Some contributions to the study of population dynamics models and coagulation-fragmentation

Romain Yvinec

HABILITATION À DIRIGER DES RECHERCHES

Discipline: Mathématiques

Année universitaire: 2020 / 2021

présentée et soutenue publiquement par :

M. Romain YVINEC

le 02 décembre 2020

JURY :
(Par ordre alphabétique)

Prénom NOM Grade Établissement d'exercice

- Mme Béatrice LAROCHE, Directrice de Recherche, INRAE Jouy-en-Josas

- Mme Eva LÖCHERBACH, Professeure des universités, Université Paris 1

- M. Florent MALRIEU, Professeur des universités, Université de Tours

- Mme Magali RIBOT, Professeure des universités, Université d'Orléans

- M. Philippe ROBERT, Directeur de Recherche, INRIA Paris

- Mme Angélique STEPHANOU, Chargée de Recherche HDR, CNRS, Université Grenoble-Alpes

Quelques contributions à l'étude de modèles de dynamique de populations et de coagulation-fragmentation

Romain Yvinec

Résumé

Ce manuscrit est le mémoire de mon dossier de candidature à l'habilitation à diriger des recherches. À ce titre, il contient le bilan de mes travaux de recherche depuis ma thèse soutenue en 2013. L'objectif du manuscrit est de présenter sous forme concise les principaux modèles mathématiques que j'étudie, les questions théoriques que ceux-ci soulèvent, et leurs applications en biologie. L'ensemble de mes articles est disponible sur ma page web personnelle [1].

Depuis mon recrutement au laboratoire Physiologie de la Reproduction et des Comportements (INRAE, CNRS, Université de Tours), au sein de l'équipe Biologie & Bioinformatique des Systèmes de Signalisation (BIOS), j'ai consacré une partie importante de mes activités de recherches à la modélisation de la folliculogenèse ovarienne, en particulier en décrivant précisément les différentes dynamiques de populations cellulaires et tissulaires en jeu dans ce systèmes. Je me suis attaché à construire, analyser et simuler des modèles probabilistes adaptés à chaque problématique biologique, jusqu'à la calibration des paramètres des modèles et l'identification de prédictions pouvant être testées expérimentalement. Ces travaux font l'objet de la partie 1 de ce manuscrit.

Ce travail de modélisation ne pourrait se faire sans une bonne compréhension des enjeux et connaissances actuels de la biologie de la reproduction et des travaux de modélisation qui s'y rapportent. J'ai ainsi contribué à trois articles de synthèse dans ce domaine :

- M. A. Ayoub, R. Yvinec, P. Crépieux, and A. Poupon. Computational modeling approaches in gonadotropin signaling. *Theriogenology*, 86(1):22–31, 2016
- R. Yvinec, P. Crépieux, E. Reiter, A. Poupon, and F. Clément. Advances in computational modeling approaches of pituitary gonadotropin signaling. *Expert Opin. Drug. Discov.*, 13(9):799–813, 2018
- F. Clément, P. Crépieux, R. Yvinec, and D. Monniaux. Mathematical modeling approaches of cellular endocrinology within the hypothalamo-pituitary-gonadal axis. *Mol. Cell. Endocrinol.*, 518:110877, 2020

Une partie de ces travaux a nourri la présentation du premier chapitre 1.1 de ce manuscrit, qui donne une nécessaire introduction à la biologie de la reproduction femelle.

Au cours de mes travaux sur la modélisation de la folliculogenèse ovarienne, je me suis d'abord intéressé à la modélisation de la croissance d'un follicule ovarien, en modélisant précisément la cinétique de différentiation et prolifération cellulaire au sein d'un follicule. On s'intéresse à caractériser le comportement de la population de cellule somatiques au sein d'un follicule ovarien, et à préciser l'influence des différents paramètres des modèles qui s'écrivent sous forme de processus de Markov en temps continu. Ces modèles caractérisent les lignages cellulaires et sont parcimonieux dans leur description de phénomènes biologiques complexes, afin de pouvoir les calibrer de manière rigoureuse à l'aide de données quantitatives. Cette problématique a fait l'object du co-encadrement de thèse de Frédérique Robin (2016-2019), co-encadrée par Frédérique Clément (Centre de recherche Inria Saclay-Île-de-France), et a nourri deux articles scientifiques :

- F. Clément, F. Robin, and R. Yvinec. Analysis and calibration of a linear model for structured cell populations with unidirectional motion : Application to the morphogenesis of ovarian follicles. *SIAM J. Appl. Math.*, 79(1):207–229, 2019
- F. Clément, F. Robin, and R. Yvinec. Stochastic nonlinear model for somatic cell population dynamics during ovarian follicle activation. Accepted in J. Math. Biol., 2020

Le chapitre 1.2 présente ces travaux.

Je me suis également intéressé à décrire l'ensemble de la population folliculaire au cours de la vie reproductive d'une individue, et à modéliser en particulier les interactions entre populations folliculaires qui résultent de rétro-actions de nature paracrine et endocrine. En prenant en compte les différentes échelles de temps en jeu dans ce système, nous obtenons un problème de type perturbation singulière que nous pouvons résoudre. Le modèle limite est plus adapté à une caractérisation détaillée du comportement transitoire et en temps long de la population de follicules, ainsi qu'à sa calibration aux données quantitatives. Ce travail a fait l'objet d'un co-encadrement du stage de recherche de Céline Bonnet et Keltoum Chahour (CEMRACS 2018), co-encadrées par Frédérique Clément et Marie Postel (Sorbonne Université, Université Paris-Diderot SPC, CNRS, Laboratoire Jacques-Louis Lions, LJLL), et d'un co-encadrement en cours de la thèse de Guillaume Ballif (2019–.), co-encadrée par Frédérique Clément. Le chapitre 1.3 présente ces travaux, publiés partiellement dans

1. http://yvinec.perso.math.cnrs.fr/publications.html

- C. Bonnet, K. Chahour, F. Clément, M. Postel, and R. Yvinec. Multiscale population dynamics in reproductive biology: singular perturbation reduction in deterministic and stochastic models. *ESAIM: ProcS*, 67:72 – 99, 2020

et dans un article en préparation. Je conclue cette première partie par quelques perspectives et réflexions sur la modélisation de la folliculogenèse ovarienne dans le chapitre 1.4.

Une partie de mes travaux de thèse a nourri des questions théoriques sur l'étude d'un modèle particulier de coagulation-fragmentation, le modèle de Becker-Döring, et sont présentées dans la partie 2 de ce manuscrit. Ce modèle est très utilisé en physique des matériaux pour décrire des phénomènes de nucléation et de transition de phase, mais également depuis quelques années dans divers domaines de la biologie cellulaire et moléculaire. Je présente dans le chapitre introductif 2.1 la définition du modèle de Becker-Döring et ses variantes, ainsi que des résultats clés de la littérature, essentiellement basé sur mon article de synthèse

- E. Hingant and R. Yvinec. Deterministic and Stochastic Becker-Döring Equations: Past and Recent Mathematical Developments. In *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*, pages 175–204. Springer, Cham, 2017

Mes travaux, en collaboration avec Erwan Hingant (Universidad del Bio-Bio, Chili) et Juan Calvo (Universidad de Granada, Espagne), ont porté notamment sur l'établissement de théorèmes limites mettant en lien différents modèles entre eux, par des procédés de renormalisation et d'étude de compacité de suites de solutions. Ce travail m'a amené à définir des notions de convergences subtiles dans des espaces fonctionnels en dimension infinie, et pointe les propriétés particulières du modèle de Becker-Döring. Je présente ce travail dans le chapitre 2.2, basé sur les publications :

- J. Deschamps, E. Hingant, and R. Yvinec. Quasi steady state approximation of the small clusters in Becker-Döring equations leads to boundary conditions in the Lifshitz–Slyozov limit. *Commun. Math Sci*, 15(5):1353–1384, 2017
- E. Hingant and R. Yvinec. The Becker–Döring process: law of large numbers and non-equilibrium potential. *J. Stat. Phys.*, 177:506–527, 2019
- J. Calvo, E. Hingant, and R. Yvinec. Initial-boundary value problem to the lifshitz-slyozov equation with non-smooth rates at the boundary. Submitted. arXiv:2004.01947, 2020

Un deuxième axe de travail sur le modèle de Becker-Döring porte sur l'établissement des propriétés en temps long de sa version stochastique, en s'inspirant de travaux analogues en réseaux de réactions biochimiques ou en dynamiques de populations. La compréhension fine de l'état stationnaire et des phénomènes de métastabilité reste un important problème ouvert. Nous avons tout d'abord obtenu un principe de grande déviation sur la mesure stationnaire, en le reliant à la fonction de Lyapounov connue pour ce système. Puis nous avons caractérisé le phénomène de métastabilité dans une version linéaire du modèle de Becker-Döring, grâce à l'établissement d'une convergence exponentielle vers une distribution quasi-stationnaire dont on connaît une expression analytique. Je présente ma contribution dans ce domaine au chapitre 2.3, basé sur les publications

- E. Hingant and R. Yvinec. The Becker–Döring process: law of large numbers and non-equilibrium potential. *J. Stat. Phys.*, 177:506–527, 2019
- E. Hingant and R. Yvinec. Quasi-stationary distribution and metastability for the stochastic Becker-Döring model. Submitted. arXiv: 2008.02544, 2020

et je conclue sur quelques perspectives au chapitre 2.4.

Certains travaux post-thèse ne sont pas abordés dans ce manuscrit. Cela concerne notamment la thématique de la modélisation dynamique des réseaux de signalisations des gonadotrophines, au cœur des aspects cellulaires de la biologie de la reproduction. Sur cette thématique, je juge mes travaux publiés non encore suffisamment aboutis pour être présentés dans ce manuscrit, et ma contribution personnelle a été pour l'essentiel mineure [A2, A8, A13, A14, A15, A16, A17]. Cette thématique n'en reste pas néanmoins un axe important dans mon projet de recherche futur.

Le chapitre introductif 1.1 est écrit en français, les autres chapitres sont écrits en anglais.

Table des matières

1 Modeling ovarian folliculogenesis : population dynamics

1.1 Introduction

La présentation qui suit est librement inspirée de revues de modélisation dans le domaine de la reproduction [81, 32, A19], du manuscrit de thèse de Frédérique Robin [94] que nous avons co-encadrée avec Frédérique Clément, ainsi que des excellents articles et revues de notre collaboratrice biologiste Danielle Monniaux et ses co-autrices et co-auteurs [35, 75, 76, 78, 79]. Je me reporte également à quelques endroits au "textbook" [101]. Le but de cette introduction n'est pas d'être exhaustif dans la présentation de la biologie de la reproduction femelle, mais plutôt de présenter certains aspects qui motivent l'introduction des modèles mathématiques de dynamiques de populations que j'étudie par la suite.

1.1.1 L'axe hypothalamo-hypophyso-gonadique

La fonction de reproduction est principalement assurée par l'axe hypothalamo-hypophyso-gonadique (HHG) (voir Figure 1), qui comprend les trois composantes anatomiques majoritairement responsables de la production des hormones de la reproduction. Le système de reproduction femelle chez les mammifères (post-puberté et hors phase de gestation) subit des changements cycliques d'environnement hormonal mais également de structure et de fonction des ovaires, au cours de ce qui est communément appelé les cycles menstruels ou cycles ovariens. Cependant, ce comportement cyclique cache des mécanismes et dynamiques qui s'opèrent à des échelles très différentes et tout au long de la vie des individues.

Au sein de l'ovaire, le développement et la maturation des gamètes (ovocytes ou cellules œufs) est un processus qui s'initie dès la période embryonnaire et continue tout au long de la vie de l'individue, et donne lieu aux ovulations à chaque cycle ovarien, sous le contrôle principal de certaines hormones hypophysaires (dites gonadotrophines), l'hormone folliculo-stimulante (FSH) et l'hormone lutéinisante (LH). Ces hormones sont sécrétées par l'hypophyse sous stimulation par une neuro-hormone, l'hormone de libération des gonadotrophines hypophysaires (GnRH), elle-même sécrétée au niveau de l'hypothalamus. Ces hormones ont des profils spatio-temporels très particuliers. La GnRH et la LH sont sécrétées de manière fortement intermittente, pulsatile, alors que la sécrétion de FSH est plus régulière. Au moment de l'ovulation, ces signaux ont également des variations temporelles très importantes (on parle de "pic") sur une courte période de temps (heure/jour). Ces hormones induisent des signaux intra-cellulaires au niveau de leurs cellules cibles qui se produisent à une toute autre échelle temporelle, typiquement de l'ordre de la minute ou de la seconde.

1.1.2 Ovogenèse et folliculogenèse ovarienne

Les cellules cibles des gonadotrophines FSH et LH sont les cellules somatiques qui constituent, avec la cellule germinale, le follicule ovarien. Dès le stade embryonnaire, la lignée des cellules germinales se sépare des cellules somatiques. Initialement proliférantes, les cellules germinales débutent leur processus de différenciation dans l'ovaire embryonnaire et, parallèlement, elles recrutent et s'associent avec des cellules somatiques avoisinantes pour constituer l'unité anatomique et fonctionnelle de base des ovaires, le *follicule ovarien*. À l'issue de cette phase qui a lieu intégralement aux stades embryonnaire et fœtal (dans la majorité des espèces de mammifères), l'ensemble de tous les follicules constituent une réserve (dite "statique" [79]) qui ne sera plus renouvelée et assurera la fonction de reproduction pendant la vie adulte qui suivra. Les follicules ainsi formés sont appelés *follicules primordiaux*, et sont initialement dans une phase de quiescence, les ovocytes dans ces follicules étant maintenus à l'arrêt dans une phase particulière de la méiose, la prophase I. Cette réserve statique s'amenuise lentement tout au long de la vie des individues (jusqu'à presque épuisement chez la femme, après la ménopause). Le processus de développement des follicules ovariens constitué par l'ensemble des étapes de sortie de quiescence (activation), croissance, morphogenèse et maturation des follicules ovariens s'appelle la *folliculogenèse ovarienne*. Ce processus débute donc dès le stade embryonnaire et persiste à l'âge adulte. La croissance d'un follicule, depuis sa sortie du pool primordial jusqu'à l'ovulation (ou la dégénérescence par atrésie), s'étend sur plusieurs cycles ovariens (plusieurs mois dans l'espèce humaine et chez les mammifères domestiques de rente). On distingue classiquement deux grandes phases de croissance, qui sont caractérisées par des critères morphologiques des follicules mais également par les processus de régulation qui sont mis en jeux (avec une certaine conservation entre espèces) : (i) la croissance dite *basale* est contrôlée par des facteurs sécrétés localement au sein de l'ovaire, indépendamment de la dynamique de l'axe HHG ; (ii) la croissance

dite *terminale*, strictement dépendante de la fourniture en hormones hypophysaires. La croissance basale peut démarrer dès le stade fœtal ou périnatal alors que la croissance terminale ne peut avoir lieu qu'après la puberté [2]. À chaque cycle ovarien, un relativement petit nombre de follicules ayant atteint une certaine maturité (fin de croissance basale) démarre une phase de croissance terminale : on parle de recrutement cyclique. Une ou plusieurs vagues de recrutement folliculaire par cycles sont observées selon les espèces. Il est important de préciser que la croissance basale des follicules est asynchrone (ainsi tous les types folliculaires peuvent être observés à un instant donné au sein d'un ovaire) alors que la croissance terminale est plus synchrone et coordonnée (et donc une corrélation avec la dynamique du cycle ovarien peut-être observée pour les follicules en croissance terminale, un follicule en croissance terminale pouvant néanmoins être recruté à un cycle ovarien antérieur). Les follicules en fin de croissance basale constituent ce qui est appelé une réserve "dynamique" : celle-ci est continuellement alimentée par l'activation des follicules primordiaux et la croissance basale des follicules, et est vidée régulièrement par le recrutement de follicule en phase terminale. Notons que la taille de cette réserve dynamique est mesurée en routine cliniquement (appelée "Antral Follicle Count") et est d'importance cruciale pour les biotechnologies de la reproduction assistée, humaine et animale.

La folliculogenèse ovarienne est finement régulée par l'intermédiaire d'un certain nombre de molécules messagères. Les différents types de cellules somatiques du follicule ovarien reçoivent les signaux hypophysaires et sont en dialogue (direct ou indirect) permanent avec la cellule germinale qu'elles confinent. Ce dialogue est nécessaire pour le bon déroulement de la croissance et de la maturation de l'ovocyte et du follicule ovarien qui le contient (abordé plus loin en section 1.1.3), mais donne également lieu à des rétro-actions sur une partie ou l'ensemble des follicules ovariens. En particulier, l'expression de l'hormone Anti-Mullërienne par les cellules somatiques d'un follicule en croissance basale a pour effet d'inhiber l'activation de nouveaux follicules primordiaux et de ralentir la croissance des petits follicules en croissance basale. Également, la synthèse des stéroïdes sexuels (progestérone, testostérone, estradiol) par les cellules somatiques du follicule ovarien en croissance terminale, en réponse aux stimulations des gonadotrophines, modulent en retour la sécrétion des gonadotrophines et de la GnRH, terminant la boucle de rétro-action. Cette modulation résulte en particulier par une diminution du niveau circulant d'hormone FSH, associée aux vagues de croissance terminale, et mène à la sélection des follicules dits "dominants" (avec une certaine part d'aléatoire, ceux-ci devant être à la "bonne" maturité au "bon" moment [35, 97]) qui sont seuls capables de continuer leur croissance et deviendront les follicules ovulatoires.

Un dernier aspect primordial, quoique peu compris à ce jour, de l'ovogenèse mérite d'être mentionné. Durant la phase embryonnaire et fœtale, un très grand nombre de cellules germinales est produit alors qu'une infime minorité sera ovulée : ainsi, chez la femme, sur les 6 à 7 millions de cellules germinales présentes dans l'ovaire fœtal, seules environ 500 seront ovulées. Il y a vraisemblablement beaucoup de perte de cellules germinales lors du processus de formation des follicules ovariens, mais également tout au long de la folliculogenèse ovarienne. L'atrésie folliculaire est le mécanisme d'arrêt de croissance et de dégénérescence du follicule, et concerne environ 99% des follicules qui entrent en croissance [80]. La perte la plus importante se produit à la naissance, où le pourcentage de follicules (quiescent) atrétiques seraient de l'ordre de 50% selon [52]. Le déclenchement du programme d'atrésie d'un follicule peut survenir à tous les stades de croissance, et est également finement régulé.

Pour terminer ce bref panorama de la folliculogenèse ovarienne, nous donnons quelques ordres de grandeurs de taille de populations (voir Tables 1-2) et de durée moyenne (voir Table 3) de chaque stade de croissance pour différentes espèces. [3] En particulier, selon [101, Chapitre 15], les follicules primordiaux représentent plus de 95% de la population folliculaire ovarienne. Il faut garder à l'esprit cependant qu'il y a une grande hétérogénéité dans le nombre de follicules entre individues de même espèce et même âge (d'un facteur 5 à 10 selon [39])

Dans le chapitre 1.3, nous mettrons en musique (ou en équations) la dynamique de populations des follicules ovariens. Ces modèles sont du type "naissance-migration-mort" et modélisent l'évolution de la population de follicules structurée selon une variable de maturité, discrète, qui ne peut qu'augmenter au cours du temps. On peut penser à des classes de taille par exemple, qui sont fréquemment utilisées par les biologistes pour dénombrer les follicules, ou d'autres critères morphologiques ou fonctionnels. Les différences

2. Les repères temporels des phases de constitution de la réserve statique et de la folliculogenèse pré et post pubère chez différentes espèces de mammifères peuvent être visualisés dans la figure 1 de l'article de synthèse [79].

3. *Nb : La durée moyenne de la phase de quiescence est une estimation relativement "grossière", basée sur une estimation du taux d'activation par jour et de l'effectif total. Le taux d'activation journalier est lui-même issu de mesure d'effectif folliculaire à différents âges (voir par exemple [53]).*

d'ordre de grandeur et d'échelle de temps en jeu chez certaines espèces motivent l'introduction de facteurs d'échelle et la considération de problèmes de type perturbation singulière. Les différentes régulations de la folliculogenèse au sein de l'ovaire et au sein de l'axe HHG, abordées plus haut, guident l'introduction de termes d'interaction entre les différentes sous-populations de follicules ovariens.

TABLE 1 – Ordre de grandeur du nombre de cellules germinales et follicules primordiaux à différentes périodes de la vie. Voir [101, Chapitre 15] et les références mentionnées par espèce ci-dessous.

Espèces	Cellules germinales (embryon)	Follicules primordiaux (naissance)	Follicules primordiaux (puberté)	Follicules primordiaux (Jeune adulte)	Follicules primordiaux (Fin de vie reproductive)
Femmes [53]	$5 \cdot 10^6$	0.2 à $1 \cdot 10^6$	1 à $5 \cdot 10^5$ (10 ans)	$5 \cdot 10^4$ (25 ans)	0.15 à $2.4 \cdot 10^4$ (45 ans)
Brebis [73, 72, 95]	$5 \cdot 10^5$	10^5	2 à $8 \cdot 10^4$ (2 ans)		10^4 (5 ans)
Souris [42, 44, 26, 45, 103]	$1.3 \cdot 10^4$	0.7 à $1 \cdot 10^4$	2.5 à $5 \cdot 10^3$ (50j)	1 à $2.5 \cdot 10^3$ (100j)	0 à 500 (500j)

TABLE 2 – Ordre de grandeur du nombre de follicules en croissance pendant la vie reproductive. Voir [101, Chapitre 15], [39] et les références mentionnées par espèce ci-dessous.

Espèces	Follicules en croissance basale	Follicules en croissance antrale	Follicules en croissance terminale	Follicules pré-ovulatoire	Nombres d'ovulations par cycles
Femmes (à 20-30 ans)	1 à $10 \cdot 10^4$	100	10 à 20	1 à 10	1
Brebis (1.5 ans) [30, 95]	$5 \cdot 10^4$	40-60	10 à 30		1 à 10
Souris (100j) [42, 44, 26, 45]	500 à 1000	50 à 100		10	10

TABLE 3 – Ordre de grandeur de la durée moyenne d'une phase de croissance de la folliculogenèse ovarienne. Voir [101, Chapitre 15] et les références mentionnées par espèce ci-dessous.

Espèces	Quiescence (Jours)	Croissance basale (Jours)	Croissance antrale (Jours)	Croissance terminale (Jours)
Femmes	$6.6 \cdot 10^4$	180	45	15
Brebis[30]	0.66 à 8.10^4	130	40	2 à 3
Souris (100j) [42, 44, 26, 45, 70]	200 à 300	30 à 140	3	

1.1.3 Croissance morphodynamique d'un follicule ovarien

Nous revenons maintenant sur la description plus fine de la morphologie d'un follicule ovarien, et de son évolution au cours de sa croissance. Ces considérations mènent également à l'étude de modèles mathématiques de dynamique de populations (abordés au chapitre 1.2) pour caractériser l'évolution du nombre de cellules somatiques au sein d'un follicule, et (plus ou moins finement) la géométrie d'un follicule ovarien. Suivant les modèles, ces cellules somatiques seront structurées en âge, en type cellulaire et/ou en position spatiale relativement à l'ovocyte.

