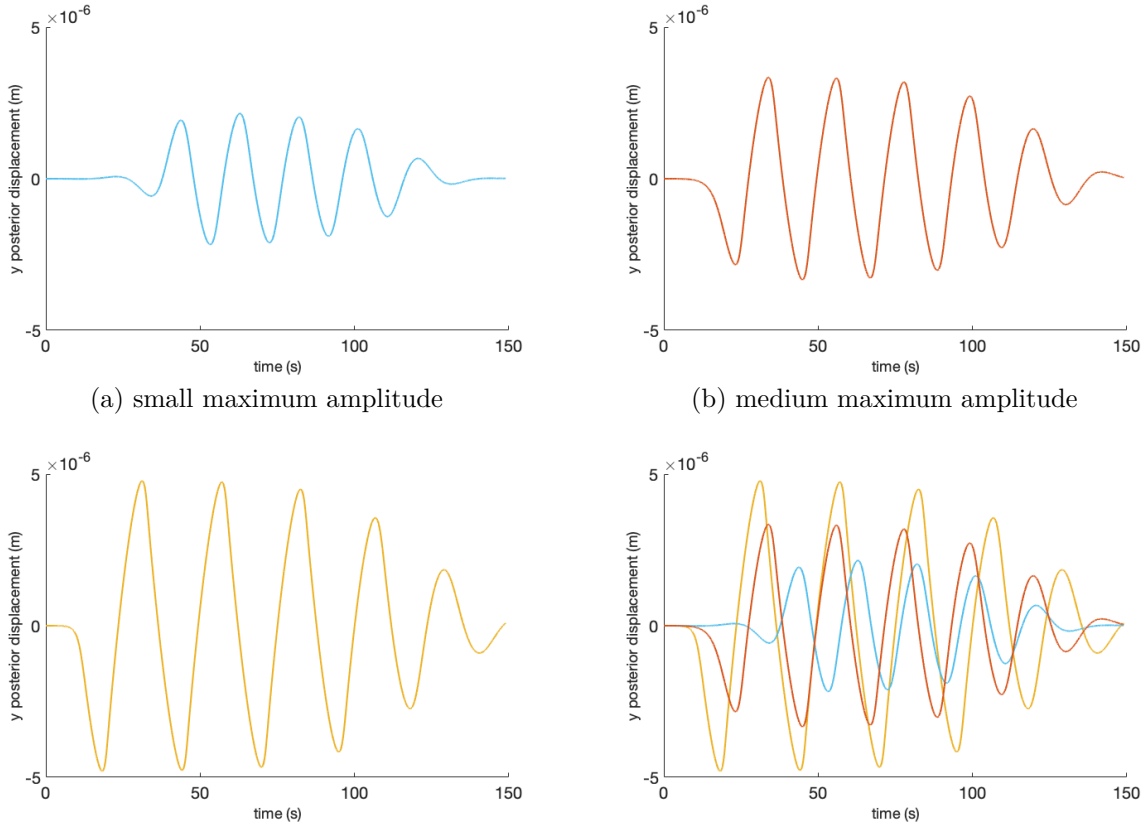
We first describe how our model adapts to variations in oscillations and, in particular, to three characteristics of interest observed on wild-type embryos. We then illustrate the influence of key parameters on these characteristics. Finally, we show how our model behaves for limit values of these parameters and we deduce necessary ranges of these parameters for oscillations to still be generated.

3.2.3.1 Simulations and phenotypic variability

Oscillations, considered as a phenotypic trait, show a great variability in wild-type embryos as well as in RNAi embryos [85]. Here, we show that this variability is well accounted for by our model, and in particular, the three main types of variation given by the model will be discussed :

- variation in maximum amplitude of oscillations;
- variation in frequency of oscillations;
- variation in number of oscillations.

Variation in maximum amplitude A large spectrum of values for the maximum amplitude is possible with our model. Simulations with varying maximum amplitude are



(a) small maximum amplitude

(b) medium maximum amplitude

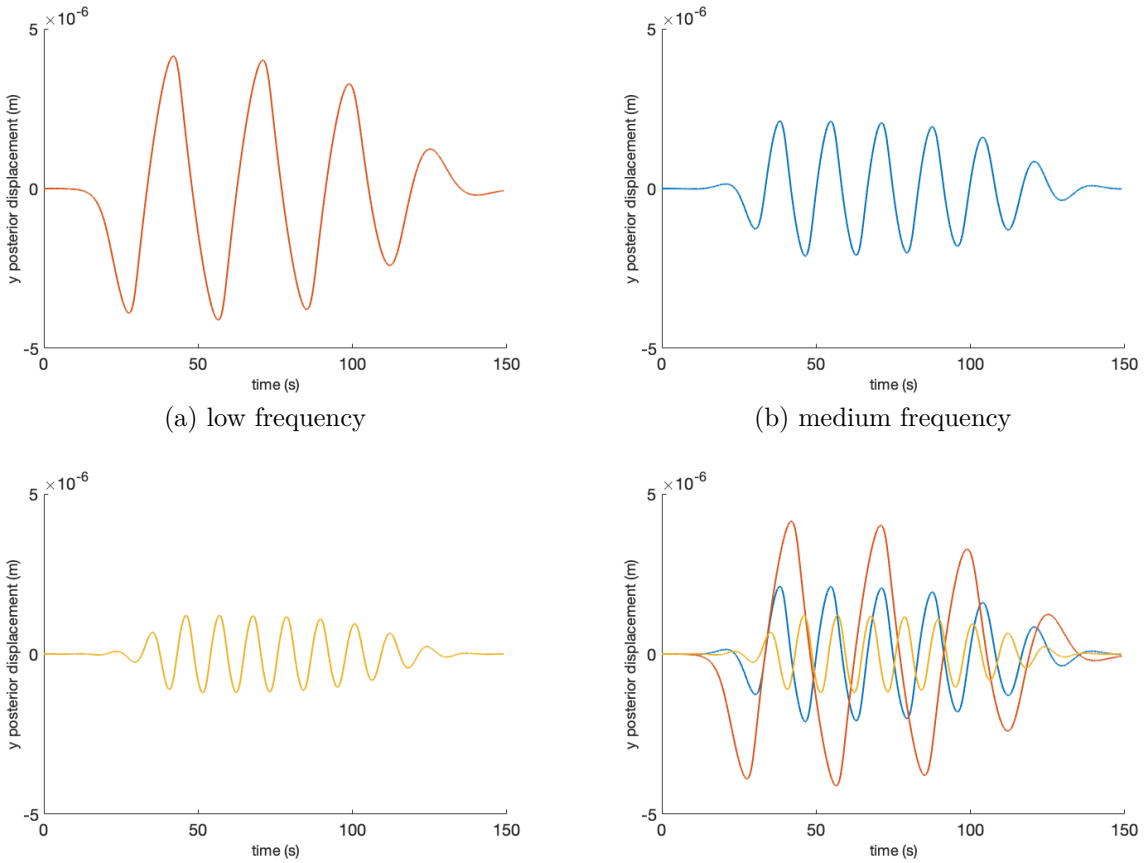
(c) large maximum amplitude

(d) all simulations

Figure 3.8 – **Oscillation variability illustrated by our model.** Variation of maximum amplitude across simulations.

given in Fig. 3.8. The regular range of maximum amplitude in wild-type embryos is in the order of micrometers. The provided simulation examples go from 2 μm in Fig. 3.8a, which corresponds to a rather small maximum amplitude, to 3.5 μm in Fig. 3.8b, which is around the mean maximum amplitude observed in wild-type embryos, and up to approximately 5 μm in Fig. 3.8c, which is a rather large maximum amplitude in wild-type embryos. Fig. 3.8d shows the three simulations together, for a better appreciation of the variation in maximum amplitude. In RNAi embryos, the mean maximum amplitude varies significantly depending on the condition, ranging from 1 μm to 7 μm. Our model accounts for variation in maximum amplitude for all embryos, wild-type and perturbed alike.

Variation in frequency Although less important than those on amplitude, variations of frequency of oscillations in wild-type embryos of C.elegans are observed. Our model



(a) low frequency (b) medium frequency

(c) high frequency (d) all simulations

Figure 3.9 – **Oscillation variability illustrated by our model.** Variation of frequency across simulations.

allows for some variation in frequency, with a limited impact for most of the parameters. In Figure 3.9, several scenarios for the frequency are given : an example of low frequency of 0.035 *Hz* in Fig. 3.9a, a regular frequency of 0.06 *Hz* in Fig. 3.9b, and a high frequency of 0.09 *Hz* in Fig. 3.9c. In wild-type embryos, the typical frequency is between 0.05 and 0.06 *Hz*. Thus, our model takes into account the variability in the frequency of oscillations, and allows for simulations with frequency values ranging from 0.02 *Hz* to 0.1 *Hz*.

Variation in number of oscillations Our model also allows for variations in the number of oscillations. In wild-type and perturbed embryos, there is significant variability in the number of oscillations. As an example, Fig. 3.10 shows several simulations of our model with different numbers of oscillations, from 3 oscillations and up to 6 oscillations. In wild-type embryos, the number of oscillations typically ranges between 3 and 5, but in

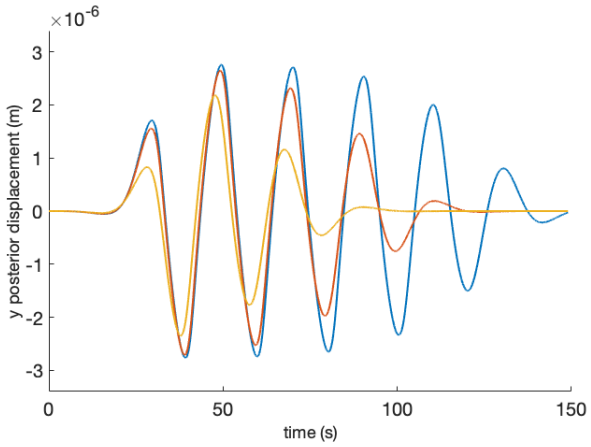


Figure 3.10 – **Oscillation variability illustrated by our model.** Variation of the number of oscillations across simulations: 3 oscillations (in yellow), 4 oscillations (in red) and 6 oscillations (in blue).

RNAi embryos, this number can vary substantially. In some conditions, the oscillations are not present at all, or are highly impacted with only 1 or 2 oscillations, while in other conditions, the number of oscillations goes up to 6 or 7 oscillations. Our model is able to take into account all these scenarios.

Other types of variations Variability in oscillations is very clear during anaphase in the C.elegans embryo. We already showed that out model produces simulations that take into account the main sources of variability: maximum amplitude, frequency and number of oscillations. However, there are other secondary sources of variability in oscillations that can be identified, and that are not handled by our model. Specifically, our model does not account for aborted oscillations that can sometimes be seen in experimental data, and it does not model trajectory drifts (i.e., cases in which the mean value of the vertical position of the pole is not constant). The aborted oscillations are probably due to resynchronization mechanisms between the two poles. Our model focuses on the movement of a single pole, and thus, cannot take these mechanisms into account.

Other types of variations, concerning the total time of oscillations, or the build-up and die-down times, are covered by our model, and will be discussed, when appropriate, later in this chapter.

3.2.3.2 Model analysis: how parameters influence the oscillations

In order to better understand the results of the parameter estimation (discussed in the next chapter), it is helpful to analyze the relation between the variation in parameter values and the variation in the characteristics of the oscillations. In this section, we explicit the influence of the variation of each parameter (taken individually) on the oscillations. The default parameters values are set in accordance to values estimated or computed in previous papers [62, 35, 30], and such that a typical wild-type oscillatory behavior is observed:

$$\begin{cases} N = 80 \\ \bar{f} = 7 \times 10^{-12}\ N \\ f_p = 1.2 \times 10^{-6}\ N \cdot s/m \\ f_c = 7 \times 10^{-13}\ N \\ K_0 = 50 \times 10^{-6}\ N/m \\ \Gamma = 400 \times 10^{-6}\ N \cdot s/m \\ k_{on} = 0.6\ s^{-1} \\ a = 1.2 \\ b = 7 \\ d = 0.165 \\ \tilde{k}_{end} = 0.2 \end{cases} \tag{3.26}$$

Parameters $\bar{f}$, f_p and f_c are related to the forces generated by the motors. Several studies suggest that a single type of motor is responsible for the forces generated at the cortex. Thus, these parameters are fixed to the default values provided by the literature [83, 62].

Regarding the parameters related to $\overline{k_{off}}$, the value of $\tilde{k}_{end}$ can be fixed without loss of generality: the impact of $\overline{k_{off}}$ on the oscillations directly stems from its variation over time, which can be controlled by the parameters a, b and d only.

Finally, parameters K_0 and Γ are physical parameters whose values are hard to interpret at a molecular level, which is why we leave them out of the following analysis.

For these reasons, we only study here the influence of parameters N, k_{on}, a, b and d on the oscillations. Appendix D completes this analysis, in order to check the correct

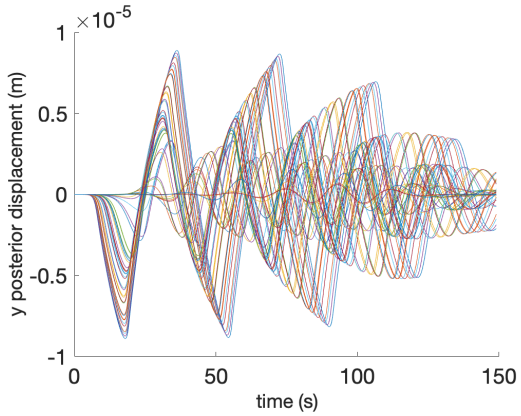
behavior of the model.

The considered characteristics of the oscillations are the maximum amplitude, the number of oscillations and the frequency. In particular, specific computations are required in order to compute the latter. We rely on a discrete Fourier transform of the signal via FFT; the component of the Fourier transform with the greatest amplitude approximates the fundamental frequency of the signal, i.e., the frequency of the oscillations. All signals are subject to the same time discretization. As a result, all approximated frequencies lie in the same finite set of possible frequencies provided by the FFT. Hence, plotting the computed frequencies with respect to the values of a parameter results in a step-wise approximation of the actual behavior. Such plots provide us with valuable information about the frequency range and general trend. More precise results could be obtained by increasing the sampling rate.

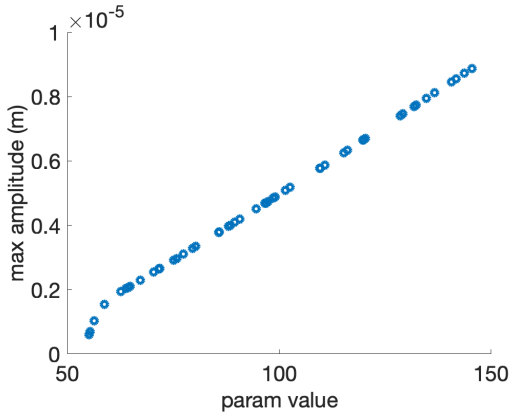
In what follows, we summarize the results of this analysis in a table format : the first column defines the parameter of interest and gives the tested range of values, and the second column reports the most important observations with respect to the influence of the parameter on the characteristics of the oscillations.

Parameter range and definition	**Influence on the oscillations**
$N : [50; 150]$ Number of molecular motors generating pulling forces.	*Maximum amplitude* : increases linearly with the value of N (see Fig. 3.11b). *Frequency* : decreases with the value of N, but remains close to typical values (see Fig. 3.11c). *Build up and die-down* : the timing of the beginning of the oscillations is fairly stable, whereas the end time varies considerably (see Fig. 3.11a). *Number of oscillations* : varies slightly.
$k_{on} : [0.5; 1.6]\ s^{-1}$ Binding rate.	*Maximum amplitude* : decreases linearly with the value of k_{on} (see Fig. 3.12b). *Frequency* : increases with the value of N, but remains close to typical frequency values (see Fig. 3.12c). *Build up and die-down* : the timing of the beginning of the oscillations is fairly stable, whereas the end time varies considerably (see Fig. 3.12a). *Number of oscillations* : varies slightly.
$a : [0.38; 1.5]$ Starting value of the unbinding rate.	*Maximum amplitude* : increases with the value of a, then reaches a plateau at $a \approx 1.2$ (see Fig. 3.13b). *Frequency* : mostly stable (see Fig. 3.13c). *Build up and die-down* : the timing of the beginning and ending of the oscillations varies largely; this variation seems to follow a step-wise pattern and might be correlated to the addition/subtraction of a new oscillation. (see Fig. 3.13a). *Number of oscillations* : increases step-wise.

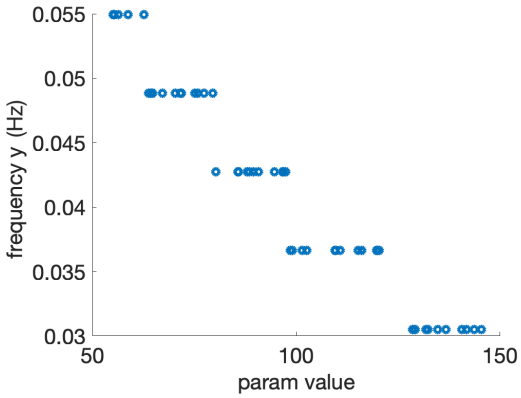
$b : [3; 10]$ Delay after which the unbinding rate starts to decay.	*Maximum amplitude* : moderately impacted; it increases with the value of b, but quickly reaches a plateau (see Fig. 3.14b). *Frequency* : almost constant (see Fig. 3.14c). *Build up and die-down* : the timing of the beginning of oscillations is not influenced by b. The build-up of oscillations follows the same pattern for all simulations, but the die-down pattern and timing change in a slow monotonic fashion (see Fig. 3.14a). *Number of oscillations* : increases monotonically with b.
$d : [0.05; 0.22]$ Slope of the unbinding rate.	*Maximum amplitude* : moderately impacted; it decreases with the value of d, starting from a relative plateau (see Fig. 3.15b). *Frequency* : fairly stable (see Fig. 3.15c). *Build up and die-down* : the timing of the beginning and the pattern of the build-up are the same for all simulations, with only the amplitude changing slightly. The die-down timing and the number of oscillations decrease rapidly (after the plateau values) with increasing values of d (see Fig. 3.15a). *Number of oscillations* : monotonically decreases while d increases.



(a) y vs t

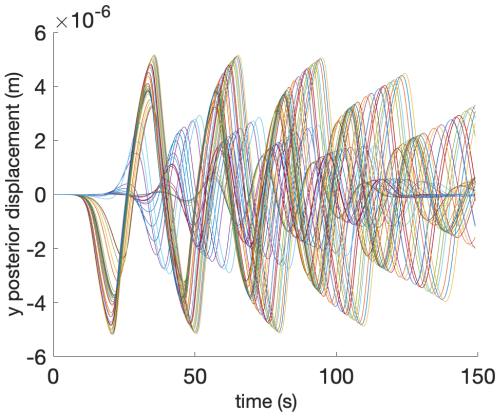


(b) max ampl vs param

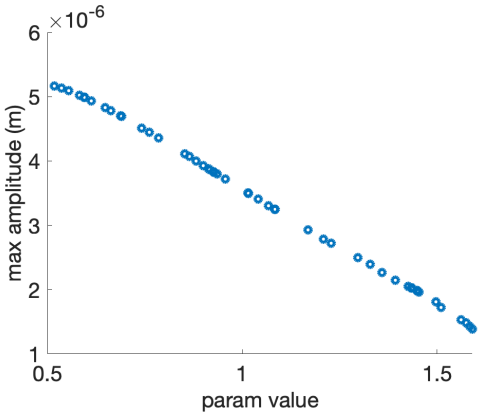


(c) freq vs param

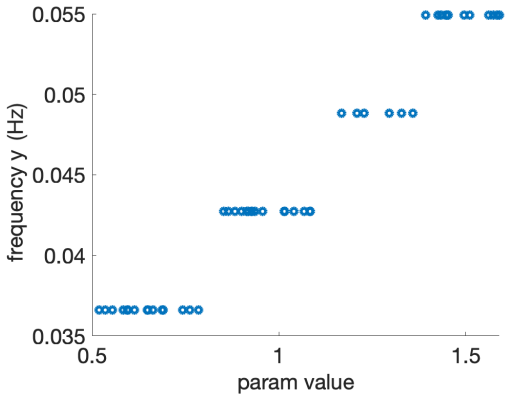
Figure 3.11 – **Influence on the oscillations of the values of the parameter** N**.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.



(a) y vs t

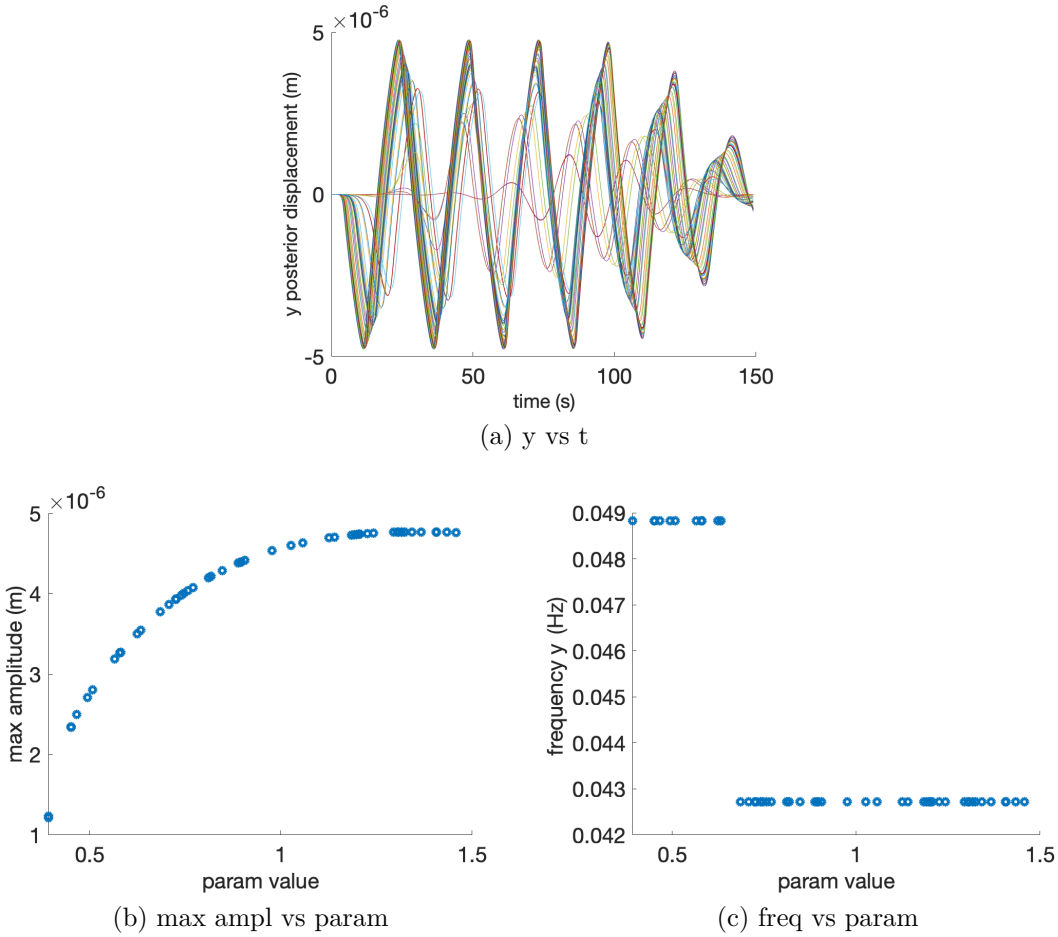


(b) max ampl vs param



(c) freq vs param

Figure 3.12 – **Influence on the oscillations of the values of the parameter k_{on}.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.

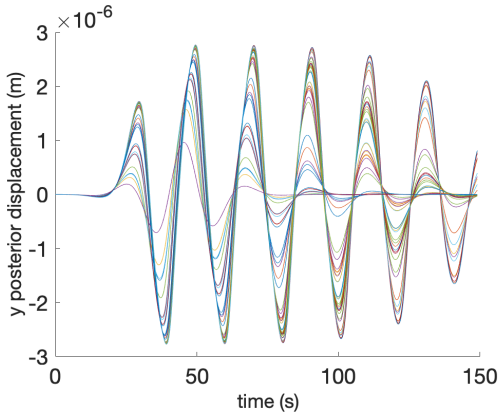


(a) y vs t

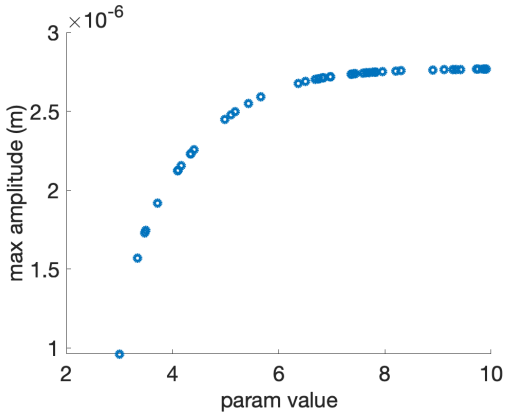
(b) max ampl vs param

(c) freq vs param

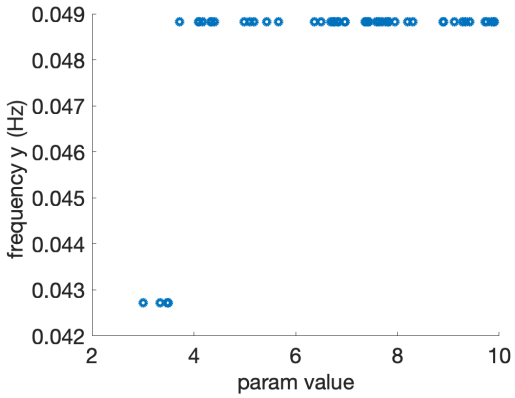
Figure 3.13 – **Influence on the oscillations of the values of the parameter** a**.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.



(a) y vs t

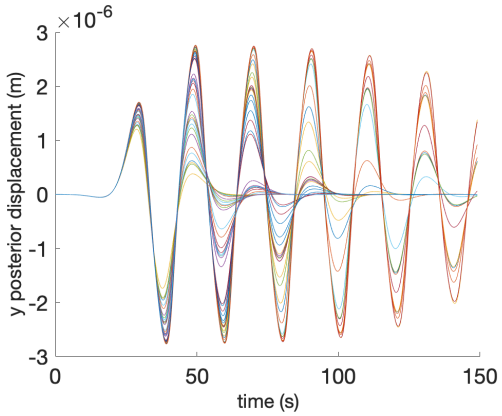


(b) max ampl vs param

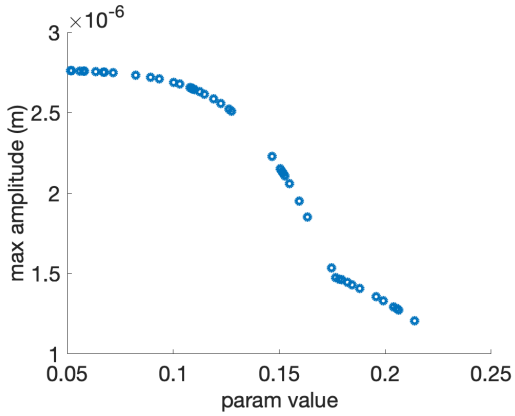


(c) freq vs param

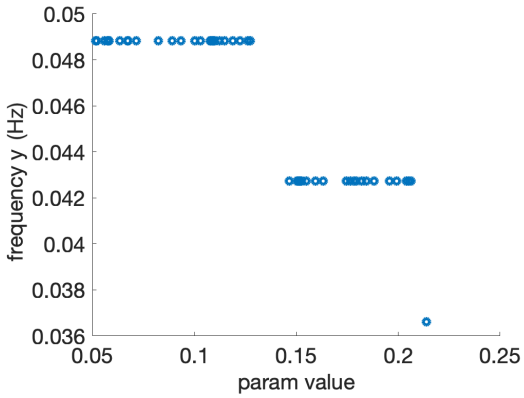
Figure 3.14 – **Influence on the oscillations of the values of the parameter** b**.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.



(a) y vs t



(b) max ampl vs param



(c) freq vs param

Figure 3.15 – **Influence on the oscillations of the values of the parameter** d**.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.

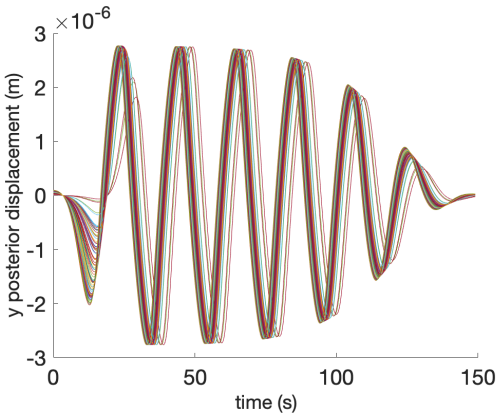
Robustness of the model to parameter Y_0 : Although not an explicit model parameter, the influence of Y_0 – the initial value of the vertical position of the pole, required for model simulation – is of particular interest. Indeed, this initial value is typically set by default, as measuring it from the acquired real microscopy data is not possible; thus, robustness to this value is crucial in order to avoid significant numerical errors. The influence of the parameter Y_0 was thus tested here, with Y_0 varying from 0.01 nm to 0.1 μm, the other parameters being fixed to the default values given above, and 80 simulations of the model were generated.