Chez tous les mammifères, les premières étapes du développement folliculaire sont comparables au sens où les tailles de follicules et les changements morphologiques sont remarquablement conservés entre des espèces

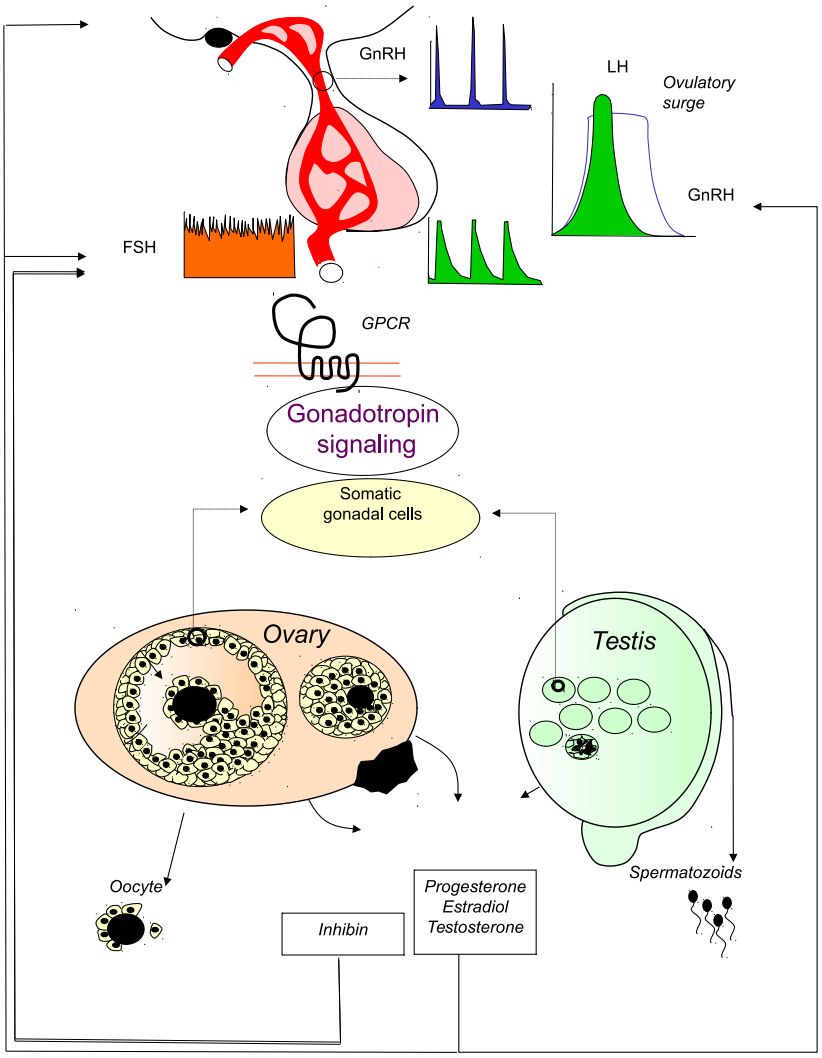

FIGURE 1 – Illustration de l'axe hypothalamo-hypophyso-gonadique. La GnRH est relarguée sous forme de pulses par les "neurones à GnRH" au niveau de l'hypothalamus. Cette hormone induit la sécrétion de la LH, sous forme pulsatile, et de la FSH, de manière plus continue, toutes les deux au niveau de l'hypophyse. Ces hormones hypophysaires ciblent les cellules somatiques des follicules au sein des ovaires (ou au sein des testicules chez les mâles). Lors de la croissance des follicules ovariens, la synthèse des stéroïdes sexuels et hormones peptidiques modulent en retour les hormones GnRH, LH et FSH. Cette boucle de rétro-action est en particulier responsable du "pic ovulatoire" de LH observé lors de l'ovulation. *Figure extraite de [A19].*

de poids très différents. Le follicule primordial est constitué d'une part de la cellule germinale, l'ovocyte en quiescence (processus de méiose arrêté au stade diplotène de la prophase I, qui ne reprendra qu'à l'ovulation) situé en position centrale, et d'autre part d'une simple couche de cellules somatiques qui entourent l'ovocyte. Ces cellules somatiques sont appelées les cellules de la *Granulosa*. A ce stade, l'ovocyte a un diamètre de l'ordre de 20 à 30μm et le follicule a un diamètre compris entre 30 et 50 μm. Les cellules de la Granulosa ont alors une forme aplatie, et sont en nombre très variable entre follicules, généralement de l'ordre de la dizaine (5 à 20 cellules d'après [51]). Cette variabilité est supposée provenir du processus complexe de formation des follicules primordiaux. Ce processus est marqué initialement par la prolifération des cellules germinales, leur migration puis le "recrutement" de cellules somatiques par les cellules germinales au sein de structure appelée cordon ovigère [75]. Les follicules primordiaux se forment enfin à l'issue de la fragmentation de ces cordons ovigères, et entrent pour la plupart dans un état de "dormance", inactif.

L'activation des follicules primordiaux est ensuite caractérisée par trois processus principaux [91] :

(i) Une transition irréversible du phénotype des cellules somatiques (cellules quiescentes puis prolifératives), caractérisée par un changement concomitant de leur forme (d'abord aplatie puis cuboïde)) ;

(ii) Une augmentation du nombre de cellules somatiques par division cellulaire ;

(iii) Une augmentation de l'activité métabolique et l'élargissement associé de l'ovocyte.

On considère que la phase d'activation se termine lorsque toutes les cellules somatiques ont changé de forme, moment où le stade de développement monocouche est terminé [47].

La phase de croissance suivante est appelé croissance *basale*, et comprend notamment les stades communément appelés *primaires*, *secondaires* et *pré-antraux*. Au cours de ces stades, la croissance folliculaire est le résultat d'une augmentation conjointe de la taille de l'ovocyte et du nombre de cellules somatiques environnantes [4]. L'ovocyte occupe toujours la position centrale du follicule, et les cellules somatiques se répartissent en couches successives autour de l'ovocyte. Au cours de la phase basale du développement du follicule, du stade primaire au stade pré-antral, le follicule ovarien se développe ainsi sous la forme d'un agrégat cellulaire compact avec une structure sphérique. Chez les ovins, la population de cellules somatiques double environ 10-12 fois et le nombre de couches de cellules somatiques passe de une à quatre à six couches [34]. Pendant ce temps, le diamètre de l'ovocyte passe de quelques dizaines de μm à une centaine de μm. De manière cruciale, la croissance des ovocytes et la prolifération cellulaire sont couplées par un dialogue moléculaire étroit établi entre les ovocytes et les cellules somatiques. Les facteurs de croissance dérivés des cellules somatiques (ligand KIT) favorisent la croissance de l'ovocyte et, à leur tour, ceux de l'ovocyte (BMP15, GDF9) influencent la prolifération cellulaire somatique. L'évolution morphologique d'un follicule à ce stade résulte donc d'un équilibre finement régulé entre le taux de croissance de l'ovocyte et le taux de prolifération des cellules folliculaires déterminant, pour un diamètre folliculaire donné, la taille de l'ovocyte, et le nombre de cellules et de couches somatiques (voir Figure 2). Cet équilibre est compromis dans le cas de certaines mutations génétiques naturelles (observées et étudiées en particulier chez les ovins) ou induites expérimentalement (Knock-Out (KO) chez la souris). En fonction de la cible moléculaire de ces mutations, le déséquilibre entre croissance et prolifération conduit à la formation de gros ovocytes entourés de moins de cellules (exemple de la mutation FecB chez les brebis Booroola, associée à augmentation du nombre d'ovulations), ou au contraire de petits ovocytes enfouis dans une masse dense de cellules folliculaires (exemple de l'inhibine KO chez la souris) [79].

Dans le follicule multicouches, d'environ 200 μm diamètre (taille relativement conservée entre espèces), des cavités remplies de liquide apparaissent et fusionnent pour former une seule grande cavité appelée l'*antrum*. Jusqu'à ce stade, le développement folliculaire est principalement contrôlé par des facteurs ovariens locaux. On rappelle qu'au stade de développement terminal, le follicule devient dépendant de l'approvisionnement en hormones hypophysaires (FSH et LH). Le passage de la phase basale à la phase terminale correspond donc à un changement fonctionnel du follicule : pour continuer sa croissance, le follicule a besoin de fourniture en hormones gonadotrophines. Chez certaines espèces de mammifères, ce changement fonctionnel apparaît après l'apparition de la cavité antrale alors que chez d'autres, comme les rongeurs, il apparaît au moment de la formation de l'antrum. La phase de croissance après apparition de l'antrum est aussi caractérisée par une importante prolifération cellulaire des cellules de Granulosa, ainsi qu'une différenciation des cellules proliférantes en cellules appelées cellules murales, situées à la périphérie du follicule. L'ovocyte quant à lui ne grossit presque plus, et occupe désormais une position excentrique dans le follicule. La cavité antrale

4. La taille des cellules somatiques évolue certainement au cours du cycle cellulaire mais est considérée relativement constante tout au long du développement folliculaire, avec un diamètre de l'ordre 10μm.

croît jusqu'à occuper la majeure partie du volume du follicule pré-ovulatoire. L'ovulation est caractérisée par une rupture de la *membrane basale* qui entoure les cellules somatiques, et l'expulsion de l'ovocyte. Le rôle exact de l'antrum n'est pas complètement connu, mais plusieurs hypothèses tendent à le considérer comme une stratégie pour maintenir un certain niveau de "ressources" accessibles pour l'ovocyte et les cellules somatiques du follicule (réservoir). La vascularisation du follicule ovarien s'arrête au niveau de la fine couche a-cellulaire constituant la membrane basale, à l'extérieur de l'amas cellulaires constitué par les cellules de la Granulosa et de l'ovocyte. Cependant, pour atteindre une production de stéroïdes suffisante, capable d'influencer la dynamique de l'axe HHG, le follicule en croissance doit accroître considérablement son nombre de cellules somatiques. Ainsi, à partir d'une certaine taille, sans formation de l'antrum mais avec toujours une prolifération cellulaire importante, le follicule n'aurait pas accès à suffisamment de facteurs de croissance et/ou à l'oxygène. L'antrum apparaît alors comme une stratégie possible pour accroître le nombre de cellules somatiques tout en maintenant relativement constant l'épaisseur de cette couche de cellules, qui serait le facteur limitant la diffusion de "ressources" depuis l'extérieur de l'amas cellulaires vers l'intérieur [28]. Notons également que l'antrum a certainement un rôle clé au moment de l'ovulation, en donnant une force considérable d'écoulement pour permettre à l'ovocyte de rejoindre les tubes utérins.

Nous terminons ce chapitre avec également quelques ordres de grandeurs de morphologie de follicules et de nombre de cellules somatiques, donnés en Table 4. Là aussi, il faut garder à l'esprit qu'il y a une grande hétérogénéité dans le nombre de cellules pour deux follicules de même taille (d'un facteur 5 selon [39]).

TABLE 4 – Ordre de grandeur du diamètre d'un follicule, du diamètre de l'ovocyte et du nombres de cellules somatiques aux différents stades de développement (Primordial, Antral, Terminal, Ovulatoire). Voir [101, Chapitre 15], [52] et les références mentionnées par espèce ci-dessous.

	Diamètre du follicule				Diamètre de l'ovocyte				Nbre de cellules somatiques			
	Prim.	Ant.	Term	Ov.	Prim.	Ant.	Term	Ov.	Prim.	Ant.	Term	Ov.
Espèces	(mm)				(μm)				-			
Femmes [54]	0.03 à 0.06	0.1 à 0.3	2	20	19	120	120	120	15	10^4	10^6	5.10^7
Brebis [30, 72]	0.03 à 0.06	0.1 à 0.3	2	6 à 8	10 à 40	100	100	100	15	10^4	$2 \cdot 10^6$	$3.5 \cdot 10^6$
Souris [87, 86, 54, 45]	0.007 à 0.025	0.2 à 0.35	0.35 à 0.6		4 à 6	80 à 120	85 à 180		<20	10^3 à 10^4	$1.5 \cdot 10^4$	$3 \cdot 10^5$

1.2 Modeling an individual follicle growth

In this chapter, we go through two works on the modeling of the growth of an ovarian follicle [A6, A5]. This work was carried out as part of the thesis of Frédérique Robin [94], which we co-supervised with Frédérique Clément. I start by presenting the datasets used during these two works in the section 1.2.1. Then I detail in section 1.2.2 a model of somatic cell transition during activation of a primordial follicle [A6], and in section 1.2.3 a model of somatic cell proliferation during basal follicle growth [A5]. Finally, I discuss some modeling and statistical perspectives in section 1.2.4.

1.2.1 Data set presentation

This section is mostly taken from [94] and is included for the reader convenience. The dataset used throughout this work was provided to us by Kenneth McNatty and is partially published in [72, 104]. The dataset consists of morphological measurements of follicles performed by histology on ovaries from 120 and 135 days old sheep fetuses (cf Figure 2) : follicle and oocyte diameter, cell number, layer number, presence or absence of an antrum. The sheep used in this study are of the Romney strain. The dataset is subdivided into two subsets corresponding to two different genotypes : the "wild-type" genotype (WT) and the "mutant" Booroola (BB) genotype. Morphological and timing differences are observed between these two genotypes. During fetal development, all follicles stages appear at a later age in BB compared to WT. Also, at the end of the compact phase, oocytes are larger and there are fewer cells in BB. The alteration of follicle development observed in Booroola genotype comes from a natural mutation affecting the receptor to growth factor BMP15 [79].

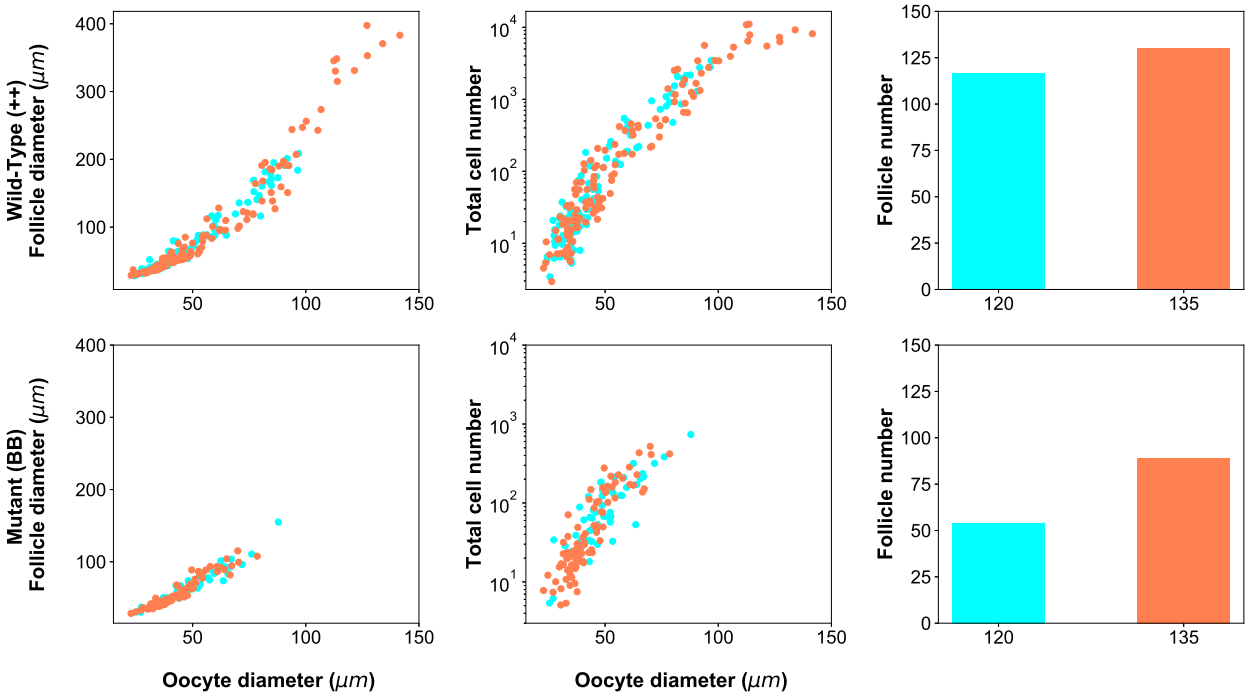

FIGURE 2 – **Dataset presentation**. Follicle diameter, oocyte diameter and somatic cell numbers in sheep fetuses from Wild-Type and Mutant phenotype, taken from [72]. The blue points represent the 120–days old fetuses while the orange ones correspond to the 135–days old fetuses.

The measurement of morphological characteristics of ovarian follicles is an invasive procedure. The ovaries are removed from individuals, then sectioned (slices of $20\mu m$ thickness) and fixed with a chemical agent. To perform the measurements (diameter, number of layers, etc.), experimentalists select the optical plane passing through the oocyte nucleolus (see for instance Figure 3). Cells are counted from this optical plane. This 2D cell number can be used to assess the whole 3D cell number from stereological techniques (see details in

FIGURE 3 – **Histological slice of an ovary with different types of follicles**. Courtesy of Danielle Monniaux.

[72]). The only temporal information associated with the morphological measures is the fetus age, which, as explained in the introduction, is not in principle correlated with the time spent by each follicle since its activation (which is asynchronous).

Although the dataset presented in Figure 2 can look consistent as a whole, it hides very distinct biological processes at play within the follicle growth. In particular, there a mixture of quiescent follicle and actively growing follicles. To segregate between each other, we dwell on a particular feature of our dataset, the follicle type. Indeed, follicles (in basal growth phase) were classified by the expert according to the following six categories (see details in [94]) :

- Type 1 : primordial quiescent follicles (one layer of flattened somatic cells),
- Type 1A : transitory follicles (one layer with a mixture of flattened and cuboidal somatic cells),
- Type 2 : primary follicles (from one to less than two complete layers of cuboidal somatic cells),
- Type 3 : small preantral follicles (from two to less than four complete layers of cuboidal somatic cells),
- Type 4 : large preantral follicles (from four to less than six complete layers of cuboidal somatic cells),
- Type 5 : small antral follicles (more than five layers of cuboidal somatic cells and a fully formed antrum).

Figure 4 presents the number of follicles of different types observed in our dataset. As mentioned, there are some differences in the distribution of follicle types according to the genotype. In the Mutant genotype, there are only very few primordial follicles already formed at age 120 days, and newly formed follicles have not reached the Type 4 and 5 at age 120 neither 135 days. In the sequel, we have used the data for Type 1, 1A and 2 to model the activation phase (section 1.2.2), and Type 2 to 4 to model the compact growth phase (section 1.2.3).

Furthermore, we applied additional criteria, gathering information from other studies, in order to really select follicles that can be termed as "viable" (see details in [94]). Only properly formed follicles will be able to get activated, and only follicles with a good balance between the oocyte diameter and the cell layer number will be able to pursue their development.

Finally, some partial kinetics information can be inferred from the first times of appearance of the different follicular types during the fetal life (see details in in [94]). From these information, we roughly deduce the transit times : it takes 20 days to go from a Type 2 to a Type 3 follicle, and 15 more days to go from a Type 3 to a Type 4. These orders of magnitude have to be taken with cautious given the time resolution of the data (10 to 15 days !).

The dataset used for the calibration of the follicle activation model finally consists of 90 data points for the Wild-Type dataset, and 81 data points for the BB dataset. The dataset used for the calibration of the model dedicated to the compact growth phase consists of 101 follicles, all from the Wild-Type phenotype. These datasets are presented in Figure 5.

We close this presentation on available data by mentioning that our experimentalist collaborators are able to follow the growth of ovarian follicle *in-vitro* during a specific window of follicle development, namely from pre-antral to large antral follicles [29]. This experimental device presents a very rich opportunity to

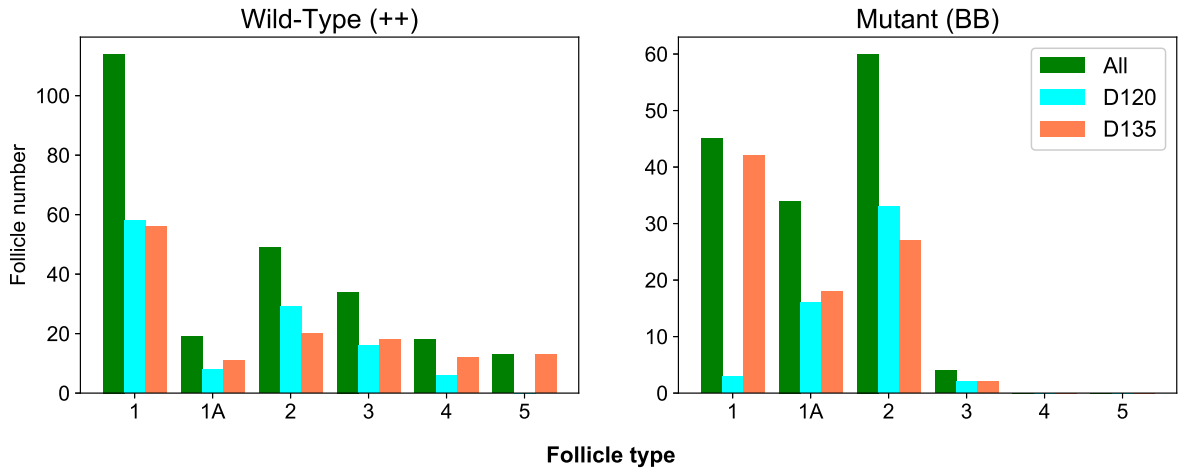

FIGURE 4 – **Follicle types**. Follicle distribution according to follicle types (same dataset as presented in Fig 2).