Figure 3.16 shows the influence of the parameter Y_0 on the oscillations. The maximum amplitude (see Fig. 3.16b) and frequency (see Fig. 3.16c) are almost fixed. The only noticeable effect is on the beginning of the oscillations, with a half-oscillation being added for larger values of Y_0 (see Fig. 3.16a). A slight shift in time can also be observed. Overall, the influence of the parameter Y_0 on the oscillations is very limited, which means that our model is robust to this parameter and that its value can be set by default as convenient.

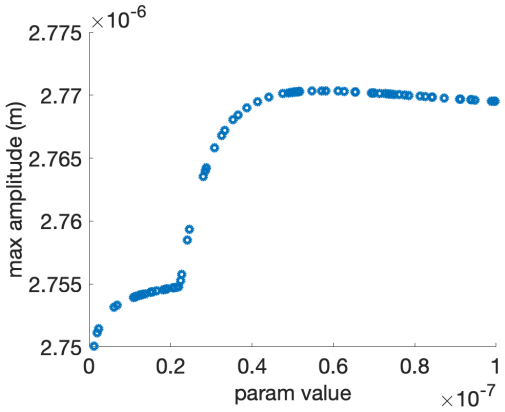
3.2.3.3 Model analysis: limits of parameter values

In this section, we explore the limit oscillatory behavior given by our model and discuss the extremal parameter values with respect to the characteristics of the oscillations. In addition to testing the limits of the model and comparing its behavior to observations on wild-type embryos, this analysis provides bounds to drive oscillations, and the transitory behavior of the model from oscillatory to non-oscillatory. For this purpose, we proceed as in the previous section: the parameters are set to default values such that the simulation reproduces the behavior of a typical wild-type embryo. Then each parameter is tested individually. The minimal and maximal ranges of values are empirically obtained by performing large numbers of simulations.

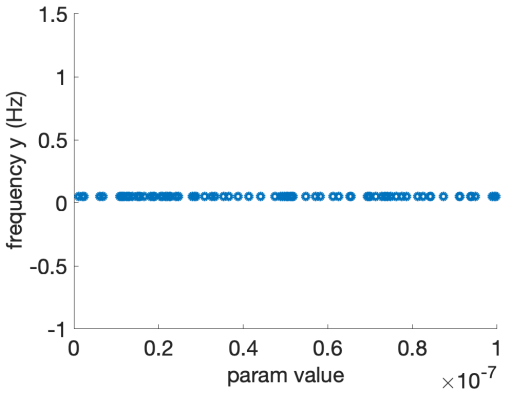
Limit behavior for the parameter N The lower limit of N was tested in the range $[20, 55]$, and the upper limit in the range $[150, 200]$. As seen in the previous section, the maximum amplitude increases with the values of N. When the value of N is around 50, oscillations are still present, but their amplitude decreases with smaller values, as shown in Fig. 3.17b. When N gets even lower, i.e., $N < 40$, oscillations do not build up anymore; instead, they die out immediately, starting from the initial position $Y_0 = 1$ nm, or even become nonexistent when N gets closer to 20, in which case the trajectory of Y



(a) y vs t



(b) max ampl vs param

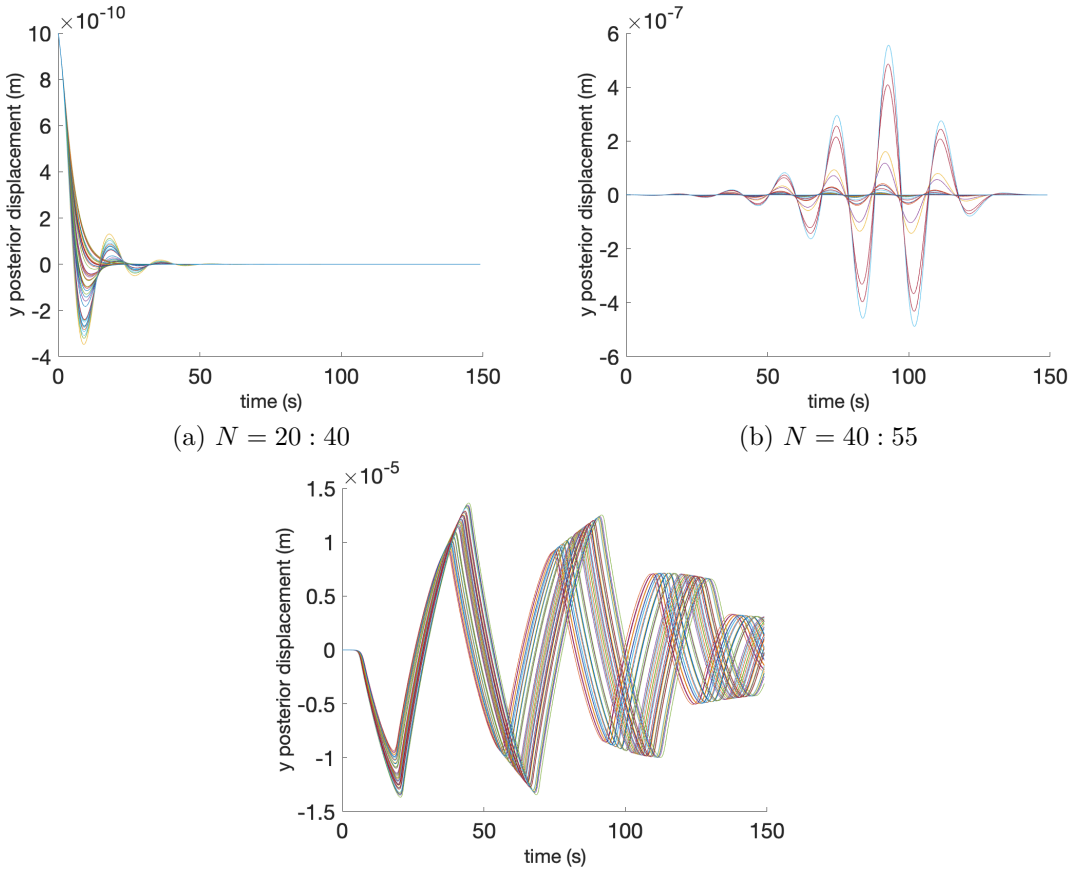


(c) freq vs param

Figure 3.16 – **Influence of the value of parameter Y_0 on the oscillations.** (a) Oscillations of the pole for different values. (b) Maximum amplitude with respect to the parameter value. (c) Frequency of oscillations with respect to the parameter value.

follows a simple decay to 0, as shown in Fig. 3.17a. In contrast, for high values of N, the oscillations continue to be present, as shown in Fig. 3.17c, but the pattern is not realistic. The amplitude continues to increase with increasing values of N, while the frequency and number of oscillations decrease.

Limit behavior for the parameter k_{on} The lower limit of k_{on} was tested in the range $[0.4, 0.7]$ s^{-1}, and the upper limit in the range $[1.7, 2.1]$ s^{-1}. For low values of k_{on} (see Fig. 3.18a), the oscillations occur, the maximum amplitude is high, the frequency is low, and the number of oscillations is stable (this last part is actually the case for all values of k_{on}, as this parameter has a very limited impact on the number of oscillations). For high values of k_{on} (see Figs. 3.18b and 3.18c), the oscillations decrease in maximum amplitude until



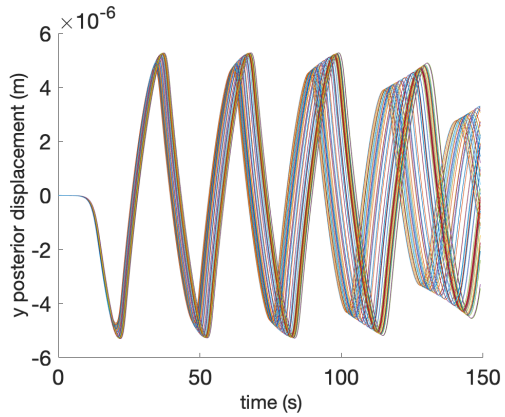
(a) $N = 20 : 40$ (b) $N = 40 : 55$

(c) $N = 150 : 200$

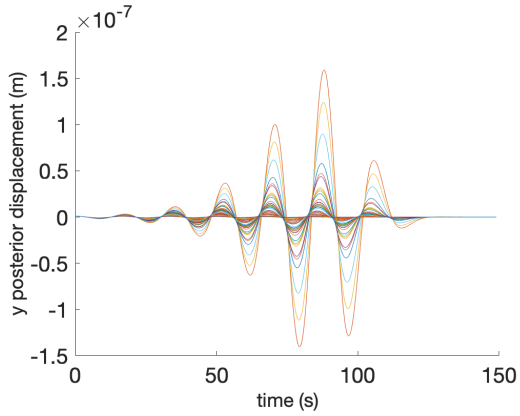
Figure 3.17 – **Limit behavior of the model for extremal values of the parameter N.**

there is no build-up, and finally disappear. The maximum amplitude decreases steadily. The frequency and the number of oscillations are stable for this range of values.

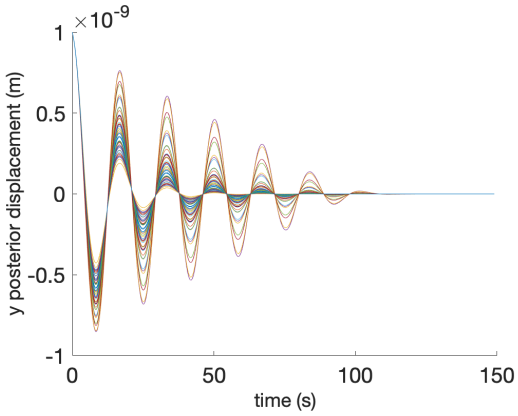
Limit behavior for the parameters related to k_{off} The limit behavior of the model for extremal values of the parameters a, b, and d were tested individually. The lower limit of a was tested in the range $[0.34, 0.38]$, and the upper limit in the range $[1.40, 1.80]$. The results for low values of a are given in Figs. 3.19a and 3.19b. The oscillations decrease steadily in maximum amplitude, whereas the frequency and number of oscillations are stable. For high values of a, the maximum amplitude is high but almost steady, and the frequency and number of oscillations are stable, as shown in Fig. 3.19c.



(a) $k_{on} = 0.4 - 0.7\ s^{-1}$



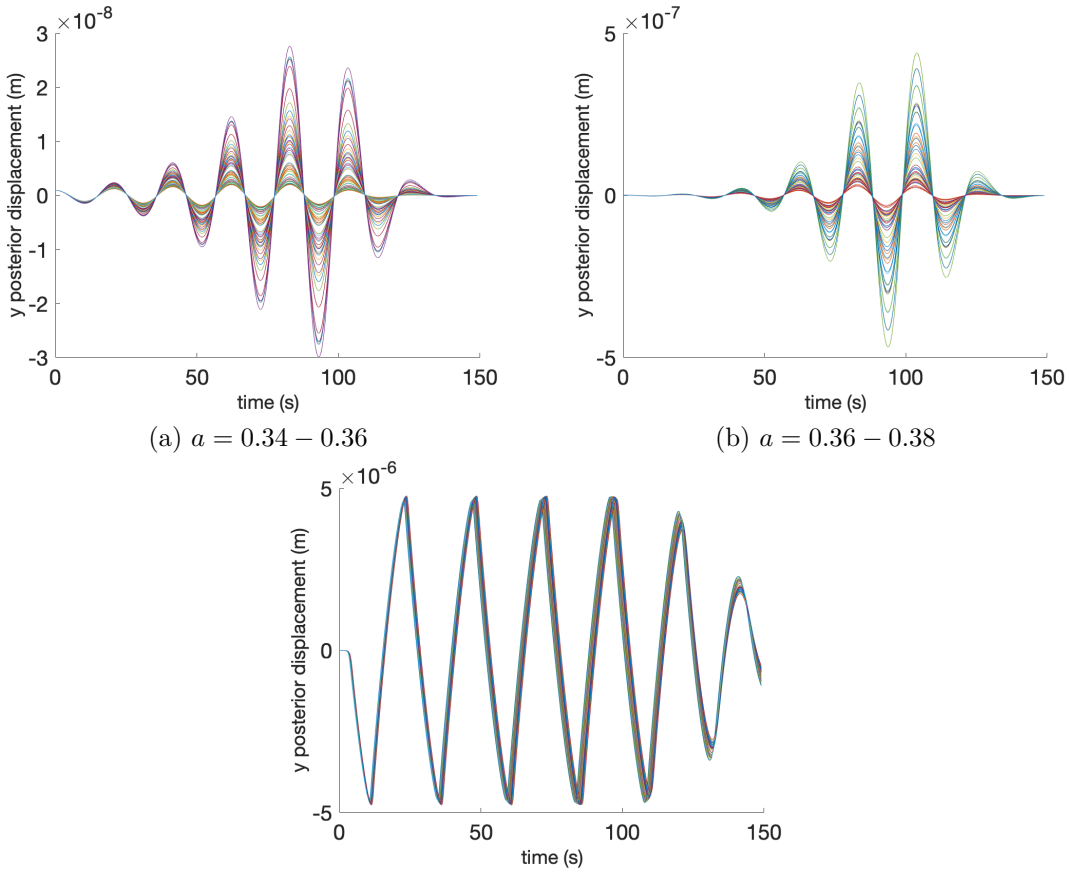
(b) $k_{on} = 1.7 - 1.9\ s^{-1}$



(c) $k_{on} = 1.9 - 2.1\ s^{-1}$

Figure 3.18 – **Limit behavior of the model for extremal values of the parameter k_{on}.**

The limit behavior for extremal values of the parameters b and d mirror each other: the behavior for low values of b corresponds to the behavior for high values of d, and the other way around. The parameter b was tested in the range $[0.30, 2.50]$ for low values, and in the range $[9, 12]$ for high values. The parameter d was tested in the range $[0.04, 0.06]$ for low values, and in the range $[0.20, 0.40]$ for high values. For low values of b (respectively, high values of d), the oscillations decrease in maximum amplitude, but the frequency is stable, as shown in Figs. 3.20a and 3.20b (resp., Figs. 3.21b and 3.21c). For high values of b (respectively, low values of d), the oscillations get to a stable pattern that is realistic, with high amplitude, high number of oscillations, and stable frequency, as can be seen in Fig. 3.20c (resp., Fig. 3.21a). The beginning and build-up of oscillations are not impacted at all by the values of these parameters.

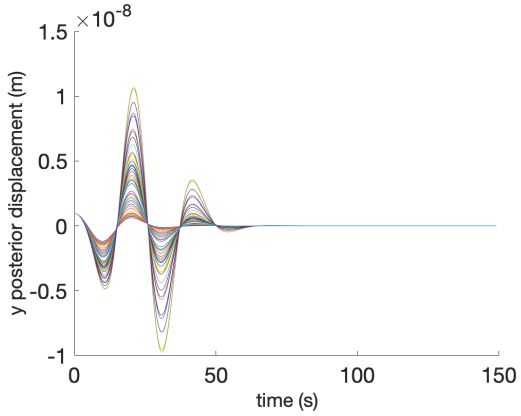


(a) $a = 0.34 - 0.36$

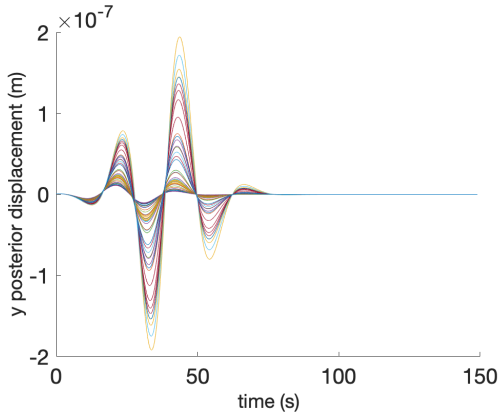
(b) $a = 0.36 - 0.38$

(c) $a = 1.40 - 1.80$

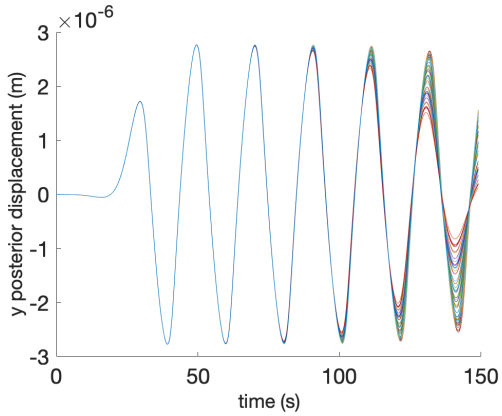
Figure 3.19 – **Limit behavior of the model for extremal values of the parameter a.**



(a) $b = 0.30 - 1.50$

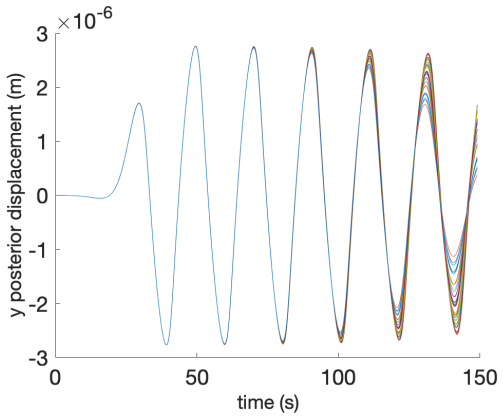


(b) $b = 1.50 - 2.50$

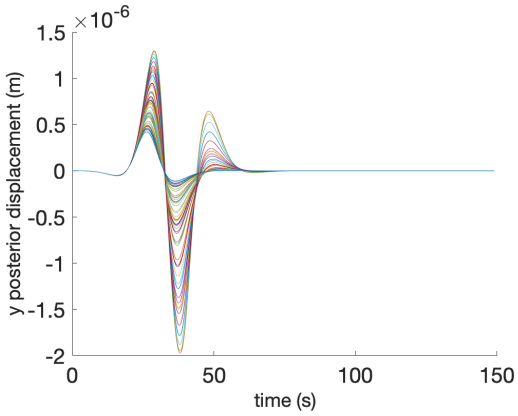


(c) $b = 9 - 12$

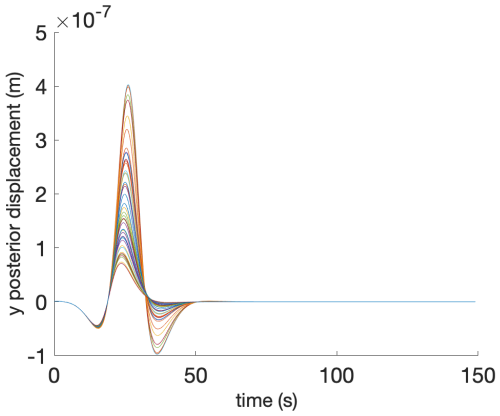
Figure 3.20 – **Limit behavior of the model for extremal values of the parameter b.**



(a) $d = 0.04 - 0.06$



(b) $d = 0.20 - 0.30$



(c) $d = 0.30 - 0.40$

Figure 3.21 – **Limit behavior of the model for extremal values of the parameter d.**

Discussion The general patterns observed in the previous section with respect to the characteristics of the oscillations were confirmed for the extremal values for each parameter; no sudden change of behavior is observed in the simulations when a parameter transitions from characteristic values to extremal ones. Increasing values for parameters N, a and b yields an increase in maximum amplitude, whereas parameters k_{on} and d have the opposite effect on amplitude. In all cases, the pattern is monotonic and the rate of increase/decrease is fairly constant. Only the parameter N displays, for high values, a behavior that might not correspond to real data, as the oscillations are too sharp for values close to 200. However, the maximum amplitude for such values is higher than 10 μm which is never observed in real data, where the maximum amplitude reaches, at most, 8 μm. Thus, this type of behavior, although allowed by the model, will not be attained

when estimating the parameters by fitting the real data. The frequency also follows a consistent pattern with the one observed in the previous section. Increasing extremal values for parameter N yields a decrease in frequency, whereas increasing parameters k_{on}, a, b and d results in a constant or almost constant frequency. Meanwhile, similar patterns in terms of number of oscillations occur for all the parameters: for values that result in low maximum amplitudes, the number of oscillations varies monotonically, in a way consistent with the previously observed pattern; for values generating high maximum amplitudes, the number of oscillations is constant or nearly so between simulations.

The limit dynamic behavior is complex to observe in real data, especially on wild-type embryos under normal conditions, that show a rather typical behavior. Yet, we showed that all observed behaviors (wild-type and RNAi embryos) can be accounted for by our model, with only the exceptions discussed in Section 3.2.3.1, in particular, cases of aborted oscillations and trajectory drifts. Furthermore, as described by our model, the system can transition from non-oscillatory to oscillatory mode, in a smooth way, without changing the underlying dynamics. This is in agreement with previous conclusions in [32, 33, 62], stating that a single process is responsible for the global behavior of the spindle poles during anaphase. The analysis of our model also shows that the transition to oscillatory mode is achieved by a change in parameter value only; this is in line with [62], showing that a threshold in active force generators is required for the beginning of the oscillations. Finally, our analysis shows that the parameters N, k_{on}, a, b and d can all be responsible for initiating or abolishing the oscillations, when their effect is considered individually. Table 3.1 gives approximates of the boundary values needed for oscillations to exist, for the five mentioned parameters.

Parameter	$\boldsymbol{N}$	$\boldsymbol{k_{on}}$ (s^{-1})	$\boldsymbol{a}$	$\boldsymbol{b}$	$\boldsymbol{d}$
Upper boundary	–	1.8	–	–	0.3
Lower boundary	45	–	0.35	0.8	–

Table 3.1 – **Approximate boundary values required for oscillations to exist.** All other parameter values are set according to (3.26).

3.3 Estimation of model parameters

3.3.1 ABC-like estimation framework

3.3.1.1 Parameter sensitivity analysis with the Morris method

Our model involves several correlated parameters (11 parameters), that can have non-linear effects on the outcome. Before estimating the parameters, we first performed a parameter sensitivity analysis in order to assess the sensitivity of the model to each parameter, which in turn, will allows us to categorize the parameters depending on their respective influence on the model outcome. This step is crucial for building an effective estimation method, as the basic parameters estimation can result in large errors because of the interaction and non-linear effects of the parameters.

There are several methods that can be used to perform a sensitivity analysis. For a review on global sensitivity analysis methods we refer the reader to [37]. Here, only a qualitative assessment of the parameters sensitivity is needed, so we chose to use a screening technique, from the family of OAT (One At a Time) methods. In particular, we focus on the Morris method [55], which allows to classify the parameters into three categories :

- Parameters that have a low influence on the outcome : these parameters can be either set to default values or let within large intervals.
- Parameters that have a linear effect on the outcome : these parameters have an easy, predictable influence on the model outcome.
- Parameters that have a non-linear effect on the outcome, with or without interaction with the other parameters : these parameters have the largest influence on the model, and must be well constrained in the estimation process, because a small variation in these parameters can result in a complex variation in the model outcome.

Screening techniques use a discretization of the parameter space in order to explore the model behavior. The aim is to find the non-influential and the most influential parameters from a reasonable number of evaluations of the model, and with realistic hypotheses on the model.

From a practical point of view, the Morris method can be implemented in four steps as

follows:

1. Building of random trajectories. The parameter space is scaled to the hypercube $[0,1]^N$ (where N is the number of parameters), then discretized with a fixed step size for each parameter. A trajectory is then defined as a sequence of points in the discretized parameter space. Random trajectories are built from the following rules:

— each parameter varies only once per trajectory;
— the order of variation of the parameters in a trajectory is random;
— the starting point of the trajectory is chosen randomly;
— the direction of variation of each parameter is also random.

2. Cost function for the evaluation of the model outcome. The cost function involves important characteristics of the model outcome.

3. Computing of elementary effect of each parameter. For a given parameter X_i, a trajectory j and a cost function F, the elementary effect is given by :

$$E_i^{(j)} = \frac{F(\dots, X_i + \alpha\delta, \dots) - F(\dots, X_i, \dots)}{\delta} \tag{3.27}$$

where δ is the step size (that is the same for all parameters), and $\alpha = \pm 1$ is the direction of variation of the parameter (randomly chosen when the trajectory is built).

4. Computing of the two Morris indices. For each parameter X_i, the first index is the mean μ_i^* of the absolute values of the elementary effects, and the second index is the standard deviation σ_i of the elementary effects. A plot of σ_i as a function of μ_i^* gives the needed information to classify the model parameters according to the three categories.
Specifically, the two Morris indices can be interpreted as follows :

— μ_i^* measures the influence of the parameter X_i on the outcome of the cost function. The larger μ_i^*, the larger the influence of this parameter to the dispersion or the outcome.
— σ_i measures the non-linear or interactions effects of the parameter X_i on the outcome. If σ_i is small, a small variation in the parameter results in a small variation in the outcome, which in turn means that a linear effect of the parameter on the outcome is very likely. If σ_i is large, than a small perturbation in the parameter results in a large variation in the outcome, which means that there are non-linear

effects and/or interaction effects on the outcome.

Thus, the three categories are the following ones:

- Parameters with low influence : small σ_i and small μ_i^*;
- Parameters with linear effect : small σ_i and larger μ_i^*;
- Parameters with non linear effect : large σ_i and large μ_i^*.

Application of the Morris method to our model. Our aim is to classify the model's parameters depending on their influence on the model outcome, that can then be used to design the estimation method. Our model is controlled by 11 heterogeneous parameters. The most obvious characteristic of our model outcome is the maximum amplitude of the oscillations. The parameters space was discretized into 20 levels (20 values for each parameter), and 50 trajectories were used to compute the two Morris indices. A clustering algorithm, using the K-means method, was then applied, yielding three parameter groups :

— parameters with low influence: a, b, d, $\tilde{k}_{end}$, k_{on}, N;
— parameters with linear effects: f_p, K_0, Γ;
— parameters with non-linear effects: $\bar{f}$, f_c.

These groups were further used in the design of the hierarchical estimation method described below, from which all results given in the remainder of this chapter were obtained.

A first implementation of our parameter estimation method without parameter grouping yielded unsatisfactory results. Using the parameter groups given by the sensitivity analysis, on the other hand, made it possible to obtain very promising results, as shown in the remainder of this chapter.

Nevertheless, implementation issues for the sensitivity analysis were detected at the end of the writing of this manuscript. Better choices of parameter groups could potentially result in better fitting of the experimental data, but it appears that the choice of an optimal grouping is harder than initially thought. This perspective of our works is presented in Appendix F.

3.3.1.2 Hierarchical estimation framework

In this section, an estimation framework is proposed. In Chapter 2, a simple implementation of the ABC method was introduced in order to estimate the two model parameters. Here, the model involves 11 parameters that interact in a non-linear way and for which the *a priori* knowledge is weak. Below, we propose a three step procedure for the estimation of the model parameters. This hierarchical estimation method was developed in Matlab 9.4.

Step 1 : Perform a parameter sensitivity analysis in order to group the parameters into 3 categories that impact differently the model outcome. This step corresponds to a model reduction procedure, as it will allow for a blockwise estimation, each block estimating a fraction of the parameters. In turn, the accuracy of the estimation should be improved. In our study, only a qualitative assessment of the parameters sensitivity is required and the Morris method investigated.

Step 2 : Define specific characteristics that the model simulations should respect. The aim is to constrain the model such that the simulations are close to the experimental data. This can be achieved by imposing a range of acceptable values for the most important features of the model outcome, or by defining common parameter values for all generated simulations if this is appropriate and in line with the model's assumptions. In our case, two features are essential: the amplitude and the frequency of oscillations, that were set to be at most 50% higher or lower than the values computed on the experimental data. In addition, 4 parameters were set to default values in all simulations : the parameters related to the force per motor $\bar{f}$, f_c and f_p, as well as one of the parameters of the unbinding coefficient $\tilde{k}_{end}$. The first three parameters can be considered fixed because the pulling forces at the cortex are supposed to be generated by only one type of molecular motor, and the mechanics of the motor makes it difficult to allow for significant variation in such characteristic values. The last parameter, related to the unbinding coefficient, is supposed to be set to a default value for similar reasons : the underlying mechanics of unbinding is probably constrained in order to yield coherent outcomes. We thus chose to set one parameter out of the four parameters governing the behavior of the unbinding coefficient.

Step 3 : Estimate the parameters using the Approximate Bayesian Computation method. If several parameter blocks were identified in the first step, then the estima-

tion should be done block by block, starting with the parameters that do not need to be constrained because their effects on the outcome are small, and ending with the parameters that need to be strongly constrained because their effects on the outcome are strong and non-linear. For estimating the parameters of a given block, all other parameters are fixed to their previously estimated values or, if no such values were computed, to their default values. This whole estimation pass is then iterated until global convergence is obtained.

The characteristic features are used to select the model samples that are close enough to the experimental data, with respect to these features. Our method proposes three criteria to decide whether a model sample can be accepted: the amplitude as well as the frequency of the simulation must be close to those of the experimental curve (maximum twice the value for the maximum amplitude of the experimental data, and one and a half times the frequency of the experimental data), and the $L2$ distance error between the simulation curve and the experimental curve has to be in the lowest 1% of all computed simulations (with around 200 000 simulations generated for each embryo). The distance between the experimental data and the simulation is weighted such that the difference for middle oscillations weight more in the computed distance, than the beginning and the end of the oscillations, as our model does not capture some of the mechanisms involved in the behavior of the first and last oscillations. This means that the frequency is slightly perturbed at the beginning and at the end of the simulation, but this difference is not significant for the purpose of the method.

The current implementation relies on a temporal registration to speed up this process. This registration amounts to shifting each simulated data in time so that the maximum amplitude of the oscillations coincides (in time) with that of the experimental data. In the model, the time shift is entirely encoded in the relationship between parameters b and d, and it can be taken into account by only changing the value of b. One can see temporal registration as a way of "correcting" the value of parameter b from the experimental data. This results in more model samples being selected, thus reducing the global computation time.

3.3.2 Parameter estimation on simulated data

The aim of this section is to evaluate our parameter estimation method on simulated data of pole oscillations. This will allow to evaluate the performance of the method given ground truths, before testing it on experimental real data.

First, a description of the data is provided, together with information about the parameters and how the artificial data was obtained. The second part presents the convergence error (in terms of $L2$ error), and the error of the estimation of the parameters, for all the tested data. Finally, the last part shows the complete distributions of the parameters estimations.

3.3.2.1 Description of the experimental simulated data

In order to evaluate our estimation method, 14 experimental data were created by simulating the model with specific parameters. The values of the parameters were chosen such that, the associated experimental curves are close to the ones obtained from real data, and that are coherent with the order of magnitude given in the literature. The parameters were also chosen such that a wide range of oscillatory behaviors are tested.

As explained in the previous section, several parameters are set to constant values as follows : $\bar{f} = 7 \times 10^{-12}\ N$, $f_c = 7 \times 10^{-13}\ N$, $f_p = 1.2 \times 10^{-6}\ N \cdot s/m$, $\tilde{k}_{end} = 0.2$. Thus, our method is applied to estimate the following parameters : N, K_0, Γ, k_{on} and the parameters related to the unbinding coefficient a, b and d. The sets of parameters for the 14 experimental data are repported in Appendices (Section G).

The selected sets of parameters yield a variety of oscillations features. The 14 cases are illustrated in Fig. 3.22. Some cases show low amplitude of oscillations as in Figs. 3.22a and 3.22c, while other cases show high amplitude as in Fig. 3.22j. The number of oscillations also varies significantly, between 2 oscillations such as displayed in Fig. 3.22e, to 5 or more as shown in Figs. 3.22j and 3.22k. As for the frequency of oscillations, it varies also, but in a smaller range than the two other features.

3.3.2.2 Convergence and estimation error assessment

Estimating the parameters on simulated data allows for a proper evaluation of the proposed method. For this purpose, the two most important aspects to investigate are the convergence of the algorithm and the parameter estimation error. Figures 3.23 shows the relative error of the MMSE estimate for each parameter. The relative estimation error is low for all tested case, for the parameters N, K_0, Γ and k_{on}, with a maximum relative error of 20%. The parameters related to the unbinding coefficient a, b and d seem to

be slightly less well estimated. This can be explained by the fact that a registration is preliminary performed. Nevertheless, even if a few cases show higher relative error for these parameters, in most cases the relative error is lower than 20%. Thus, our estimation framework produces reliable results for a wide variety of oscillatory behaviors, and a wide range of parameters values.

3.3.2.3 Estimation of parameter distribution

In the previous section, the relative errors of the MMSE estimates were presented. Here we focus on the distributions of the parameters.

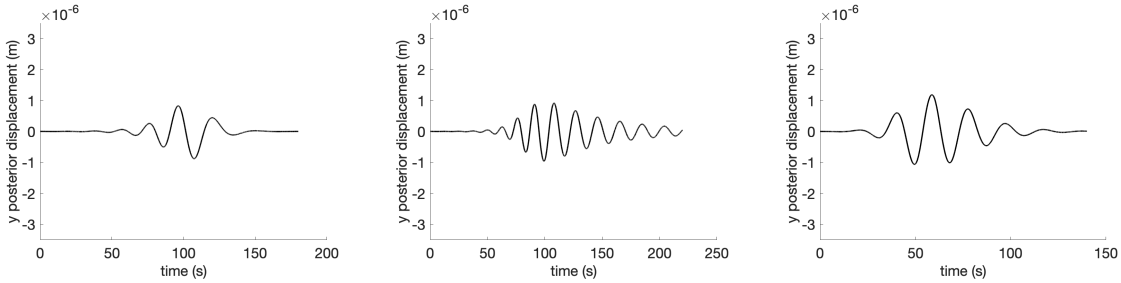
First, Fig. 3.24 shows the estimated distribution for each given case, and every figure shows the results for one particular parameter. The theoretical (known) value is labeled with a black dot, and the MMSE estimate is labeled with a magenta dot, for each case.

Figure 3.24a shows the results for the number of active motors N, with theoretical values between 40 and 90. The MMSE estimates are close to the theoretical values, as explained in the previous sub-section, but a certain variability is observed in the standard deviation of the estimation. This result is consistent with the parameter sensitivity analysis, that showed that the parameter N does not have to be particularly constrained in order to obtain a good fitting solution. From a biological point of view, this might mean that the number of active motors can vary in a significant way between embryos, and still achieve oscillations during anaphase.

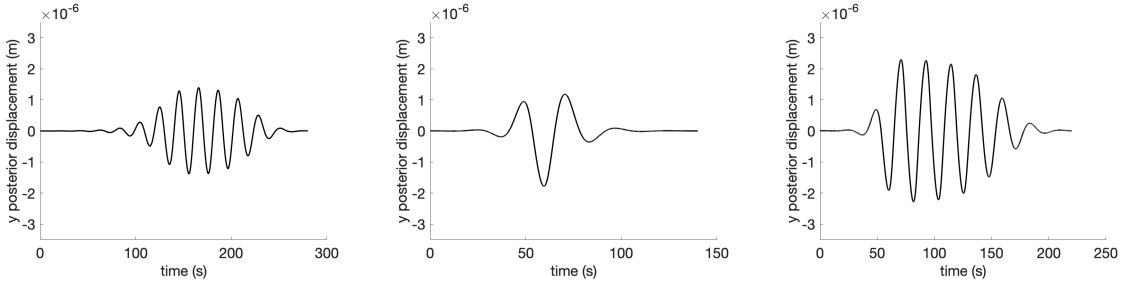
The estimation of the physical parameters K_0 and Γ is illustrated in Figs. 3.24b and 3.24c. In both cases, the MMSE estimates are close to the theoretical values. The standard deviation shows some variability between the different tested cases. Nevertheless, the results are highly homogeneous in the case of the viscosity coefficient Γ, which is to be expected as the viscosity of the cell does not vary significantly, and only other parameters that might yield a viscous-like outcome can make the viscous coefficient vary (such as the microtubules that are growing but don't touch the cortex). The elasticity coefficient K_0 shows more variability in standard deviation because this component is not well understood, and it is likely due to dynamic components and not only physical properties of the microtubules/mitotic spindle.

Figure 3.24d shows the estimated distribution for the binding coefficient k_{on} for all tested cases. This parameter shows similar characteristics to the parameter N and they are probably correlated (which is to be expected from the mathematical expression of the model) : when N is high, k_{on} is also high, and the other way around. The MMSE estimates for k_{on} are fairly good, and the standard deviation varies between the tested cases. In the sensitivity analysis, the parameter k_{on} was also in the group of parameters that do not need to be strongly constrained, so the variation in standard deviation is to be expected.

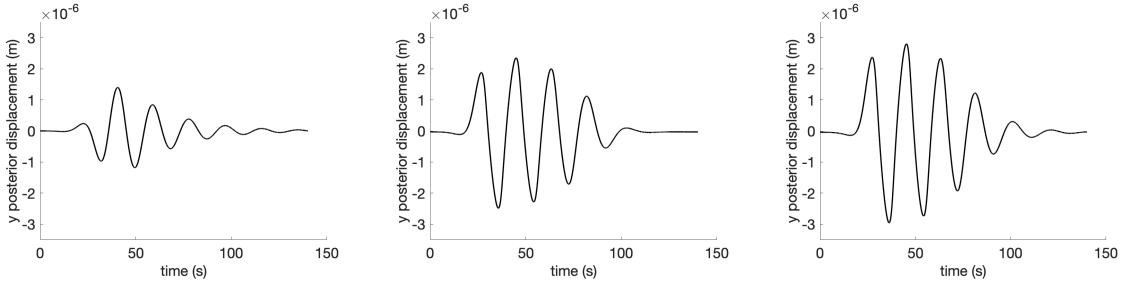
The estimation of the parameters related to the unbinding coefficient is illustrated in Figs. 3.24e, 3.24f and 3.24g. The parameter a which sets the beginning of the unbinding coefficient evolution, is mostly homogeneous for the tested cases, and the standard deviation is low, except in 4 cases. Parameters b and d, which drive the timing of the decrease of the unbinding coefficient and the slope of the decrease respectively, display a large variability, in both mean values and standard deviation. This can be explained by the variety of oscillatory behaviors observed in experimental data in terms of increase and decrease timing, and the parameters that influence the most these aspects are b and d.



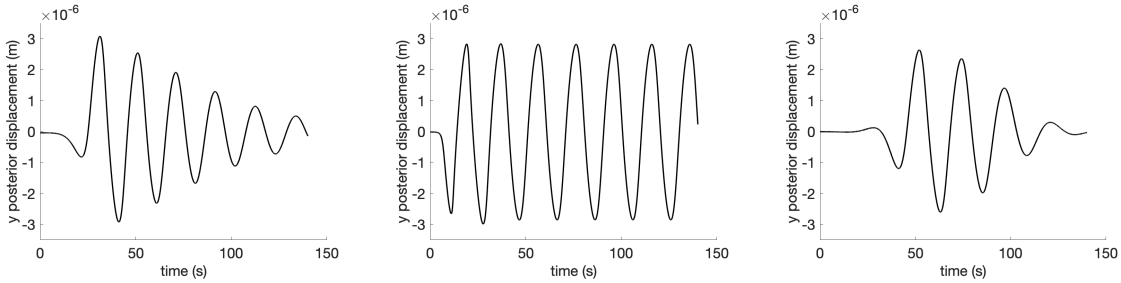
(a) case 1 (b) case 2 (c) case 3



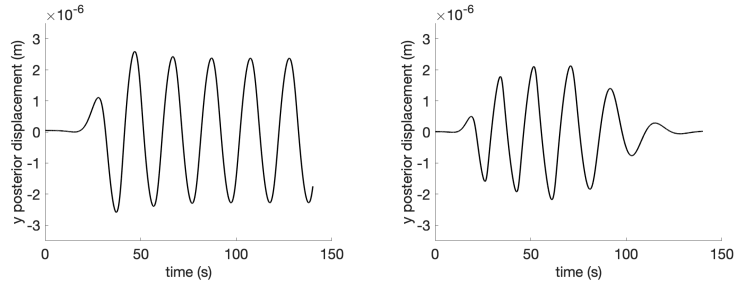
(d) case 4 (e) case 5 (f) case 6



(g) case 7 (h) case 8 (i) case 9

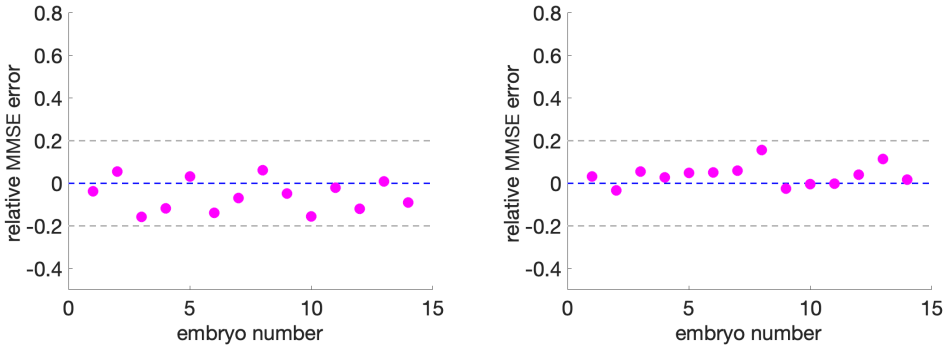


(j) case 10 (k) case 11 (l) case 12



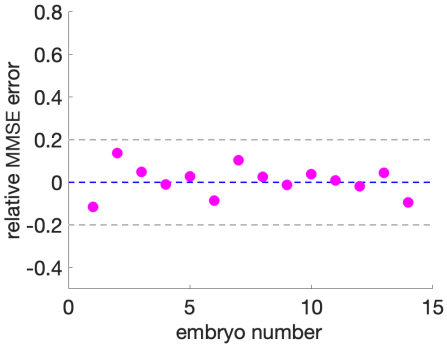
(m) case 13 (n) case 14

Figure 3.22 – **Experimental simulated data:** oscillations (posterior pole position y vs time) for 14 cases simulated by our model with the parameter values given in Appendix G.

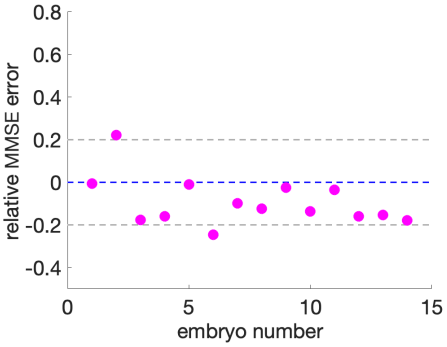


(a) N

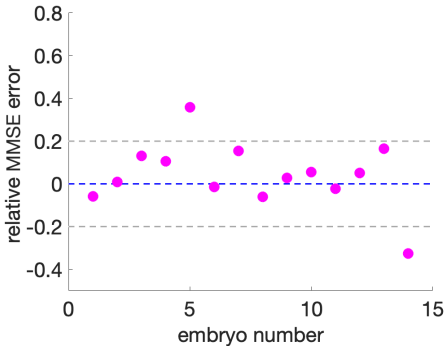
(b) K_0



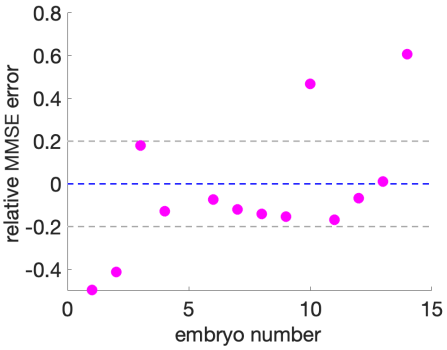
(c) Γ



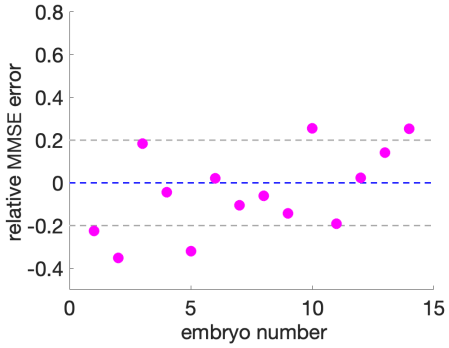
(d) k_{on}



(e) a

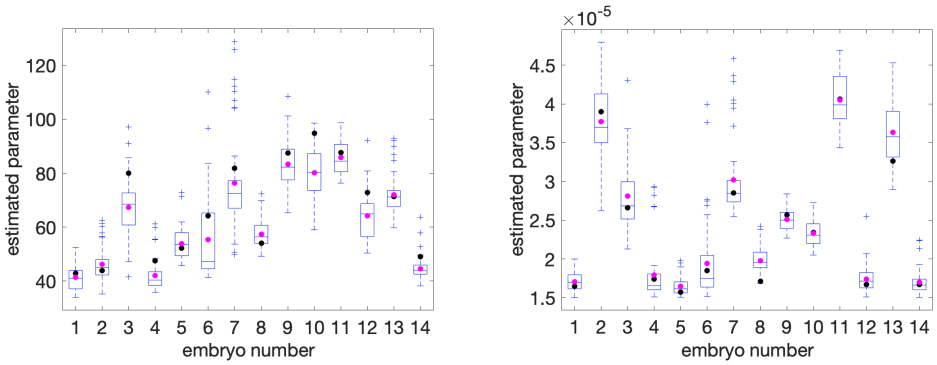


(f) b



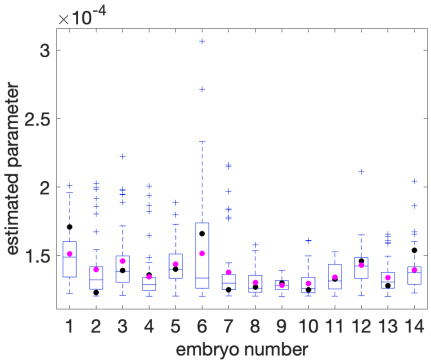
(g) d

Figure 3.23 – **Relative error of MMSE parameter estimates for the 14 experimental data illustrated in Figure 3.22.**

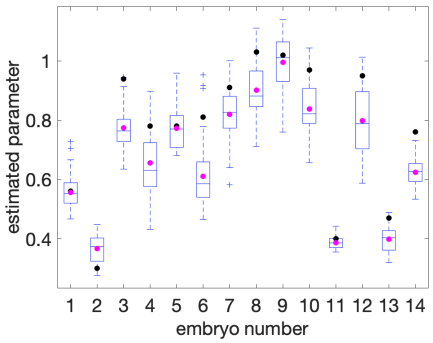


(a) N

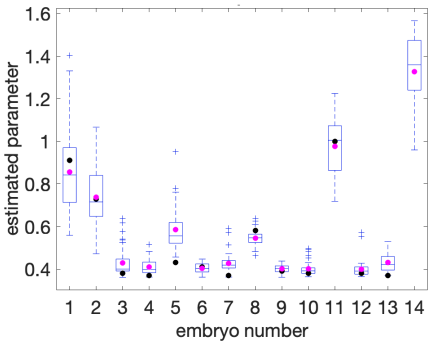
(b) K_0



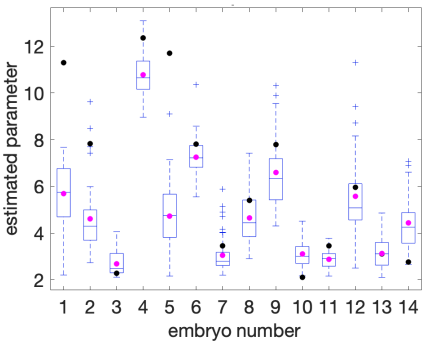
(c) Γ



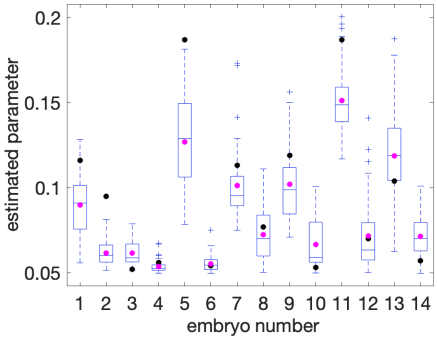
(d) k_{on}



(e) a



(f) b



(g) d

Figure 3.24 – **Distribution of the estimated parameters for the** 14 **experimental curves displayed in Fig. 3.22.** Black dots represent the real parameter values; magenta dots represent the MMSE estimates of the parameters.

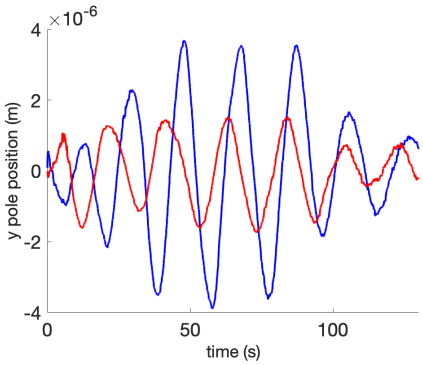
3.3.3 Parameter estimation on real data

In this section, the proposed estimation framework is applied to experimental real data, that is, to tracks obtained from images of the spindle pole oscillations in the first division of the C.elegans embryo. A total of fourteen tracks were used, that were generated from seven embryos, for both the anterior and the posterior poles. The experimental data is briefly described in the first part, then the results in terms of fitting and error between the real data and the selected model samples are illustrated. Finally, a discussion concerning the parameters estimation and a comparison between the anterior and posterior poles is proposed.

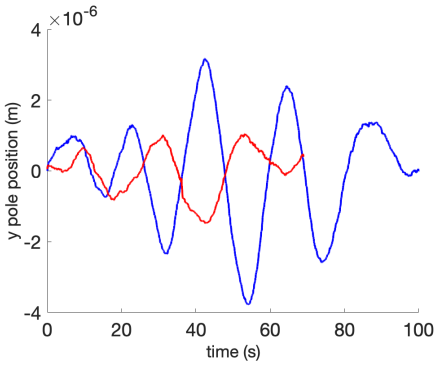
3.3.3.1 Description of the experimental real data

The data used in this section was obtained from seven wild-type embryos, from both anterior and posterior spindle poles. The embryos were observed at a temperature of 23 degrees Celsius. The corresponding oscillations tracks from the fourteen poles are illustrated in Fig. 3.25. Each figure shows the tracks for one embryo, for the two poles : blue track for posterior, and red track for anterior.

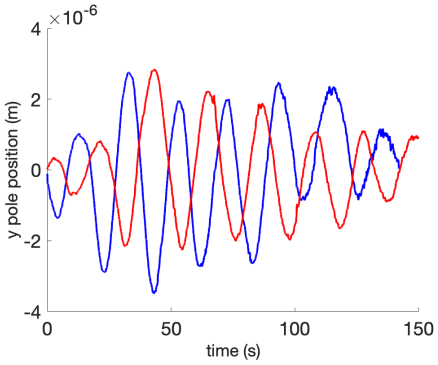
The behavior between the different embryos, as well as the behavior between the anterior and posterior poles show a large diversity in the number of oscillations, the amplitude and frequency of oscillations, the build-up/die-down times etc. Our model is able to exhibit all these behaviors, as shown in the previous section. Notably, the maximum amplitude of the oscillations of the anterior pole is lower or at most equal to the maximum amplitude of the posterior pole. This can be due to different parameters : the number of active force generators has a strong influence on the maximum amplitude, but other parameters can also influence the maximum amplitude (as discussed in the model analysis section). The general behavior for the posterior pole is more homogeneous than the anterior pole, with well marked build-up and die-down periods, and a distinct oscillation corresponding to the maximum amplitude.



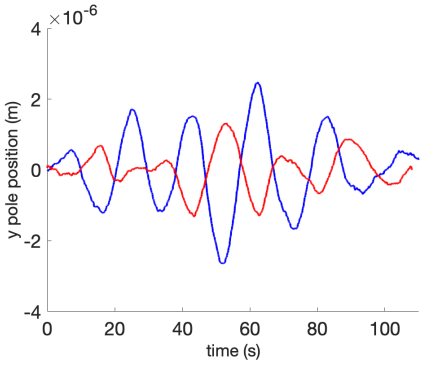
(a) embryo $E1$



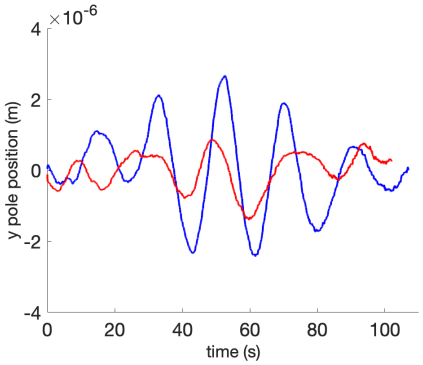
(b) embryo $E2$



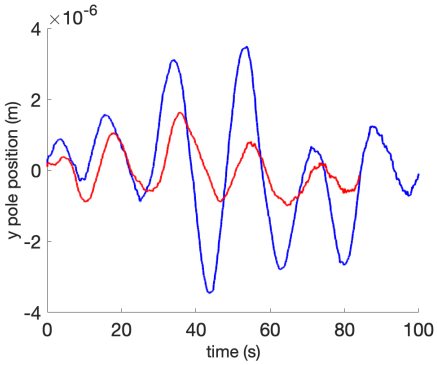
(c) embryo $E3$



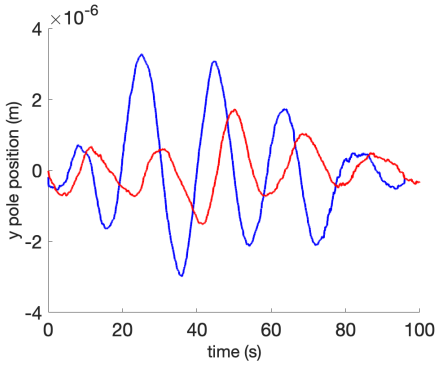
(d) embryo $E4$



(e) embryo $E5$



(f) embryo $E6$



(g) embryo $E7$

Figure 3.25 – **Experimental real data:** each sub-figure illustrates the oscillations of the anterior pole (in red) and the posterior pole (in blue) of one wild-type embryo.

3.3.3.2 Fitting the real data : example of the posterior pole

The fitting between the experimental data and the given solution, as well as the $L2$ distance errors, are discussed for the set of posterior poles described above. This is a crucial step in the proposed framework, as it constitutes the main criterion to decide whether a model sample must be selected or discarded. We will see here that our method is able to fit a variety of behaviors, even on real data, that can present a large spectrum of features, but also "flaws/artifacts" due to the dynamic nature of the oscillations.

The fitting of the selected model samples for each embryo is illustrated in Fig. 3.27 (for the fitting on the anterior pole see Fig. H.1). The beginning of the oscillations is usually less well fitted, with sometimes, small oscillations that are being missed by the simulation, or the amplitude for one oscillation not corresponding to the experimental data. This can be due to the model itself because it does not capture the whole dynamics of the build-up (because of the model being only temporal, and not having a spacial component that drives some of the dynamics of the beginning of oscillations). The ending of the oscillations is sometimes less well fitted as well, as can be seen for the posterior pole, for the embryos $E1$ and $E3$ for example. One can notice that the experimental data can have a partially missed oscillation (embryos $E1$, $E5$, $E6$, $E7$), a small drift (embryo $E3$), or sustained oscillations (embryos $E1$, $E3$), and these can also slightly change the frequency for the second part of the oscillations. This is probably due to a local event interfering with the general dynamics of the pole, such as a synchronicity event between the anterior and the posterior pole. As the model does not include these events, these perturbations will have some impact on the fitting and the value of the distance error, but will very slightly, or not at all, influence the parameter estimation. In general, our model yield very good fitting for the main oscillations, which confirms that our model is well suited to describe the global behavior of the pole during anaphase.

The $L2$ distance error between the experimental data and the proposed solutions is reported in Fig. 3.26. The embryo with the largest mean distance error is the embryo $E6$ which has a partially missed oscillation in the real data, so it is expected to have a consequence on the distance error. The smallest distance errors are for the embryos $E4$ and $E5$, which have the best fitting, as seen previously. The other embryos have similar distance errors, between 5×10^{-6} m and 6×10^{-6} m. Thus, the distance error and the fitting are consistent. Although there is some variability between the seven embryos, the

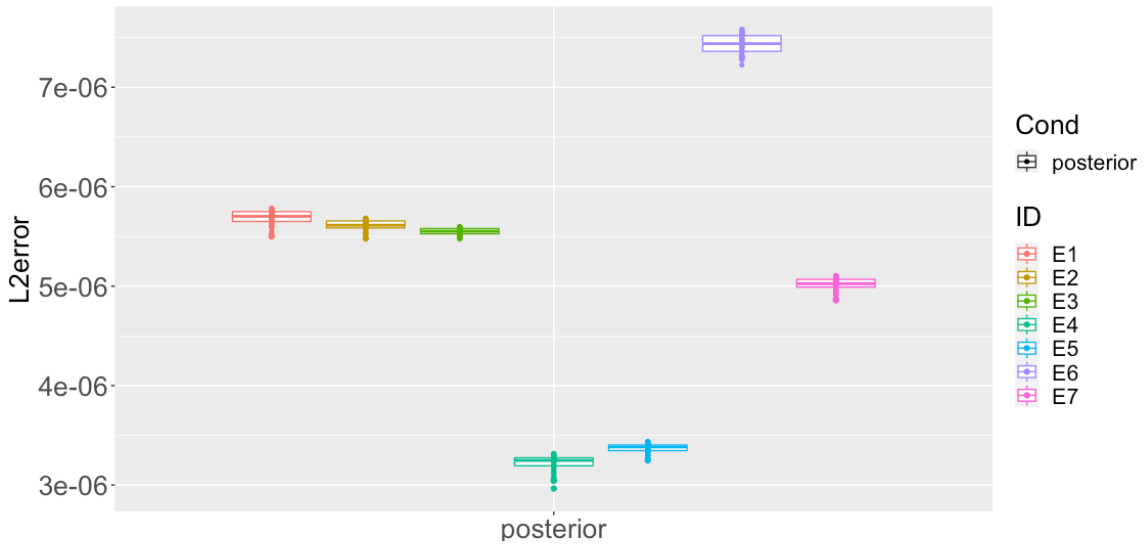
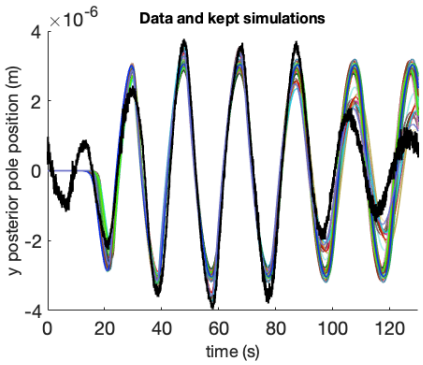
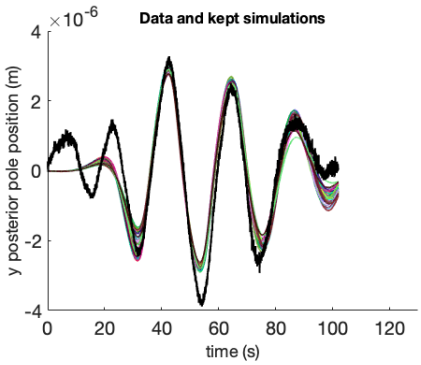


Figure 3.26 – $L2$ **errors between the experimental data and the solution given by our method for the posterior pole of the wild-type embryos illustrated in Figure 3.25.**

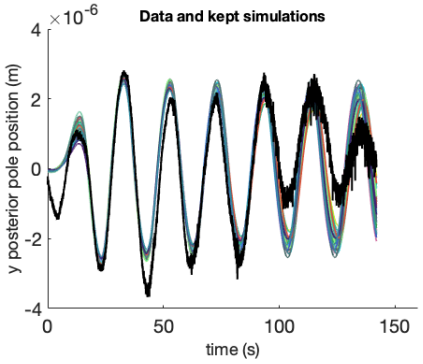
internal variability for each embryo is very small. This means that the distance error can not be used by itself as a criterion for the proposed solution, so it must be used in conjunction with other criteria, but it does provide homogeneous final solutions.