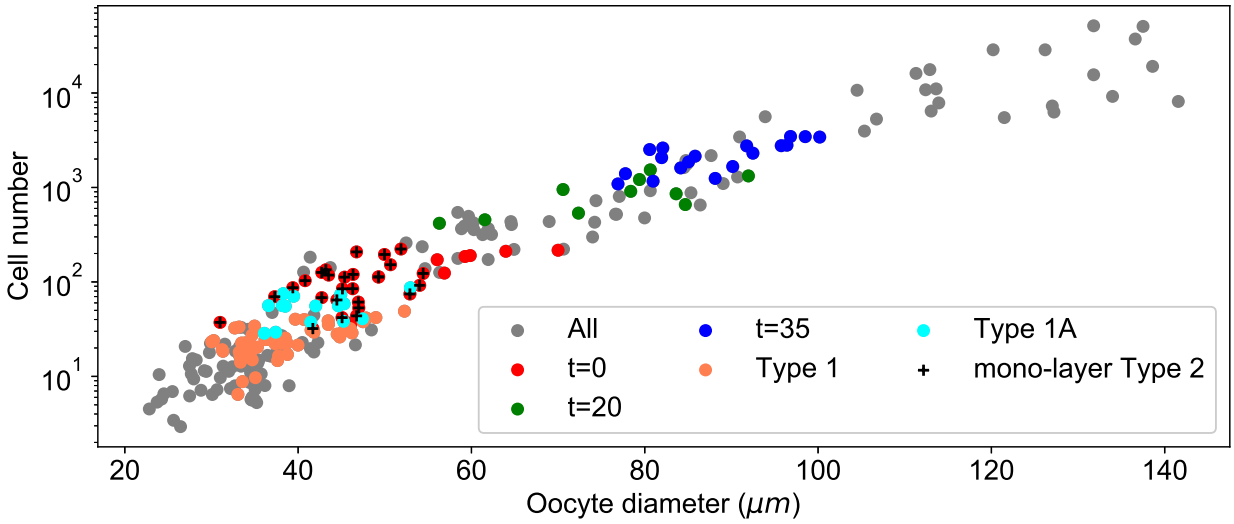

FIGURE 5 – **Wild-Type extracted datasets.** We extract from the original dataset (see Fig 2), two datasets : one dedicated to the compact growth phase (red, green and blue points, corresponding respectively to Type 2, 3 and 4) and one dedicated to the activation phase (coral and cyan points, and cross sign +). The gray points represent the follicles of the original dataset that we did not select based on other criteria (see details in in [94]).

obtain real kinetic information on this process (with of course all possible bias due to the *in-vitro* settings). This and other recent improvements on experimental devices (such as CLARITY [45]) constitute interesting perspectives for quantitative dynamical modeling of follicle growth.

1.2.2 Activation phase

This chapter is taken from [A6].

We model the follicle activation by focusing on the number of somatic cells in one follicle, from its initial quiescent state (primordial follicle) to its fully activated state (primary follicle). The somatic cells are structured in two types : flattened (here after named *precursor*) cells and cuboidal (here after named

proliferative) cells. The model aims to study the joint dynamics of the precursor cells F and proliferative cells C within a single follicle, whose populations are ruled by four types of cell events. In the absence of specific information, we used the simplest formulation as possible for all event rates, according to Occam's razor principle. Two cell events occur at the expense of the precursor cells, which are consumed during their transition : (i) $\mathcal{R}_1$ is the spontaneous transition of precursor cells into proliferative cells, whose rate $\alpha_1 F$ is linearly proportional to the number of precursor cells ; (ii) $\mathcal{R}_2$ is the auto-amplified transition of precursor cells into proliferative cells, which occurs at rate $\beta \frac{FC}{F+C}$. This event represents the feedback of proliferative cells onto the transition of the precursor cells. Two other cell events increase the proliferative cell population without affecting the precursor cell population : (i) $\mathcal{R}_3$ is an asymmetric division of precursor cells F (giving rise to one precursor cell and one proliferative cell), which occurs at rate $\alpha_2 F$; (ii) $\mathcal{R}_4$ is a symmetric division of the proliferative cells C (giving rise to two proliferative cells), which occurs at rate γC.
These four cell events are the building blocks of model $\mathcal{M}_{FC}$:

		Cell events	Rate	
Spontaneous transition	$\mathcal{R}_1$:	$(F,C) \to (F-1, C+1)$,	$\alpha_1 F$,	
Auto-amplified transition	$\mathcal{R}_2$:	$(F,C) \to (F-1, C+1)$,	$\beta \frac{FC}{F+C}$,	$(\mathcal{M}_{FC})$
Asymmetric division of F	$\mathcal{R}_3$:	$(F,C) \to (F, C+1)$,	$\alpha_2 F$,	
Symmetric division of C	$\mathcal{R}_4$:	$(F,C) \to (F, C+1)$,	γC .	

Cell events $\mathcal{R}_1$ and $\mathcal{R}_4$ constitute the fundamental ingredients involved in the activation process. We also consider two additional cell events, $\mathcal{R}_2$ and $\mathcal{R}_3$, which are not only intended to enrich the model behavior, but are also substantiated by biological observations. We refer to [A6] for a deeper discussion on the model design and the choice of rate events.

The model ($\mathcal{M}_{FC}$) is mathematically formulated as a Continuous time Markov chain (CTMC), whose matrix transition can be directly deduced from ($\mathcal{M}_{FC}$). This stochastic description is especially appropriate when dealing with a small number of cells. As it can be easily seen from ($\mathcal{M}_{FC}$), the proliferative C population grows monotonously as the precursor F population decreases until extinction. The number of precursor cells remains constant whenever there is a division event ($\mathcal{R}_3$ or $\mathcal{R}_4$), and decreases by one whenever there is a transition event ($\mathcal{R}_1$ or $\mathcal{R}_2$). If we are only given the sequence of jump events ($(-1,1)$ vs $(0,1)$), we cannot discriminate $\mathcal{R}_1$ from $\mathcal{R}_2$, neither $\mathcal{R}_3$ from $\mathcal{R}_4$. The biological data will however be slightly richer by providing snapshot data with the number of flattened and cuboidal cells in distinct follicles. The fact that each event has distinct rate events in ($\mathcal{M}_{FC}$) gives *a priori* some possibilities to infer parameter values from these observations (e.g. $\mathcal{R}_1$ and $\mathcal{R}_3$ intensities decrease along the activation phase, while $\mathcal{R}_2$ and $\mathcal{R}_4$ increases). We present these data in Figure 6 (see [A6] for a detailed presentation).

Starting with no proliferative cells and given (random) number of precursor cells F_0, we have mainly focused in our study [A6] on characterizing the extinction time of the precursor cell population (which represents the total follicle activation time),

$$\tau := \inf\{t \geq 0; \quad F_t = 0 | F_0\}\,, \tag{1}$$

and the number of proliferative cells C at that time,

$$C_\tau\,. \tag{2}$$

We may notice that the proliferative cells undergo an exponential growth due to the cell event $\mathcal{R}_4$ in ($\mathcal{M}_{FC}$). It is also clear that the precursor cells F cannot decrease faster than the maximal transition rate $(\alpha_1+\beta)F$, so that τ is lower bounded by a hypo-exponential distribution, whose tail distribution is exponentially decreasing at rate $(\alpha_1+\beta)$. This suggests that the maximal transition rate $(\alpha_1+\beta)$ must be greater than the proliferative rate γ to control the mean number of proliferative cells at the extinction time of precursor cells. This is indeed true, and proved in the following main result :

Theorem 1. *For any integer $p \in \mathbb{N}^*$, we have*

$$\mathbb{E}\left[(C_\tau)^p\right] < \infty\,, \tag{3}$$

if and only if

$$p\gamma < \alpha_1 + \beta\,. \tag{$\mathcal{M}_{FC}$-H1}$$

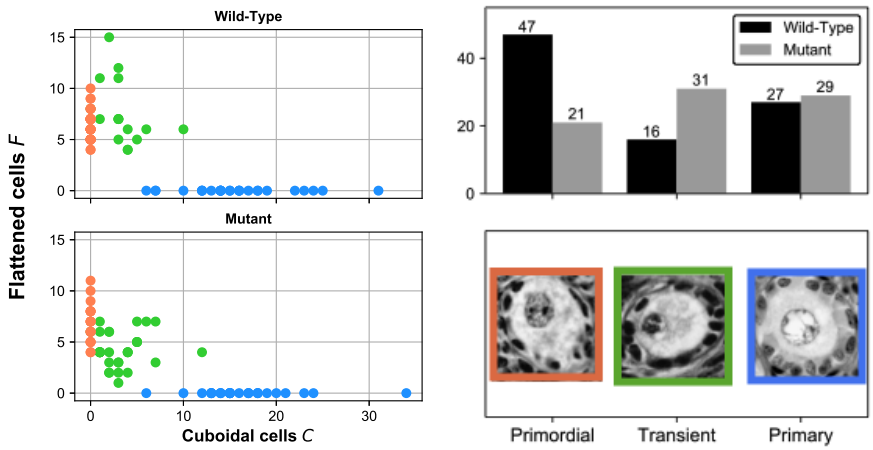

FIGURE 6 – **Experimental dataset for the activation phase**. Left panel : experimental data points of the number of Flattened cells F and the number of Cuboidal cells C, for both the Wild-Type and Mutant datasets. Red points : primary follicles, green points : transient follicles, blue points : primary follicles. Right panel : histological slices illustrating the different steps of activation (from left to right : primordial, transitory and primary follicles). Experimental dataset : courtesy of Ken McNatty ; Histological images : courtesy of Danielle Monniaux.

The proof of Theorem 1 is obtained by a coupling argument, finding appropriate lower and upper bounds of C_τ, for which analytical approaches are feasible. As discussed above, the lower bound is simply given by a Yule process $\tilde{C}$ (with only binary division event $\mathcal{R}_4$, that is cell division at rate $\gamma\tilde{C}$) stopped at an exponential random variable of rate $(\alpha_1+\beta)$. The upper bound is slightly more technical and goes along the following simple verbal argument. The proliferative cell population C may be bounded by an immigration-birth process $\hat{C}$ (immigration at rate $\alpha_2 F_0$, birth at rate $\gamma\hat{C}$). If C increases without bounds, the rate of transition of precursor cells will approach $(\alpha_1+\beta)$ (or the precursor cells extinct before), so that C_τ will be upper bounded by $\hat{C}$ stopped at a hypo-exponential distribution whose tail decreases with a rate arbitrary close to $(\alpha_1+\beta)$. Hypothesis ($\mathcal{M}_{FC}$-H1) is thus optimal to guarantee (3).

Theorem 1 provides strong theoretical argument to argue that the division rate of proliferative cells must be lower than the precursor transition rate, otherwise transient follicle with high number of proliferative cells will be observed, which is not the case (see Figure 6). The asymmetric transition rate α_2 is however not constrained by such consideration. The proof of Theorem 1 also provides an interesting upper-bound, that is analytically tractable *(at the price of some efforts not shown here !)*.

This upper-bound motivates the use of a finite projection method [82, 62] to numerically solve the infinite backward Kolmogorov equation that governs $\mathbb{E}[\tau]$ and $\mathbb{E}[(C_\tau)]$. Instead of trying to calculate the mean of (1) and (2), the algorithm computes (exactly) the mean of :

$$\tau_r := \inf\{t \geq 0; \quad F_t = 0\,, \text{or } C_t \geq r | F_0\}\,, \quad C_{\tau_r}.$$

The justification of the algorithm's accuracy is based on the idea that it is unlikely that $C_{\tau_r} \geq r$, for large r, so that $\tau_r = \tau$ (and $C_{\tau_r} = C_\tau$) with high probability. As C_τ may potentially have infinite moments, some care must actually be taken to calculate those moments with a truncation argument. As a consequence, we require $\gamma < \alpha_1 + \beta$ to calculate $\mathbb{E}[\tau]$ with an arbitrary precision, and we require $2\gamma < \alpha_1 + \beta$ to calculate $\mathbb{E}[C_\tau]$ with an arbitrary precision (we need $\mathbb{E}\left[C_\tau^2\right] < \infty$). See [A6, Proposition 4] for more details. We note that Monte-Carlo procedures may actually provide estimates for $\mathbb{E}[\tau]$ without parameter restriction, since the second moment of τ is finite for any parameter combination.

Parameter calibration requires a statistical model to attest the distance between the model to the data. A remarkable feature of ($\mathcal{M}_{FC}$) is that the proliferative cell population C increases by one at any cell event. Hence the embedded discrete Markov Chain $(F_n, C_n)_{n\in\mathbb{N}}$ given by the population at any cell event is such that C_n is deterministic with $C_n = n$. Also, the iteration F_n may be deduced from F_{n-1}, or, say differently, we deduce the law of F_c at the "pseudo-time" $C = c$ from the law of F_{c-1} at the "pseudo-time" $C = c-1$ as

follows : for all $(f,c) \in \mathcal{S}$,

$$\mathbb{P}\left[F_c = f\right] = \underbrace{q_{f+1,f}(c-1)\mathbb{P}\left[F_{c-1} = f+1\right]}_{\text{transition}} + \underbrace{q_{f,f}(c-1)\mathbb{P}\left[F_{c-1} = f\right]}_{\text{asymmetric/symmetric division}}, \tag{4}$$

where

$$q_{f+1,f}(c) = \frac{\alpha_1(f+1) + \beta\frac{(f+1)c}{f+1+c}}{(\alpha_2+\alpha_1)(f+1) + \gamma c + \beta\frac{(f+1)c}{f+1+c}}, \quad q_{f,f}(c) = \frac{\alpha_2 f + \gamma c}{(\alpha_2+\alpha_1)f + \gamma c + \beta\frac{fc}{f+c}}. \tag{5}$$

Hence $(F_c)_{c\in\mathbb{N}}$ is a non-homogeneous discrete time Markov chain, its law $(\mathbb{P}\left[F_c = f\right])_{f,c\geq 0}$ gives the probability to get f precursor cells when there is c proliferative cells. We use this discrete time Markov chain to build a likelihood for our data.

Remark 1. Besides neglecting measurement error and inter-individual variability, we note that doing so, we implicitly assume that we have a complete observation of the successive cell events, or equivalently, that the time spent in each state through the follicle activation process is roughly of the same order and do not bias any snapshot observation from the point of view of the discrete time Markov chain. However, a look at the data in Figure 6 clearly shows that there is a gap in our observation : we do not have data points for $F \in [1,4]$ in the Wild-Type strain, suggesting that the end of the follicle activation was fast enough to be barely visible from a snapshot data of a subset of the follicle population present in the ovaries. We disregard this consideration for the moment, and will come back to that in chapter 1.4.

To complement Eqs. (4), we take as initial law for F_0 a shifted Poisson random variable (on $\mathbb{N}^*$) of parameter μ, for all $f \geq 1$,

$$\mathbb{P}\left[F_0 = f\right] = \frac{\mu^f}{(e^\mu - 1)f!}. \tag{6}$$

We compute the maximum likelihood (MLE) associated to Eqs (4)-(5)-(6) (see details in [A6]). The best fit results are shown in Figure 7, where we included two submodels, with only $(\mathcal{R}_1, \mathcal{R}_4)$ events or only $(\mathcal{R}_1, \mathcal{R}_3)$ events. For both the Wild-Type and Mutant subsets, a visual inspection shows that submodel $(\mathcal{R}_1, \mathcal{R}_4)$ leads to a "direct" transition, followed by prolonged cell proliferation after precursor cell extinction, while with submodel $(\mathcal{R}_1, \mathcal{R}_3)$, there is a higher probability that the total number of cells increases before precursor cell extinction (as cell proliferation is not anymore possible after cell precursor extinction). We also observe from the lower panels of Figure 7 that the full model lies somehow in between, with trajectories that display an intermediate level of cell proliferation before precursor cell extinction.

To analyze the parameter identifiability, we follow the practical approach based on the profile likelihood estimate (PLE), see for instance [92]. Note that we have fixed $\alpha_1 = 1$ as the time scale cannot be inferred from these data. The results are shown in Figure 8. In short (see [A6] for details), we could faithfully estimate the initial condition parameter μ, the proliferative cell division rate can only be upper bounded, there is a very large confidence interval for α_2 in the complete model and β cannot be inferred. Consistently with Theorem 1, we could nevertheless show that the self-amplification transition rate β is constrained to be grater than the proliferation rate γ. We conclude that the best fit trajectories indeed favors transition over proliferation.

Comparing the results between WT and Mutant, we found that there is around one more flattened cells initially in WT follicles. No clear timescale separation between the Wild-type and Mutant dynamics could however be revealed, although some parameter combinations are compatible with a faster transition in the Wild-Type case than in the Mutant case. This is compatible with monitoring studies, which observed that the times of apparition of both the first primordial and primary follicles are shifted compared with wild-type animals (they appear a little later). The predicted mean number of proliferative cells at the extinction time lies between 8 and 10 cells in both phenotype, while the number of divisions during the follicle activation is smaller in the Wild-Type than in the Mutant subset ($\mathbb{E}[C_\tau - F_0] \approx 2$ in Wild-Type, $\mathbb{E}[C_\tau - F_0] \approx 4$ in Mutant). In overall, we conclude from our extensive data-fitting analysis that the Wild-Type subset exhibits a clearer separation of dynamics during follicle activation (first cell transition, then cell proliferation), while in the Mutant cell proliferation could occur at a substantial rate before precursor cell extinction. We note that this conclusion has to be tempered by the sparse character of our experimental dataset. With the available experimental dataset, we have yet not managed to make a clear distinction between, on one side, a progressive transition with a steady net flux from flattened to cuboidal cells, and, on the other side, an auto-catalytic transition with an ever increasing flux all along the activation phase.

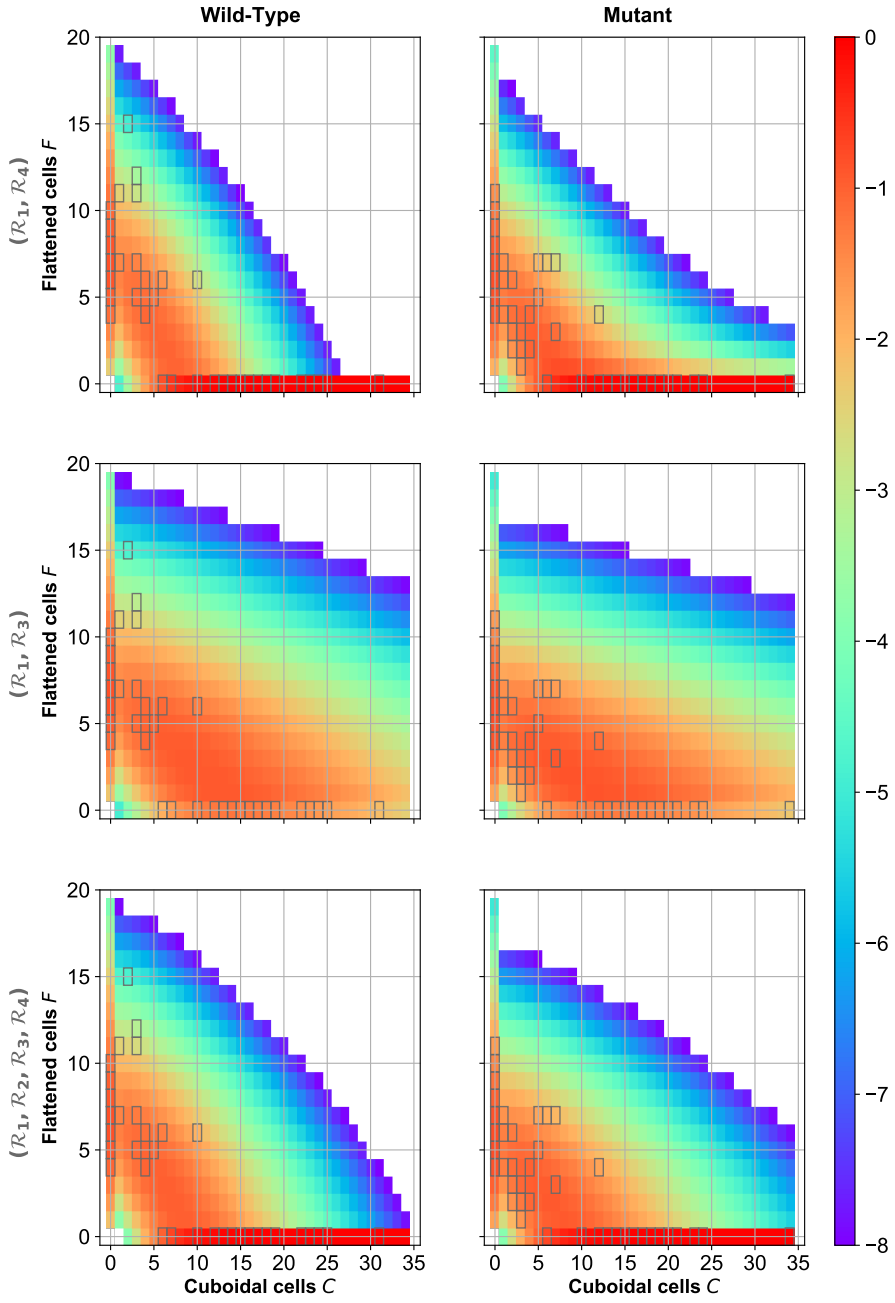

FIGURE 7 – **Two-events and full submodels : best fit trajectories.** Using Formula 4, we compute each probability $\mathbb{P}[F_c = f]$ for submodels $(\mathcal{R}_1, \mathcal{R}_4)$, $(\mathcal{R}_1, \mathcal{R}_3)$ and the complete model $(\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4)$ with their respective maximum likelihood parameter estimate for Wild-Type data set (left column) and Mutant dataset (right column). Each dark gray square corresponds to a data point. The colormap corresponds to the probability values $\mathbb{P}[F_c = f]$ in log10 scale.

1.2.3 Basal growth phase

This section is taken from [A5].

The objective of this study is to model the compact growth phase of small follicles, focusing on describing the somatic cell population evolution. We study a multi-type age dependent branching process to represent the dynamics of a finite population of proliferative cells distributed into successive layers around the oocyte.

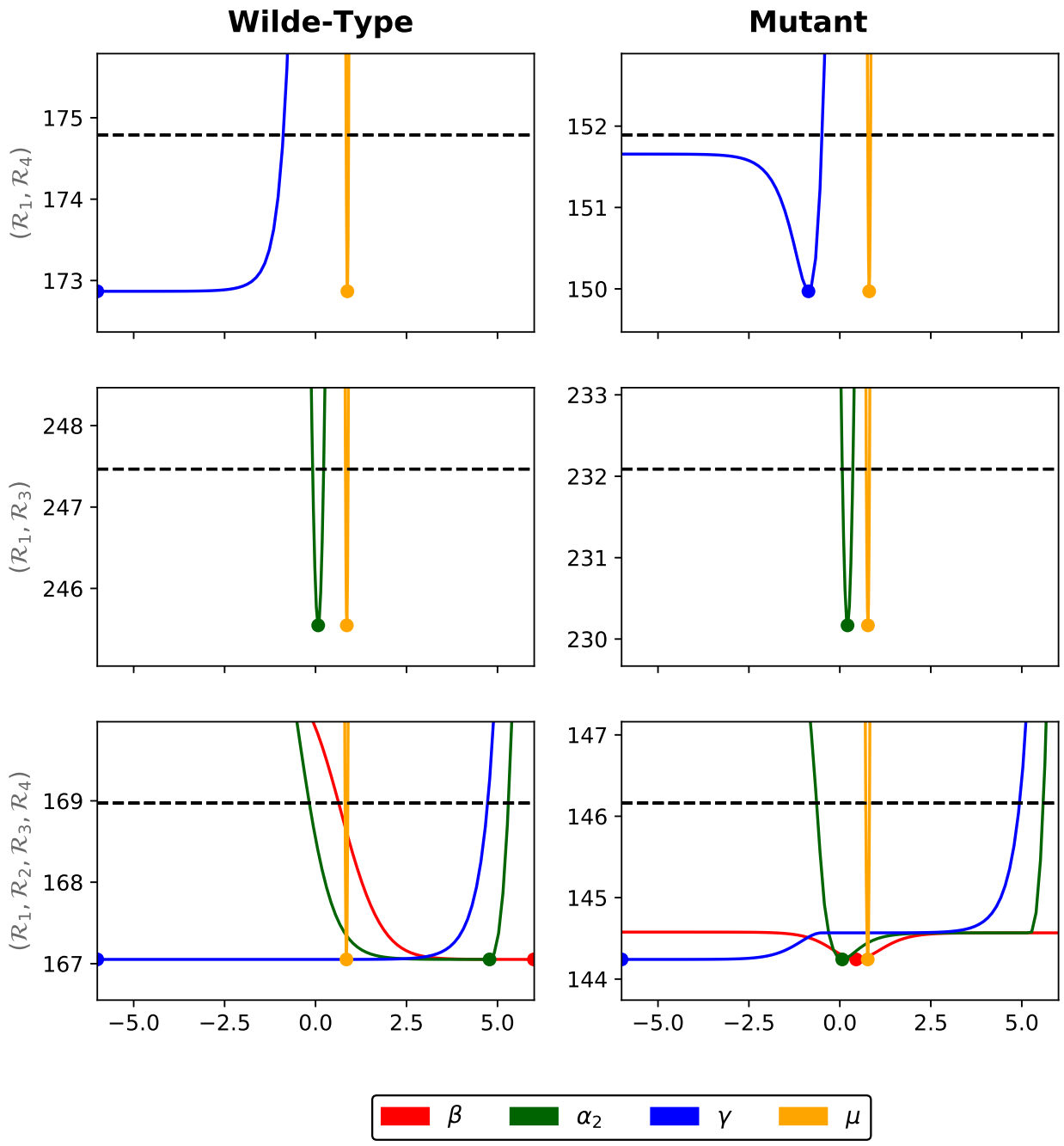

FIGURE 8 – **Two-events and full submodels : PLE**. Each panel represents the PLE, in log10 scale, obtained from the experimental datasets, and either submodel $(\mathcal{R}_1, \mathcal{R}_4)$ (top panels), $(\mathcal{R}_1, \mathcal{R}_3)$ (center panels), or $(\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4)$ (down panels). The dashed black line represents the 95%-statistical threshold. Orange solid lines : PLE values for the initial condition parameter μ ; blue solid lines : PLE values for the symmetric cell proliferation rate γ ; green solid lines : PLE values for the asymmetric cell division rate α_2 ; red solid lines : PLE values for the self-amplification transition rate β. The colored points represent the MLE.

The cell population has two structuring variables : a continuous age variable $a \in \mathbb{R}^+$ and a discrete layer index variable $j \in [\![1, J]\!]$. All cells behave independently from each other, and a cell of age a and index j may divide and give birth to two new cells according to the following transitions

$$\begin{array}{lcl} & \text{Cell events} & \text{Rate} \\ \text{No migration :} & (j,a) \to (j,0),(j,0) & b_j(a)p_{2,0}^{(j)}, \\ \text{One migration :} & (j,a) \to (j+1,0),(j,0), & b_j(a)p_{1,1}^{(j)}, \\ \text{Two migrations :} & (j,a) \to (j+1,0),(j+1,0), & b_j(a)p_{0,2}^{(j)}\,. \end{array} \qquad (\mathcal{M}_C)$$

In model ($\mathcal{M}_C$), $b = b_j(a)$ is an age-and-layer-dependent instantaneous division rate, and $p_{2,0}^{(j)}$ is the probability that both daughter cells remain on the same layer, $p_{1,1}^{(j)}$ and $p_{0,2}^{(j)}$, the probability that a single or both daughter cell(s) move(s) from layer j to layer $j+1$, respectively, with $p_{2,0}^{(j)} + p_{1,1}^{(j)} + p_{0,2}^{(j)} = 1$. There is a fixed maximal layer J, with $p_{2,0}^{(J)} = 1$. We note the probability that a cell taken randomly among both daughter cells remains on the same layer by $p_S^{(j)} = \frac{1}{2}p_{1,1}^{(j)} + p_{2,0}^{(j)}$, and $p_L^{(j)} := 1 - p_S^{(j)}$ is the probability that the cell moves, for any $j \in [\![1, J-1]\!]$.

Hence, cells divide and the two daughter cells of age 0 may move irreversibly from one layer to the next. The cell population may be represented for each time $t \geq 0$ by a measure $Z_t \in \mathcal{M}_P$, the set of point measures on $\mathcal{E} := [\![1, J]\!] \times \mathbb{R}^+$:

$$Z_t = \sum_{k=1}^{N_t} \delta_{I_t^{(k)}, A_t^{(k)}}, \quad N_t := \ll Z_t, \mathbb{1} \gg = \sum_{j=1}^{J} \int_0^{+\infty} Z_t(dj, da)\,,$$

where N_t is the total number of cells at time t. In the sequel, we take for simplicity $Z_0 = \delta_{1,0}$. Yet, a typical biological condition would be the outcome of the activation process, see model ($\mathcal{M}_{FC}$), that is the number of proliferative cells of a primordial follicle.

The mathematical question is to understand the long time behavior of Z_t. Clearly, an exponential growth is expected, like in standard branching processes or finite-dimensional linear systems.

On a layer j, the cell division time is given by a random variable τ_j, whose probability distribution is given by

$$d\mathcal{B}_j(x) = b_j(x)e^{-\int_0^x b_j(a)da}\,.$$

Looking for the time being a layer j as an isolated system (without influx from layer $j-1$), it is classical that the growth rate of the population of cells on that layer is dependent of the sign of $(2p_S^{(j)} - 1)$ and governed by the rate λ_j defined by :

Definition 1. The intrinsic growth rate λ_j of layer j is (if it exists) the unique solution of

$$d\mathcal{B}_j^*(\lambda_j) := \int_0^\infty e^{-\lambda_j s} d\mathcal{B}_j(s)ds = \frac{1}{2p_S^{(j)}}\,. \qquad (7)$$

Note that a solution of (7) does not necessarily exist, which may lead to technical complications (essentially because cells would divide too fast, and leads to infinitely many cells in age 0, or too slow, and leads to cells never dividing of age infinity). For this reason, we assume the (non-optimal) hypotheses :

Assumption 1. *For all $j \in [\![1, J-1]\!]$, $p_S^{(j)}, p_L^{(j)} \in (0, 1)$.*

Assumption 2. *For each layer j, b_j is continuous bounded below and above :*

$$\forall j \in [\![1, J]\!], \quad \forall a \in \mathbb{R}_+, \quad 0 < \underline{b}_j \leq b_j(a) \leq \bar{b}_j < \infty\,.$$

Assumption 3. *For all $j \in [\![1, J]\!]$, $\lambda_j > -\liminf_{a\to+\infty} b_j(a)$.*

Assumptions (1)-(2) guarantee the existence and uniqueness of λ_j with $\lambda_j > -\bar{b}_j$, and we recall that $\lambda_j < 0$ when $p_S^{(j)} < \frac{1}{2}$, $\lambda_j > 0$ when $p_S^{(j)} > \frac{1}{2}$ and $\lambda_j = 0$ when $p_S^{(j)} = \frac{1}{2}$. In particular, $\lambda_J > 0$ as $p_S^{(J)} = 1$. Assumption 3 implies additional regularity for $t \mapsto e^{-\lambda_j t} d\mathcal{B}_j(t)$:

$$\forall j \in [\![1, J]\!], \forall k \in \mathbb{N}, \int_0^\infty t^k e^{-\lambda_j t} d\mathcal{B}_j(t)dt < \infty\,.$$

Thus each layer has a proper intrinsic rate of growth (or degrowth). However, this is not enough to obtain an exponential growth for the whole population. Just like a matrix may have multiple eigenvalue, one has to exclude some degenerate cases that would lead to polynomial corrections. This consideration motivate the :

Assumption 4. *There is an unique maximal element taken among the intrinsic growth rates (λ_j, $j \in [\![1, J]\!]$) (defined in Definition 1). This unique element is called the Malthus parameter, λ_c, and the layer such that the index $j = c$ is called the leading layer.*

Note that it is clear that λ_c is positive. The precise long time results requires the introduction of an eigenproblem (P), analogous to finite-dimensional linear systems :

$$\begin{cases} \mathcal{L}^P \rho(a) = \lambda \rho(a),\ a \geq 0 \\ \rho(0) = \displaystyle\int_0^\infty K(a)\rho(a)da \\ \ll \rho, \mathbb{1} \gg = 1 \text{ and } \rho \geq 0\,, \end{cases}, \quad \mathcal{L}^P \rho(a) = \partial_a \rho(a) - B(a)\rho(a), \tag{P}$$

and its dual problem (D)

$$\begin{cases} \mathcal{L}^D \phi(a) = \lambda \phi(a),\ a \in \mathbb{R}_+^* \\ \ll \rho, \phi \gg = 1 \text{ and } \phi \geq 0 \end{cases}, \quad \mathcal{L}^D \phi(a) = \partial_a \phi(a) - B(a)\phi + K(a)^T \phi(0)\,, \tag{D}$$

where

$$B(a) = \mathrm{diag}(b_1(a), ..., b_J(a)), \quad [K(a)]_{i,j} = \begin{cases} 2p_S^{(j)} b_j(a), & i = j, & j \in [\![1, J]\!]\,, \\ 2p_L^{(j-1)} b_{j-1}(a), & i = j-1, & j \in [\![2, J]\!]\,. \end{cases}$$

We now state the existence of maximal solutions to this eigenproblem.

Theorem 2 (Eigenproblem)**.** *Under Assumptions 1, 2, 3 and 4, there exists a first eigenelement triplet (λ, ρ, ϕ) solution to equations (P) and (D) where $\rho \in \mathbf{L}^1(\mathbb{R}_+)^J$ and $\phi \in \mathcal{C}_b(\mathbb{R}_+)^J$. Moreover, $\lambda = \lambda_c$ is the Malthus parameter given in Definition 4, and ρ and ϕ are unique.*

Due to the specific form of the transition between layers, we may actually obtain analytical formulas for the eigenfunctions, so that the hypotheses of Theorem 2 could be weakened. In any case, from these eigenelements, using martingale techniques [56], we deduce directly the following result of convergence for the stochastic process Z :

Theorem 3. *Under Assumptions 1, 2, 3 and 4, $W_t^\phi = e^{-\lambda_c t} \ll \phi, Z_t \gg$ is a square integrable martingale that converges almost surely and in $\mathbf{L}^2$ to a non-degenerate random variable W_∞^ϕ.*

An analytical characterization of the law of W_∞^ϕ is not always possible. Instead, using generating function methods developed for multi-type age dependent branching processes (see [55], Chap. VI), we write a system of renewal equations and obtain analytical formulas for the first asymptotic moment *(up to some integral calculations, some effort needed !)*. We define $Y_t^{(j,a)} := \langle Z_t, \mathbb{1}_{j,\leq a} \rangle$ as the number of cells on layer j and of age less than or equal to a at time t, and $m_i^a(t)$ its mean starting from one mother cell of age 0 on layer 1 :

$$m_j^a(t) := \mathbb{E}[Y_t^{(j,a)} | Z_0 = \delta_{1,0}]\,.$$

Theorem 4. *Under Assumptions 1, 2, 3 and 4, for all $a \geq 0$,*

$$\forall j \in [\![1, J]\!], \quad m_j^a(t) e^{-\lambda_c t} \to \widetilde{m}_j(a), \quad t \to \infty, \tag{8}$$

where $\widetilde{m}_j(a) =$

$$\begin{cases} 0, & j \in [\![1, c-1]\!], \\ \dfrac{\int_0^a \rho^{(c)}(s)ds}{2p_S^{(c)} \rho^{(c)}(0) \int_0^\infty s d\mathcal{B}_c(s) e^{-\lambda_c s} ds}, & j = c, \\ \dfrac{\int_0^a \rho^{(j)}(s)ds}{2p_S^{(c)} \rho^{(c)}(0) \int_0^\infty s d\mathcal{B}_c(s) e^{-\lambda_c s} ds} \displaystyle\prod_{k=1}^{c-1} \frac{2p_L^{(k)} d\mathcal{B}_k^*(\lambda_c)}{1 - 2p_S^{(k)} d\mathcal{B}_k^*(\lambda_c)}, & j \in [\![c+1, J]\!]. \end{cases}$$

At the price of some assumptions, the theoretical results thus predict an asymptotic exponential growth, with a stable repartition according to age/layer (we refer to [A5] for analogous results from the deterministic point of view on the same model). It can be seen from our analytical results that under assumption 4 there is a leading (or driving) layer, the one that has the highest intrinsic growth rate. Because cells are not allowed to come back to previous layer, the cell population in layers below the leading one are growing at an exponential rate strictly lower than the Malthus parameter, and thus vanishes in the renormalization (8).

Interestingly, our data-fitting approach on the data presented in Figure 5 showed that it is plausible that the leading layer is the first one and that the somatic cell population in a basal follicle undergo an exponential growth, driven by the intrinsic exponential rate of the first layer and subsequent migration towards outer layers. Due to the very sparse data compared to our parametrization in model ($\mathcal{M}_C$), we had to simplify quiet a lot the model calibration approach. We then considered age-*independent* birth rates of the form

$$b_j = \frac{b_1}{1 + (j-1) \times \alpha}, \, j \in [\![1, 4]\!], \, \alpha \in \mathbb{R}. \tag{9}$$

These relations (9) are motivated by biological specifications and were primarily used in [34] : somatic cell division is supported by growth factors secreted by the oocyte, thus division is slower and slower when the somatic cells get further away of the oocyte. Hence the law (9) reflects that (b_j) is decreasing, which does *not* necessarily imply that $\lambda_j = (2p_s^j - 1)b_j$ does so as well. We finally focused only on the mean number of cells in each layer. We are then lead to a very simple finite dimension linear system

$$\left\{ \begin{array}{l} \frac{d}{dt}M(t) = AM(t) \\ M(0) = (N, 0, ..., 0) \in \mathbb{R}^J \end{array} \right. , \quad [A]_{i,j} := \left\{ \begin{array}{lll} (2p_S^{(j)} - 1)b_j, & i = j, & j \in [\![1, J]\!], \\ 2p_L^{(j-1)}b_{j-1}, & i = j-1, & j \in [\![2, J]\!]. \end{array} \right. \tag{10}$$

where we considered an initial condition with $N \in \mathbb{N}^*$ cells in the first layer.

We proved the structural identifiability of the parameter set $\mathbf{P} := \{N, b_j, p_S^{(j)}, j \in [\![1, J]\!]\}$ when we observe the vector $M(t; \mathbf{P})$ at each time t.

Theorem 5. *Under Assumption 1, and complete observation of system (10), the parameter set* $\mathbf{P}$ *is identifiable.*

We then perform the estimation of the parameter set $\mathbf{P}$ from experimental cell number data retrieved on four layers and sampled at three different time points.

We estimate the parameter set $\mathbf{P}_{exp} = \{N, b_1, \alpha, p_S^{(1)}, p_S^{(2)}, p_S^{(3)}\}$ using deterministic optimization of a likelihood with an additive Gaussian noise model, using the D2D software [93] (see Figure 9 and Table 5). An analysis of the profile likelihood estimate showed that all parameters except $p_S^{(2)}$ are practically identifiable (see details in [A5]).

TABLE 5 – **Estimated values of the parameters** of model (10), associated to the best fit in Fig 9.. Only black parameters were free parameters. The blue ones are given by Eq. (9), and the orange ones given by $\lambda_j = (2p_s^j - 1)b_j$.

Layer j	$p_S^{(j)}$	b_j	λ_j
1	0.6806	0.1146	0.0414
2	0.4837	0.0435	-0.0014
3	0.9025	0.0354	0.0285
4	1	0.0324	0.0324

Going back to Theorem 4, the predicted long time repartition with the parameter set of Table 5 is a cell population with same order of magnitude in all layers ($\lambda_c = \lambda_1$), such that layer 3 has the more cells among the four layers included in the study. This certainly points out the limitation of our approach.

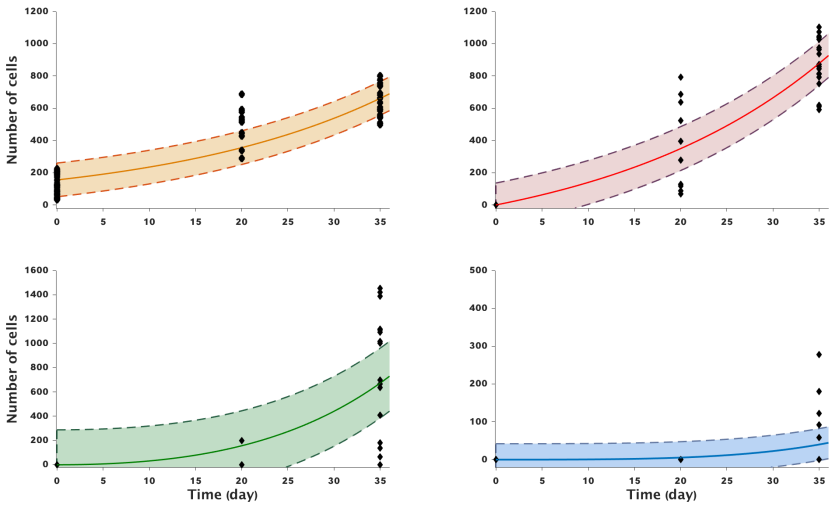

FIGURE 9 – **Data fitting with model** (10)**.** Each panel illustrates the changes in the cell number in a given layer (top-left : Layer 1, top-right : Layer 2, bottom-left : Layer 3, bottom-right : Layer 4). The black diamonds represent the experimental data, the solid lines are the best fit solutions of (10) and the dashed lines are drawn from the estimated variance.

1.2.4 Perspectives

We have studied two relatively simple population dynamics model, which allow to shed light on the cell kinetic mechanism at play during the early phase of ovarian follicular growth. These model were somehow adapted to the given set of data we dispose of, in order to extract the maximum of information from it (see chapter 1.4).

Regarding the activation phase, one may wonder if the spatial location of the first proliferative cells versus the remaining flattened cells play a role. This would require a refined model with some geometrical structure. The introduction of auto-amplification rate was motivated by some spatial interpretation, with two possible (and non exclusive) underlying mechanisms. First, the very first cell transitions could awake the oocyte and settle a positive feedback loop between the somatic cells and the oocyte [77] that would in turn secrete stimulatory factors reaching the surrounding somatic cells by diffusion (global amplification). Second, communications between adjacent somatic cells could help propagate activation step by step, from one (or a few) originally activated cell (local amplification). The later local amplification might be detected in the data by recording the location of cuboidal cells and checking whether cluster of spatially related cuboidal cells can be detected. Homogeneous repartition of cuboidal cells from static histological data would rather favor the the global amplification scenario.

For the compact growth phase, we are well aware that our model lacks several crucial ingredient in the dynamics of follicle growth, namely the cell crowding on each successive layers around the oocyte, and the dialog that settle between the growth of the oocyte and the proliferation of somatic cells. The interplay between the growth and proliferation processes was at the heart of the first model designed by Frédérique Clément and co-authors in [34]. Although this study has shown the necessary balance between the two processes, the detailed geometrical structure makes it very hard to analyze extensively the model outputs and to fully appreciate the role of each parameters in the transient and long term evolution of the follicle. The mean-field limit obtained recently in [74] may provide an interesting intermediate between somehow too simplistic and hardly tractable models.

Last but not least, the current perspective in our team is to design a satisfactory model of antrum formation and growth, to close the loop of a series of models on ovarian follicle growth. The difficulty there is to capture the essential biophysical processes that explain the apparition of this cavity, at a precise follicle diameter, and its subsequent growth. Our current working hypothesis is that competing forces are exerting on the antrum. On one side, an osmotic gradient pressure tends to enlarge it, as a consequence of the secretion by somatic cells of some large osmotically-active molecules (proteoglycans and glycosaminoglycans, see [33])

which are "trapped" inside the follicle. On the other side, a more classical mechanical gradient pressure, as a consequence of cell proliferation and overcrowding. We have formulated this model in a continuous PDE approach, based on advection-diffusion equations. We obtain a moving-boundary problem, which is a variant of the porous medium equation with non trivial boundary condition. The analysis of such models is a *work in progress...*

1.3 Modeling the whole follicle population

In this chapter, we turn to models that takes into account the whole follicle population on a lifespan time scale. We first describe briefly in section 1.3.1 the typical dataset we are interested in. In section 1.3.2 we present a multi-time scale model together with a singular perturbation analysis. We close this chapter with some perspectives in section 1.3.3.

1.3.1 Data set and working hypothesis

Typical data on follicle counts are based on invasive measures of follicles that are manually counted and classified into distinct morphological classes adapted to the given species. Repeated measures on distinct individuals at different ages then results in unpaired time series of follicles according to age. These data are mostly available on mouse models [42, 44] (see Figure 10), sheep [95] or women [37], with variable granularity on particular sub-classes of follicles according to studies. This is little to say that exhaustive counting data are scarce. Yet, as mentioned in introductory chapter 1.1, recent biotechnological improvements may lead to a gain of interest for more quantitative data in a near future.

There have been quiet of interests in statistical analysis of those data, trying most of the time through regression analysis to assess the difference between different genotype or even different species [37]. From the modeling side, pioneering studies (see e.g. [42]) have focused on fitting the parameters of a linear compartmental model, with one compartment by follicle class. This approach was clearly motivated by the apparent exponential decay of the total follicle numbers through age. A closer look at the data actually revealed some limitations of simple exponential models. Generalizations have thus considered time-varying (piecewise constant) rates [43], where the timing of the rate changes is motivated by global anatomical changes (closely after birth, at puberty, close to menopause, etc...).

These studies remain rather descriptive, and do not take into account follicle interactions, which is now clearly established through paracrine and endocrine feedback loops (see introductory chapter 1.1). Starting from previous linear compartmental studies, we thus have extended in [A3] the population dynamic approach accounting for interactions between follicles. For instance, we have in mind the possibility to represent the population-level modulation exerted by growing ovarian follicles on the activation of quiescent, primordial follicles, and/or the competition between terminally growing follicles [35].

We may observe in Tables 1-2-3 that various order of magnitude and time scales are involved in the whole folliculogenesis. In particular, it is striking to note that the level of the population of primordial follicles is 10 to 100 times higher than small growing follicles, and that the estimated rate of activation is also 10-100 times slower than the growth rate of basal follicles, according to the species. We make then the hypothesis that two distinct time scales are present, which are also related to two distinct orders of magnitude :

- The "static" reserve of the quiescent follicles is initially very abundant, of order $\frac{1}{\varepsilon}$, and each quiescent follicle has a slow basal activation rate, of order ε. The overall rate of apparition of new small growing follicles is thus or order 1.
- The "dynamic" reserve of growing follicles is of order one, and their growth rate is of order 1.
- The timescale of interest, the lifespan, is of order $\frac{1}{\varepsilon}$.

We take advantage of the timescale difference between the growth and activation processes to apply model reduction techniques in the framework of singular perturbations (slow/fast systems). The mathematical model and theoretical results are presented in the next section.

1.3.2 Slow-Fast model

This section is taken from [A3] and from a joint work in preparation with Guillaume Ballif (Ph.D student, co-supervised with and Frédérique Clément).