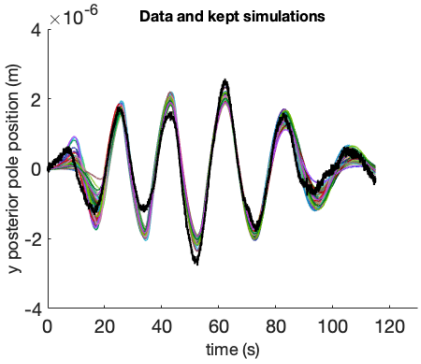
(a) embryo $E1$



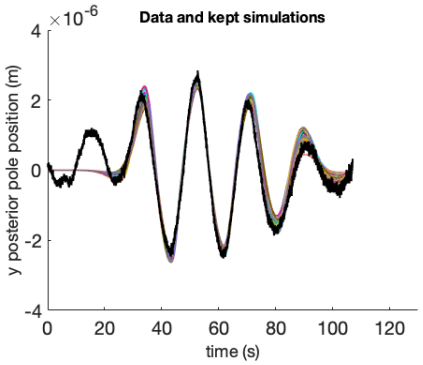
(b) embryo $E2$



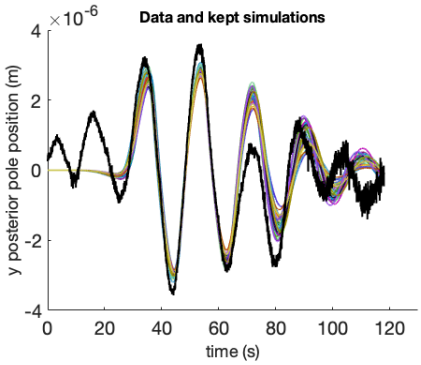
(c) embryo $E3$



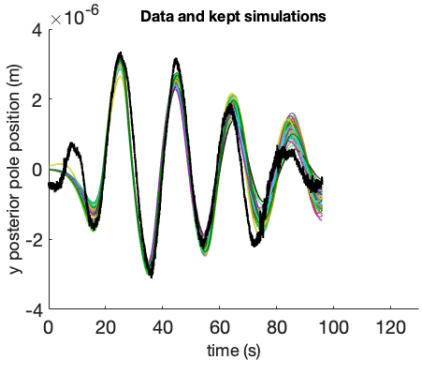
(d) embryo $E4$



(e) embryo $E5$



(f) embryo $E6$



(g) embryo $E7$

Figure 3.27 – **Fitting the posterior pole on wild-type embryos.** In each sub-figure, the black curve represents the experimental data (tracking of the posterior pole position in time) and the colored curves represent the simulations corresponding to the solution given by our method.

3.3.3.3 Parameter estimation on real data : anterior vs posterior poles

In this section, the results of the estimation of the parameters for the seven described embryos are discussed. In particular, the estimation of the following parameters are detailed : the number of molecular motors N, the binding coefficient k_{on} and the parameters related to the unbinding coefficient $\overline{k_{off}}$. Figures 3.29 to 3.33 below display the estimation of each parameter for the seven anterior poles and seven posterior poles. Each figure shows the estimation of one particular parameter, with the estimation for the anterior poles being illustrated on the left-hand side of the figure, and the estimation for the posterior poles on the right-hand side.

The anterior pole has a smaller maximum amplitude than the posterior one. This means that the solution given by our method probably has a smaller maximum amplitude for the anterior pole than for the posterior pole. Figure 3.28 shows the maximum amplitude of the proposed solution, for all the seven tested embryos. It confirms that our method yields solutions that are coherent in terms of amplitude with the real data values. For each embryo, the maximum amplitude for the anterior pole is lower than the posterior pole, but the difference anterior-posterior varies between the embryos. In particular, for the embryo $E3$ the difference is very small, which is to be expected as the difference from real anterior and posterior data is also small. This could imply that some parameters are of similar average value between the anterior and the posterior pole for this embryo. Another interestingly observation is that the maximum amplitude for the proposed solution shows a certain internal variability for some embryos. Some embryos show little variability, such as for the embryo $E3$ (anterior and posterior poles), while others show larger internal variability, such as the anterior pole for the embryo $E2$ and the posterior pole for the embryo $E1$. Are these differences linked to particular parameters values, or is internal variability naturally appearing because of the variability in amplitude given by the experimental data ?

The estimation of the number of molecular motors N is shown in Fig. 3.29. The estimated values for N vary between 30 to 125 approximately. The MMSE estimates for the anterior poles are all lower than the MMSE estimates for the posterior poles, which is expected on wild-type embryos. In addition, for each individual embryo, the MMSE estimate for the anterior pole is also lower than the MMSE estimate for the posterior pole. There is one exception in the tested embryos, the embryo $E4$, which has a higher

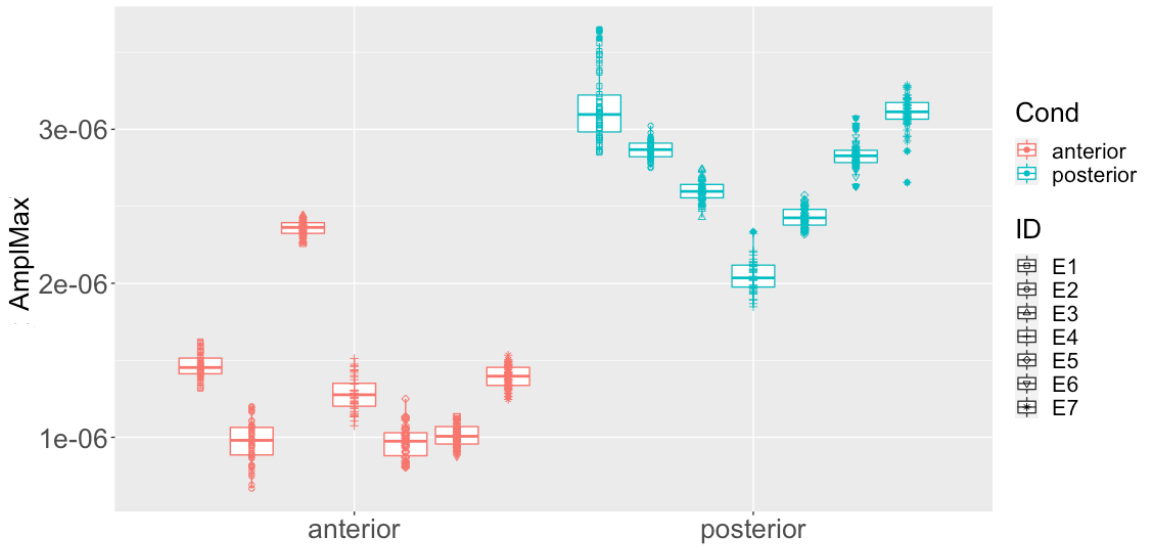


Figure 3.28 – **Maximum amplitude of the simulations corresponding to the solution given by our method for the posterior and anterior poles of the wild-type embryos illustrated in Fig. 3.25.** The results for the anterior poles are illustrated on the left-hand side; the results for the posterior poles are illustrated on the right-hand side.

mean N for the anterior pole than for the posterior pole. This is probably due to several factors. First, the posterior pole shows a relatively low maximum amplitude, similar to the one of the embryo $E5$, and their MMSE estimates are the lowest, around 50, among all posterior poles. The result for the posterior pole is thus coherent with the remaining parameters of the estimation. For the anterior pole, the estimate is on the high range, around 80, which goes against the general trend for the anterior vs posterior behavior. However, the MMSE estimate for this embryo is close to the MMSE estimate for the anterior pole of embryo $E7$ (also around 80), and their oscillatory behavior is similar. Another factor that can influence the estimation here is the fact that, the anterior pole for the embryo $E4$, has two partially missed oscillations, on each side of the maximum amplitude oscillation. This could have as a consequence that the model is not able to capture the real dynamics of the oscillations, and thus, over-estimated N. Another possible scenario is that, the missed oscillations are generated by a dynamic biological event that interferes with the global oscillatory behavior, which forces the number of molecular motors to go up.

The estimation of the binding coefficient k_{on} is shown in Fig. 3.30. The mean estima-

tion is between 0.3 s^{-1} and 1.2 s^{-1} and there does not seem to be a significant difference between the anterior and the posterior poles. For both groups, there are two poles (not the same between the two groups) that have a low estimation for k_{on}, in the range $[0.3; 0.5]$ s^{-1}, and the others show a mean binding coefficient in the range $[0.6; 1.2]$ s^{-1}. Also, there are 4 embryos that follow the same trend as for N, that is, the estimation is lower for the anterior pole than for the posterior pole, and 3 embryos for which the anterior pole has a higher MMSE estimate than the posterior pole. Although no general trend emerges, the estimation for the binding coefficient is relatively homogeneous with most of the estimation values being between 0.6 s^{-1} and 1.2 s^{-1}. Thus the binding coefficient seems to has a moderate variability, but without a strong influence from the features of the oscillations.

The estimation of the parameters related to the unbinding coefficient, i.e., a, b and d, are respectively shown in Figs. 3.31, 3.32 and 3.33. The parameter a, corresponding to the beginning of the oscillations is stable for most poles (except two posterior poles), with a vary small individual variability. This is to be expected as the timing of the beginning of the oscillations, influenced by a, should not vary too much between the simulations in order to fit in the best way possible the experimental data. The parameter d is also quite stable, although with a larger variability between the different poles. This parameter sets the timing for the die-down of the oscillations, and, as explained before, it can have a strong influence on the number of oscillations and total time of the die-down.The MMSE estimates, between 0.05 and 0.15, correspond to the range where this parameter has a weak influence on the features of the oscillations; the stability of the estimation is thus consistent with what is expected theoretically. Finally, the parameter b is the one that varies the most both individually for each pole, and between the poles. The parameter b influences also the timing of the die-down of the oscillations, as well as the number of oscillations, but in a weaker manner than d. It is thus to be expected to have a larger variability. However, the individual variability is very large for some poles, and in particular for the anterior poles.

Another interesting result is the estimation of the parameter N with respect to the maximum amplitude of the oscillations. Figure 3.34 shows the plot of the estimated N versus the maximum amplitude of the solution, for all the tested embryos. For the anterior poles (in red), no correlation is observed between N and the maximum amplitude, whereas, for the posterior poles, there seems to be a positive correlation between these

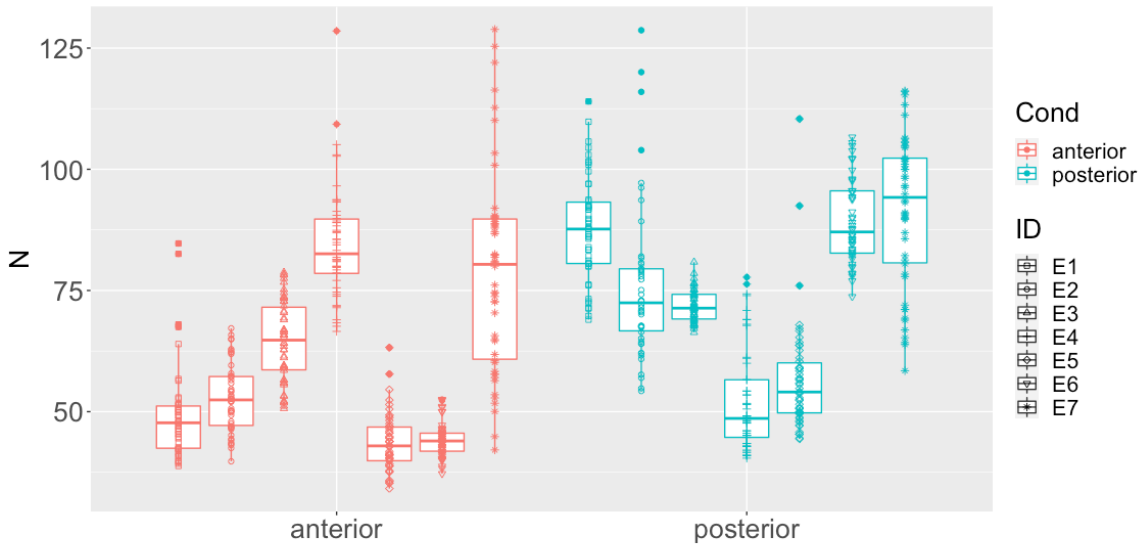


Figure 3.29 – **Estimated distribution of the parameter N for the wild-type embryos illustrated in Fig. 3.25.** The results for the anterior poles are illustrated on the left-hand side; the results for the posterior poles are illustrated on the right-hand side.

two parameters. From the model analysis, it was noticed that, when taken independently, N increases with the maximum amplitude. This might mean that, for the posterior pole, the parameter N is not much influenced by other parameters (or that they are correlated in the same way with the maximum amplitude), or that N is the parameter that leads the behavior with respect to the maximum amplitude. This does not seems to be the case for the anterior pole, either because there is a stronger correlation with other parameters, or because the value of the parameter N is not large enough to have the strongest influence with respect to the maximum amplitude.

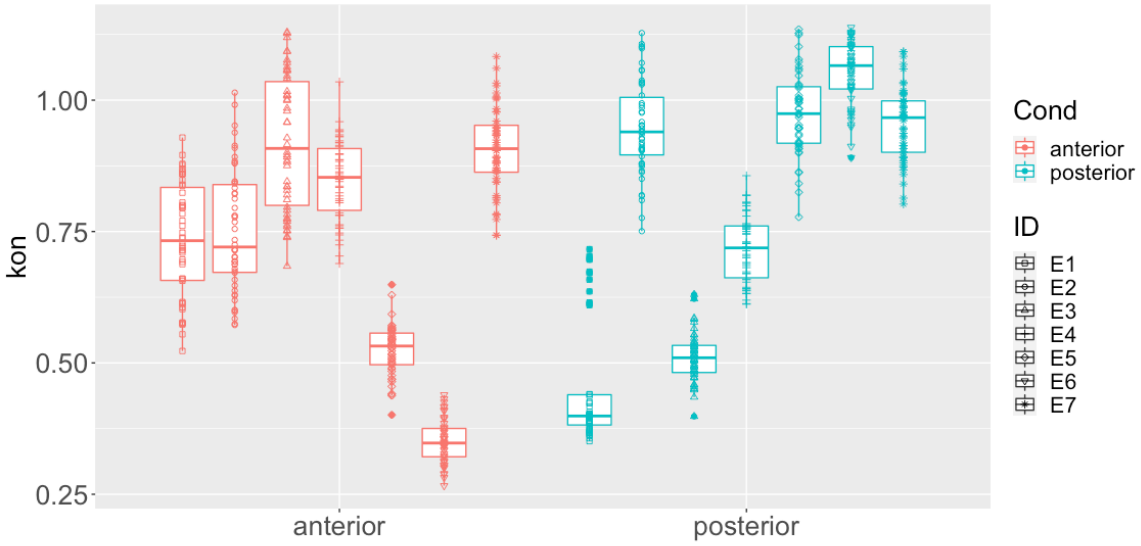


Figure 3.30 – **Estimated distribution of the parameter k_{on} for the wild-type embryos illustrated in Fig. 3.25.** The results for the anterior poles are illustrated on the left-hand side; the results for the posterior poles are illustrated on the right-hand side.

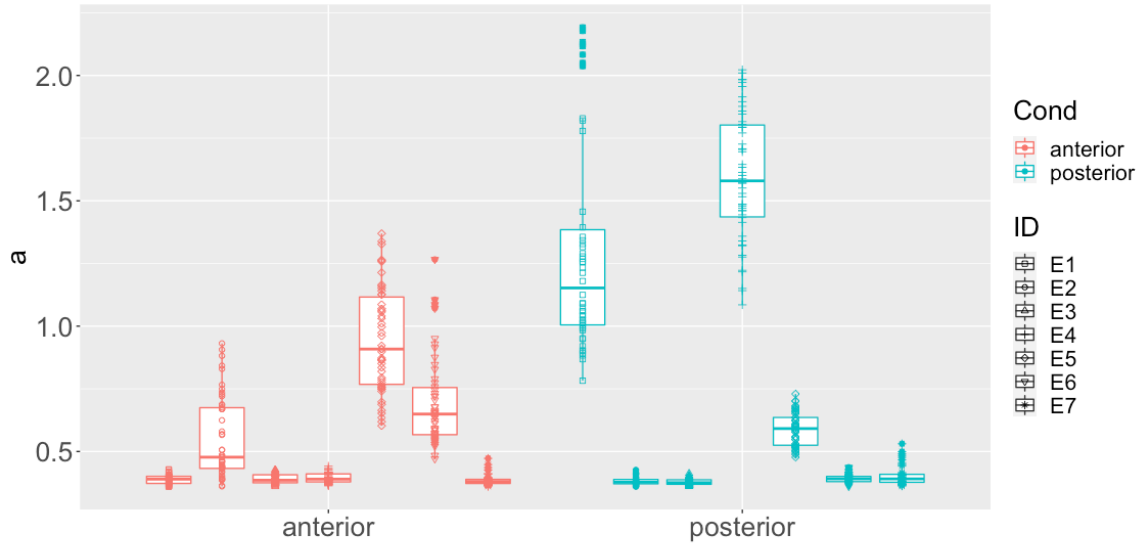


Figure 3.31 – **Estimated distribution of the parameter a for the wild-type embryos illustrated in Fig. 3.25.** The results for the anterior poles are illustrated on the left-hand side; the results for the posterior poles are illustrated on the right-hand side.

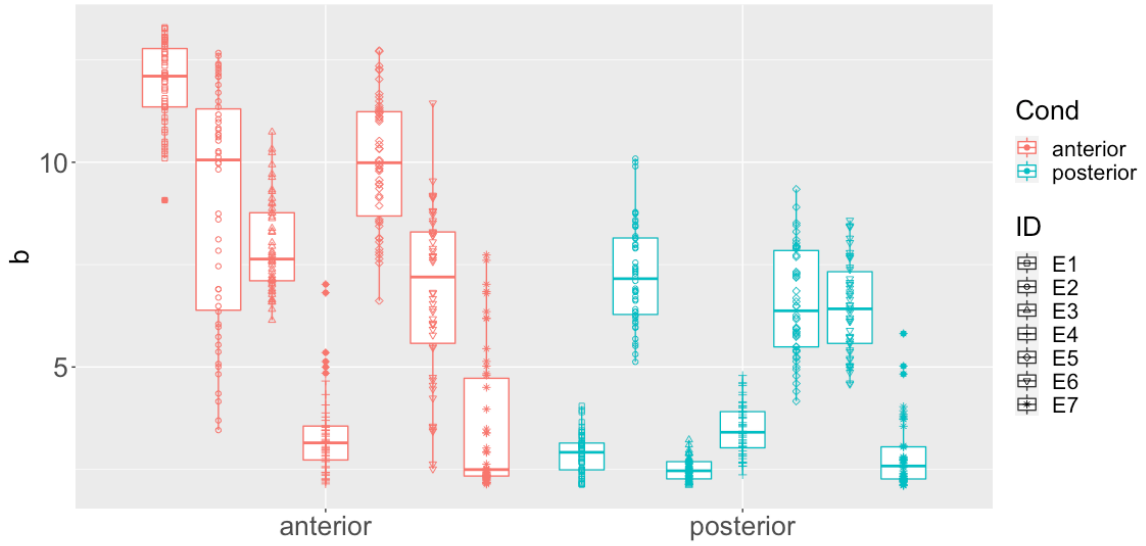


Figure 3.32 – **Estimated distribution of the parameter b for the wild-type embryos illustrated in Fig. 3.25.** The results for the anterior poles are illustrated on the left-hand side; the results for the posterior poles are illustrated on the right-hand side.

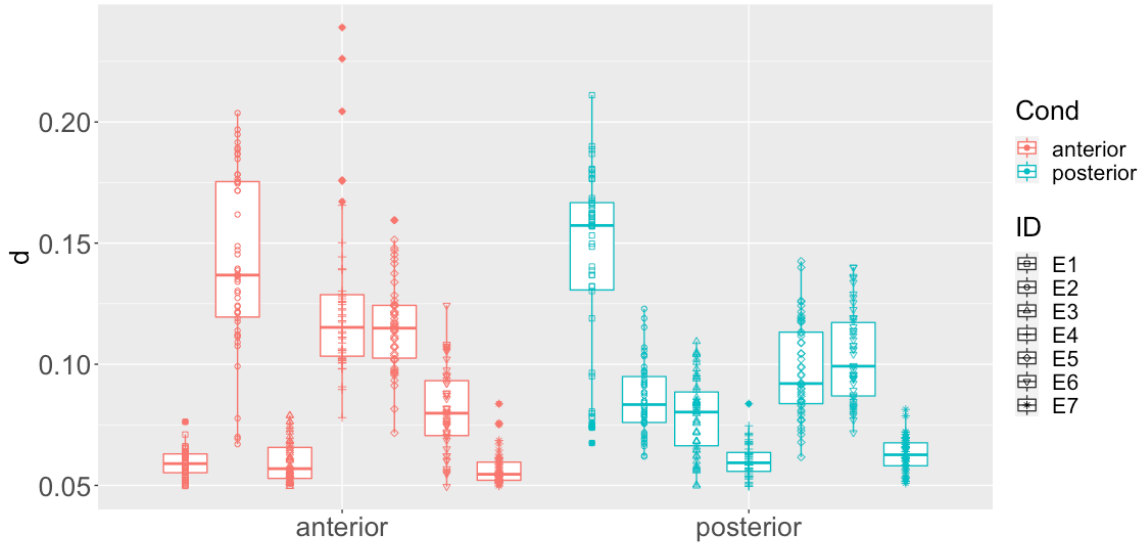


Figure 3.33 – **Estimated distribution of the parameter d for the wild-type embryos illustrated in Fig. 3.25.** The results for the anterior poles are illustrated on the left-hand side; the results for the posterior poles are illustrated on the right-hand side.

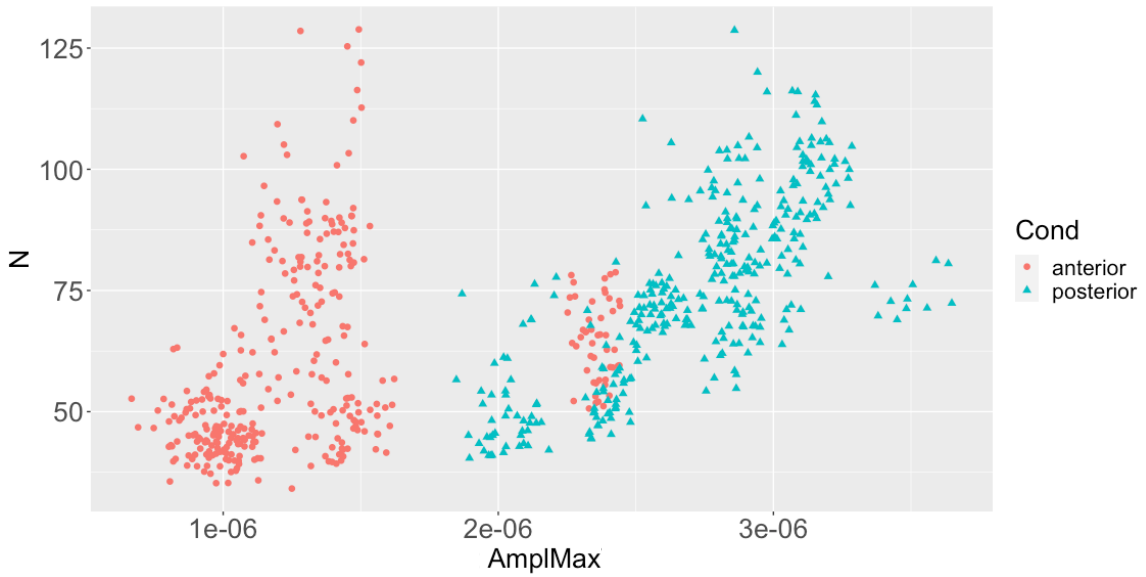


Figure 3.34 – **Maximum amplitude vs N for all anterior poles (in red) and all posterior poles (in blue).** The red dots on the right-hand side correspond to the solution for the anterior pole of embryo E7, that has a maximum amplitude higher than the other anterior poles.

3.4 Discussion and conclusion

In this chapter, we focused on the study of spindle pole oscillations during anaphase. Transversal oscillations of the mitotic spindle poles have been shown to be a crucial process in asymmetric cell division, and this behavior is influenced by a large number of molecular actors. The existing literature provides both a very thorough global understanding of the process and specific information about some of its aspects, but the interactions between the actors involved in the oscillations, or even in the whole asymmetric division process, are still mostly unknown. To address these issues, we introduced a new biophysical model of spindle pole oscillations and a new estimation method that provide a link between the parameters influencing the oscillations, as well as a way of assessing the variability of these parameters between embryos or populations of embryos.

Section 3.1 introduced the biological context of the study of C.elegans, and provided a general description of the first division of the C.elegans embryo. The main modeling concepts already validated in the literature were discussed.

In Section 3.2, we proposed a new biophysical model for the spindle pole oscillations and performed a model analysis that provided both validation of the model and further information about the conditions under which oscillations can occur, by making it possible to compute parameters thresholds above/below which oscillations are inhibited. We showed that this model is able to reproduce a very wide range of oscillating behaviors that are observed on real data, and that the characteristic values of these computed behaviors, especially the frequency and amplitude of the oscillations, are realistic.

In Section 3.3, we presented a hierarchical parameter estimation framework based on a Bayesian method. We first described the successive steps of the estimation framework, then we applied this framework to the estimation of the parameters of the aforementioned model of oscillations, on both simulated and experimental data. Our method was validated on synthetic data, as we showed that it yields estimates that are very close to the expected values on data generated from the model. Applying our method on the anterior and posterior poles of wild-type embryos suggested that the difference in maximum amplitude and number of oscillations between the poles can be explained by a difference in the parameter related to the number of active force generators, in accordance with results from the literature. We also showed that some parameters are stable between the anterior and posterior poles, while others vary significantly between the two types of poles, or even between embryos of the same type. This suggests that some parameters have a certain internal variability, that might contribute to the robustness of the oscillations during anaphase.

Estimating the parameters of this model from real data proved to be a very complex task for several reasons, including the large number of parameters and their nonlinear interactions. Moreover, as mentioned above, preexisting methods are not fit for variability studies. We thus designed a novel parameter estimation framework; important features of this framework are the use of sensitivity analysis in order to group parameters with strong interactions, and the implementation of an Approximate Bayesian Computation method that yields posterior distributions of parameter for every embryo. This framework is versatile, and can easily be adapted to studies other than the one detailed in this chapter.

General conclusion

This thesis deals with the estimation of the biophysical parameters of cellular mechanisms, with a special focus on two applications: diffusion events linked to exocytosis, and spindle pole oscillations during anaphase in asymmetric cell divisions. In both cases, fluorescence microscopy techniques coupled with mathematical modeling have been widely used for studying the mechanisms of interest; however, existing approaches are typically unable to provide information about both general and local behaviors.

In order to address these shortcomings, we focused on models and estimation methods that would make it possible to gain insights into the variability of the studied mechanisms, be it the local variability of the diffusion events inside a cell, or the population-wise variability of the spindle poles behavior during asymmetric division.

We first introduced a novel mathematical model for local diffusion at the cell membrane during exocytosis. Although this model is based on a standard correlation method, it does not rely on any homogeneity assumption; thus, its use in a Bayesian framework yields information about the spatial variability of diffusion along the cell membrane. An estimate for the local diffusion can be computed from very small regions of interest in an image sequence. The model has proved to be robust to noise and spot positions in the frame; it enables the evaluation of local diffusion in various regions of interest in the sequence, including cases in which several diffusion events occur next to each other.

Possible short- and medium-term improvements were listed in the conclusion of the corresponding chapter; they include outputting diffusion maps, incorporating flow terms in the model for dealing with "kiss and run" fusions, and extending the approach to 3D sequences. On the longer term, the method could be used for other proteins than the one on which our works focused. Several transmembrane proteins have been identified, for which diffusion behaviors are not yet understood; they include Rab11, that has been shown to play a role in protein trafficking among others, and Langerin, which has been linked to the protection of cells against HIV-1 infection.