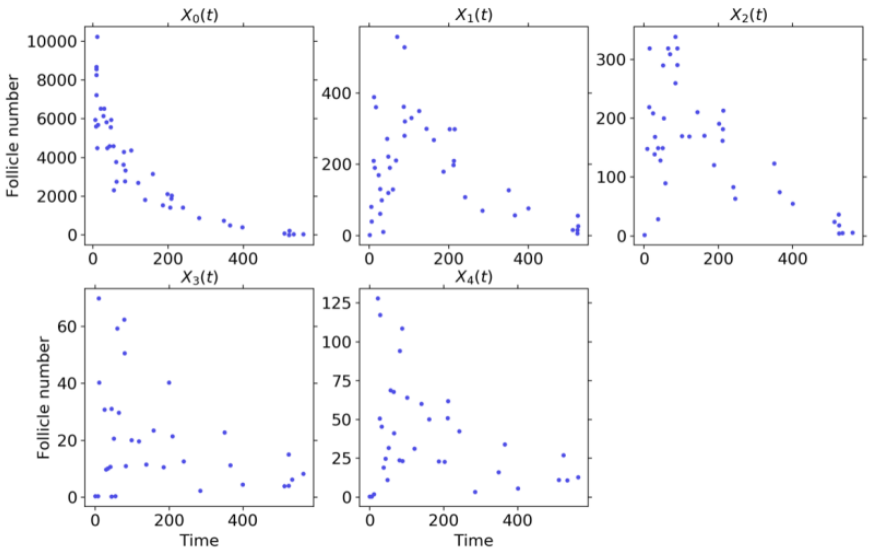

FIGURE 10 – **Population of Follicles in mice according to age (in days).** The data is extracted from [42].

We model through a continuous time Markov chain the population of Follicles structured in $d+1$ compartments. The first compartment represents the number of quiescent ("static reserve") follicles X, and the subsequent d compartments the number of growing ("dynamic reserve") follicles, $Y = (Y_1, Y_2, \cdots Y_d)$. After scaling by a small parameter ε, the transitions of the chain obey the following rules :

$$
\begin{array}{lcl}
 & \text{Follicle events} & \text{Rate} \\
\text{self-renew :} & (X,Y) \to (X+\varepsilon, Y), & \frac{1}{\varepsilon} r_0(X) X\,, \\
\text{atresia :} & (X,Y) \to (X-\varepsilon, Y), & \frac{1}{\varepsilon} \mu_0(Y) X\,, \\
\text{activation :} & (X,Y) \to (X-\varepsilon, Y+e_1), & \frac{1}{\varepsilon} \lambda_0(Y) X\,, \\
\text{growth :} & (X,Y) \to (X, Y+e_{i+1}-e_i), & \frac{1}{\varepsilon} \lambda_i(Y) Y_i\,, i = 1..d-1\,, \\
\text{atresia :} & (X,Y) \to (X, Y-e_i), & \frac{1}{\varepsilon} \mu_i(Y) Y_i\,, i = 1..d\,,
\end{array}
\qquad (\mathcal{M}_{pop})
$$

and where e_i denotes the unit vector on $\mathbb{R}^d$ such that $e_{i,j} = 1_{j=i}$, X is valued in $\varepsilon\mathbb{N}$, and Y is valued in $\mathbb{N}^d$.

A few comments about model ($\mathcal{M}_{pop}$) are in order. The ε-dependency of rates in ($\mathcal{M}_{pop}$) is due to a time scaling to look the process at longer time scale than the individual growth time scale. The jump size of ε for the X compartment reflects its large initial size after abundance scaling. Rate functions $r_0, \mu_0, \lambda_0, \lambda_i, \mu_i$ provide a compact formalism to resume the nonlinear interaction between follicles. The rate r_0 is included to take into account specific capacity of germ cells to self-renew in some species like in fish species. We could think for instance of logistic-like growth rate

$$r_0(x) = \frac{h_0}{c_0 + x}\,.$$

For most mammals species, this self-renew capacity is believed to be null. For the remaining rates, we typically have in mind the following choices, for $i \in \{0, \ldots, d\}$,

$$\lambda_i(y) = m_i + \frac{f_i}{1 + K_{1,i} \sum_{j=1}^{d} \omega^1_{i,j} y_j}, \quad \mu_i(y) = g_i \left(1 + K_{2,i} \sum_{j=1}^{d} \omega^2_{i,j} y_j\right)\,, \qquad (11)$$

with non-negative parameter constants, f_i ($f_d = 0$), $K_{1,i}$, g_i, $K_{2,i}$ and $\omega^k_{i,j} \in [0,1]$. The formulation of the growth and death rates in Eq. (11) is based on the following, biologically-grounded principles. Parameters $m_i + f_i$ and g_i set the "basal levels" of growth or death rates, as they would be ideally observed in standardized

situations where follicle dynamics would be uncoupled from one another. Such a situation is rather well approximated by innovative devices of *in vitro* culture of isolated ovarian follicles. From this basal situation, switching the coupling on would result in lowering the growth rate and increasing the death rate. Parameters $K_{1,i}$ (resp. $K_{2,i}$) tune the sensitivity of the growth rate (resp. death rate) of a given developmental stage i to the other follicles' feedback. Parameters $\omega_{i,j}^k$ are used to weight, or even filter (when their values are set to zero) the contribution of follicles from any developmental stage $j \geq 1$ to the control of the maturation ($k = 1$) and death rates ($k = 2$) of follicles in developmental stage $i \geq 0$. For instance, the maturation rates of the earliest stages are slowed down by AMH, which is secreted by the small growing follicles. On the other end, the survival of the terminally developing follicles is highly sensitive to the sufficient supply of FSH, whose levels are in turn down regulated by hormones secreted by the cohort of terminally developing follicles itself.

The limits $\varepsilon \to 0$ of model ($\mathcal{M}_{pop}$) is intuitively deduced from the following consideration. The first compartment X will make more and more jumps of size ε at rate $\frac{1}{\varepsilon}$. This is the typical scaling for a functional law of large numbers, and X will then be expected to converge to a continuous deterministic function solution of some ordinary differential equation, provided one can prove sufficient regularity and compactness property for X. The growing compartment population Y keeps doing jumps of size "± 1", at an ever and ever increasing rate $\frac{1}{\varepsilon}$. This process will thus be a fast varying process and cannot converge in a standard functional sense. It is tempting to describe the time dependence of Y as a function of t/ε, $Y \approx Y(\frac{t}{\varepsilon})$. Formally letting $\varepsilon \to 0$ leads to $Y(\infty)$. Indeed, providing Y is ergodic for any "frozen" X value, it turns out that Y will converge in some sense towards its X-dependent (and thus time-dependent) steady-state probability distribution. At the limit $\varepsilon \to 0$, the time dependent function Y does not have a proper meaning any more, and any function of Y are rather to be replaced by their averaged value against a limiting time-dependent measure. Probabilistic literature usually refer to averaging theorem, while more applied literature refers to quasi steady-state approximation (QSSA).

We now assume the following hypotheses on functions λ_i, μ_i and r_0 :

- Upper bound of the renewal rate r_0 :

$$\exists R_0 > 0,\ \forall x \in \mathbb{R}^+,\ r_0(x) \leq R_0\,, \tag{$\mathcal{M}_{pop}$-H1}$$

  and regularity :

$$x \mapsto x r_0(x) \text{ is locally Lipschitz continuous on } \mathbb{R}^+\,. \tag{$\mathcal{M}_{pop}$-H2}$$

- Upper bound of the activation rate λ_0 :

$$\exists B_0 > 0,\ \forall y \in \mathbb{N}^d,\ \lambda_0(y) \leq B_0\,, \tag{$\mathcal{M}_{pop}$-H3}$$

- Positivity of activation/maturation rates :

$$\forall y \in \mathbb{N}^d\,, \forall i \in [\![0, d-1]\!]\,, \lambda_i(y) > 0\,. \tag{$\mathcal{M}_{pop}$-H4}$$

- Lower bound of the total transit rates from each compartment

$$\forall i \in [\![0, d-1]\!],\ \exists \alpha_i > 0,\ \forall y \in \mathbb{N}^d,\ \alpha_i \leq \lambda_i(y) + \mu_i(y)\,. \tag{$\mathcal{M}_{pop}$-H5}$$

$$\exists \alpha_d > 0,\ \forall y \in \mathbb{N}^d,\ \alpha_d \leq \mu_d(y)\,. \tag{$\mathcal{M}_{pop}$-H6}$$

Our main result is

Theorem 6. *Under the assumptions ($\mathcal{M}_{pop}$-H4)-($\mathcal{M}_{pop}$-H6), we assume furthermore that*

$$\exists p_0 \in \mathbb{N},\ \exists D_0 \in \mathbb{R}^+ \text{ such that } \forall y \in \mathbb{N}^d,\ \mu_0(y) \leq D_0 \left(1 + \sum_{i=1}^{d} y_i^{p_0}\right)\,. \tag{$\mathcal{M}_{pop}$-H7}$$

Then for a sequence of relatively compact initial condition $X^{\varepsilon,\text{in}} \to x^{\text{in}}$, the process $(X^\varepsilon, Y^\varepsilon)$ is relatively compact in $\mathcal{D}_{\mathbb{N}}[0,\infty[\times \mathcal{L}_m(\mathbb{N}^d)$ and has a unique limit process $(x, \pi) \in \mathcal{C}^1(\mathbb{R}^+) \times \mathcal{L}_m(\mathbb{N}^d)$ such that :

$$\left\{ \begin{array}{rcl} \frac{dx}{dt}(t) & = & \Big[r_0\big(x(t)\big) - \Lambda_0\big(x(t)\big)\Big] x(t)\,, \qquad x(0) = x^{\text{in}}\,, \\ \Lambda_0\big(x(t)\big) & = & \displaystyle\sum_{y \in \mathbb{N}^d} \Big(\lambda_0(y) + \mu_0(y)\Big) \pi_{x(t)}(y)\,. \end{array} \right. \tag{12}$$

with for $x \in \mathbb{R}^+$, π_x such that for all ψ bounded on $\mathbb{N}^d$,

$$\sum_{y\in\mathbb{N}^d} L_x\psi(y)\pi_x(y) = 0\,, \tag{13}$$

where

$$\forall y \in \mathbb{N}^d,\ L_x\psi(y) = \lambda_0(y)x\Big[\psi(y+e_1)-\psi(y)\Big] + \sum_{i=1}^{d-1}\lambda_i(y)y_i\Big[\psi(y+e_{i+1}-e_i)-\psi(y)\Big]$$
$$+\sum_{i=1}^{d}\mu_i(y)y_i\Big[\psi(y-e_i)-\psi(y)\Big]\,. \tag{14}$$

The proof of Theorem 6 is based on the following averaging methodology from [65, 66]. We first identify a limit candidate. To that, we start from the generator associated to the chain ($\mathcal{M}_{pop}$), given by $\forall (x,y) \in \varepsilon\mathbb{N}\times\mathbb{N}^d$:

$$L^\varepsilon\psi(x,y) = \frac{1}{\varepsilon}\Bigg(r_0(x)x\Big[\psi(x+\varepsilon,y)-\psi(x,y)\Big] + \mu_0(y)x\Big[\psi(x-\varepsilon,y)-\psi(x,y)\Big]$$
$$+\lambda_0(y)x\Big[\psi(x+\varepsilon,y+e_1)-\psi(x,y)\Big] + \sum_{i=1}^{d-1}\lambda_i(y)y_i\Big[\psi(x,y+e_{i+1}-e_i)-\psi(x,y)\Big]$$
$$+\sum_{i=1}^{d}\mu_i(y)y_i\Big[\psi(x,y-e_i)-\psi(x,y)\Big]\Bigg)\,. \tag{15}$$

We define the occupation measure π^ε associated to the fast process Y^ε,

$$\forall K \subset \mathbb{N}^d,\ \pi^\varepsilon([0,t]\times K) = \int_0^t \mathbf{1}_K(Y^\varepsilon(s))ds\,.$$

Then, we split the generator (15) in two pieces for each variable X^ε and π^ε. On one hand, for test functions that depends only of x, the generator (15) is close to

$$L^\varepsilon\psi(x,y) \approx \mathcal{A}\psi(x,y) := x\left(r_0(x)-\mu_0(y)-\lambda_0(x)\right)\psi'(x)\,,$$

so that we deduce

$$\psi(X^\varepsilon(t)) - \psi(X^\varepsilon(0)) - \int_{[0,t]\times\mathbb{N}^d} A\psi(X^\varepsilon(s),y)\pi^\varepsilon(ds\times dy) + R^\varepsilon_\psi(t) \tag{16}$$

is a martingale, where $R^\varepsilon_\psi(t)$ can be shown to converge to 0, almost surely. On the other hand, for test function depending only of y, we have that

$$\varepsilon\left[\psi(Y^\varepsilon(t)) - \psi(Y^\varepsilon(0))\right] - \int_{[0,t]\times\mathbb{N}^d} L_{X^\varepsilon(s)}\psi(y)\pi^\varepsilon(ds\times dy) \tag{17}$$

is a martingale. Finally, with appropriate compactness criteria, we may extract a converging sub-sequence of $(X^\varepsilon,\pi^\varepsilon)$ and pass to the limit in martingales (16)-(17) to recover Eqs. (12)-(13)-(14).

The uniqueness part requires to prove that a single quasi steady-state distribution exits (namely a single solution π_x of (13) for each x), together some moment control and x-dependent regularity, which is simplified here by the specific x-dependence of the generator (14) of the fast process.

Hypothesis ($\mathcal{M}_{pop}$-H1) is used to obtain straightforward compactness property for X^ε, and ($\mathcal{M}_{pop}$-H2) is a standard assumption for uniqueness of the limit model. Hypotheses ($\mathcal{M}_{pop}$-H3) with ($\mathcal{M}_{pop}$-H5)-($\mathcal{M}_{pop}$-H6) ensure compactness of the fast process Y^ε while ($\mathcal{M}_{pop}$-H4) ensure its irreductibility. The lower bounds ($\mathcal{M}_{pop}$-H5)-($\mathcal{M}_{pop}$-H6) also give (non-optimal) criteria that guarantee exponential ergodicity of the fast process, which is required to prove uniqueness of solutions of the limit model. The last hypothesis ($\mathcal{M}_{pop}$-H7), although not very restrictive, is purely technical to ensure integrability of π_x against μ_0, required to define Λ_0 in the limit model (12).

1.3.3 Perspectives

The limit model obtained in Theorem 6 is a first step to analyze the effect of nonlinear interactions between follicles on key output of folliculogenesis, such as the decay rate of the follicle pool, its robustness to genetic or external perturbations, and so on. The limit model is also adequate for data fitting of follicle count data such as those shown in Fig 10 (away from time origin, in order for the fast process to get time to reach the QSSA). However, this requires to be able either to solve analytically this model (see below), or to simulate it numerically in an efficient way. This is not completely trivial, and two strategies are currently in evaluation. A first one is based on a direct solution of (12) with a truncation argument (like in section 1.2.2) to find π_x and compute the right hand-side of (12). This strategy appears to be restricted with respect to the dimension of the state-space, and thus the number of compartments. A second strategy tries instead to simulate sample path of ($\mathcal{M}_{pop}$) with a cost that is independent of ε. The X^ε process is simulated using the τ-leap method (fixed time-step simulation), and the fast process Y^ε is replaced by an auxiliary "nested" chain that approximates the quasi steady-state distribution.

We also note that the solution of (13) can be computed analytically when the rate functions are linear with respect to x, y (constant λ_i, μ_i), see [A3]. The solution is a product of Poisson laws, with compartment-dependent parameter given by $\lambda_i/(\mu_{i+1} + \lambda_{i+1})$. This particular case shows that the full set of parameters is not identifiable with data like in Fig 10, where we cannot distinguish the growth rate from the death rate. Further data/information is thus needed, like the number of atretic follicles (which could potentially be accessed with specific bio-markers), or transit time between each compartment (which have been estimated using various methods, including data fitting of similar models, see table 3).

1.4 Perspectives on combining individual morphological models with population models

In this last chapter on modeling the ovarian folliculogenesis, we discuss how to combine both approaches presented in chapter 1.2 and 1.3, namely individual-scale morphological models with population-level models.

To that, we come back to a comment on the likelihood used for calibrating the activation model, see remark 1. The biological data of chapter 1.2 provides snapshot on possible morphological characteristics of follicles, namely number of cells, diameter of oocyte, antrum, and so on. We have used a model of the growth of a single follicle with the requirement that its possible trajectories resemble as closely as possible to the observed follicles in the snapshot data. Doing so, we are not able to capture neither absolute nor relative kinetics requirements.

In contrast, the quasi steady-state distribution π_x, obtained in theorem 6 from chapter 1.3 contains dynamical information that are build in its infinitesimal generator : the relative proportion of follicle in each compartment is dependent on the growth rates λ_i's (and death rates μ_i's).

The snapshot morphological data are certainly more satisfactory interpreted as (a sampling of) *a quasi steady state distribution of a dynamical morphological process*. Furthermore, this point of view gives an unified way to gather our different models on a single modeling framework to decipher the various hidden interactions and kinetic processes at play within the ovarian folliculogenesis.

To be more specific, the structuring d fast compartment of model ($\mathcal{M}_{pop}$) can be replaced by the cell content $(f, c) \in \mathbb{N}^2$ of follicles of model ($\mathcal{M}_{FC}$). Thus, a possible inclusion of model ($\mathcal{M}_{FC}$) within the multiscale population framework of model ($\mathcal{M}_{pop}$), leads to the following model, by analogy with the limit model (12),

$$\left\{ \begin{array}{rcll} \dfrac{dx}{dt}(t) & = & -\Lambda_0(x(t))x(t)\,, & x(0) = x^{\text{in}}\,, \\ \Lambda_0(x(t)) & = & \displaystyle\sum_{y \in \ell_1(\mathbb{N}^2,\mathbb{N})} (\lambda_0(y) + \mu_0(y))\,\pi_{x(t)}(y)\,, & \end{array} \right. \tag{18}$$

with for $x \in \mathbb{R}^+$, π_x is a probability distribution on the set of summable integer-valued sequences indexed by $\mathbb{N}^2$,

$$\ell_1(\mathbb{N}^2, \mathbb{N}) := \left\{ y = (y_{f,c})_{f,c\in\mathbb{N}^2}\,, y_{f,c} \in \mathbb{N}\,, \sum_{(f,c)\in\mathbb{N}^2} y_{f,c} < \infty \right\}\,,$$

such that for all ψ with finite support on $\ell_1(\mathbb{N}^2, \mathbb{N})$,

$$\sum_{y \in \ell_1(\mathbb{N}^2,\mathbb{N})} L_x \psi(y) \pi_x(y) = 0\,, \tag{19}$$

where for any $y \in \ell_1(\mathbb{N}^2, \mathbb{N})$,

$$\begin{aligned} L_x\psi(y) = \lambda(y)x \sum_{f\in\mathbb{N}^*} \Big[\psi(y + e_{f,0}) - \psi(y)\Big] \mathbb{P}\left[F_0 = f\right] \\ + \sum_{(f,c)\in\mathbb{N}^2} \left(\alpha_2(y)f + \gamma(y)c\right) y_{f,c} \Big[\psi(y + e_{f,c+1} - e_{f,c}) - \psi(y)\Big] \\ + \sum_{(f,c)\in\mathbb{N}^2} \left(\alpha_1(y)f + \beta(y)fc/(f+c)\right) y_{f,c} \Big[\psi(y + e_{f-1,c+1} - e_{f,c}) - \psi(y)\Big] \\ + \sum_{(f,c)\in\mathbb{N}^2} \mu(y,f,c) y_{f,c} \Big[\psi(y - e_{f,c}) - \psi(y)\Big]\,. \end{aligned} \tag{20}$$

In Eqs. (18)-(19)-(20), x is the (macroscopic) number of quiescent follicles and $y = ((y_{f,c})_{f,c\in\mathbb{N}^2})$ represents the population of growing follicles, with $y_{f,c}$ the number of follicles with given flattened cells f and cuboidal cells c. The measure π_x is a probability distribution on those follicles. The parameter λ is the rate of activation of new growing follicles, μ_0 (resp. μ) are death rate of quiescent (resp. growing) follicles, and $\alpha_1, \alpha_2, \beta, \gamma$ are a population-dependent version of the cell kinetic rates of model ($\mathcal{M}_{FC}$). A perspective would be to use $\pi_{x(t)}(y)$ as a building block of a likelihood to observe a snapshot y of follicles with given content of flattened and cuboidal cells, in an individual of age t with (known or hidden) $x(t)$ quiescent follicles. We note that the mathematical description of the model is slightly more involved, however its interpretation and (importantly) its parametrization has not drastically changed. The validity of Eqs. (18)-(19)-(20) essentially depends on the exponential ergodicity of the fast process Eqs. (20). Requirement like $\mu > \gamma$ should be sufficient to ensure that the number of cells c in a growing follicle stays finite (in mean) before atresia occurs (see similarity with hypothesis $\mathcal{M}_{FC}$-H1). It is clear that further individual morphological models can be "added" within (20). Model extensions should hence consider to handle more structuring variable, in order to better represent growing follicles with respect to several important bio-markers (size, number of cells, maturity etc...).

Relatively few modeling works have been attempted in this field, yet general trends strike out from this studies : linear models are predominant, and parameter identifiability is often overlooked. Nonlinear multiscale models like Eqs. (18)-(19)-(20), close to biological data, are thus expected to be valuable in the reproductive biology field.

2 Becker-Döring model

Natural mathematical formalism to model the kinetics of polymer formation, or more generally cluster formation, are coagulation-fragmentation processes which describe the dynamics of association and dissociation of particles in clusters. Among them, the Becker-Döring (BD) model is perhaps the simplest kinetic model describing a number of properties of the dynamics of phase transitions, such as **metastability, nucleation and coarsening** [99]. Those three notions are very large. Roughly speaking, in the context of the BD model, they mean the following.

- **Metastability** refers to the existence of a particular point in the phase space which is not stationary for the dynamical system, yet has the following property : it attracts some solutions that stay arbitrary close the metastable state for a very large time, before converging towards the true steady-state.
- **Nucleation** refers to the formation of the most thermodynamically unstable cluster, the nucleus. Clusters bigger than the nucleus tend to growth, while clusters shorter than the nucleus tend to shrink. The energy barrier is related to the metastability time.
- **Coarsening** refers to two-phase systems, in which the interface between the two phases progressively reduces due to energy minimization constraints : the pattern formed by the two phases grows ("coarsens") with time.

The BD model goes back to the seminal work *"Kinetic treatment of nucleation in supersaturated vapors"* by [25]. Since then, it has found many different applications ranging from physics to biology. From the mathematical point of view, this model has received much more attention in the deterministic literature than the probabilistic one. We refer to our review [A10] for historical comments and detailed literature review on theoretical results from the deterministic side.

The model was initially designed to explain critical phase condensation phenomena where macroscopic droplets self-assemble and segregate from an initially supersaturated homogeneous mixture of particles, at a rate that is exponentially small in the excess of particles. This led to important applications in kinetic nucleation theory [98]. Mathematical studies in the 90's showed that (in the deterministic context), departing from certain initial conditions, the size distribution of clusters quickly reaches a metastable configuration composed of "small" clusters, and remains arbitrary close to that state for a very large time, before it converges to the true stationary solution that leads to "infinitely large" clusters (interpreted as droplets) [60, 88]. This notion will be investigated in a probabilistic context in chapter 2.3.

The BD model consists in describing the cluster distribution according to their size $i \geq 1$, *i.e.* the number of particles that composed them. Clusters belong to a "solvent" in much smaller proportion and are assumed to be spatially homogeneously distributed. Along their motion, clusters give rise to two types of reactions, namely the *Becker-Döring rules* :

1. A cluster of size 1, commonly called *monomer* or *elementary particle*, may encounter a cluster of size $i \geq 1$ to coalesce and give rise to a cluster of size $i+1$.
2. A cluster of size $i \geq 2$ may release spontaneously a monomer resulting in a cluster of size $i-1$.

These can be summarized by the set of kinetic reactions, for each $i \geq 1$,

$$C_1 + C_i \underset{b_{i+1}}{\overset{a_i}{\rightleftharpoons}} C_{i+1}\,, \tag{21}$$

where C_i denotes clusters consisting of i particles. Coefficients a_i and b_{i+1} stand, respectively, for the rate of lengthening and shortening. These may depend on the size of clusters involved in the reactions and typical coefficients are derived from physical principles by [89] and [83] :

$$a_i = i^\alpha\,, \quad b_{i+1} = a_{i+1}\left(z_s + \frac{q}{(i+1)^\gamma}\right)\,, \qquad i \geq 1\,.$$

for $0 \leq \alpha < 1$, $z_s > 0$, $q > 0$ and $0 < \gamma < 1$. This choice is in agreement with original derivation where $a_i \approx i^{2/3}$, $b_i \approx a_i \exp(i^{-1/3})$. In particular, the diffusion-limited case of monomers clustering into sphere corresponds to

- $\alpha = 0$, $\gamma = 1/2$ in 2D
- $\alpha = 1/3$, $\gamma = 1/3$ in 3D

while the interface-reaction-limited case corresponds to

— $\alpha = 1/2$, $\gamma = 1/2$ in 2D.
— $\alpha = 2/3$, $\gamma = 1/3$ in 3D

In its mean-field (deterministic) version, the BD model is an infinite set of ordinary differential equations for the time evolution of the concentration of clusters structured according to their size. In its stochastic version, the BD model is a continuous time Markov chain, on a finite state space. We detail below the necessary material for introducing our results, and refer to our review [A10] for more details.

2.1 Introduction

2.1.1 Deterministic Becker-Döring model (BD)

The general formulation of the deterministic BD equations, as studied today, seems to go back to [27] and was popularized among mathematicians by [90]. It assumes the system behaves homogeneously in space with a high number of clusters, and considers concentrations $c_i(t)$ of clusters with size $i \geq 1$ at time $t \geq 0$. Assuming classical Law of Mass Action, the lengthening is considered as a second order reaction while the shortening is a first-order (linear) reaction. The flux associated to the kinetic scheme (21) is thus given, for each $i \geq 1$, by

$$J_i = a_i c_1 c_i - b_{i+1} c_{i+1} \,. \tag{22}$$

Considering all the fluxes involved in the variation of the concentration of each c_i entails the infinite system of differential equations, namely the BD equations :

$$\begin{aligned} \frac{d}{dt} c_1 &= -J_1 - \sum_{i \geq 1} J_i \,, \\ \frac{d}{dt} c_i &= J_{i-1} - J_i \,, \end{aligned} \tag{23}$$

for every $i \geq 2$. The system considered here has no source nor sink. Consequently, for the total amount of particles, we should have, for all $t \geq 0$,

$$\sum_{i \geq 1} i c_i(t) = \rho \,. \tag{24}$$

The infinite sums in Eqs. (23)-(24) are at the heart of some mathematical difficulties. Therefore, a natural functional space to study the BD model is

$$X^+ = \{ c \in (\mathbb{R}^+)^{\mathbb{N}} \,:\, \|c\| := \sum_{i=1}^{\infty} i c_i < \infty \} \,. \tag{25}$$

The mathematical foundations of the BD equations have been laid down by [24]. Any solution of the BD equations defined by [24] satisfies the balance of mass given by (24) at all times, and avoids the so-called gelation phenomenon *in finite time* which can occur in general coagulation-fragmentation equations. Essential, a linear bound on the lengthening rate a is needed to control the infinite sum $\sum a_i c_i$ in Eq. (23) by the mass $\sum_{i \geq 1} i c_i$. Up to my knowledge, the weakest condition for well-posedness needs to suppose the following growth condition on rate a, b (see [68]) :

$$a_i - a_{i-1} \leq K \,, \quad b_i - b_{i+1} \leq K \,, \tag{BD-H1}$$

where the second condition on b is purely technical and is needed to obtain uniqueness of solutions.

The most interesting behavior of the BD equations is their long-time behavior. It is not difficult to see that any steady state solutions of (23) should satisfy the detailed balance condition $J_i = 0$, for all $i \geq 1$, which leads to

$$c_i = Q_i z^i \,, \quad \text{where} \quad Q_i = \frac{a_1 a_2 \cdots a_{i-1}}{b_2 b_3 \cdots b_i} \,, \tag{26}$$

with $z > 0$ a free parameter. Due the conservation of mass in Eq. (24), it is natural to wonder if one (or several) candidate equilibrium belongs to the stoichiometric subspace of X^+, given by

$$X^+_\rho = \{ c \in X^+ \,:\, \|c\| := \sum_{i=1}^{\infty} i c_i = \rho \} \,.$$

To that, we define the following critical quantities

$$z_s := \left(\limsup_{i\to\infty} Q_i^{1/i} \right)^{-1}, \quad \text{and} \quad \rho_s := \sup_{\{z<z_s\}} \sum_{i\geq 1} iQ_i z^i.$$

The long-time behavior of the BD solution is shown to obey the following dichotomy (under some technical conditions) :

1) if the initial condition has a sub-critical density $\rho \leq \rho_s$, then its associated solution converges (in strong norm on X^+) as time goes to infinity towards the unique equilibrium of BD equations of same density ρ ;
2) if the initial condition has a super-critical density $\rho > \rho_s$, then its associated solution converges (in a weak sense) as time goes to infinity towards the unique equilibrium of strictly inferior density ρ_s. The difference $\rho - \rho_s$ is linked to the formation of particles with infinite size that are interpreted as a different phase, leading to a *phase transition*.

These long time results are proved with an analogous of a H-theorem with (relative) entropy given by

$$\mathcal{H}_z(c) = \sum_{i=1}^{+\infty} \left\{ c_i \left(\ln \frac{c_i}{Q_i z^i} - 1 \right) + Q_i z^i \right\}. \tag{27}$$

However, the super-critical case is not completely understood, and no convergence rate towards equilibrium is known, in contrast to the sub-critical case which is known to be exponentially ergodic [31]. Using time-independent upper-bound on solutions of BD equations, the existence of a metastable state has been proved in the limit of vanishing excess of density $\rho - \rho_s$ [88, 60, 40]. This metastable state has been related to extremely small nucleation rates in classical nucleation theory [98]. Our work in chapter 2.3 (section 2.3.1) revisits some of those results for the stochastic (linear) Becker-Döring model.

Further, the large cluster dynamics in the Becker-Döring model have been related to solutions of various version of Lifshitz-Slyozov equations (LS) [89, 84, 102] (see section 2.1.3 below), for which some coarsening and self-similarity properties are known, and related to the theory of Ostwald ripening *(roughly speaking : the large clusters grow at the expense of the small ones, which is the case for instance for the oil droplets in pastis mixed with water – the ouzo effect)*. Our work in chapter 2.2 (section 2.2.2) adds a connection between the BD and LS models, in an intermediate regime where nucleation and Ostwald ripening take place at comparable rates.

2.1.2 Stochastic Becker-Döring model (SBD)

A stochastic version of the Becker-Döring model may be defined as a continuous time Markov chain analog of the set of ordinary differential equations (23), for which transition are given by the same set of kinetic reactions (21), but modeling *discrete numbers* of clusters instead of continuous concentrations. Precisely, given a positive integer n, we define

$$X_n := \left\{ C = (C_i)_{i\geq 1} \in \mathbb{N}^{\mathbb{N}} \, : \, \sum_{i=1}^{n} iC_i = n \right\},$$

and the state space

$$X_\infty = \bigcup_{n\in\mathbb{N}^*} X_n = l_1(\mathbb{N})$$

of all summable integer-valued sequences. On X_∞, we introduced the following operators defined by, for any configuration C on X_∞,

$$\begin{aligned} R_1^+ C &= (C_1 - 2\,, C_2 + 1\,, \cdots, C_i\,, \cdots)\,, \\ R_2^- C &= (C_1 + 2\,, C_2 - 1\,, \cdots, C_i,\,, \cdots)\,, \end{aligned}$$

and, for any $i \geq 2$,

$$\begin{aligned} R_i^+ C &= (C_1 - 1\,, C_2\,, \cdots, C_i - 1\,, C_{i+1} + 1\,, \cdots)\,, \\ R_{i+1}^- C &= (C_1 + 1\,, C_2\,, \cdots, C_i + 1\,, C_{i+1} - 1\,, \cdots)\,. \end{aligned}$$

Given non-negative kinetic rates $(a_i)_{i\geq 1}$, $(b_i)_{i\geq 2}$, the stochastic Becker-Döring model (SBD) is defined as the continuous time Markov chain on X_∞ with transition rates

$$\left\{\begin{array}{lcll} q(C, R_1^+ C) & = & a_1 C_1 (C_1 - 1)\,, & \\ q(C, R_i^+ C) & = & a_i C_1 C_i\,, & i \geq 2\,, \\ q(C, R_i^- C) & = & b_i C_i\,, & i \geq 2\,. \end{array}\right.$$

which clearly leaves invariant each finite subspace X_n. We consider a given fixed $n \in \mathbb{N}^*$ for now. For an initial configuration $C^{\rm in} \in X_n$ (deterministic or random), the configuration $C(t)$ defined by the SBD may alternatively be represented as the solution of the following system of stochastic equations

$$\left\{\begin{array}{lcll} C_1(t) & = & C_1^{\rm in} - 2J_1(t) - \displaystyle\sum_{i\geq 2} J_i(t)\,, & \\ C_i(t) & = & C_i^{\rm in} + J_{i-1}(t) - J_i(t)\,, & i \geq 2\,, \end{array}\right. \tag{28}$$

with

$$J_i(t) = Y_i^+\Big(\int_0^t a_i C_1(s)(C_i(s) - \delta_{1,i})ds\Big) - Y_{i+1}^-\Big(\int_0^t b_{i+1} C_{i+1}(s) ds\Big)\,, \quad i \geq 1\,,$$

where $\delta_{1,i} = 1$ if $i = 1$ and $\delta_{1,i} = 0$ if $i > 1$ and Y_i^+, Y_{i+1}^- for $i \geq 1$ are independent standard Poisson processes. Analogy between Eq. (28) and Eq. (23) is clear. The number of clusters of size $i \geq 2$ evolves according to the differences between two (stochastic) cumulative counts J_{i-1} and J_i. Finally, we may also identified the SBD with the help of its infinitesimal generator L, defined by, for any bounded functions f on X_n,

$$Lf(C) = \sum_{i=1}^{n-1} \big[f(R_i^+ C) - f(C)\big]\, a_i C_1 (C_i - \delta_{1,i}) + \big[f(R_{i+1}^- C) - f(C)\big]\, b_{i+1} C_{i+1}\,. \tag{29}$$

Although the well-posedness of the SBD model is of course standard on each X_n (as a pure-jump Markov process on a finite state-space), a first non trivial question arises with respect to the precise description of the state space, and in particular to its cardinality. In fact, the state space X_n is given by all possible partitions of the integer n, a well-known problem in combinatorics [46, chap I.3]. In particular, we have the asymptotic as $n \to \infty$,

$$|\, X_n \,| \propto \frac{1}{4n\sqrt{3}} \exp\left(\pi\sqrt{\frac{2n}{3}}\right)\,.$$

Perhaps surprisingly, the stationary solution associated to Eq. (29) has a relatively simple form, namely a product-form [23]. Indeed, the (unique) stationary probability Π on X_n of Eq. (29) is given by [59, Theorem 8.1]

$$\Pi(C) = B_n \prod_{i=1}^n \frac{(Q_i)^{C_i}}{C_i!}\,, \tag{30}$$

where B_n is a normalizing constant and Q_i is defined by Eq. (26). One may verify simply that the following detailed balance condition holds [59, Theorem 1.2]

$$\Pi(C) q(C, R_i^+ C) = \Pi(R_i^+ C) q(R_i^+ C, C)\,.$$

Note also that, for all $z > 0$, with $B_z := B_n / z^n$, the expression (30) may be rewritten $\Pi(C) = B_z \prod_{i=1}^n \frac{(Q_i z^i)^{C_i}}{C_i!}$, which has a clearer analogy with the deterministic equilibrium of the BD equation. Finally, the distribution Π has the following probabilistic meaning : let Z_i, $i = 1, \cdots, n$, be independent Poisson random variables with respective means Q_i, then it is easily seen that, for all $C \in X_n$,

$$\Pi(C) = \mathbf{P}\left\{Z_1 = C_1, \cdots, Z_n = C_n \mid \sum_{i=1}^n i Z_i = n\right\}\,.$$

In section 2.2.1, we study the (finite-time) functional law of large numbers that allows to make the rigorous link between Eq. (28) and Eq. (23). In section 2.3.1, we study metastability properties for a linear version the SBD model, when c_1 is a fixed parameter. Finally, we come back to the study of some properties of the equilibrium distribution in chapter 2.3.2, related to the formation of "infinitely large" clusters and the phase transition in BD model.

2.1.3 Lifshitz-Slyozov model (LS)

The Lifshitz–Slyozov (LS) system [71] describes the temporal evolution of a mixture of monomers and aggregates, where individual monomers can attach to or detach from already existing aggregates, the later being described by a *continuous size variable*. The aggregate distribution follows a transport equation with respect to that size variable, and the transport rates are coupled to the dynamics of monomers through a mass conservation relation. The initial-boundary value problem for the Lifshitz–Slyozov model reads

$$\begin{cases} \dfrac{\partial f(t,x)}{\partial t} + \dfrac{\partial[(a(x)u(t) - b(x))f(t,x)]}{\partial x} = 0\,, \quad t > 0\,,\ x \in (0,\infty)\,, \\ u(t) + \displaystyle\int_0^\infty x f(t,x)\,dx = \rho\,, \quad t > 0\,, \end{cases} \tag{31}$$

for some given $\rho > 0$, subject to the initial condition

$$f(0,x) = f^{in}(x)\,, \qquad x \in (0,\infty)\,, \tag{32}$$

and the boundary condition

$$\lim_{x\to 0^+} (a(x)u(t) - b(x))f(t,x) = \mathfrak{n}(u(t))\,, \tag{33}$$

at any times $t > 0$ where $u(t) > \lim_{x\to 0^+} \frac{b(x)}{a(x)}$. Here $f(t,x)$ is a non-negative distribution of aggregates according to their size x and time t, $u(t)$ is the monomer concentration and ρ is interpreted as the total mass of the system. The kinetic rates $a(x)$ and $b(x)$ determine how fast do attachment (a given monomer attaches to a given aggregate) and detachment (a monomer detaches from a given aggregate) reactions take place. Aggregates change their size over time according to the quantity of monomers that they gain or lose through the previous reactions. Note that the attachment process is a second order kinetics whereas detachment is a first order kinetics, as reflected in the transport term in (31).

The Lifshitz–Slyozov model has been traditionally used to describe late stages of phase transitions, where Ostwald ripening phenomena take place : large aggregates grow larger at the expense of smaller ones, in which case the flux $J(x,t) := (a(x)u(t) - b(x))$ is negative for small x near 0 and the boundary condition (33) is not needed. Classical Lifshitz–Slyozov rates are indeed given by $a(x) = x^{1/3}$ and $b(x) = 1$, see e.g. [85]. The BD model is rather used to describe the initial stage of phase transition, where the nucleation process is the dominant one. Recently, the intermediate stage has been considered in the physics literature [21], where the growth of large aggregates and the ongoing nucleation rate are of equal importance, leading to equations like (31)–(33) or variants of it. Indeed, some sets of kinetic rates for Eq. (31) may lead to Ostwald ripening phenomena only after a certain transient period, where the dynamics of the Lifshitz–Slyozov model are driven by boundary effects at very small sizes, and for which the boundary term (33) becomes important. Moreover, recent applications of this framework in biologically oriented contexts make use of different sets of kinetic rates for which a boundary condition becomes mandatory in order to make sense of the model. A growing literature can be found on applications to protein polymerization phenomena and neurodegenerative diseases, starting from the so-called prion model and some of its variants (see e.g. [96] and references therein), whose different versions come as modifications of the standard Lifshitz–Slyozov equations. Inflow boundary conditions are there used to describe nucleation processes. Our work in section 2.2.2 shows that a suitable scaling of the discrete BD models lead to this scenario, where nucleation and Ostwald ripening take place at rates of equal order of magnitude. We also study in section 2.2.2 the well-posedness of Eq. (31)–(33).

2.2 Links between SBD/BD/LS models

In this chapter, we present two sections making the link between time-dependent solutions of the models SBD, BD and LS. Those results are quiet technical and have required adaptations to adequate functional framework in infinite dimensional space of somehow standard limit theorem strategies. A first important step, closer to a modeling step, requires re-scaling of parameters/state space to define a sequence of solutions indexed by a small parameter that tends to 0. Two main strategies are then developed to obtain the limit of that sequence. A generic one is based on compactness (or tightness) arguments in suitable topology, and identification of the limit. The second one is more model-dependent and aims to prove a contraction

property, that often relies on an underlying Lipschitz property of the dynamical system which imposes a strong constraint. For BD/LS models, monotonicity arguments and maximum principles, systematically developed by [69], allows to overcome this drawback.

2.2.1 SBD->BD

The link between the SBD process and the BD equations is expected as $n \to \infty$, up to an appropriate re-scaling procedure. Here, n stands for a total number of particles in Eq. (28), while ρ stands for a particle density, like in Eq. (23)-(24). It is thus natural to consider a volume parameter V such that

$$V = \frac{n}{\rho} \to \infty \,.$$

For $n \geq 1$ and $\rho > 0$, we define the state space

$$X_\rho^n = \left\{ (c_i)_{i\geq 1} \in \mathbb{R}^{\mathbb{N}} \ : \ \forall i \geq 1 \,, \tfrac{n}{\rho} c_i \in \mathbb{N}, \ \sum_{i=1}^{+\infty} i c_i = \rho \right\}.$$

An important fact is that X_ρ^n is a finite state space that can be embedded into X^+ (defined in Eq. (25)). On a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, we define the SBD process as the pure jump Markov process with value in X_ρ^n, and having infinitesimal generator L^n given by, for all Borel function $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ and finite on X_ρ^n,

$$L^n f(c) = \frac{n}{\rho} \sum_{i=1}^{+\infty} \left(a_i c_i (c_1 - \delta_{i1} \tfrac{\rho}{n}) [f(c + \tfrac{\rho}{n} \Delta_i) - f(c)] + b_{i+1} c_{i+1} [f(c - \tfrac{\rho}{n} \Delta_i) - f(c)] \right), \tag{34}$$

where $\Delta_i = e_{i+1} - e_i - e_1$ with $(e_1, e_2, \ldots)$ the canonical basis of $\mathbb{R}^{\mathbb{N}}$, that is $e_{ik} = 1$ if $k = i$ and 0 otherwise. This process corresponds to a rescaled version of the original SBD with n particles defined by the generator in Eq. (29). The parameter $\frac{n}{\rho}$ can be seen as a volume scaling parameter, ρ being the total concentration, and the rescaled SBD process with generator (34) has the form of a *classical scaling* of a reaction network model in large volume, see *e.g.* [22]. As a finite state-space continuous time Markov chain, given an initial law $c^{\mathrm{in},\,n} \in X_\rho^n$, there exists a unique (in law) SBD process c^n with $c^n(0) = c^{\mathrm{in},\,n}$. By construction of X_ρ^n, this yields the mass conservation,

$$\sum_{i=1}^{+\infty} i c_i^n(t) = \rho \,.$$

We are ready to state the pathwise convergence of the SBD to the BD equations in the next theorem.

Theorem 7. *Under hypothesis* (BD-H1) *on the rate constants, let a sequence* $\{c^{\mathrm{in},\,\mathrm{n}}\}$ *in* X_ρ^n *being deterministic and strongly converging in* $(X^+, \|\cdot\|)$ *toward* c^{in}, *namely* $\lim_{n\to+\infty} \|c^{\mathrm{in},\,\mathrm{n}} - c^{\mathrm{in}}\| = 0$. *If* $\{c^n\}$ *is the sequence of SBD processes with* $c^n(0) = c^{\mathrm{in},\,n}$ *and* c *the unique solution of the BD equations satisfying* $c(0) = c^{\mathrm{in}}$ *then, for all* $T > 0$, *we have*

$$\lim_{n\to+\infty} \sup_{t\in[0,T]} \|c^n(t) - c(t)\| = 0 \,, a.s.$$

Functional law of large number is somehow classical for reaction network model [63]. Nevertheless, within coagulation-fragmentation models, and in particular for BD model, most of the results, e.g. [57], are obtained from tightness arguments. Here, we obtained a proof of pathwise convergence of the SBD to BD model in the natural space X^+ associated to mass conservation, rather than in some Hilbert space as in [57]. Furthermore, our proof differs from classical results, see [63, Theorem 2.11] and [64, Theorem 2.2], in the sense that this model lacks of a Lipschitz property. In [100], the authors also proved a pathwise law of large numbers for the SBD process, but in the case of bounded rates (which makes the BD model having a Lipschitz property). In Theorem 7, the class of kinetic rates allowed is more general and naturally adapted with up-to-date results on the BD model. Here, we take advantage of monotonicity properties within the BD model (linked to hypothesis (BD-H1) on the rate constants), together with fine control of large-sized clusters, taking inspiration from deterministic results in [68].

In particular, besides standard moment estimates, the few key ideas are :

— Control a superlinear moment like $\sum i^{1+\alpha} c_i$ for $\alpha > 0$ in order to control the mass of large clusters $\sum_{i \geq N} i c_i$. Any good control will essentially allow to deal with finite sums instead of infinite ones, easier to handle than *ad-hoc* limit-sums inversion. The "magic" of functional analysis here is that we do not even have to hypothesize finite superlinear moment for the initial condition. A consequence of the Dunford-Pettis theorem, known as the de la Vallée Poussin theorem [69], guarantees the existence of a superlinear moment for a sequence of uniformly integrable functions in X^+.