We then presented a new biophysical model for spindle pole oscillations during anaphase, as well as a novel parameter estimation framework based on both sensitivity analysis and an Approximate Bayesian Computation approach. The model has proved able to reproduce a wide range of observed behaviors, and provided valuable information about the conditions under which pole oscillations, that are essential for asymmetric cell division, can occur. It was then successfully used in our framework, yielding satisfactory estimates on synthetic data and making it possible to identify interactions between some parameters, hinting at interactions between the molecular actors themselves. The introduction of spatial components could be used to improve the current model.

A major perspective of these works is the application of our method to different populations, in order to further study the impact of several factors, including local genetic mutations. In many cases, one would like to determine whether or not the same mechanism takes place, but with different parameters, and if so, the molecular causes for the observed alterations might be traced back from the variations in the parameters.

The biophysical parameters of the model could also be expressed in terms of molecular actors and mechanisms, according to the various hypotheses available in the literature. Thorough testing would then make it possible to partially confirm or infirm these hypotheses according to the corresponding behavior of the model and its similarity to the observed behaviors in real image sequences.

APPENDICES

A Local variability - Expected autocorrelation function for varying parameter values

Expected autocorrelation function with model G_1

Figure A.1a shows the autocorrelation function G for varying values of the parameter D, using 100 values that were randomly chosen in the range $[0.1, 1.5]$, all the other parameters being set to their default values. For this first model, the parameter D has a global impact on the autocorrelation curve. In particular, D influences the slope of G, as well as its starting and ending values.

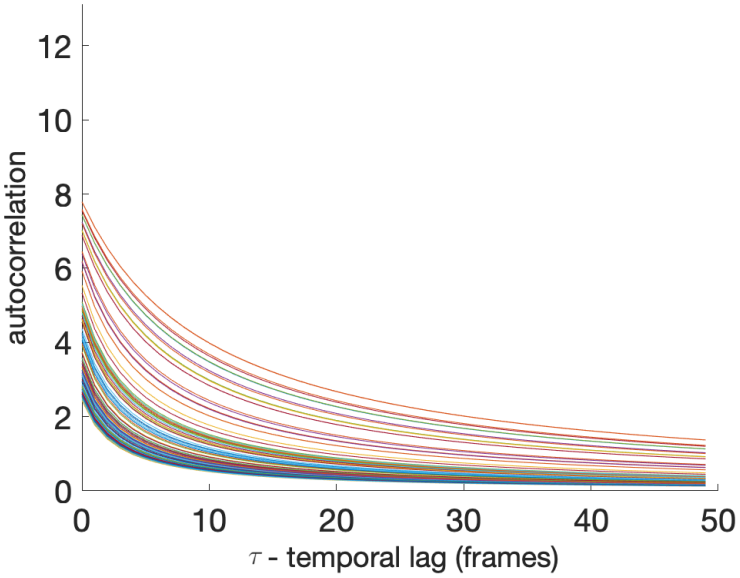
Figure A.1b shows the autocorrelation function G for varying values of the parameter σ^2_{PSF}, for which 100 values were randomly chosen in the range $[0.5, 2]$. All the other parameters were set to the default values. For this first model, the parameter σ^2_{PSF} influences the starting value of G, but not its dynamics, as expected: in Equation (2.6), parameter σ^2_{PSF} is not tied to τ, so its influence is restricted to the starting point.

Expected autocorrelation function with model G_2

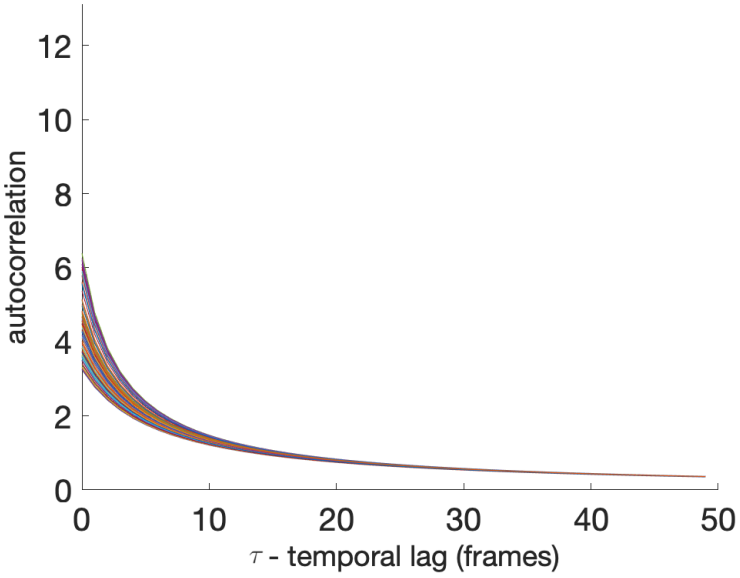
The influence of the parameter D alone was studied by computing the autocorrelation function g for varying values of D in the range $[0.1, 1.5]$, and fixed values for the other parameters. For this second model, D also has a global impact on the autocorrelation curve, but this impact is lower than for the first model, as showed in Fig. A.2a. The beginning and ending of the autocorrelation curve are influenced by the value of D, but the range of g is smaller than for the first model. In particular, its starting values are closer to one another. The parameter D still influences the slope of the autocorrelation curve, as for the previous model.

The influence of σ^2_{PSF} alone, was studied by computing the autocorrelation function g

for varying values in the range $[0.5, 2]$, and fixed values for the other parameters. Figure A.2b shows that the influence of σ^2_{PSF} is similar to the previous model. The main impact is on the beginning point of the autocorrelation curve. In addition, σ^2_{PSF} has a slight influence on the ending point, but no particular impact on the dynamics of g.

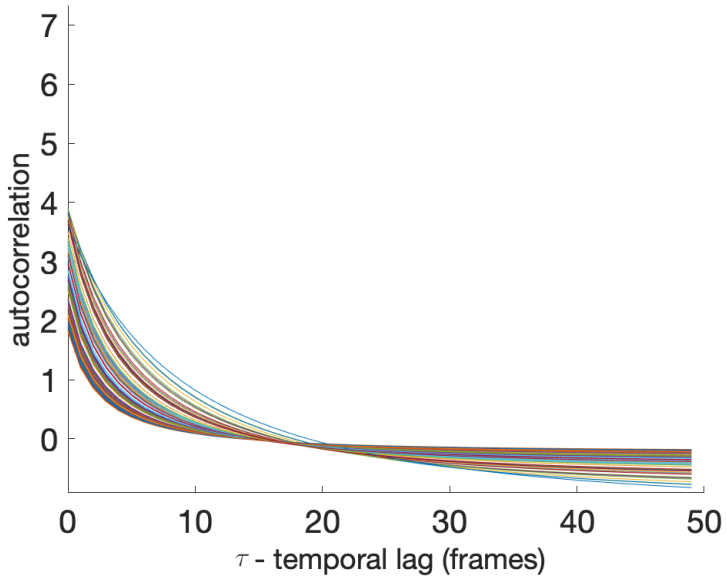


(a) Influence of D

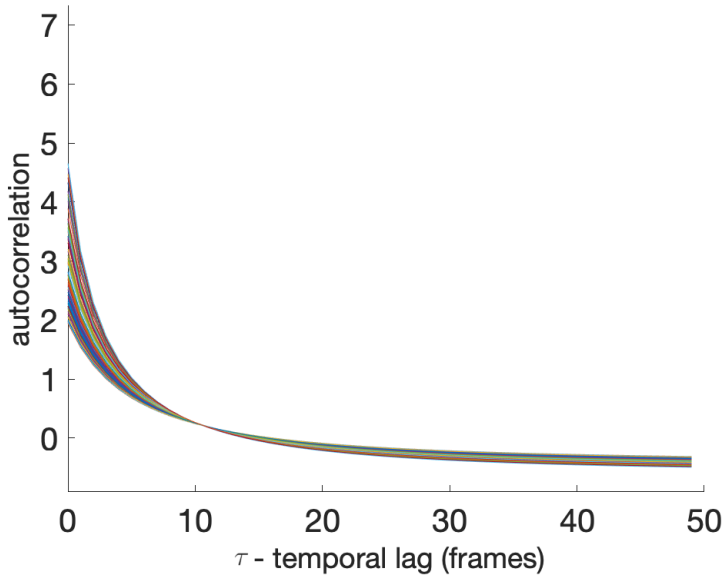


(b) Influence of σ^2_{PSF}

Figure A.1 – **Influence of model parameters on the autocorrelation function for model G_1.** Each curve displays the simulation of model G_1 for a specific value of D (a) or σ^2_{PSF} (b), the other parameters being set to the default values.



(a) Influence of D



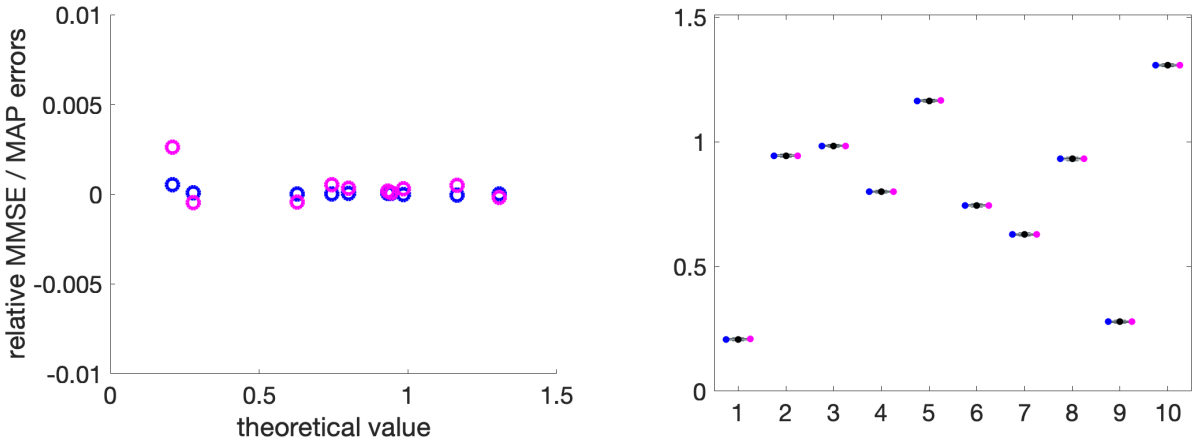
(b) Influence of σ^2_{PSF}

Figure A.2 – **Influence of model parameters on the autocorrelation function for model G_2.** Each curve displays the simulation of model G_1 for a specific value of D (a) or σ^2_{PSF} (b), the other parameters being set to the default values.

B Local variability - Validation of the parameter estimation method for each parameter

Parameter estimation with model G_1

The diffusion coefficient D was first estimated, and the results are reported in Figure B.1. Figure B.1a plots the relative errors on the estimation of D. The relative error is very low for all tested values: it is below 0.5% in all cases. This confirms that both the MMSE and MAP estimates for D alone are precise. Figure B.1b shows the variability of the estimator in a boxplot form. The real value is labeled a black filled circle, the MMSE estimate with a magenta filled circle, and the MAP estimate by a blue filled circle. Here, the three circles fully overlap for all tested values and the standard deviation is very low: the estimator for D alone is thus very accurate.



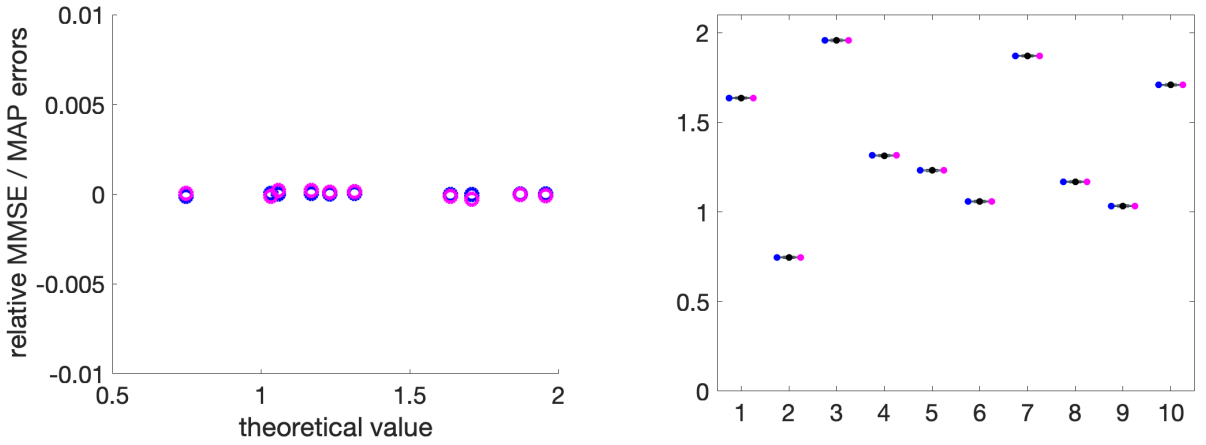
(a) D relative error estimation

(b) D posterior distribution

Figure B.1 – **Estimation of D alone with model G_1.** (a) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the theoretical value of D. (b) Estimated distribution of the parameters, for the ten tested cases. The true value of D is represented by a black circle.

The standard deviation for the PSF, σ^2_{PSF}, was then estimated alone; the results are shown in Fig. B.2. As for the parameter D, both the MMSE and MAP estimates are accurate, as shown in Fig. B.2a, with both the MMSE and MAP relative errors being close to 0 (below 0.5%). In Fig. B.2b, the standard deviation of the estimator is very low for all tested values of σ^2_{PSF}. The estimation of the parameter σ^2_{PSF} alone is also very

accurate with model G_1.



(a) σ^2_{PSF} relative error estimation

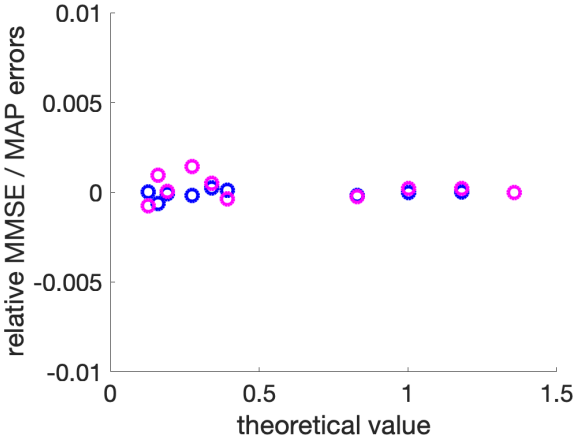
(b) σ^2_{PSF} posterior distribution

Figure B.2 – **Estimation of σ^2_{PSF} alone with model G_1.** (a) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the theoretical value of σ^2_{PSF}. (b) Estimated distribution of the parameters, for the ten tested cases. The true value of σ^2_{PSF} is represented by a black circle.

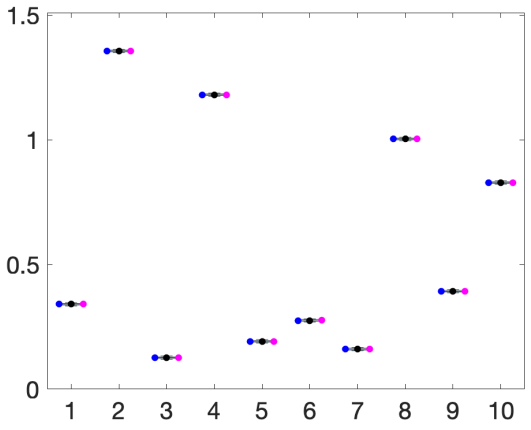
Parameter estimation with model G_2

The estimation of the diffusion coefficient D alone, with model G_2, is shown in Figure B.3. The MMSE and MAP estimates are very accurate, with relative errors of less than 0.5% for both (see Fig. B.3a). Figure B.3b confirms these results, as the standard deviation of the estimation of D is also low. The estimation of D alone with model G_2 is thus very accurate.

The estimation of σ^2_{PSF} alone with model G_2 is also accurate. In Figure B.4a, one can see that the relative error of both MMSE and MAP estimates are very low, i.e., below 0.5%. The boxplots in Figure B.4b show that the standard deviation of the estimator is also low for all tested real values. The estimation of σ^2_{PSF} alone is thus very accurate with model G_2.

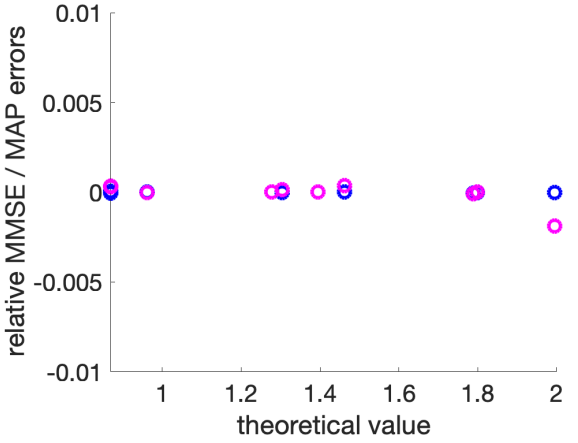


(a) D relative error estimation

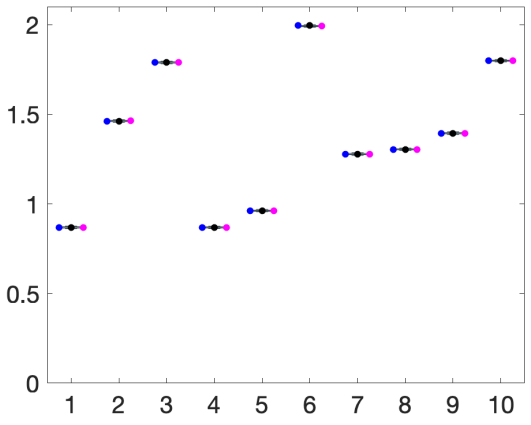


(b) D posterior distribution

Figure B.3 – **Estimation of D alone with model G_2.** (a) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the theoretical value of D. (b) Estimated distribution of the parameters, for the ten tested cases. The true value of D is represented by a black circle.



(a) σ^2_{PSF} relative error estimation



(b) σ^2_{PSF} posterior distribution

Figure B.4 – **Estimation of σ^2_{PSF} alone with model G_2.** (a) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the theoretical value of σ^2_{PSF}. (b) Estimated distribution of the parameters, for the ten tested cases. The true value of σ^2_{PSF} is represented by a black circle.

C Local variability - Influence of parameters $N_{samples}$ and R on the estimation

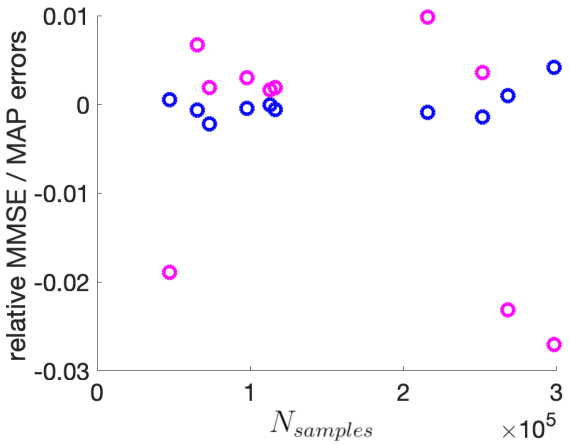
The influence of the parameters $N_{samples}$ (the algorithm parameter that defines the total number of generated simulations of the autocorrelation) and R (the algorithm parameter that sets the number of best simulations that are kept for the distribution estimation) is evaluated in what follows with respect to both model G_1 and model G_2.

Influence of the parameters $N_{samples}$ and R on the estimation with model G_1

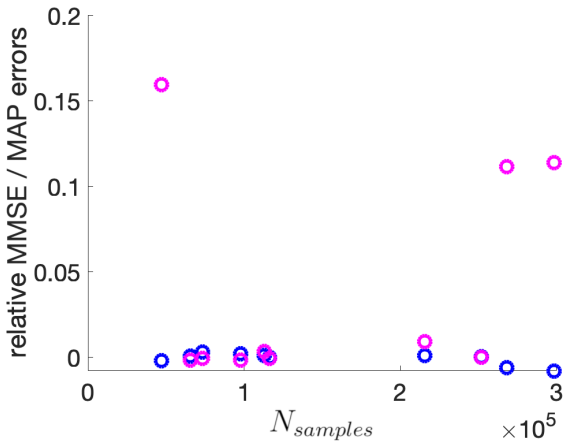
The influence of the parameters $N_{samples}$ and R for model G_1 is illustrated in Fig. C.3 and C.4. The influence of t is very small on the estimation of the parameter D with a relative estimation error of less than 2% for most cases, whereas, for the parameter σ^2_{PSF}, the influence is significant, but reasonable, with a relative error of $10\% - 20\%$ for most cases. In addition, the standard deviation of the estimated σ^2_{PSF} is much larger than for the estimation of D. A similar trend is observed for the parameter $N_{samples}$, that has a low influence on the estimation of the parameter D (relative estimation error of less than 3% and small standard deviation) but a slightly higher influence on the parameter σ^2_{PSF} (relative estimation error of less than 5% for most cases, and a larger standard deviation). Finally, for the parameter R, we notice a significant influence on the estimation of both parameters D and σ^2_{PSF} with, again, a higher influence for the parameter σ^2_{PSF}.

Influence of the parameters $N_{samples}$ and R on the estimation with model G_2

The influence of the parameters $N_{samples}$ and R for model G_2 is illustrated in Fig. C.7 and C.8. The results are very similar to those for model G_1 : the estimation of the parameter D is barely influenced by these parameters (slightly for R), whereas the estimation of the parameter σ^2_{PSF} is significantly influenced by these parameters, but remains in reasonable ranges. In all cases, the MAP estimation is very accurate.

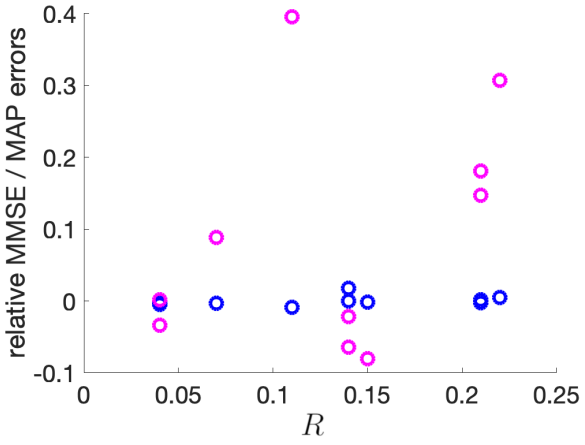


(a) D relative error estimation

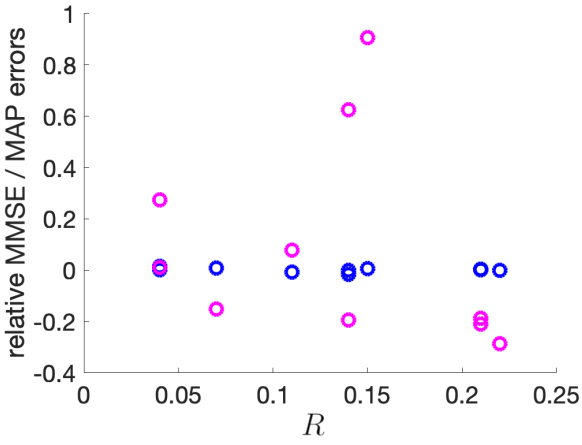


(b) σ^2_{PSF} relative error estimation

Figure C.1 – **Influence of $N_{samples}$ on parameter estimation with model G_1.** (a,b) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the value of $N_{samples}$.

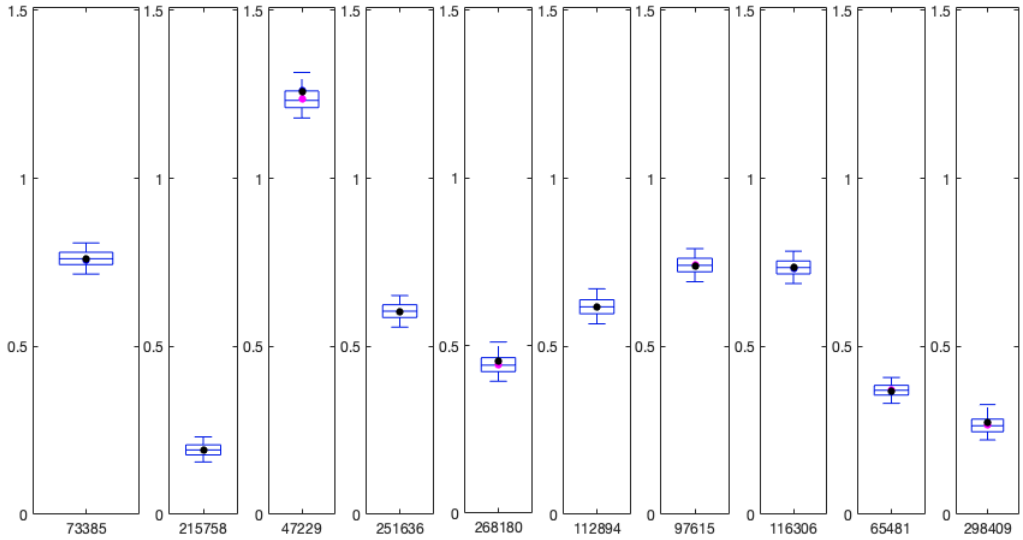


(a) D relative error estimation

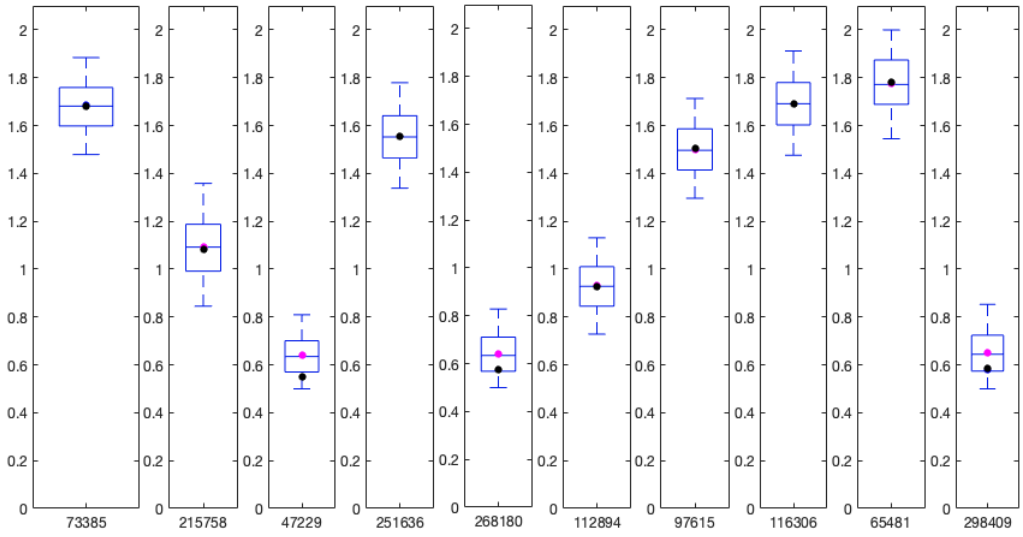


(b) σ^2_{PSF} relative error estimation

Figure C.2 – **Influence of R on parameter estimation with model G_1.** (a,b) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the value of R.

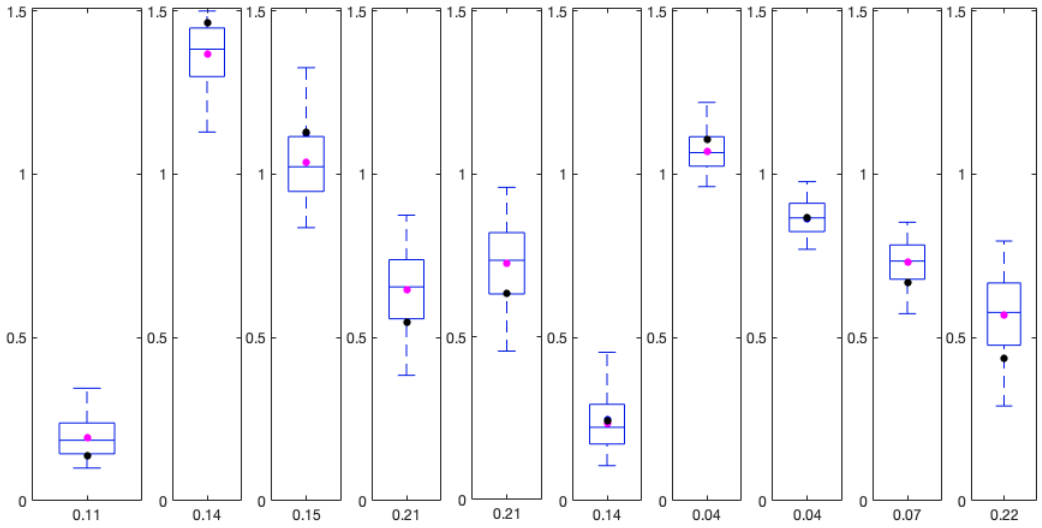


(a) D posterior distribution

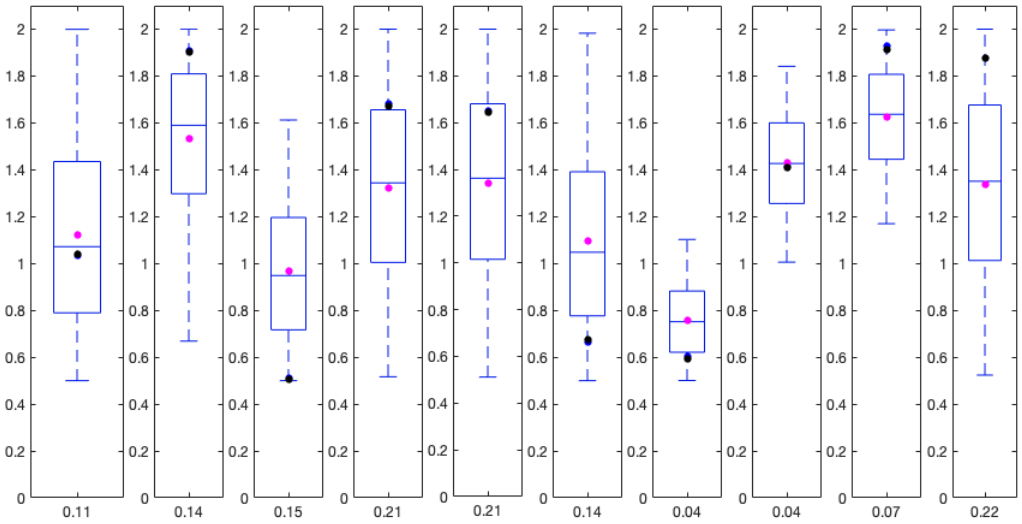


(b) σ^2_{PSF} posterior distribution

Figure C.3 – **Influence of $N_{samples}$ on parameter estimation with model G_1.** (a,b) Estimated distribution of parameters D and σ^2_{PSF}, for the ten tested cases corresponding to ten values of $N_{samples}$. The true values of D and σ^2_{PSF} are represented by black circles.

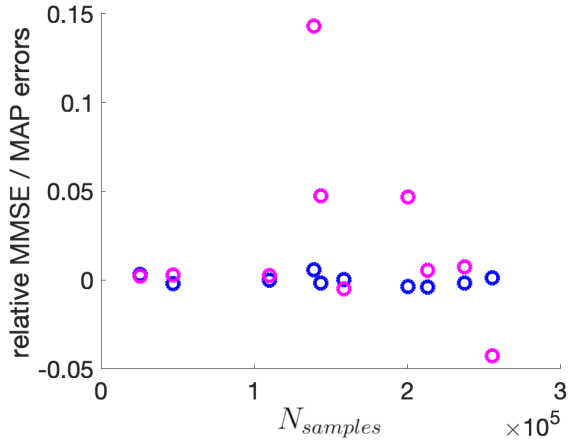


(a) D posterior distribution

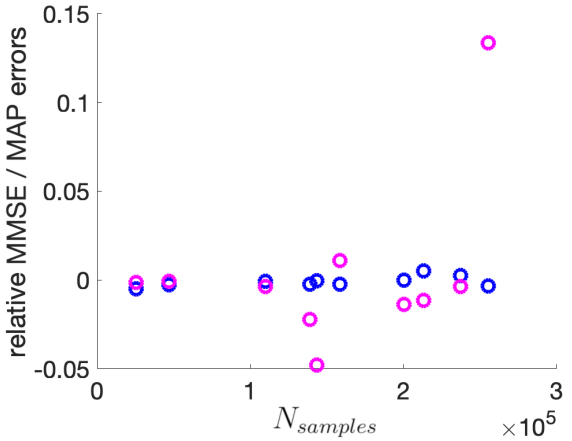


(b) σ^2_{PSF} posterior distribution

Figure C.4 – **Influence of R on parameter estimation with model G_1.** (a,b) Estimated distribution of parameters D and σ^2_{PSF}, for the ten tested cases corresponding to ten values of R. The true values of D and σ^2_{PSF} are represented by black circles.

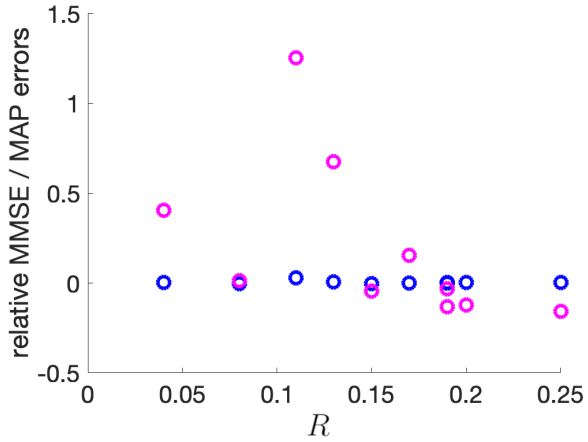


(a) D relative error estimation

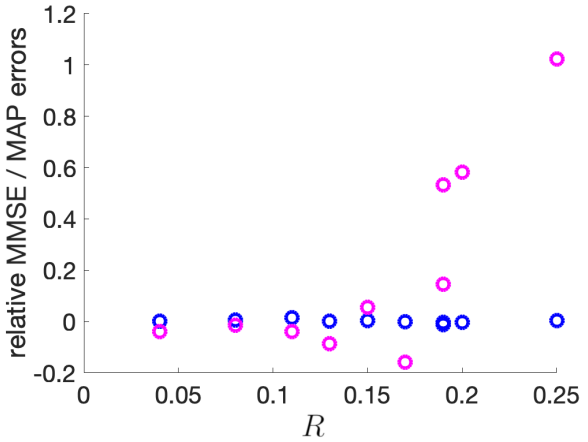


(b) σ^2_{PSF} relative error estimation

Figure C.5 – **Influence of $N_{samples}$ on parameter estimation with model G_2.** (a,b) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the value of $N_{samples}$.

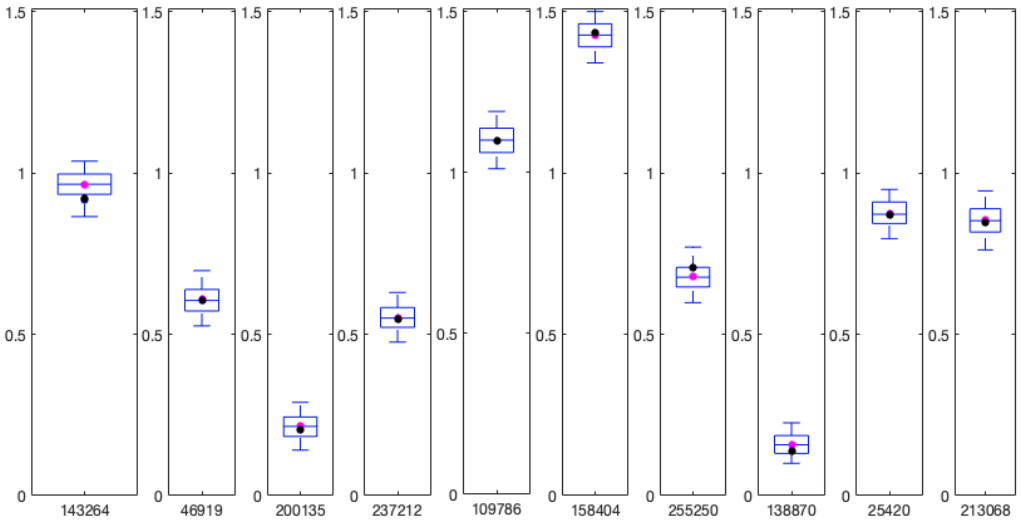


(a) D relative error estimation

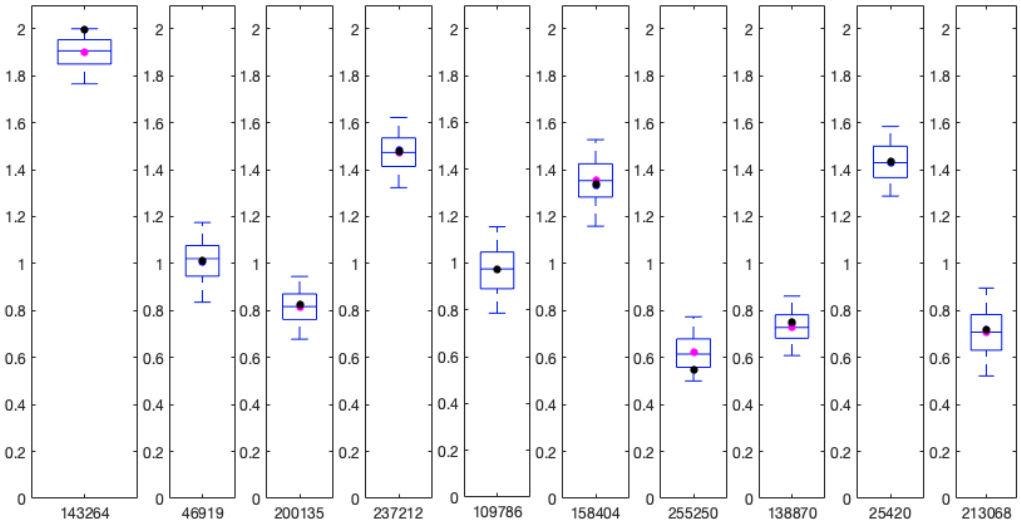


(b) σ^2_{PSF} relative error estimation

Figure C.6 – **Influence of R on parameter estimation with model G_2.** (a,b) Relative estimation error for the MAP (blue) and MMSE (magenta) estimates with respect to the value of R.

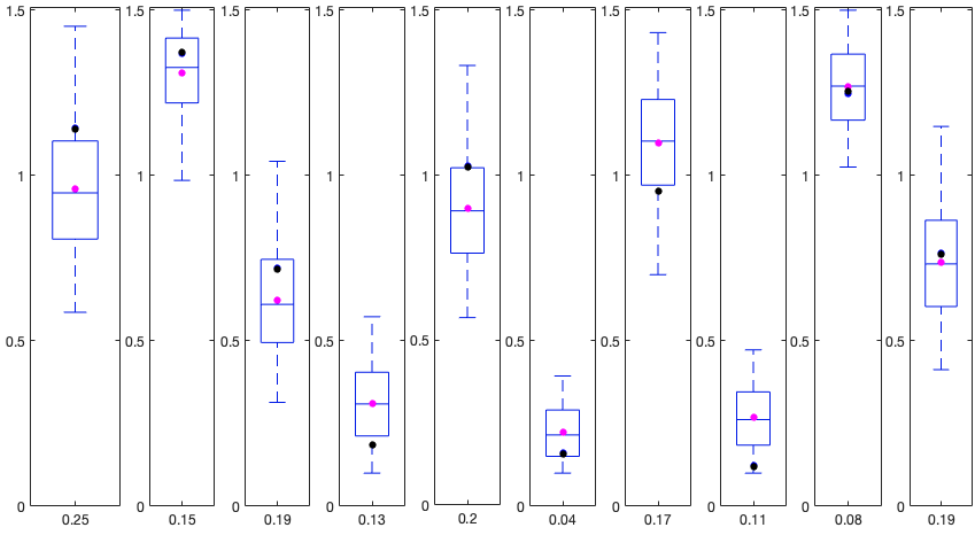


(a) D posterior distribution

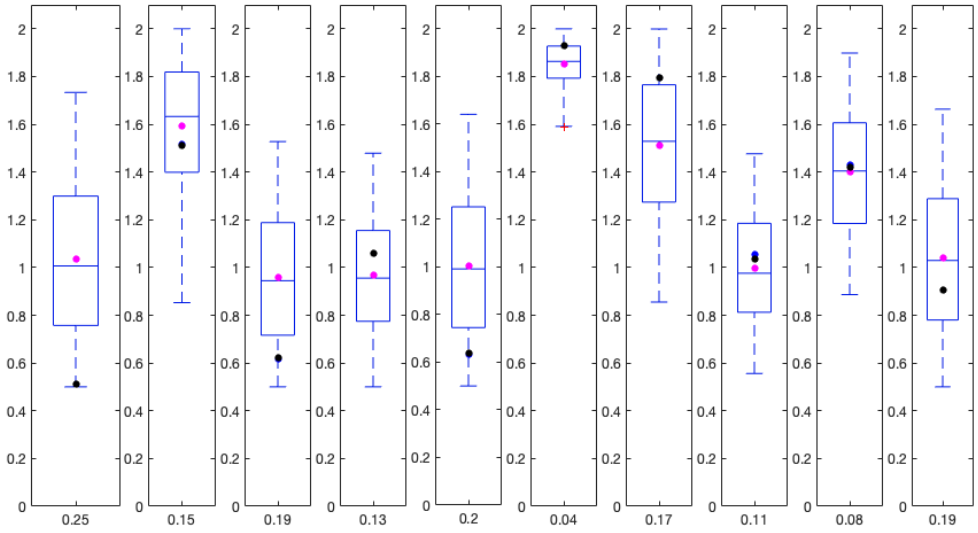


(b) σ^2_{PSF} posterior distribution

Figure C.7 – **Influence of $N_{samples}$ on parameter estimation with model G_2.** (a,b) Estimated distribution of parameters D and σ^2_{PSF}, for the ten tested cases corresponding to ten values of $N_{samples}$. The true values of D and σ^2_{PSF} are represented by black circles.



(a) D posterior distribution



(b) σ^2_{PSF} posterior distribution

Figure C.8 – **Influence of R on parameter estimation with model G_2.** (a,b) Estimated distribution of parameters D and σ^2_{PSF}, for the ten tested cases corresponding to ten values of R. The true values of D and σ^2_{PSF} are represented by black circles.

D Cell population variability - influence on the oscillations of the other model parameters

Influence of the parameter $\bar{f}$ on the oscillations Figure D.1 shows the influence of the parameter $\bar{f}$ on the oscillations. Here, $\bar{f}$ varies in the range $[6 \times 10^{-12}, 10 \times 10^{-12}]$ N, the other parameters being set to default values. We considered 50 simulations of the oscillations corresponding to 50 values of $\bar{f}$ that were randomly chosen in the given interval.

Figure D.1a shows that $\bar{f}$ has also a global impact on the oscillations; amplitude, frequency, shape and number of oscillations change with the value of $\bar{f}$. Figure D.1b shows the maximum amplitude as a function of the value of this parameter. In the same way as for N, the maximum amplitude depends almost linearly on the value of $\bar{f}$: the higher the value of $\bar{f}$, the higher the maximum amplitude of the oscillations. The maximum amplitude goes from 1×10^{-6} m for $\bar{f} = 6 \times 10^{-12}$ N to 7×10^{-6} m for $\bar{f} = 10 \times 10^{-12}$ N. Here, a small variation in $\bar{f}$ creates a large variation in the maximum amplitude of the oscillations. The frequency of oscillations is given in Figure D.1c. As for N, the frequency goes from approximately 0.03 Hz to 0.05 Hz, and it decreases with the value of $\bar{f}$.

Influence of the parameter f_c on the oscillations Figure D.2 shows the influence of the parameter f_c on the oscillations. Here, f_c varies in the range $[4 \times 10^{-13}, 8 \times 10^{-13}]$ N, the other parameters being set to default values. We considered 50 simulations of the oscillations corresponding to 50 values of f_c that were randomly chosen in the given interval.

In Fig. D.2a, one can notice that the parameter f_c has a global impact on the oscillations, similar to the previous parameters. The maximum amplitude, frequency and number of oscillations vary with the values of f_c. The maximum amplitude is shown in Figure D.2b; its value decreases with higher values of f_c in an almost linear fashion, going from 6×10^{-6} m for a value of $f_c = 4 \times 10^{-13}$ N, to 2×10^{-6} m for a value of $f_c = 8 \times 10^{-13}$ N. As for $\bar{f}$, the variation in f_c has a large impact on the maximum amplitude. The frequency of oscillations is given in Fig. D.2c. As for previous parameters, the frequency ranges from approximately 0.03 Hz to 0.05 Hz, and it increases with the value of f_c.

Influence of the parameter f_p on the oscillations Figure D.3 shows the influence of the parameter f_p on the oscillations. Here, f_p varies in the range $[0.9 \times 10^{-6}, 3 \times 10^{-6}]$ $N \cdot s/m$, the other parameters being set to default values. We considered 50 simulations of the oscillations corresponding to 50 values of f_p that were randomly chosen in the given interval.

Figure D.3a shows that the parameter f_p has also a global impact on the oscillations. The maximum amplitude, frequency and number of oscillations vary with the values of f_p, although the variation in f_p has a slightly lower impact on the oscillations than that in $\bar{f}$ and f_c. Indeed, the maximum amplitude, given in Figure D.3b, goes from 1×10^{-6} m to 5×10^{-6} m while the value of f_p more than triples, going from 0.9×10^{-6} $N \cdot s/m$ to 3×10^{-6} $N \cdot s/m$. In contrast, studies of the respective influences of $\bar{f}$ and f_c yielded maximum amplitude that were the same or even greater, but for a maximum parameter value that was, as most, double the initial value. As for the frequency of the oscillations, it decreases with the value of f_p, as showed in Fig. D.3c, and goes from 0.055 Hz to 0.025 Hz, so that f_p has approximately the same impact on the frequency as the previous parameters.

Influence of the parameter K_0 on the oscillations Figure D.4 shows the influence of the parameter K_0 on the oscillations. Here, K_0 varies in the range $[6 \times 10^{-6}, 60 \times 10^{-6}]$ N/m, the other parameters being set to default values. We considered 50 simulations of the oscillations corresponding to 50 values of K_0 that were randomly chosen in the given interval.

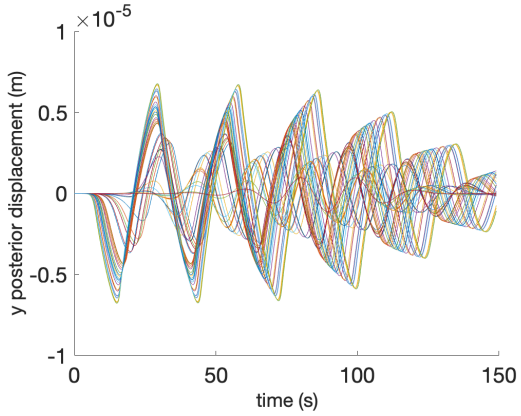
Figure D.4a shows the impact of the parameter K_0 on the oscillations. As for the other parameters, the impact is global, but in a different manner: the beginning and the end of the oscillations occur at almost the same time for all values of the parameter K_0; this was not the case for the previous parameters N, $\bar{f}$, f_c and f_p. Thus, K_0 does not affect the timing of the oscillations. Nevertheless, the amplitude, frequency and number of oscillations do vary with the values of K_0. The maximum amplitude, shown in Fig. D.4b, decreases with the value of the parameter, going from 5×10^{-6} m to 1×10^{-6} m as K_0 goes from 6×10^{-6} N/m to 60×10^{-6} N/m. The frequency of oscillations, shown in Fig. D.4c, increases from 0.03 Hz to 0.1 Hz. Hence, the parameter K_0 has a moderate global impact on the oscillations, similar to f_p.

Influence of the parameter Γ on the oscillations Figure D.5 shows the influence of the parameter Γ on the oscillations. Here, Γ varies in the range $[50 \times 10^{-6}, 200 \times 10^{-6}]$ $N \cdot s/m$, the other parameters being set to default values. We considered 50 simulations of the oscillations corresponding to 50 values of Γ that were randomly chosen in the given interval.

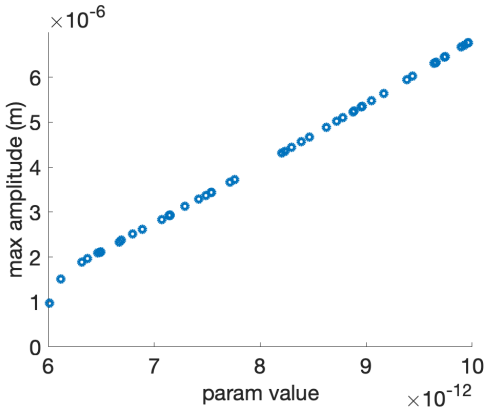
Figure D.5a shows that Γ has a global impact on the oscillations; in particular, it has a strong influence on the amplitude and the number of oscillations. Interestingly, Γ seems to have a higher impact on the timing for the beginning of the oscillations than the other parameters. The maximum amplitude is given in Fig. D.5b: it decreases in a linear manner with the values of Γ, going from 5×10^{-6} m for $\Gamma = 50 \times 10^{-6}$ $N \cdot s/m$, to less than 1×10^{-6} m for $\Gamma = 200 \times 10^{-6}$ $N \cdot s/m$. In contrast to the other parameters, Γ has almost no impact on the frequency of oscillations. Figure D.5c shows the frequency as a function of the value of the parameter; it shows that, for almost all simulations, the frequency is constant to approximately 0.049 Hz.

Influence of the parameter $\tilde{k}_{end}$ on the oscillations Figure D.6 shows the influence of the parameter $\tilde{k}_{end}$ on the oscillations. Here, $\tilde{k}_{end}$ varies in the range $[0.01, 1]$, the other parameters being set to default values. We considered 50 simulations of the oscillations corresponding to 50 values of $\tilde{k}_{end}$ that were randomly chosen in the given interval.

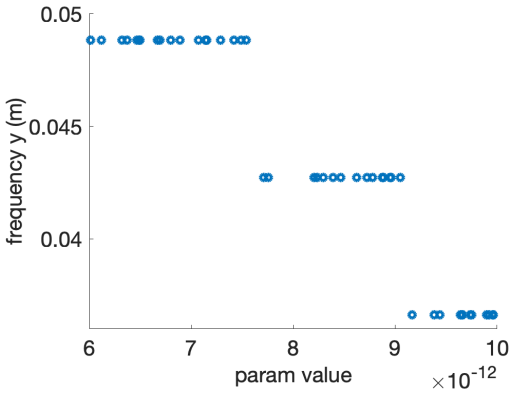
Figure D.6a shows the impact of the parameter $\tilde{k}_{end}$ on the oscillations. The parameter $\tilde{k}_{end}$ has an influence that is even more restricted than that of parameter d, although the impact is similar to that of b and d. The timing of the beginning and the build-up are exactly the same for all simulations. In addition, the amplitude during build-up does not change either from one simulation to another. Only the second part of the oscillations changes. Depending on the value of $\tilde{k}_{end}$ relative to a, the oscillations either die out (if $\tilde{k}_{end} < a$) or get amplified (if $\tilde{k}_{end} > a$). The maximum amplitude is shown in Fig. D.6b. For values of $\tilde{k}_{end}$ in the interval $[0.01, 0.5]$, the maximum amplitude does not change significantly, whereas for $\tilde{k}_{end} > 0.5$ (and, here, $a = 0.5$), the amplitude increases with $\tilde{k}_{end}$. In contrast, the frequency of oscillations does not change at all for varying $\tilde{k}_{end}$, as shown in Fig. D.6c: its value is stable at 0.05 Hz.



(a) y vs t

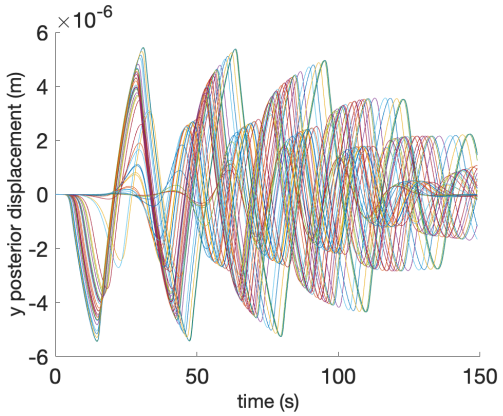


(b) max ampl vs param

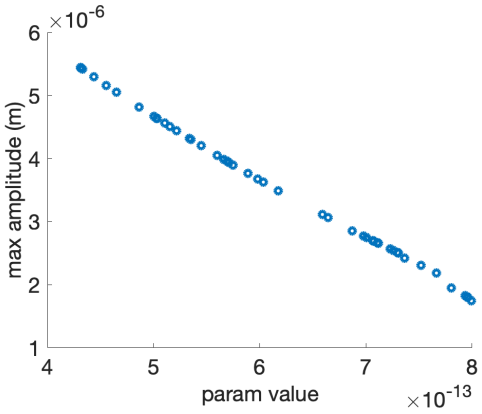


(c) freq vs param

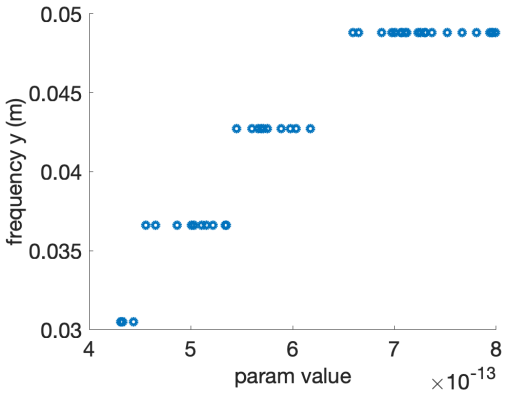
Figure D.1 – **Influence of parameter $\bar{f}$ on the oscillations.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.



(a) y vs t

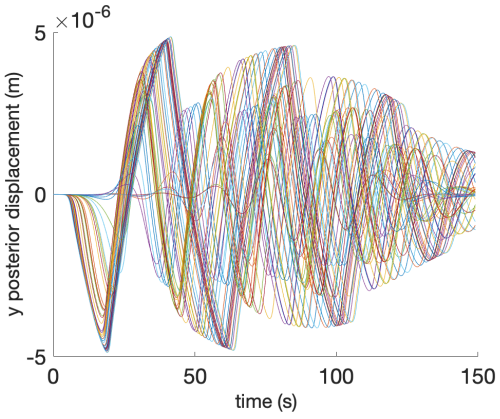


(b) max ampl vs param

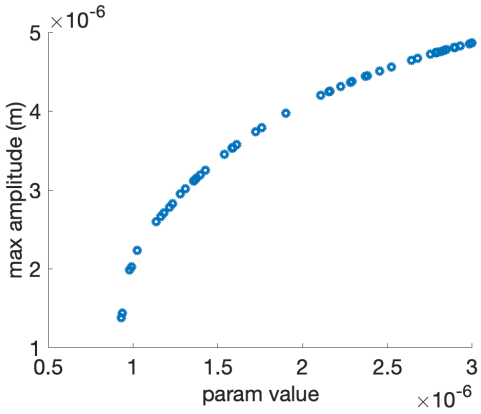


(c) freq vs param

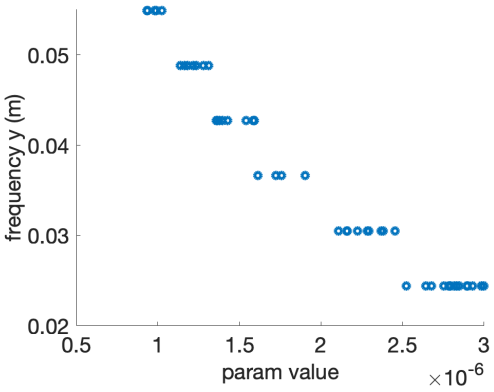
Figure D.2 – **Influence of parameter f_c on the oscillations.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.



(a) y vs t

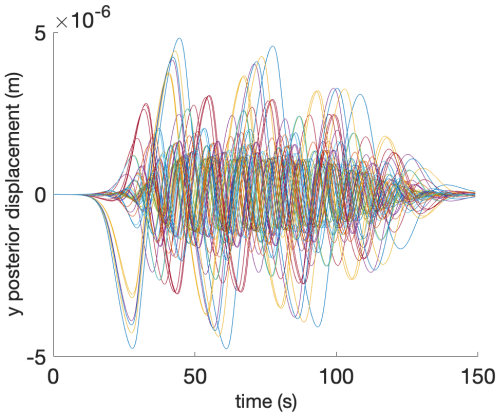


(b) max ampl vs param

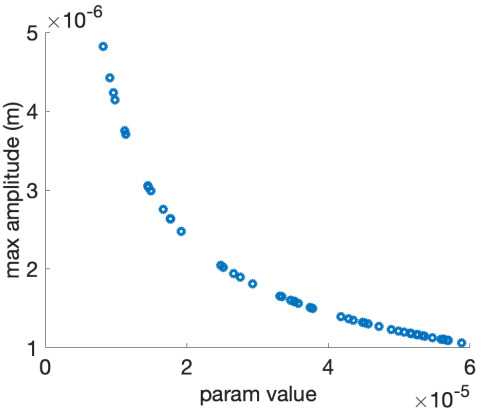


(c) freq vs param

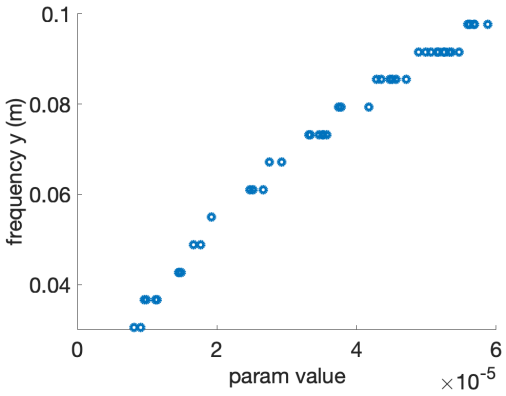
Figure D.3 – **Influence of parameter f_p on the oscillations.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.