— Trying to compute directly the difference between c_i^n and c_i is difficult, and it turns way easier to estimate

$$\left\| \sum_{j=i}^{\infty} c_j^n(t) - c_j(t) \right\| .$$

Deep reasons for that is the fact that the differential equation on the tail function $F_i := \sum_{j=i}^{\infty} c_j$ satisfy "almost" a maximum principle, which allows much finer estimates. Indeed, a straightforward computation gives, for $i \geq 2$,

$$\frac{dF_i}{dt} = a_{i-1} c_1(t) F_{i-1}(t) - (a_{i-1} c_1(t) + b_i) F_i(t) + b_i F_{i+1}(t) ,$$

which, ignoring the variable $c_1(t)$, looks like an infinite tridiagonal linear systems with non-negative off-diagonal entries (Metzler matrix). It turns out that any control on c_1 can be lifted to the tail variables F_i if they are ordered for two different initial conditions, that is $F_i(0) \leq F_i'(0)$ for each $i \geq 2$. The later is too much to ask at the level of individual c_i's for two solutions of same density. This observation was at the heart of the improvement on well-posedness for BD equations in [68], and also leads recently to uniform-in-time moment estimates in [38], for sub-critical solution of the BD equations. It is also connected to a very simple probabilistic interpretation : if $c_1(t) \leq c_1'(t)$ for all times, the inhomogeneous Birth-Death process with birth rate $a_i c_1(t)$ and death rate b_i will be stochastically dominated by the inhomogeneous Birth-Death process with birth rate $a_i c_1'(t)$ and death rate b_i.

Some perspectives The author in [100] has obtained a functional central limit theorem, providing an infinite dimensional Gaussian process that quantifies the second order approximation between c^n and c in Theorem 7. The author uses a convenient Hilbert space representation, in order to properly defined the Gaussian process and to handle technical estimates. We note that the Hilbert space used there leads to restrictive hypothesis on kinetic rates a_i, b_i, which had to be bounded with respect to i.

We envision two strategies to overcome this limitation. Either used a Hilbert space more adapted to BD equations, like the one used in [60] (see also section 2.3.1), which has the powerful advantage to yield a self-adjoint operator. A second alternative, *probably technically simpler*, would be to quantify the fluctuations between c^n and the *truncated* versions of the BD system, which boils down to a finite n-dimensional system [24, 41], for which Gaussian processes are simpler to handle.

Also, whether or not uniform in time approximations in Theorem 7 are possible for the rescaled SBD process is an open question. We may anticipate that the subcritical case is easier to deal than the supercritical case, using in particular uniform in time moment propagation proved in BD equations [38].

2.2.2 BD->LS

This chapter is taken from [A9] and [A4].

Starting from the BD equations (22)-(23)(24), we may recover the LS equations (31).

This connection is somehow classical and has been proved in the context of outgoing characteristic at the boundary $x = 0$ for the LS model when small clusters tend to shrink. Two main approaches are used.

The first one considers the large time behavior of the supercritical BD model, with specific size-dependence of the rate functions, and relates the dynamics of large clusters to solutions of various version of LS equations. It is related to the so-called theory of Ostwald ripening, see [89, 84, 102].

The second one assumes that the initial condition of the BD equations as well as the rate coefficient satisfy some particular scaling hypotheses, and relates the finite-time solution of this rescaled model with the solution of the LS equations in [36, 68]. We rather follow the second approach.

The main novelty of this work resides in a new estimate on the growth of small clusters, which behave at a fast time scale. Through a rigorous quasi steady state approximation, we derive boundary conditions for the incoming characteristic case, when small clusters tend to grow.

Here we start from a dimensionless (and already rescaled) BD model that involves a small parameter $\varepsilon > 0$. We refer to the standard scaling procedure in [A9]. A few notations has changed from the BD equations (22)-(23)(24) to highlight the specificity of the monomer variable and the first flux J_1 in the BD equations. Denote by $c_i^\varepsilon(t)$ the concentration at time $t \geq 0$ of clusters consisting of $i \geq 2$ particles and $u^\varepsilon(t)$ the concentration of free particles (clusters of size 1), where we make explicit the dependence on $\varepsilon > 0$. The dimensionless system reads :

$$\begin{aligned} \frac{d}{dt}u^\varepsilon &= -\varepsilon J_1^\varepsilon - \varepsilon \sum_{i\geq 1} J_i^\varepsilon \,, \qquad\qquad t \geq 0\,, \\ \frac{d}{dt}c_i^\varepsilon &= \frac{1}{\varepsilon}\Big[J_{i-1}^\varepsilon - J_i^\varepsilon\Big]\,, \qquad i \geq 2\,,\ t \geq 0\,, \end{aligned} \tag{35}$$

where fluxes are defined by :

$$J_1^\varepsilon = \alpha^\varepsilon (u^\varepsilon)^2 - \varepsilon^\eta \beta^\varepsilon c_2^\varepsilon\,, \quad \text{and} \quad J_i^\varepsilon = a_i^\varepsilon u^\varepsilon c_i^\varepsilon - b_{i+1}^\varepsilon c_{i+1}^\varepsilon\,, \quad i \geq 2\,.$$

Here, coefficients a_i^ε and b_{i+1}^ε, for $i \geq 2$, denote respectively the rates of lengthening and shortening (ε-dependent), while α^ε and β^ε denote respectively the first rate of lengthening ($i = 1$) and shortening ($i = 2$). Finally, η is an exponent standing for the strength of the first shortening rate, in which the results strongly depend (see discussion in [A9]). Observe that such model (at least formally) preserves the total number of particles (no source nor sink), that is

$$u^\varepsilon(t) + \sum_{i\geq 2} \varepsilon^2 i c_i^\varepsilon(t) = \rho^\varepsilon\,, \quad \forall t \geq 0\,.$$

The constant ρ^ε is entirely determined by the initial conditions at $t = 0$ given by $u^{\mathrm{in},\varepsilon}$ and $(c_i^{\mathrm{in},\varepsilon})_{i\geq 2}$, non-negative and ε-dependent.

The first key step to obtain the limit of $(u^\varepsilon, c^\varepsilon)$ is to embed the infinite system of differential equations (35) into a partial differential equation with a suitable function framework. Accordingly, the size of each cluster is now represented by a continuous variable $x > 0$, and we let, for all $\varepsilon > 0$,

$$f^\varepsilon(t,x) := \sum_{i\geq 2} c_i^\varepsilon(t) \mathbf{1}_{\Lambda_i^\varepsilon}(x)\,, \quad x \geq 0\,,\ t \geq 0\,, \tag{36}$$

where for each $i \geq 2$, we defined $\Lambda_i^\varepsilon = [(i-1/2)\varepsilon, (i+1/2)\varepsilon)$. We denote for the remainder $f^{\mathrm{in},\,\varepsilon} := f^\varepsilon(0,x)$. Hence, each cluster of (discrete) size $i \geq 2$ is seen as a cluster of size roughly $i\varepsilon \in \mathbb{R}_+$. The scaling used in the dimensionless BD equations (35) consists in an acceleration of the fluxes (by $1/\varepsilon$), so that a cluster can reach an *asymptotically infinite* size $i = x/\varepsilon$ in finite time. The scaling is perhaps made clearer with a weak form version of (35) using the function f^ε constructed in (36). For all φ sufficiently regular, we have (at least formally), for all $t \geq 0$,

$$\begin{aligned} &\int_0^{+\infty} f^\varepsilon(t,x)\varphi(x)\,dx \\ &\qquad = \int_0^{+\infty} f^{in,\varepsilon}(x)\varphi(x)\,dx + \int_0^t [\alpha^\varepsilon u^\varepsilon(s)^2 - \beta^\varepsilon \varepsilon^\eta c_2^\varepsilon(s)] \left(\frac{1}{\varepsilon}\int_{\Lambda_2^\varepsilon} \varphi(x)\,dx\right) ds \\ &\qquad\qquad + \int_0^t \int_0^{+\infty} [a^\varepsilon(x) u^\varepsilon(s) f^\varepsilon(s,x) \Delta_\varepsilon \varphi(x) - b^\varepsilon(x) f^\varepsilon(s,x) \Delta_{-\varepsilon}\varphi(x)]\,dx\,ds\,, \end{aligned} \tag{37}$$

where $\Delta_h \varphi(x) = (\varphi(x+h) - \varphi(x))/h$, for $h \in \mathbb{R}$, and

$$u^\varepsilon(t) + \int_0^\infty x f^\varepsilon(t,x)\,dx = \rho^\varepsilon. \tag{38}$$

Then, an appropriate scaling of the rate functions together with the initial conditions entails that the solution of Eq. (37) converges to a solution of the weak form of the LS model (31). Compare to previous work, the main difficulty is to deal with the limit of the first flux J_1^ε for $\varphi(0) \neq 0$, in order to recover the boundary flux condition.

We now detail our precise (and somehow technical) assumptions.

Assumption 5. (Well posedness of BD equations). *The rates α^ε, β^ε, $(a_i^\varepsilon)_{i\geq 2}$ and $(b_i^\varepsilon)_{i\geq 3}$ are positives and, for each $\varepsilon > 0$, there exists a constant $K(\varepsilon) > 0$ such that*

$$\begin{aligned} a_{i+1}^\varepsilon - a_i^\varepsilon &\leq K(\varepsilon)\,, \quad i \geq 2\,, \\ b_i^\varepsilon - b_{i+1}^\varepsilon &\leq K(\varepsilon)\,, \quad i \geq 3\,. \end{aligned} \tag{LS-H1}$$

Assumption 6. (Convergence of the rates). *Let α and β be two positive numbers, and let a and b be two non-negative continuous functions on $[0,+\infty)$ that are positive on $x \in (0,+\infty)$. As $\varepsilon \to 0$, we suppose that*

$$\{\alpha^\varepsilon\} \text{ converges towards } \alpha\,. \tag{LS-H2}$$

$$\{\beta^\varepsilon\} \text{ converges towards } \beta\,. \tag{LS-H3}$$

$$\begin{aligned} &\{a^\varepsilon(.)\} \text{ converges uniformly on any compact set of } [0,+\infty) \text{ towards } a(.) \text{ and} \\ &\qquad \exists K_a > 0 \text{ s.t. } a^\varepsilon(x) \leq K_a(1+x),\ \forall x \in \mathbb{R}_+ \text{ and } \forall \varepsilon > 0\,. \end{aligned} \tag{LS-H4}$$

$$\begin{aligned} &\{b^\varepsilon(.)\} \text{ converges uniformly on any compact set of } [0,+\infty) \text{ towards } b(.) \text{ and} \\ &\qquad \exists K_b > 0 \text{ s.t. } b^\varepsilon(x) \leq K_b(1+x),\ \forall x \in \mathbb{R}_+ \text{ and } \forall \varepsilon > 0\,. \end{aligned} \tag{LS-H5}$$

Assumption 7. (regularity of the LS rate functions and Behavior of the rate functions near 0).

$$a,\, b \in \mathcal{C}^0([0,\infty)) \cap \mathcal{C}^1(0,\infty)\,, \tag{LS-H6}$$

$$a' \text{ and } b' \text{ are bounded on } (1,\infty)\,, \tag{LS-H7}$$

$$a(x) > 0 \text{ for all } x > 0\,. \tag{LS-H8}$$

We define the quantity

$$\Phi(x) := \frac{b(x)}{a(x)}\,, \Phi_0 := \lim_{x\to 0^+} \frac{b(x)}{a(x)}\,. \tag{39}$$

We suppose that

$$\Phi(x) := \frac{b(x)}{a(x)} \geq \Phi_0\,, \tag{LS-H9}$$

$$\Phi' \in L^1(0,1)\,, \tag{LS-H10}$$

$$\text{there exists } x^* > 0 \text{ such that } \Phi \text{ is monotone on } [0,x^*)\,, \tag{LS-H11}$$

and there exist $r_a \in [0,1)$, $r_b \geq r_a$, $\overline{a} > 0$, $\overline{b} > 0$ such that

$$\begin{aligned} a(x) &\sim_{0^+} \overline{a} x^{r_a}\,, \\ a^\varepsilon(\varepsilon i) &= a(\varepsilon i) + o((\varepsilon i)^{r_a})\,, \end{aligned} \quad \left| \quad \begin{aligned} b(x) &\sim_{0^+} \overline{b} x^{r_b}\,, \\ b^\varepsilon(\varepsilon i) &= b(\varepsilon i) + o((\varepsilon i)^{r_b})\,. \end{aligned} \right. \tag{LS-H12}$$

Finally, we assume some control on the initial conditions.

Assumption 8. (Initial conditions). *We assume there exists $\rho > 0$, and a non-negative function f^{in}, with*

$$f^{\text{in}} \in L^1((0,\infty),(1+x+x^2)\,dx)\,, \tag{LS-H13}$$

$$u^{\text{in}} := \rho - \int_0^\infty x f^{\text{in}} > \Phi_0\,, \tag{LS-H14}$$

such that ρ^ε *converges to* ρ, $u^{\mathrm{in},\,\varepsilon}$ *converges to* u^{in} *in* $\mathbb{R}_+$ *and* $\{f^{\mathrm{in},\,\varepsilon}\}$ *converges to* f^{in}, *in the weak topology of* $L^1((0,\infty),(1+x)\,dx)$, *e.g. for any* $\varphi \in C_b$,

$$\int_0^\infty (1+x)\varphi(x) f^{\mathrm{in},\,\varepsilon} dx \to \int_0^\infty (1+x)\varphi(x) f^{\mathrm{in}} dx\,, \tag{LS-H15}$$

as $\varepsilon \to 0$. *Moreover, we suppose that for all* $z \in (0,1)$,

$$\sup_{\varepsilon>0} \sum_{i\geq 2} \varepsilon^{r_a} c_i^{\mathrm{in},\,\varepsilon} e^{-iz} < +\infty\,. \tag{LS-H16}$$

Some comments about our assumptions. Hypothesis 5 on initial rate functions is needed to ensure well-posedness of the solution to (35), and Hypothesis 6 states convergence of the rescaled rates to some limit rate function. It is mainly technical although not too restrictive. Hypothesis 7 states the necessary regularity of the limit rate function to have global uniqueness of solution to LS equations (31)-(32)-(33) (see our recent work in [A4]). The main difficulty is to ensure that the monomer variables u^ε and u stays above the threshold Φ_0 so that incoming boundary condition always makes sense in the limit model. Finally, Hypothesis 8 states the convergence of the initial condition, and could probably be optimized (in particular the second moment in (LS-H13)). Hypothesis (LS-H16) looks exotic, and is mostly needed to control the small clusters in the rescaled BD equations, thanks to uniform-in-time moment estimate, taking inspiration from [38].

We now state our main theorem on the link between rescaled BD equations and the LS equations with incoming boundary condition.

Theorem 8. *For all* $T > 0$, *the sequence* f^ε *converges in* $\mathcal{C}([0,T]; w - L^1(\mathbb{R}_+,(1+x)dx))$ *towards* f *that satisfies, for all* $t \in [0,T)$ *and for every real-valued functions* h *locally bounded on* $[0,\infty)$ *such that* $h' \in L^\infty(0,\infty)$,

$$\int_0^\infty h(x) f(t,x)\,dx = \int_0^\infty h(x) f^{\mathrm{in}}(x)\,dx + \int_0^t \int_0^\infty (a(x)u(s) - b(x)) h'(x) f(s,x)\,dx\,ds + \int_0^t h(0)\mathfrak{n}(u(s))\,dt\,.$$

Moreover, u^ε *converges in* $\mathcal{C}([0,T];\mathbb{R}_+)$ *to* $u := \rho - \int_0^\infty x f(t,x)$, *which is continuously differentiable on* $(0,T)$ *and*

$$\frac{du(t)}{dt} = -u(t)\int_0^\infty a(x) f(t,x)\,dx + \int_0^\infty b(x) f(t,x)\,dx\,,$$

for all $t \in (0,T)$. *In particular,* f *satisfies*

$$\lim_{x\to 0^+} (a(x)u(t) - b(x)) f(t,x) = \mathfrak{n}(u(t))\,,$$

and, we have

$$\mathfrak{n}(u) = \begin{cases} \alpha u^2\,, & if\ \eta > r_a \\ \alpha u^2 \dfrac{u}{u + \beta/(\bar{a} 2^\eta)}\,, & if\ \eta = r_a < r_b\,, \\ \alpha u^2 \dfrac{\bar{a}u - \bar{b}}{\bar{a}u - \bar{b} + \beta/2^\eta}\,, & if\ \eta = r_a = r_b\,. \end{cases}$$

Theorem 8 requires several steps that can be found in [A9]-[A4]. A standard step is to prove moment estimates and uniform equicontinuity of f^ε in some suitable functional space. In [A9], a measure-valued space was first used. Compactness arguments allows to extract a converging subsequence and we then need to pass to the limit $\varepsilon \to 0$ in Eq. (37). There, the main difficulty is to handle the term c_2^ε. Multiplying the re-scaled BD equations (35) by ε, and integrating with respect to time, the later satisfies

$$\varepsilon(c_i^\varepsilon(t) - c_i^\varepsilon(0)) = \int_0^t (J_{i-1}^\varepsilon(s) - J_i^\varepsilon(s))ds\,, i \geq 2\,. \tag{40}$$

Hence, any L^∞ bounds (in time) on c_i^ε on $[0,T]$ leads to the identity, at least formally,

$$\lim_{\varepsilon\to 0} J_{i-1}^\varepsilon(s) = \lim_{\varepsilon\to 0} J_i^\varepsilon(s)\,, \quad i\geq 2\,, \forall s\in[0,T]\,. \tag{41}$$

The identification of the limit then relies on arguments similar to the Fenichel-Tikhonov theory on singularly perturbed dynamical systems (or slow-fast systems) [61]. We find that the limit value c_i has to reach instantaneously the stationary state of the BD model with a "constant" monomer concentration $u = u(t)$, see the heuristic arguments from (40)-(41). The convergence of c_2^ε to c_2 then holds on $L^\infty(0,T)$ equipped with the weak-* topology, which is similar to the one used in stochastic averaging theorem (see section 1.3.2).

The end of the proof requires to uniquely identify possible candidate limit solutions. We proved uniqueness of density global solution f to the LS equations (31)-(32)-(33) in [A4], with suitable regularity that allows to improve *a posteriori* the convergence from f^ε to f. The regularity of the solution of the LS equation is obtained thanks to a constructive approach using the methods of characteristic. Uniqueness rely on arguments from [67] which heavily relies again on the control of tail function $F(t,x) = \int_x^\infty f(t,y)dy$, as in the BD equations.

The case $\eta < r_a$ has also been considered in [A9], where we obtain solution of LS equation (31)-(32)-(33) with $\mathfrak{n}(u) = 0$. The statement and the proof is more technical as the convergence of the function f^ε holds only on weak-* measure space, due to the lack of control of the time derivative of f^ε near $x = 0$.

Some perspectives A challenging question is to give a unified sense of solutions to LS equation (31) for both outgoing and incoming characteristic. We have proved in [A4] that indeed when conditions (LS-H9)-(LS-H11) fails, the monomer variable may reach in finite time the threshold value Φ_0, in which case it is not completely clear how to match this solution with an outgoing one on later times, raising the problem of giving a wider meaning to the solution in order to be able to extend every local solution to a global one. This is an important issue that is deeply connected with a full understanding of the long time behavior of LS equation.

Regarding the long time behavior of LS equation, preliminary results indicate that for power-law kinetic rates, $a(x) = ax^{r_a}, b(x) = bx^{r_b}$, with $r_b \geq r_a$, a generic behavior to be expected is $xf \to 0$, $u \to \rho$, as long as the nucleation rate is strong enough.

We also note that second order approximation of (35) leads to diffusion correction in (31). However, up to my knowledge, no rigorous proofs of such results are available, and the adequate boundary condition that needs to be provided, according to the specific scaling hypotheses of the rate coefficients, is still unclear. This is an important question as the boundary condition heavily impacts long time results on diffusive version of LS equations [50].

2.3 Nucleation, metastability and equilibrium in SBD model

My thesis work focused on the study of first passage time for the SBD model, with applications to the formation of PRION protein polymers, involved in neurodegenerative diseases. These works [A20, A18], not detailed here, led me to take a closer look at fine properties of the SBD model, in particular related to metastability and the creation of large polymers, in the continuity of equivalent work on deterministic BD equations.

In the section 2.3.1, I detail a recent result in collaboration with Erwan Hingant on metastability in the SBD *linear* model, variant of the SBD model (28) in which the monomer variable C_1 is assumed constant. This work constitutes a first step towards understanding metastability in the SBD model.

In section 2.3.2, I detail a result on the steady state of the SBD model, which shows that the phase transition phenomenon is also present in the stochastic version of the Becker-Döring model, suggesting the creation of at least one cluster of very large size. Here too, it is a first step to characterize the long time behavior of the SBD model and the different time scales involved, when the number of particles $n \to \infty$.

2.3.1 Metastability in the linear SBD

Note that the "monomer" variable c_1 (or C_1) is fixed in this section.

The *linear* stochastic Becker-Döring (LSBD) process is a continuous-time Markov chain on the countable state space $\mathcal{E} = \ell^1(\mathbb{N}_2, \mathbb{N})$, the space of summable integer-valued sequences indexed by $\mathbb{N}_2 = \{2, 3, \cdots\}$, with

infinitesimal generator $\mathcal{A}$, given for all ψ with finite support on $\mathcal{E}$, by

$$\mathcal{A}(\psi)(c) = \sum_{i=1}^{+\infty} \Big(a_i z c_i [\psi(c+\Delta_i) - \psi(c)] + b_{i+1} c_{i+1} [\psi(c-\Delta_i) - \psi(c)] \Big) , \tag{42}$$

with the convention $c_1 = z$, $\Delta_1 = \mathbf{e}_2$ and $\Delta_i = \mathbf{e}_{i+1} - \mathbf{e}_i$, for each $i \geq 2$, where $\{\mathbf{e}_2, \mathbf{e}_3, \ldots\}$ denotes the canonical basis of $\mathcal{E}$ namely, $e_{i,k} = 1$ if $k = i$ and 0 otherwise.

Note that the transition intensities in the generator in Eq. (42) are linear with respect to the c_i. This motivates a different interpretation of such model, by modeling explicitly the size of each cluster. Let $\mathbf{C}^{\text{in}} = (C_2^{\text{in}}, C_3^{\text{in}}, \ldots)$ a random variable on $\mathcal{E}$ and denote $N^{\text{in}} = \sum_{i=2}^{\infty} C_i^{\text{in}}$. By construction $N^{\text{in}} < \infty$ almost surely. Then, given $\mathbf{C}^{\text{in}}$, we define $X_1^{\text{in}}, X_2^{\text{in}}, \ldots$ a denumerable collection of random variables on $\mathbb{N}_2$ such that for each $i \geq 2$,

$$C_i^{\text{in}} = \#\left\{ k \in [\![1, N^{\text{in}}]\!] \;\middle|\; X_k^{\text{in}} = i \right\} , \, a.s. ,$$

and $X_k^{\text{in}} = 2$ for all $k > N^{\text{in}}$. Each X_k is interpreted as the size of a given cluster, and follows a birth and death process that can be written as

$$X_k(t) = X_k^{\text{in}} + \sum_{i=2}^{\infty} \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{s > T_{k-N^{\text{in}}}} \mathbf{1}_{X_k(s^-)=i} \big(\mathbf{1}_{u \leq a_i z} - \mathbf{1}_{a_i z \leq u \leq a_i z + b_i} \big) N_k(ds, du) , \tag{43}$$

where by convention $T_k = 0$ if $k \leq 0$, and $T_1, T_2, \ldots$ is a collection of random times such that the $T_k - T_{k-1}$ are independent exponential random variable of parameter $a_1 z^2$, and $N_1, N_2, \ldots$ a denumerable family of independent Poisson point measures with intensity the Lebesgue measure $dsdu$ on $\mathbb{R}_+^2$.