(a) y vs t

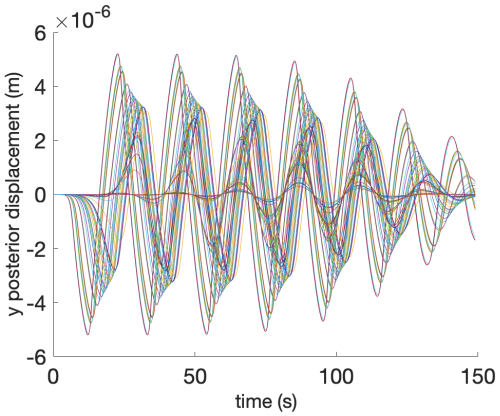


(b) max ampl vs param

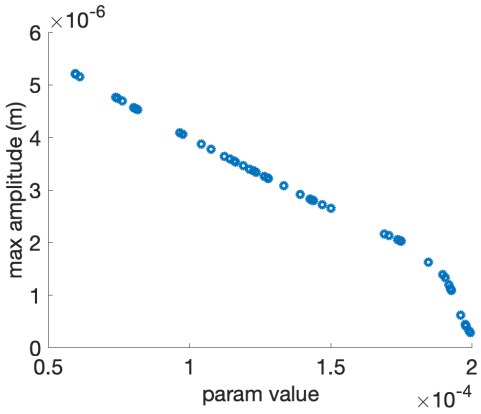


(c) freq vs param

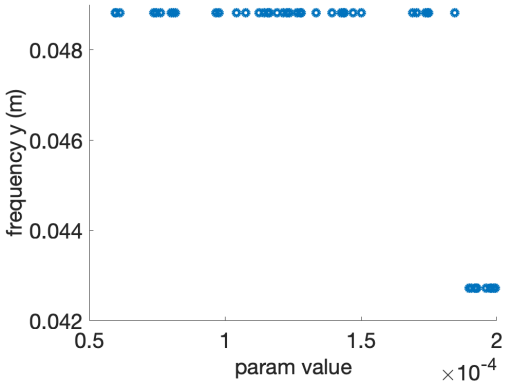
Figure D.4 – **Influence of parameter K_0 on the oscillations.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.



(a) y vs t

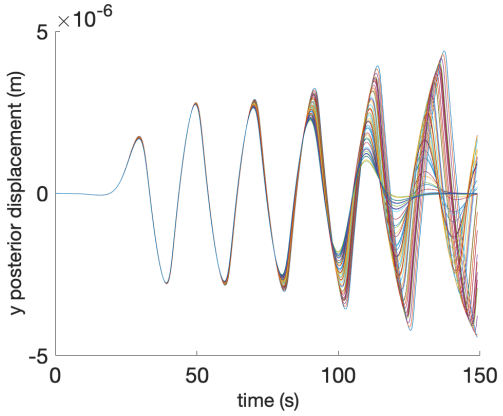


(b) max ampl vs param

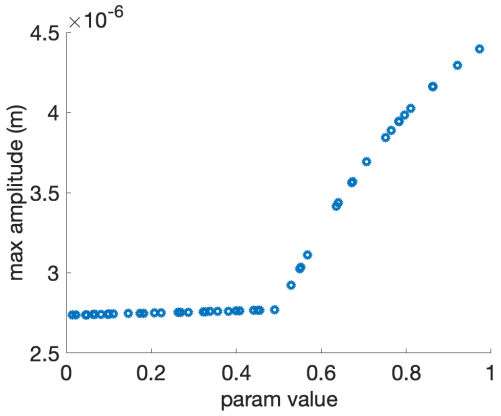


(c) freq vs param

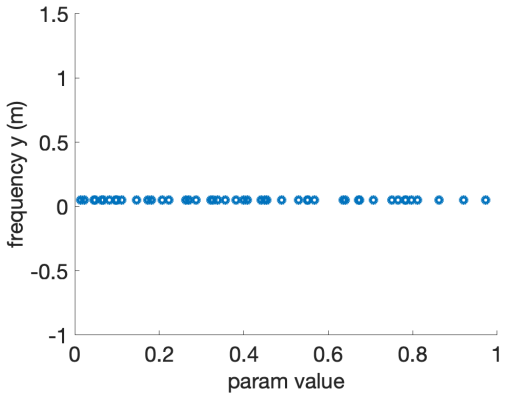
Figure D.5 – **Influence of parameter Γ on the oscillations.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.



(a) y vs t



(b) max ampl vs param



(c) freq vs param

Figure D.6 – **Influence of parameter $\tilde{k}_{end}$ on the oscillations.** (a) Transverse oscillations for different tested parameter values, providing information about the number of oscillations, the build-up and die-down timing. (b) Variation in maximum amplitude with respect to the variation in parameter value. (c) Variation in frequency with respect to the variation in parameter value.

E Cell population variability - Limit behavior for extremal values of the other model parameters

Limit behavior for the parameter $\bar{f}$ The limit behavior of the model for extremal values of the parameter $\bar{f}$ was evaluated; the results are shown in Fig. E.1. Simulations were obtained by varying $\bar{f}$, with the other parameters being set to default values. Taking into account the default values of the other parameters, the lower limit of $\bar{f}$ was studied in the range $[4\times10^{-12}, 6\times10^{-12}]$ N, and the upper limit in the range $[11\times10^{-12}, 15\times10^{-12}]$ N.

As for the parameter N, in the low range $[4\times10^{-12}, 6\times10^{-12}]$ N for $\bar{f}$, the oscillations decrease as $\bar{f}$ decreases, until they reach a point where they do not build up anymore; they then die out rapidly, and disappear completely when $\bar{f}$ is close to 4×10^{-12} N, as shown in Fig. E.1a and E.1b. When $\bar{f}$ is in the upper range, the model behaves in a way similar to the case when N is in its upper range. Figure E.1c shows the oscillations for high values of $\bar{f}$; it can be seen that, as in the case of N, the oscillations continue to exist, the maximum amplitude increases with increasing $\bar{f}$, the frequency and the number of oscillations decrease, but the pattern of oscillations is not realistic.

Limit behavior for the parameter f_c The limit behavior of the model for extremal values of the parameter f_c was evaluated; the results are shown in Figure E.2. Simulations were obtained by varying f_c, with the other parameters being set to default values. Taking into account the default values of the other parameters, the lower limit of f_c was studied in the range $[2\times10^{-13}, 3.5\times10^{-13}]$ N, and the upper limit in the range $[8.5\times10^{-13}, 11\times10^{-13}]$ N.

The oscillatory behavior for the limit values of f_c is similar to the one for the parameter $\bar{f}$, but in reverse. This is shown in Fig. E.2a. For low values of f_c, the oscillations occur as with high values for $\bar{f}$: oscillations continue to exist, but following an unrealistic pattern, with high maximum amplitude and low frequency and number of oscillations. As for high values of f_c, as shown in Fig. E.2b and E.2c, the oscillations decrease in maximum amplitude, until there is no build-up; finally, above a given value for f_c, the oscillations disappear completely.

Limit behavior for the parameter f_p The limit behavior of the model for extremal values of the parameter f_p was evaluated; the results are shown in Fig. E.3. Simulations were obtained by varying f_p, with the other parameters being set to default values. Taking into account the default values of the other parameters, the lower limit of f_p was studied in the range $[0.7 \times 10^{-6}, 0.9 \times 10^{-6}]$ $N \cdot s/m$, and the upper limit in the range $[3.5 \times 10^{-6}, 4.5 \times 10^{-6}]$ $N \cdot s/m$.

The behavior of the model for extremal values of f_p is roughly the same as the behavior for $\bar{f}$. The case of low range is shown in Fig. E.3a and E.3b, and the case of high range is given in Fig. E.3c. The main difference with the behavior for $\bar{f}$ is the impact on the maximum amplitude. Here, for decreasing values of f_p, the maximum amplitude decreases at a lower rate than for $\bar{f}$, so that the switching between the case of oscillations with build-up and the case of only die-out oscillations is clearly visible in Fig. E.3a.

Limit behavior for the parameter K_0 The limit behavior of the model for extremal values of the parameter K_0 was evaluated; the results are shown in Fig. E.4. Simulations were obtained by varying K_0, with the other parameters being set to default values. Taking into account the default values of the other parameters, the lower limit of K_0 was studied in the range $[6 \times 10^{-6}, 9 \times 10^{-6}]$ N/m, and the upper limit in the range $[120 \times 10^{-6}, 200 \times 10^{-6}]$ N/m.

Figure E.4a shows the oscillations for low values of K_0, and it can be seen that the behavior is similar to the behavior for low values of the parameter f_c: high amplitude, but low frequency and number of oscillations. For high values of K_0, the model also behaves in a way similar to the case of high values of f_c, but the frequency of the oscillations is much higher: around 0.1 Hz, that is, approximately double the frequency for high values of f_c (around 0.05 Hz). For this case, the results are shown in Fig. E.4b and E.4c. As for the previous parameter, the rate of increase/decrease of the maximum amplitude is very low, so that the point when the oscillations do not build up anymore is clearly visible, despite the very high frequency.

Limit behavior for the parameter Γ The limit behavior of the model for extremal values of the parameter Γ was evaluated; the results are shown in Fig. E.5. Simulations were obtained by varying Γ, with the other parameters being set to default values. Tak-

ing into account the default values of the other parameters, the lower limit of Γ was studied in the range $[10 \times 10^{-6}, 40 \times 10^{-6}]$ $N \cdot s/m$, and the upper limit in the range $[190 \times 10^{-6}, 240 \times 10^{-6}]$ $N \cdot s/m$.

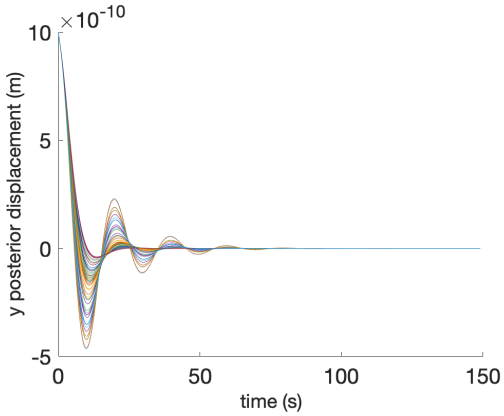
The oscillatory behavior for extremal values of Γ is similar to the behavior for extremal values of K_0. For low values of Γ, the results are given in Fig. E.5a; the main difference with the previous case (for parameter K_0) is the frequency of the oscillations, that is here constant, and not varying as for K_0. As for K_0, the oscillations continue to exist, and the maximum amplitude is high, whereas the frequency and number of oscillations stay roughly the same for the different values of Γ. The results for high values of Γ are given in Fig. E.5b and E.5c. Again, the behavior is quite similar to K_0, with the same difference concerning the frequency, which is constant for Γ while it varied for K_0. The other notable difference is in the rate of decrease for the maximum amplitude, which is much higher for Γ than for K_0. The oscillations decrease in maximum amplitude until there is no build-up, and then, until they completely disappear; this last part is not visible in Fig. E.5c.

Limit behavior for the parameter $\tilde{k}_{end}$ The behavior for limit values of $\tilde{k}_{end}$ is shown in Fig. E.6. Taking into account the default values of the other parameters, the lower limit of $\tilde{k}_{end}$ was studied in the range $[0.01, 0.10]$, and the upper limit in the range $[0.60, 0.90]$. The parameter $\tilde{k}_{end}$ only affects the die-out and ending of oscillations, so that in both low and high value cases, the oscillations continue to exist. For low values of $\tilde{k}_{end}$, the impact on the oscillations is minimal, with only the last oscillation being affected, with a lower amplitude, as shown in Fig. E.6a. For high values of $\tilde{k}_{end}$, the behavior follows one of the two following scenarios:

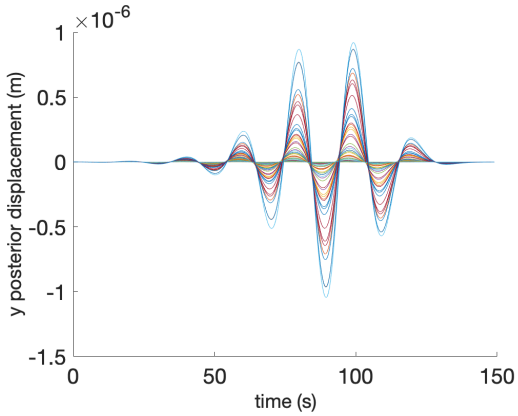
- If the value of $\tilde{k}_{end}$ is lower than the value of a, the oscillations eventually die out after a period of steady oscillations, as shown in Fig. E.6b.

- If the value of $\tilde{k}_{end}$ is higher than the value of a, the oscillations get amplified and a second build-up occurs after a period of steady oscillations, as shown in Fig. E.6c.

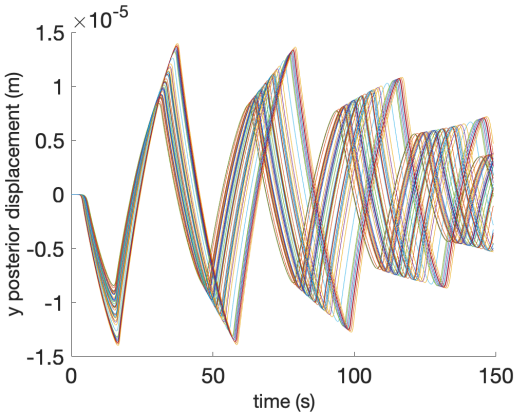
The parameter $\tilde{k}_{end}$ is the only one that cannot abolish oscillations by itself. For all cases, the frequency of oscillations is stable.



(a) $\bar{f} = 4 * 10^{-12} - 5 * 10^{-12}$

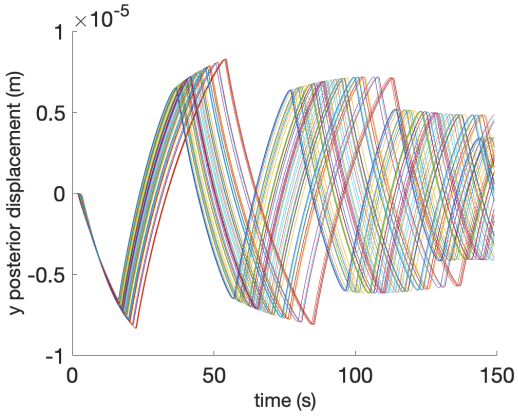


(b) $\bar{f} = 5.5 * 10^{-12} - 6 * 10^{-12}$

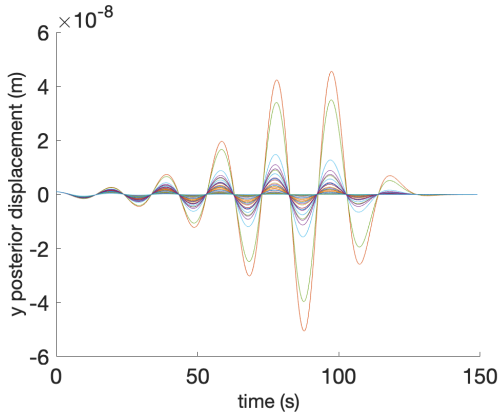


(c) $\bar{f} = 11 * 10^{-12} - 15 * 10^{-12}$

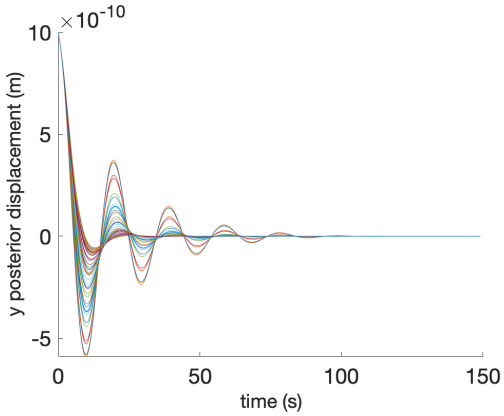
Figure E.1 – **Limit behavior of the model for extremal values of the parameter $\bar{f}$.**



(a) $f_c = 2 \times 10^{-13} - 3.5 \times 10^{-13}$

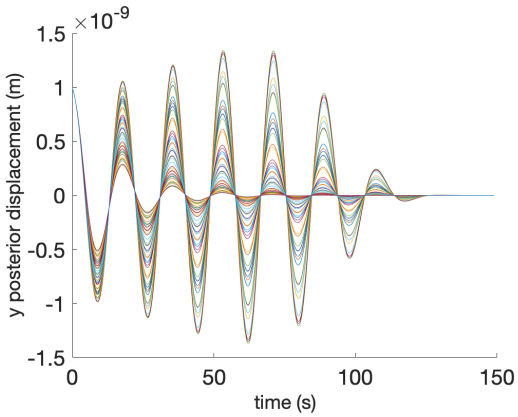


(b) $f_c = 8.5 \times 10^{-13} - 9.5 \times 10^{-13}$

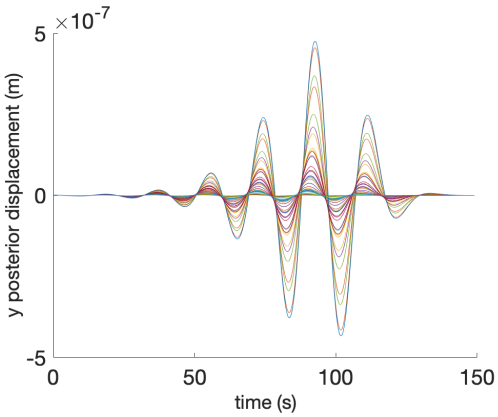


(c) $f_c = 9.5 \times 10^{-13} - 12 \times 10^{-13}$

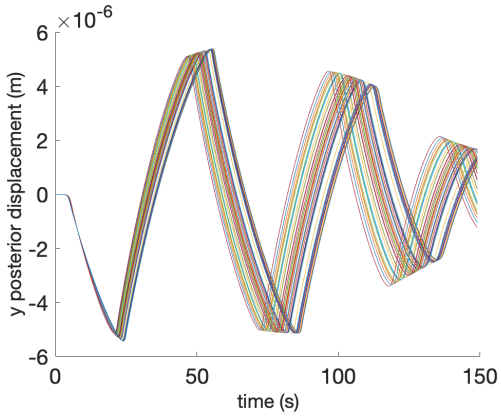
Figure E.2 – **Limit behavior of the model for extremal values of the parameter f_c.**



(a) $f_p = 0.7 \times 10^{-6} - 0.8 \times 10^{-6}$

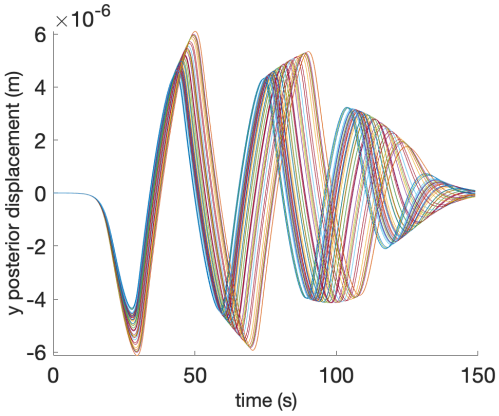


(b) $f_p = 0.8 \times 10^{-6} - 0.9 \times 10^{-6}$

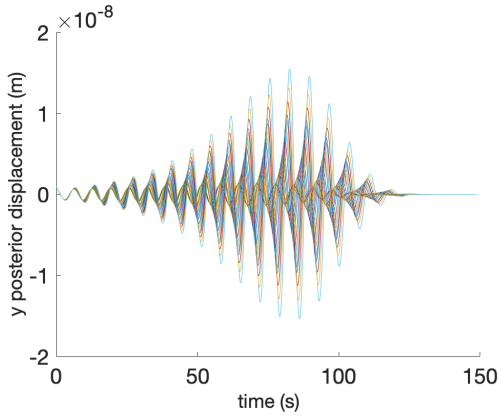


(c) $f_p = 3.5 \times 10^{-6} - 4.5 \times 10^{-6}$

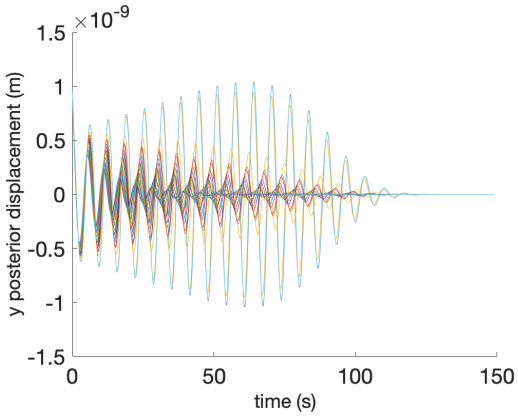
Figure E.3 – **Limit behavior of the model for extremal values of the parameter f_p.**



(a) $K_0 = 6 \times 10^{-6} - 9 \times 10^{-6}$

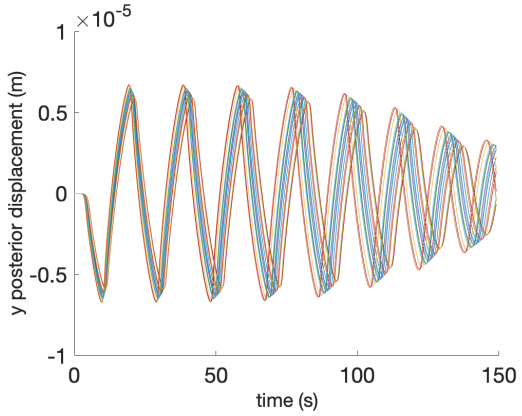


(b) $K_0 = 120 \times 10^{-6} - 150 \times 10^{-6}$

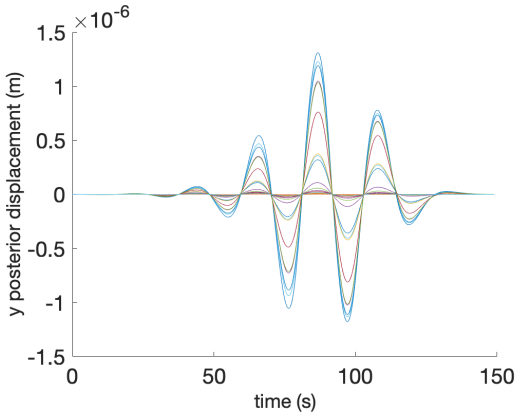


(c) $K_0 = 150 \times 10^{-6} - 200 \times 10^{-6}$

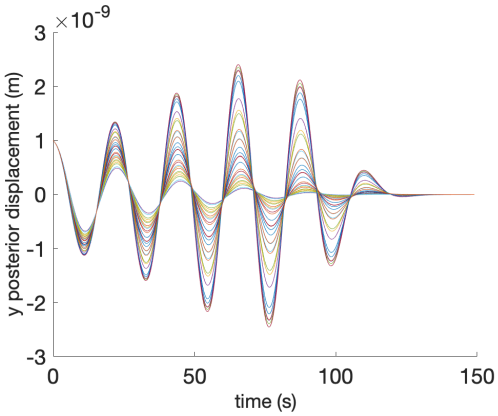
Figure E.4 – **Limit behavior of the model for extremal values of the parameter K_0.**



(a) $\Gamma = 10 \times 10^{-6} - 40 \times 10^{-6}$

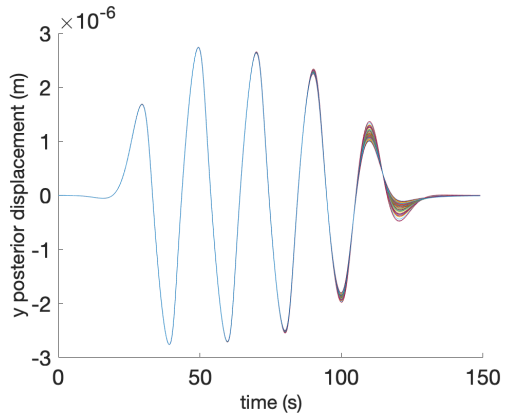


(b) $\Gamma = 190 \times 10^{-6} - 220 \times 10^{-6}$

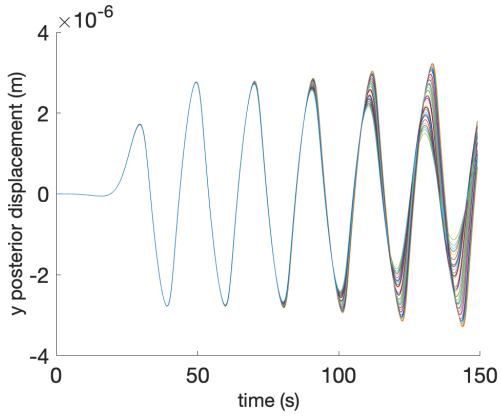


(c) $\Gamma = 220 \times 10^{-6} - 240 \times 10^{-6}$

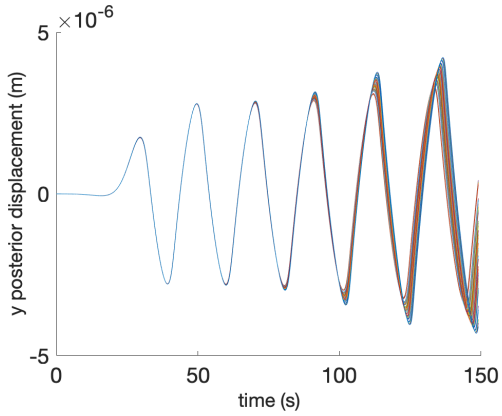
Figure E.5 – **Limit behavior of the model for extremal values of the parameter Γ.**



(a) $end_{koff} = 0.01 - 0.10$



(b) $\tilde{k}_{end} = 0.30 - 0.60$

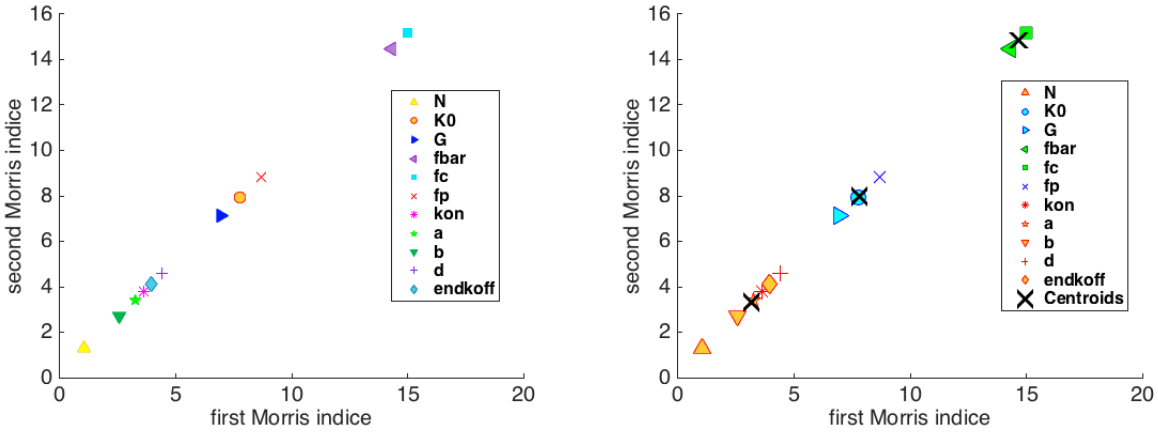


(c) $\tilde{k}_{end} = 0.60 - 0.90$

Figure E.6 – **Limit behavior of the model for extremal values of the parameter $\tilde{k}_{end}$.**

F Cell population variability - Morris sensitivity analysis

The estimation framework introduced in section 3.3 uses the Morris method to analyze the sensitivity of our model to its parameters. Our implementation of this method showed that the 11 parameters of our model could be split into three groups, describing subsets of parameters with different influences on the results. This clustering, illustrated in Fig. F.1, was used as a basis for the design of our hierarchical estimation framework.



(a) Results for the 11 parameters

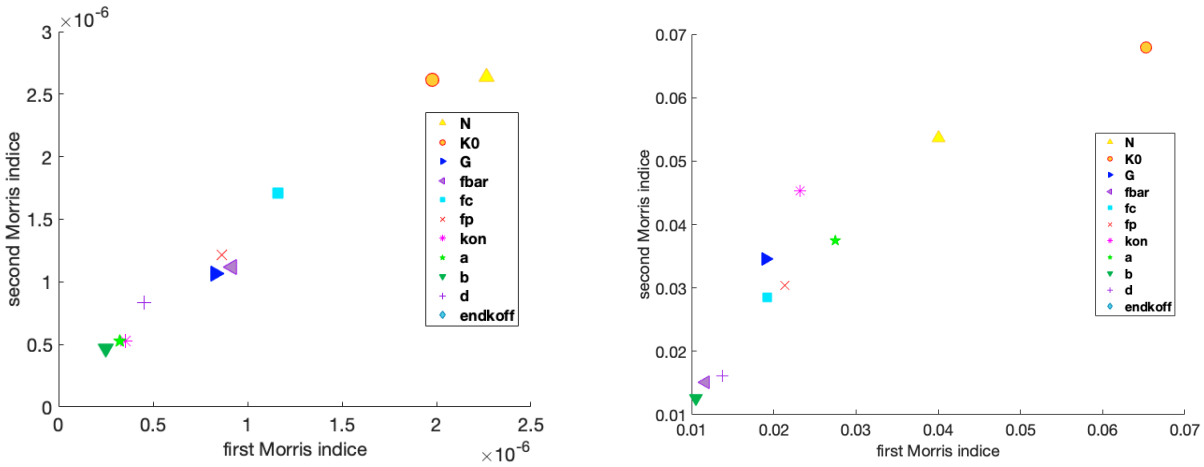
(b) Paramer Clustering

Figure F.1 – **First results of the parameter sensitivity analysis, using the maximum amplitude as a cost function (see below for corrected figure).** (a) Plot of the first and second Morris indices (log scale) for the 11 parameters of our model. (b) Clustering of the previous result.

As shown in the remainder of chapter 3, the proof-of-concept implementation of the estimation framework, based on these parameter groups, yields very promising results.

However, issues were discovered in the implementation of the Morris method during the late stages of the writing of this manuscript. More specifically, subtle bugs had been unnoticed in the computation of the elementary effects.

Fig. F.2 shows the results of the corrected analysis when two different characteristics of the model outcome are used as cost functions for the method: the maximum amplitude of the oscillations, and their computed frequency.



(a) Maximum amplitude

(b) Frequency

Figure F.2 – **Parameter sensitivity analysis with the Morris method.** Plot of the first and second Morris indices for the 11 parameters of our model using (a) the maximum amplitude and (b) the frequency of oscillations, as a cost function.

Depending on the cost function, different Morris indices are computed for the parameters of the model, and the number of apparent groups changes. In chapter 3 of the manuscript, the initial groups provided satisfactory results. Further studies would be needed in order to determine how to "optimally" split the set of parameters, and how to make this analysis fully automated.

G Cell population variability - Parameter estimation on simulated data : setting the parameters

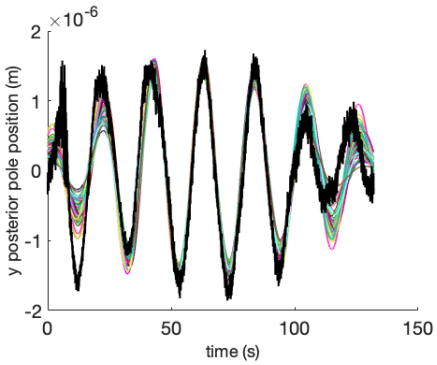
The following table provides the parameters values used to generate the 14 simulated data used in Section 3.3.2, that are close in behavior to typical wild-type embryos.

N	K_0 (N/m)	Γ $(N \cdot s/m)$	k_{on} (s^{-1})	a	b	d
42.9	1.65e-5	1.71e-4	0.56	0.91	11.30	0.116
43.8	3.90e-5	1.23e-4	0.30	0.73	7.83	0.095
80.0	2.66e-5	1.39e-4	0.94	0.38	2.28	0.052
47.6	1.74e-5	1.36e-4	0.78	0.37	12.36	0.056
52.2	1.57e-5	1.40e-4	0.78	0.43	11.70	0.187
64.2	1.85e-5	1.66e-4	0.81	0.41	7.82	0.054
81.9	2.85e-5	1.25e-4	0.91	0.37	3.46	0.113
54.1	1.71e-5	1.27e-4	1.03	0.58	5.41	0.077
87.5	2.57e-5	1.30e-4	1.02	0.39	7.79	0.119
94.9	2.34e-5	1.25e-4	0.97	0.38	2.11	0.053
87.7	4.06e-5	1.33e-4	0.40	1.00	3.45	0.187
72.8	1.67e-5	1.46e-4	0.95	0.38	5.96	0.070
71.4	3.26e-5	1.28e-4	0.47	0.37	3.10	0.104
49.0	1.67e-5	1.54e-4	0.76	1.97	2.76	0.057

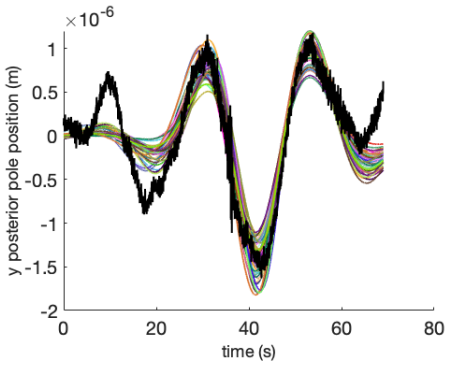
Table G.1 – **Parameter values used for the generation of the 14 simulated data used in Section 3.3.2.**

H Cell population variability - Parameter estimation on real data : fitting the anterior pole

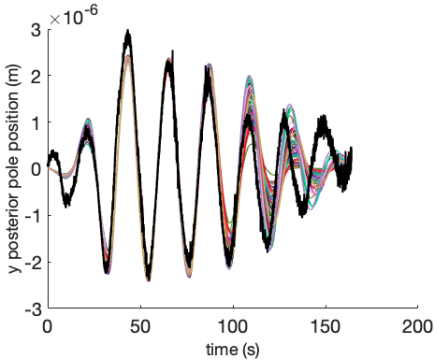
In what follows, we show the fitting of the anterior pole for the seven real embryos.



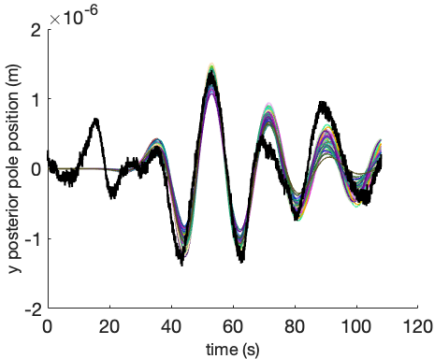
(a) embryo $E1$



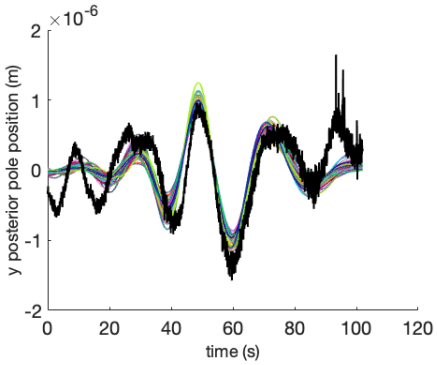
(b) embryo $E2$



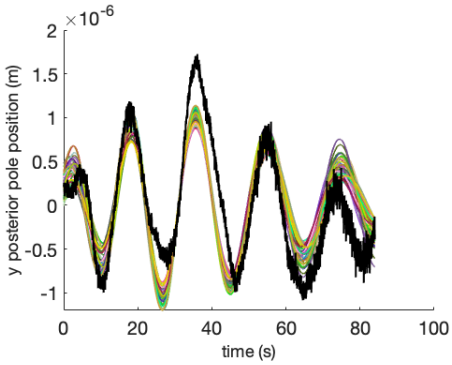
(c) embryo $E3$



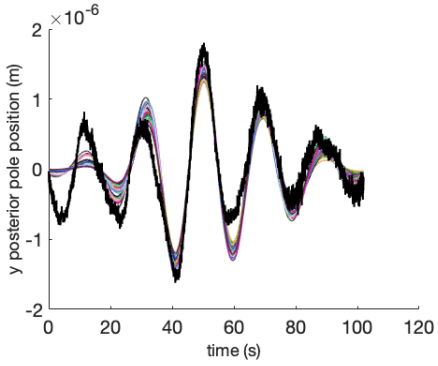
(d) embryo $E4$



(e) embryo $E5$



(f) embryo $E6$



(g) embryo $E7$

Figure H.1 – **Fitting the anterior pole on wild-type embryos.** In each sub-figure, the black curve represents the experimental data (tracking of the anterior pole position in time) and the colored curves represent the simulations corresponding to the solution given by our method.

BIBLIOGRAPHY

[1] R. Zachariah Aandahl et al., « A Model-Based Bayesian Estimation of the Rate of Evolution of VNTR Loci in Mycobacterium tuberculosis », *in*: *PLoS Computational Biology* 8 (June 2012), pp. 1–9.

[2] Murat Acar, Jerome T. Mettetal, and Alexander Van Oudenaarden, « Stochastic switching as a survival strategy in fluctuating environments », *in*: *Nature Genetics* 40.*4* (2008), pp. 471–475.

[3] Martin Ackermann, « A functional perspective on phenotypic heterogeneity in microorganisms », *in*: *Nature Reviews Microbiology* 13.*8* (2015), pp. 497–508.

[4] David G. Adams, « Heterocyst formation in cyanobacteria », *in*: *Current Opinion in Microbiology* 3.*6* (2000), pp. 618–624.

[5] Anna Akhmanova and Michel O Steinmetz, « Tracking the ends: a dynamic protein network controls the fate of microtubule tips », *in*: *Nature Reviews Molecular Cell Biology* 9.*4* (2008), pp. 309–322.

[6] Paulo F. F. Almeida and Winchil L. C. Vaz, « Lateral diffusion in membranes », *in*: *Handbook of Biological Physics*, vol. 1, Elsevier, 1995, pp. 305–357.

[7] Steven J. Altschuler and Lani F. Wu, « Cellular heterogeneity: do differences make a difference? », *in*: *Cell* 141 (2010), pp. 559–563.

[8] Markus Arnoldini et al., « Bistable expression of virulence genes in salmonella leads to the formation of an antibiotic-tolerant subpopulation », *in*: *PLoS Biology* 12.*8* (2014), e1001928.

[9] Nathalie Q. Balaban et al., « Bacterial persistence as a phenotypic switch », *in*: *Science* 305.*5690* (2004), pp. 1622–1625.

[10] Antoine Basset et al., « Detection and estimation of membrane diffusion during exocytosis in TIRFM image sequences », *in*: *Proc. of International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 695–698.

[11] Mark A. Beaumont, Wenyang Zhang, and David J. Balding, « Approximate Bayesian Computation in Population Genetics », *in*: *Genetics* 162.*4* (Dec. 2002), pp. 2025–2035.

[12] Francois Bertaux et al., « Modeling dynamics of cell-to-cell variability in TRAIL-induced apoptosis explains fractional killing and predicts reversible resistance », *in*: *PLoS Computational Biology* 10.*10* (2014), e1003893.

[13] Matthew G. Blango and Matthew A. Mulvey, « Bacterial landlines: contact-dependent signaling in bacterial populations », *in*: *Current Opinion in Microbiology* 12.*2* (2009), pp. 177–181.

[14] Sydney Brenner, « The genetics of Caenorhabditis elegans », *in*: *Genetics* 77.*1* (1974), pp. 71–94.

[15] Bryan S. Briney and James E. Crowe Jr, « Secondary mechanisms of diversification in the human antibody repertoire », *in*: *Frontiers in Immunology* 4 (2013), p. 42.

[16] James G Burchfield et al., « Exocytotic vesicle behaviour assessed by total internal reflection fluorescence microscopy », *in*: *Traffic* 11.*4* (2010), pp. 429–439.

[17] Arnaud Chastanet et al., « Broadly heterogeneous activation of the master regulator for sporulation in Bacillus subtilis », *in*: *Proceedings of the National Academy of Sciences* 107.*18* (2010), pp. 8486–8491.

[18] Yu Chen et al., « A pre-docking role for microtubules in insulin-stimulated glucose transporter 4 translocation », *in*: *The FEBS Journal* 275.*4* (2008), pp. 705–712.

[19] Kelly Colombo et al., « Translation of polarity cues into asymmetric spindle positioning in Caenorhabditis elegans embryos », *in*: *Science* 300.*5627* (2003), pp. 1957–1961.

[20] Ann K Corsi, Bruce Wightman, and Martin Chalfie, « A transparent window into biology: a primer on Caenorhabditis elegans », *in*: *Genetics* 200.*2* (2015), pp. 387–407.

[21] J. Crank and E. P. J. Crank, *The Mathematics of Diffusion*, Oxford Science Publications, Clarendon Press, 1979, ISBN: 9780198534112.

[22] Hannah Dueck, James Eberwine, and Junhyong Kim, « Variation is function: are single cell differences functionally important? Testing the hypothesis that single cell variation is required for aggregate function », *in*: *Bioessays* 38.*2* (2016), pp. 172–180.

[23] Hannah Dueck et al., « Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation », *in*: *Genome Biology* 16.*1* (2015), pp. 1–17.

[24] Michael B. Elowitz et al., « Stochastic gene expression in a single cell », *in*: *Science* 297.*5584* (2002), pp. 1183–1186.

[25] E.L. Elson and D. Magde, « Fluorescence correlation spectroscopy. I. Conceptual basis and theory », *in*: *Biopolymers: Original Research on Biomolecules* 13.*1* (1974), pp. 1–27.

[26] Felix Felmy, « Modulation of Cargo Release from Dense Core Granules by Size and Actin Network », *in*: *Traffic* 8.*8* (2007), pp. 983–997.

[27] Steven A. Frank and Marsha Rich Rosner, « Nonheritable cellular variability accelerates the evolutionary processes of cancer », *in*: *PLoS Biology* 10.*4* (2012), e1001296.

[28] Nikki E. Freed et al., « A simple screen to identify promoters conferring high levels of phenotypic noise », *in*: *PLoS Genetics* 4.*12* (2008), e1000307.

[29] Ana J Garcia-Saez and Petra Schwille, « Surface analysis of membrane dynamics », *in*: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1798.*4* (2010), pp. 766–776.

[30] Carlos Garzon-Coral, Horatiu A Fantana, and Jonathon Howard, « A force-generating machinery maintains the spindle at the cell center during mitosis », *in*: *Science* 352.*6289* (2016), pp. 1124–1127.

[31] Monica Gotta et al., « Asymmetrically distributed C. elegans homologs of AGS3/PINS control spindle position in the early embryo », *in*: *Current biology* 13.*12* (2003), pp. 1029–1037.

[32] Stephan W. Grill et al., « Polarity controls forces governing asymmetric spindle positioning in the Caenorhabditis elegans embryo », *in*: *Nature* 409.*6820* (2001), pp. 630–633.

[33] Stephan W. Grill et al., « The distribution of active force generators controls mitotic spindle position », *in*: *Science* 301.*5632* (2003), pp. 518–521.

[34] Piyush B. Gupta et al., « Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells », *in*: *Cell* 146.*4* (2011), pp. 633–644.

[35] Jonathon Howard, « Elastic and damping forces generated by confined arrays of dynamic microtubules », *in*: *Physical Biology* 3.*1* (2006), p. 54.

[36] Dann Huh and Johan Paulsson, « Non-genetic heterogeneity from stochastic partitioning at cell division », *in*: *Nature Genetics* 43.*2* (2011), pp. 95–100.

[37] Bertrand Iooss and Paul Lemaitre, « A review on global sensitivity analysis methods », *in*: *Uncertainty management in simulation-optimization of complex systems*, Springer, 2015, pp. 101–122.

[38] Rosita Ivarsson et al., « Myosin 5a Controls Insulin Granule Recruitment During Late-Phase Secretion », *in*: *Traffic* 6.*11* (2005), pp. 1027–1035.

[39] Sebastien Jaeger, Bastien Fernandez, and Pierre Ferrier, « Epigenetic aspects of lymphocyte antigen receptor gene rearrangement or when stochasticity completes randomness », *in*: *Immunology* 139.*2* (2013), pp. 141–150.

[40] Kyu-Tae Kim et al., « Single-cell mRNA sequencing identifies subclonal heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells », *in*: *Genome Biology* 16.*1* (2015), pp. 1–15.

[41] Jean-Jacques Kupiec, « A Darwinian theory for the origin of cellular differentiation », *in*: *Molecular and General Genetics* 255.*2* (1997), pp. 201–208.

[42] Thorsten Lang et al., « Role of Actin Cortex in the Subplasmalemmal Transport of Secretory Granules in PC-12 Cells », *in*: *Biophysical Journal* 78.*6* (2000), pp. 2863–2877.

[43] Ben Lehner, « Selection to minimise noise in living systems and its implications for the evolution of gene expression », *in*: *Molecular Systems Biology* 4.*1* (2008), p. 170.

[44] Oleg Lenive, Paul D. W. Kirk, and Michael P. H. Stumpf, « Inferring extrinsic noise from single-cell gene expression data using approximate Bayesian computation », *in*: *BMC Systems Biology* 10.*1* (2016), pp. 1–17.

[45] Richard Losick and Claude Desplan, « Stochasticity and cell fate », *in*: *Science* 320.*5872* (2008), pp. 65–68.

[46] Rose Loughlin, Blake Riggs, and Rebecca Heald, « SnapShot: motor proteins in spindle assembly », *in*: *Cell* 134.*3* (2008), pp. 548–548.

[47] Fabio Luciani et al., « The epidemiological fitness cost of drug resistance in Mycobacterium tuberculosis », *in*: *Proceedings of the National Academy of Sciences* 106.*34* (2009), pp. 14711–14715.

[48] Kara Lynch et al., « Synaptotagmin-1 Utilizes Membrane Bending and SNARE Binding to Drive Fusion Pore Expansion », *in*: *Molecular Biology of the Cell* 19 (Oct. 2008), pp. 5093–5103.

[49] Hedia Maamar, Arjun Raj, and David Dubnau, « Noise in gene expression determines cell fate in Bacillus subtilis », *in*: *Science* 317.*5837* (2007), pp. 526–529.

[50] D. Magde, E. Elson, and W.W Webb, « Thermodynamic fluctuations in a reacting system—measurement by fluorescence correlation spectroscopy », *in*: *Physical Rev. Letters* 29.*11* (1972), p. 705.

[51] D. Magde, E.L. Elson, and W.W. Webb, « Fluorescence correlation spectroscopy. II. An experimental realization », *in*: *Biopolymers: Original Research on Biomolecules* 13.*1* (1974), pp. 29–61.

[52] Narendra Maheshri and Erin K. O'Shea, « Living with noisy genes: how cells function reliably with inherent variability in gene expression », *in*: *Annual Reviews of Biophysics and Biomolecular Structure* 36 (2007), pp. 413–434.

[53] Sascha Martens, Michael M. Kozlov, and Harvey T. McMahon, « How Synaptotagmin Promotes Membrane Fusion », *in*: *Science* 316.*5828* (2007), pp. 1205–1208.

[54] Benjamin Mercat, « Analyse temps-frequence en mecanique cellulaire et adaptabilite du fuseau mitotique », PhD thesis, Universite Rennes 1, 2016.

[55] Max D. Morris, « Factorial sampling plans for preliminary computational experiments », *in*: *Technometrics* 33.*2* (1991), pp. 161–174.

[56] Karen Oegema and Tony Hyman, « Cell division », *in*: *WormBook: The Online Review of C. elegans Biology [Internet]* (2006).

[57] Mica Ohara-Imaizumi et al., « Transduction of MIN6 β Cells with TAT-Syntaxin SNARE Motif Inhibits Insulin Exocytosis in Biphasic Insulin Release in a Distinct Mechanism Analyzed by Evanescent Wave Microscopy », *in*: *Journal of Biological Chemistry* 277.*52* (2002), pp. 50805–50811.

[58] Martin Oheim and Walter Stuhmer, « Tracking chromaffin granules on their way through the actin cortex », *in*: *European Biophysics Journal* 29 (2000), pp. 67–89.

[59] Yu Ohsugi et al., « Lateral mobility of membrane-binding proteins in living cells measured by total internal reflection fluorescence correlation spectroscopy », *in*: *Biophysical Journal* 91.*9* (2006), pp. 3456–3464.

[60] Andrew L. Paek et al., « Cell-to-cell variation in p53 dynamics leads to fractional killing », *in*: *Cell* 165.*3* (2016), pp. 631–642.

[61] Sonali Patil et al., « Single-cell analysis shows that paracrine signaling by first responder cells shapes the interferon-β response to viral infection », *in*: *Science Signaling* 8.*363* (2015), ra16–ra16.

[62] Jacques Pecreaux et al., « Spindle oscillations during asymmetric cell division require a threshold number of active cortical force generators », *in*: *Current Biology* 16.*21* (2006), pp. 2111–2122.

[63] Jacques Pecreaux et al., « The mitotic spindle in the one-cell C. elegans embryo is positioned with high precision and stability », *in*: *Biophysical Journal* 111.*8* (2016), pp. 1773–1784.

[64] Niles O Petersen et al., « Quantitation of membrane receptor distributions by image correlation spectroscopy: concept and application », *in*: *Biophysical journal* 65.*3* (1993), pp. 1135–1146.

[65] Jean Philibert, « One and a Half Century of Diffusion: Fick, Einstein », *in*: *Diffusion Fundamentals: Leipzig 2005* 1 (2005), p. 8.

[66] J. K. Pritchard et al., « Population growth of human Y chromosomes: a study of Y chromosome microsatellites », *in*: *Molecular Biology and Evolution* 16.*12* (Dec. 1999), pp. 1791–1798.

[67] Arjun Raj and Alexander Van Oudenaarden, « Nature, nurture, or chance: stochastic gene expression and its consequences », *in*: *Cell* 135.*2* (2008), pp. 216–226.

[68] Oliver Ratmann et al., « Model criticism based on likelihood-free inference, with an application to protein network evolution », *in*: *Proceedings of the National Academy of Sciences* 106.*26* (2009), pp. 10576–10581.

[69] Oliver Ratmann et al., « Using likelihood-free inference to compare evolutionary dynamics of the protein networks of H. pylori and P. falciparum », *in*: *PLoS Computational Biology* 3.*11* (2007), e230.

[70] Peter Reuven and Avigdor Eldar, « Macromotives and microbehaviors: the social dimension of bacterial phenotypic variability », *in*: *Current Opinion in Genetics & Development* 21.*6* (2011), pp. 759–767.

[71] Lydia Robert et al., « Pre-dispositions and epigenetic inheritance in the Escherichia coli lactose operon bistable switch », *in*: *Molecular Systems Biology* 6.*1* (2010), p. 357.

[72] Jeremie Roux et al., « Fractional killing arises from cell-to-cell variability in overcoming a caspase activity threshold », *in*: *Molecular Systems Biology* 11.*5* (2015), p. 803.

[73] Alex K. Shalek et al., « Single-cell RNA-seq reveals dynamic paracrine control of cellular variation », *in*: *Nature* 510.*7505* (2014), pp. 363–369.

[74] Scott A Sisson, Yanan Fan, and Mark Beaumont, *Handbook of approximate Bayesian computation*, CRC Press, 2018.

[75] Wiep Klaas Smits, Oscar P. Kuipers, and Jan-Willem Veening, « Phenotypic variation in bacteria: the role of feedback regulation », *in*: *Nature Reviews Microbiology* 4.*4* (2006), pp. 259–271.

[76] Berend Snijder et al., « Population context determines cell-to-cell variability in endocytosis and virus infection », *in*: *Nature* 461.*7263* (2009), pp. 520–523.

[77] Sabrina L Spencer et al., « Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis », *in*: *Nature* 459.*7245* (2009), pp. 428–432.

[78] Dayalan G Srinivasan et al., « A complex of LIN-5 and GPR proteins regulates G protein signaling and spindle function in C. elegans », *in*: *Genes & development* 17.*10* (2003), pp. 1225–1239.

[79] Gurol M. Suel et al., « Tunability and noise dependence in differentiation dynamics », *in*: *Science* 315.*5819* (2007), pp. 1716–1719.

[80] Mark M. Tanaka et al., « Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data », *in*: *Genetics* 173.*3* (2006), pp. 1511–1520.

[81] Simon Tavare et al., « Inferring Coalescence Times From DNA Sequence Data », *in*: *Genetics* 145.*2* (Feb. 1997), pp. 505–518.

[82] Mukund Thattai and Alexander Van Oudenaarden, « Stochastic gene expression in fluctuating environments », *in*: *Genetics* 167.*1* (2004), pp. 523–530.

[83] Shiori Toba et al., « Overlapping hand-over-hand mechanism of single molecular motility of cytoplasmic dynein », *in*: *Proceedings of the National Academy of Sciences* 103.*15* (2006), pp. 5741–5745.

[84] Takashi Tsuboi and Mitsunori Fukuda, « Synaptotagmin VII modulates the kinetics of dense-core vesicle exocytosis in PC12 cells », *in*: *Genes to Cells* 12.*4* (2007), pp. 511–519.

[85] Aurore-Cécile Valfort et al., « Evolution of mitotic spindle behavior during the first asymmetric embryonic division of nematodes », *in*: *PLoS Biology* 16.*1* (2018), e2005099.

[86] Aniko Varadi, Takashi Tsuboi, and Guy A. Rutter, « Myosin Va transports dense core secretory vesicles in pancreatic MIN6 β-cells », *in*: *Molecular Biology of the Cell* 16.*6* (2005), pp. 2670–2680.

[87] Jan-Willem Veening et al., « Transient heterogeneity in extracellular protease production by Bacillus subtilis », *in*: *Molecular Systems Biology* 4.*1* (2008), p. 184.

[88] Brenda N. Vo et al., « Melanoma Cell Colony Expansion Parameters Revealed by Approximate Bayesian Computation », *in*: *PLoS Computational Biology* 11.*12* (Dec. 2015), pp. 1–22.

[89] Brenda N. Vo et al., « Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation », *in*: *Mathematical Biosciences* 263 (2015), pp. 133–142.

[90] Zhi Wang and Jianzhi Zhang, « Impact of gene expression noise on organismal fitness and the efficacy of natural selection », *in*: *Proceedings of the National Academy of Sciences* 108.*16* (2011), E67–E76.

[91] Christopher M. Waters and Bonnie L. Bassler, « Quorum sensing: cell-to-cell communication in bacteria », *in*: *Annual Review of Cell and Developmental Biolgy* 21 (2005), pp. 319–346.

[92] P.W. Wiseman et al., « Two-photon image correlation spectroscopy and image cross-correlation spectroscopy », *in*: *J. Microscopy* 200.*1* (2000), pp. 14–25.

[93] Da Zhou, Yue Wang, and Bin Wu, « A multi-phenotypic cancer model with cell plasticity », *in*: *Journal of Theoretical Biology* 357 (2014), pp. 35–45.

List of Figures

List of Tables

Acronyms

ABC	Approximate Bayesian Computation
AP	Anterior-Posterior
APD	Avalanche Photodiode
ATP	Adenosine Triphosphate
FCS	Fluorescence Correlation Microscopy
GLUT4	Glucose transporter type 4
GPR	G-protein regulatory motif
ICS	Image Correlation Microscopy
MAP	Maximum a posteriori
MMSE	Minimum mean square error
MT	Microtubule
OAT	One At a Time
PSF	Point Spread Function
RNA	Ribonucleic acid
RNAi	RNA interference
ROI	Region Of Interest
SNARE	Soluble N-ethylmaleimide-sensitive factor Attachment protein Receptor
TfR	Transferrin Receptor
TICS	Temporal Image Correlation Microscopy
TIRF(M)	Total Internal Reflection Fluorescence (Microscopy)
TNF	Tumor Necrosis Factor
TRAIL	TNF-Related Apoptosis-Inducing Ligand

DOCTORAT BRETAGNE LOIRE / MATHSTIC

UNIVERSITÉ DE RENNES 1

Titre : Modélisation du signal et estimation bayésienne pour l'analyse de mécanismes cellulaires : étude de la diffusion membranaire et de la division cellulaire

Mots-clés : modélisation du signal, estimation bayésienne, diffusion membranaire, division cellulaire

Résumé : La microscopie de fluorescence permet de s'appuyer sur la modélisation mathématique pour l'étude de mécanismes cellulaires complexes, mais les modèles préexistants ne permettent pas d'obtenir des informations à la fois sur les tendances globales et les comportements locaux, et ne sont pas adaptés à l'étude des interactions complexes entre les différents acteurs moléculaires. Dans cette thèse, nous proposons une approche mixte conçue pour pallier ces limitations, et l'appliquons à l'étude de deux mécanismes cellulaires. Nous proposons un premier modèle pour l'estimation de la diffusion locale en imagerie TIRF ; ce modèle permet l'évaluation de la diffusion en plusieurs spots de diffusion dans une région d'intérêt, répondant ainsi aux problèmes complexes causés par l'hétérogénéité de la membrane. Un deuxième modèle est introduit pour l'étude du comportement des centrosomes durant la division cellulaire, qui permet une meilleure compréhension des facteurs, et des interactions de ces facteurs, qui contribuent aux oscillations des centrosomes durant l'anaphase ; ce modèle fournit des moyens de prédiction pour des phénotypes liés à des divisions asymétriques erronées. Ces deux modèles sont utilisés grâce à un nouveau cadre d'estimation de paramètres, qui s'appuie sur l'analyse de sensibilité et les méthodes bayésiennes ABC. Ce framework est polyvalent et facilement adaptable à d'autres contextes que celui des études présentées dans cette thèse.

Title: Signal modeling and Bayesian estimation for cellular mechanism analysis: study of diffusion at the cell membrane and cell division

Keywords: signal modeling, Bayesian estimation, membrane diffusion, cell division

Abstract: Fluorescence microscopy makes it possible to rely on mathematical modeling for the study of complex cellular mechanisms, but pre-existing models do not yield information on both global trends and local behaviors, and are not fit for the study of the complex interactions between the molecular actors involved. In this thesis, we propose a mixed approach designed with these limitations in mind, and use this approach to study two cellular mechanisms. We propose a first model for the estimation of local diffusion in TIRF imaging, which addresses the challenging issues caused by membrane heterogeneity by allowing to evaluate the diffusion for several diffusing spots in a region of interest. A second model is introduced for the study of centrosome behavior during cell division, that allows to better understand the factors and interactions that contribute to the oscillations of centrosomes during anaphase, thus providing means of prediction for phenotypes related to misguided asymmetric division. Both models are put into operation thanks to a novel parameter estimation framework, that relies on sensitivity analysis and Approximate Bayesian Computation. This versatile framework is easily adaptable to studies beyond the context of the ones presented in this thesis.