The interpretation of (43) is clear : $X_k(t)$ denotes the size of the cluster labeled by k at time t; for $k \leq N^{\text{in}}$, clusters are initially "active" while for $k > N^{\text{in}}$ clusters are initially “inactive” at state 2, and gets “activated” at the random arrival time $T_{k-N^{\text{in}}}$. The number of "active" clusters at time $t \geq 0$ is given by the counting process

$$N(t) = N^{\text{in}} + \sum_{k \geq 1} \mathbf{1}_{t \geq T_k} ,$$

while the number of "active" cluster of size $i \geq 2$ is

$$C_i(t) = \#\left\{ k \in [\![1, N(t)]\!] \mid X_k(t) = i \right\} .$$

Let $n \geq 2$. It is remarkable on this model that we have at hand an explicit quasi-stationary distribution (QSD) before a cluster of size larger than $n+1$ is created, for the process $C(t)$. Such a distribution is given, for all $C \in \mathcal{E}_n = \{C \in \mathcal{E} \mid C_i = 0 \, , i \geq n+1\}$, by (recall that Q_i is defined in (26))

$$\Pi_n^{\text{qsd}}(C) = \prod_{i=2}^{n} \frac{(f_i^n)^{C_i}}{C_i!} e^{-f_i^n} , \qquad \text{with } f_i^n(z) = J_n Q_i z^i \sum_{k=i}^{n} \frac{1}{a_k Q_k z^{k+1}} ,$$

for $i = 2, \ldots, n$ and J_n given by

$$J_n = \left(\sum_{k=1}^{n} \frac{1}{a_k Q_k z^{k+1}} \right)^{-1} . \tag{44}$$

We define the hitting time of $\mathcal{E}_n^c$,

$$\tau_n = \inf \left\{ t \geq 0 \mid \mathbf{C}(t) \notin \mathcal{E}_n \right\} .$$

Our main assumption needed to ensure the model (43) is uniquely defined for all times (with no explosion), and thus the LSBD Markov chain too, is

$$\sum_{n=2}^{\infty} Q_n z^n \left(\sum_{k=n}^{\infty} \frac{1}{a_k Q_k z^{k+1}} \right) = \infty . \tag{LSBD-H1}$$

Theorem 9. *Under assumption (LSBD-H1). Let $z > z_s$ and $\mathbf{C}^{\text{in}}$ a random variable in $\mathcal{E}_n$ such that $\mathbb{E}\left[N^{\text{in}}\right] < \infty$. Then, for the LSBD process $\mathbf{C}$ starting at $\mathbf{C}^{\text{in}}$ in $\mathcal{E}_n$, we have*

$$\mathbf{P}_{\mathbf{C}^{\text{in}}}\left\{\tau_n > t\right\} \geq G_n^{\text{in}} e^{-J_n t}\,,$$

where

$$G_n^{\text{in}} = \mathbb{E}_{C^{\text{in}}}\Big[\prod_{i=2}^{n}\left(\frac{f_i^n}{Q_i z^i}\right)^{C_i}\Big] \geq 1 - \sum_{i=2}^{n}\mathbb{E}\left[C_i^{\text{in}}\right](1-\tfrac{f_i^n}{Q_i z^i})\,.$$

Moreover, there exists $\gamma_n > 0$ such that, for all $t \geq 0$, we have in total variation,

$$\|\mathbf{P}_{\mathbf{C}^{\text{in}}}\left\{\mathbf{C}(t) \in \cdot \mid t < \tau_n\right\} - \Pi_n^{\text{qsd}}\| \leq K_n\left(\frac{H_n^{\text{in}}}{\mathbf{P}_{\mathbf{C}^{\text{in}}}\left\{t < \tau_n\right\}} + e^{J_n t} H_n^{\text{qsd}}\right) e^{-\gamma_n t}\,,$$

where

$$K_n = (\sum_{k=2}^{n} Q_i z^i)^{\frac{1}{2}}\,,$$

and

$$H_n^{\text{in}} = \sum_{i=2}^{n}\sqrt{Q_i z^i}\frac{\mathbb{E}\left[C_i^{\text{in}}\right]}{f_i^n}\,, \qquad H_n^{\text{qsd}} = \sum_{i=2}^{n}\sqrt{Q_i z^i}\,.$$

Theorem 9 relies on the individual cluster formulation (43) and coupling methods, thanks to the fact that each clusters in (43) are independent from each others. Moreover, we make use of the results in the deterministic literature [88, 60] that give a convenient Hilbert representation for the solution of the linear BD equations, as well as important analytic characterization of birth-death processes from the seminal work by [58].

Starting from the QSD Π_n^{qsd}, the exact rate of creation of a cluster of size $n+1$ is given by J_n defined in Eq. (44). The important results in Theorem 9 comes from the fact that the pre-factor are explicitly known as a function of the initial condition. Although no optimality is claimed, these estimates are tight enough to provide a quantitative description of metastability, that we give below.

We assume standard coefficient relations, to fit with [60, 88], given by

$$A' < a_i < A i^{\alpha}\,, \quad \frac{b_{i+1}}{a_{i+1}} + \frac{\kappa}{i^{\nu}} \leq \frac{b_i}{a_i} \quad \text{and} \quad z_s e^{G i^{-\gamma}} \leq \frac{b_i}{a_i} \leq z_s e^{G' i^{-\gamma'}}\,, \tag{LSBD-H2}$$

for all $i \geq 2$, where α, $\gamma \in (0,1)$, γ', $\nu > 0$, κ, A', A, G and G' positives. Note that Hypothesis (LSBD-H2) implies (LSBD-H1). We also use the terminology of [88] namely a quantity $q(z)$ of z is : exponentially small if $q(z)/(z-z_s)^m$ is bounded for all $m > 0$ as $z \downarrow z_s$; and at most algebraically large if $(z-z_s)^{m_0} q(z)$ is bounded for some $m_0 > 0$ as $z \downarrow z_s$.

Under assumption (LSBD-H2) there exists a unique so-called critical size n^* (depending on z) such that $b_{n^*+1}/a_{n^*+1} < z < b_{n^*}/a_{n^*}$. The size n^* is interpreted as the nucleus size : for a cluster $X(t) \leq n^*$, $X(t)$ tends to shorten, while for $X(t) > n^*$, it tends to grow.

In [60, 88], n^* is proved to be at most algebraically large. Moreover, the time scale $1/\gamma_{n^*}$ (in Theorem 9) is also at most algebraically large, while the nucleation rate J_{n^*} (in Eq. (44)) is exponentially small.

We now choose an initial distribution $\mathbf{C}^{\text{in}}$ with support on $\mathcal{E}_j$ ($C_i^{\text{in}} = 0$ for all $i \geq j+1$), with j independent of z (no generality is claimed here), and such that $\mathbb{E}\left[N^{\text{in}}\right] < \infty$. We have

$$\mathbf{P}_{\mathbf{C}^{\text{in}}}\left\{\tau_{n^*} > t\right\} \geq \left(1 - \sum_{i=2}^{j}\mathbb{E}_{\mathbf{C}^{\text{in}}}\left[C_i\right](1-\tfrac{f_i^{n^*}}{Q_i z^i})\right) e^{-J_{n^*} t}\,, \tag{45}$$

which is arbitrary close to one for times $t \ll 1/J_{n^*}$ as $f_i^{n^*}/(Q_i z^i) \to 1$ when $z \searrow z_s$ (and i fixed). Moreover, we have

$$\|\mathbf{P}_{\mathbf{C}^{\text{in}}}\left\{\mathbf{C}(t) \in \cdot \mid \tau_{n^*} > t\right\} - \Pi_n^{\text{qsd}}\| \leq (\tfrac{H_j^{\text{in}}}{G_j^{\text{in}}} + H_{n^*}^{\text{qsd}}) K_{n^*} e^{J_{n^*} t - \gamma_{n^*} t}\,, \tag{46}$$

which is arbitrary small for times $1/\gamma_{n^*} \ll t \ll 1/J_{n^*}$. Indeed, note that $a_i Q_i z^i$ is decreasing up to the size n^*, thus $K_{n^*}^2 \leq \frac{a_1 z}{A'} n^*$. Hence, K_{n^*} as well as $H_{n^*}^{\text{qsd}}$ are at most algebraically large. Equations (45)-(46) show that the QSD is indeed a metastable state.

2.3.2 Phase transition in stationary state of the SBD model

We come back to the mass conservative SBD model (28) in this section. We turn to the study of the behavior of the stationary distribution of the rescaled SBD process (34), as $n \to \infty$.

We define Π^n, for any $c \in \mathcal{E}_\rho^n$, by

$$\Pi^n(c) = \frac{1}{B_n^z} \prod_{i=1}^{n} \frac{(\frac{n}{\rho} Q_i z^i)^{\frac{n}{\rho} c_i}}{(\frac{n}{\rho} c_i)!} e^{-\frac{n}{\rho} Q_i z^i},$$

where $z > 0$ is arbitrary and B_n^z is the following normalizing constant

$$B_n^z = \sum_{c \in \mathcal{E}_\rho^n} \prod_{i=1}^{n} \frac{(\frac{n}{\rho} Q_i z^i)^{\frac{n}{\rho} c_i}}{(\frac{n}{\rho} c_i)!} e^{-\frac{n}{\rho} Q_i z^i}.$$

One can easily check that Π^n satisfies the reversibility condition : for all $c \in \mathcal{E}_\rho^n$, for all $i \geq 1$,

$$A_i(c)\Pi^n(c) = B_{i+1}(c + \tfrac{\rho}{n}\Delta_i)\Pi^n(c + \tfrac{\rho}{n}\Delta_i),$$

and thus Π^n is the unique invariant distribution of the SBD process on $\mathcal{E}_\rho^n$.

We may write the invariant measure under the form

$$-\frac{\rho}{n} \ln \Pi^n(c) = \sum_{i=1}^{n} \left\{ -c_i \ln \left(\frac{n}{\rho} Q_i z^i \right) + \frac{\rho}{n} \ln \frac{n}{\rho} c_i! + Q_i z^i \right\} + \frac{\rho}{n} \ln B_n^z .$$

We clearly recognize a form closed to the relative entropy of the deterministic equations which drives its long-time behavior (see [24]), defined by Eq. (27).

Theorem 10. *Assume $0 < z_s < +\infty$. Let $\{c^n\}$ be a sequence belonging to $\mathcal{E}_\rho^n$.*

1. *If $0 < \rho \leq \rho_s$ and $\liminf_{i \to +\infty} Q_i^{1/i} > 0$ then,*

$$\lim_{n \to +\infty} -\frac{\rho}{n} \ln \Pi^n(c^n) = \mathcal{H}(c|c^{z(\rho)}),$$

   when $c^n \to c$ strongly in X^+, as $n \to \infty$.

2. *If $\rho > \rho_s$ and $\lim Q_i^{1/i}$ exists then,*

$$\lim_{n \to +\infty} -\frac{\rho}{n} \ln \Pi^n(c^n) = \mathcal{H}(c|c^{z_s}),$$

   when $c^n \rightharpoonup c$ weak $-*$ in X^+.

Due to the study of the minimum of the relative entropy (27) in X^+ and X_ρ^+ (see [24]), the stationary distribution of the SBD model thus satisfies, if $\rho \leq \rho_s$, $\Pi^n(\cdot) \to \delta_{c^{z(\rho)}}(\cdot)$ as $n \to \infty$, while if $\rho > \rho_s$, $\Pi^n(\cdot) \to \delta_{c^{z_s}}(\cdot)$ as $n \to \infty$. In the latter case, the stationary distribution thus concentrates on the deterministic state c^{z_s}, which has a mass ρ_s strictly inferior than the mass ρ of the initial condition. The quantity $\rho - \rho_s$ is interpreted as the mass which leaves the initial phase and undergoes a phase transition. Moreover, our numerical illustrations in Fig 11 indicate that the loss of mass $\rho - \rho_s$ is contained in a single giant particle, consistently with results on the Marcus-Lushnikov process [48].

Indeed, we show a numerical illustration of the law of large numbers and the phase transition phenomenon in left panels of Fig 11. We use an exact stochastic simulation algorithm of the trajectories of the SBD process defined by Eq. (34). Using classical reaction rates from literature [83] for a_i, b_i, we simulated a thousand of trajectories with $n = 1000$ and three different concentrations ρ : $\rho = 10$, $\rho = 1$ and $\rho = 0.1$. Note that with our choice of reaction rates, we have $z_s = 1/11$ and $\rho_s \approx 0.1059$. The illustration of the convergence of the SBD process towards the BD model can be seen as a single trajectory follows closely its mean value (which overlaps with a numerical simulation of the truncated BD model, not shown). For supercritical concentration $\rho > \rho_s$, we also see in left panels that the largest cluster occupies a large fraction of the total mass n.

The scaling behavior of the stationary distribution of the SBD model is illustrated with the first 10 sizes in middle panels, where the respective mean number of clusters follows closely the deterministic stationary state (little discrepancies are rather due to the finite size truncation than the stochastic fluctuations).
Finally, we illustrate the convergence to stationary state with the help of the relative entropy evaluated along the stochastic trajectories, in right panels (note that the final value of the relative entropy is not zero due to again finite size truncation). For the super-critical case with $\rho = 1$, the metastable phenomenon known for the BD (see [88]), seems to appear also in the SBD model : the relative entropy is nearly constant for a large time before converging to its final value. Interestingly, the sudden drop in the relative entropy value coincides with the formation of a giant cluster, consistently with our result on the LSBD in section 2.3.1. Thus escape of the metastable state and formation of a giant cluster appears to be concomitant. These observations will be the subject of future works.

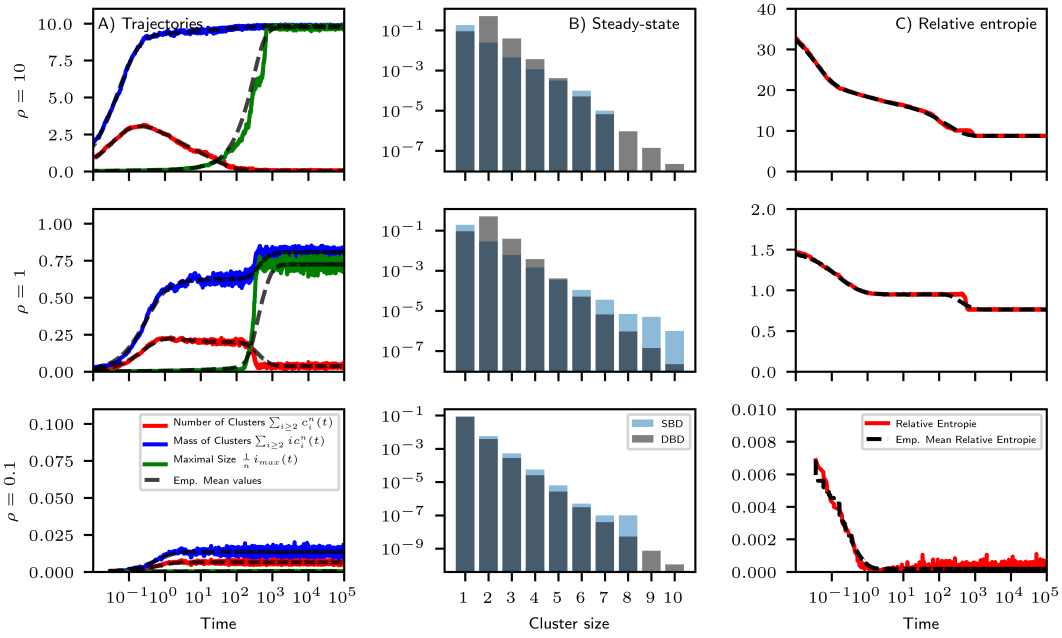

FIGURE 11 – We show results of stochastic simulation algorithm of the SBD process, with $a_i = i^{2/3}$, $b_i = i^{2/3}\left(\frac{1}{11} + \frac{10}{11 i^{1/3}}\right)$, $n = 1000$ and (top row) $\rho = 10$, (middle row) $\rho = 1$, (down row) $\rho = 0.1$. **A)** On the left column, we plot in colored lines (see legend) a single trajectory for the total number of clusters bigger than 2, their associated mass and the normalized size of the biggest cluster, *i.e.* $\frac{1}{n} i_{max}$ where $i_{max} = \max(i : c_i^n > 0)$. For the three quantities, the sampled mean on 1000 trajectories are superimposed in black dashed lines. **B)** On the middle column, we plot in blue the sampled mean (over 1000 realizations) number of clusters of size 1 to 10, at time $t = 10^5$ and in black the corresponding deterministic stationary state $c^z(\rho_s)$ (for the first two rows) and $c^z(\rho)$ (for the last row). **C)** On the right column, we plot in red plain line the relative entropy evaluated along a single stochastic trajectory, and in black dashed lines the sampled mean of this evaluation over 1000 trajectories.

2.4 Perspectives

Going further from our limit results in [A11] on the stationary distribution, we are currently looking at how to characterize the likelihood of existence of one or several giant clusters in the stationary state in the super-critical case. To that, results in limiting shapes of random combinatorial structures [49] are useful. Up to our knowledge, macroscopic scaling has not been considered in this literature, yet preliminary calculus indicates that, in the limit $n \to \infty$, a single giant cluster, of size $\lfloor n(1 - \frac{\rho_s}{\rho}) \rfloor$, is present, Π^n-almost surely. Further characterization such as second-order approximation can also be of some interests.

The main challenge in the study of the SBD process is a fine characterization of the transient and long-time dynamics in the super-critical case, with three steps :

— Timescale of super-critical clusters apparition, which tends to grow rapidly.

— Coarsening dynamics which leads to a single giant cluster emerge among the super-critical ones.

— Convergence of the single giant cluster size and the remaining small clusters to equilibrium.

As $n \to \infty$ and $\rho \searrow \rho_s$, the critical size goes to infinity and the formation of a super-critical cluster becomes a rare event, unlikely in finite time. Analytical characterization of the SBD process conditioned to the critical size not being reached, and explicit formula for the quasi-stationary distribution, seem out of reach. Yet, any estimate on the rate of creation of supercritical clusters could be important results for applications.

Références

Mes travaux de recherche

[A1] M. A. Ayoub, R. Yvinec, P. Crépieux, and A. Poupon. Computational modeling approaches in gonadotropin signaling. *Theriogenology*, 86(1) :22–31, 2016.

[A2] M. A. Ayoub, R. Yvinec, G. Jégot, J. A. Dias, S.-M. Poli, A. Poupon, P. Crépieux, and E. Reiter. Profiling of FSHR negative allosteric modulators on LH/CGR reveals biased antagonism with implications in steroidogenesis. *Mol. Cell. Endocrinol.*, 436 :10–22, 2016.

[A3] C. Bonnet, K. Chahour, F. Clément, M. Postel, and R. Yvinec. Multiscale population dynamics in reproductive biology : singular perturbation reduction in deterministic and stochastic models. *ESAIM : ProcS*, 67 :72 – 99, 2020.

[A4] J. Calvo, E. Hingant, and R. Yvinec. Initial-boundary value problem to the lifshitz-slyozov equation with non-smooth rates at the boundary. Submitted. arXiv :2004.01947, 2020.

[A5] F. Clément, F. Robin, and R. Yvinec. Analysis and calibration of a linear model for structured cell populations with unidirectional motion : Application to the morphogenesis of ovarian follicles. *SIAM J. Appl. Math.*, 79(1) :207–229, 2019.

[A6] F. Clément, F. Robin, and R. Yvinec. Stochastic nonlinear model for somatic cell population dynamics during ovarian follicle activation. Accepted in J. Math. Biol., 2020.

[A7] F. Clément, P. Crépieux, R. Yvinec, and D. Monniaux. Mathematical modeling approaches of cellular endocrinology within the hypothalamo-pituitary-gonadal axis. *Mol. Cell. Endocrinol.*, 518 :110877, 2020.

[A8] F. De Pascali, A. Tréfier, F. Landomiel, V. Bozon, G. Bruneau, R. Yvinec, A. Poupon, P. Crépieux, and E. Reiter. Chapter One - Follicle-Stimulating Hormone Receptor : Advances and Remaining Challenges. In Arun K. Shukla, editor, *International Review of Cell and Molecular Biology*, volume 338 of *G Protein-Coupled Receptors : Emerging Paradigms in Activation, Signaling and Regulation Part A*, pages 1–58. Academic Press, 2018.

[A9] J. Deschamps, E. Hingant, and R. Yvinec. Quasi steady state approximation of the small clusters in Becker-Döring equations leads to boundary conditions in the Lifshitz–Slyozov limit. *Commun. Math Sci*, 15(5) :1353–1384, 2017.

[A10] E. Hingant and R. Yvinec. Deterministic and Stochastic Becker-Döring Equations : Past and Recent Mathematical Developments. In *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*, pages 175–204. Springer, Cham, 2017.

[A11] E. Hingant and R. Yvinec. The Becker–Döring process : law of large numbers and non-equilibrium potential. *J. Stat. Phys.*, 177 :506–527, 2019.

[A12] E. Hingant and R. Yvinec. Quasi-stationary distribution and metastability for the stochastic Becker-Döring model. Submitted. arXiv : 2008.02544, 2020.

[A13] F. Landomiel, F. De Pascali, P. Raynaud, F. Jean-Alphonse, R. Yvinec, L. P. Pellissier, V. Bozon, G. Bruneau, P. Crépieux, A. Poupon, and E. Reiter. Biased Signaling and Allosteric Modulation at the FSHR. *Frontiers in Endocrinology*, 10, 2019.

[A14] E. Reiter, M. A. Ayoub, L. P. Pellissier, F. Landomiel, A. Musnier, A. Tréfier, J. Gandia, F. De Pascali, S. Tahir, R. Yvinec, G. Bruneau, A. Poupon, and P. Crépieux. β-arrestin signalling and bias in hormone-responsive GPCRs. *Mol Cell. Endocrinol.*, 449 :28–41, 2017.

[A15] E. Reiter, R. Yvinec, P. Crépieux, and A. Poupon. Coupling of recognition and effect in GPRC signaling computational modelling approaches. In *Molecular life sciences : An encyclopedic reference*. Springer-Verlag, 2018.

[A16] L. Riccetti, R. Yvinec, D. Klett, N. Gallay, Y. Combarnous, E. Reiter, M. Simoni, L. Casarini, and M. A. Ayoub. Human Luteinizing Hormone and Chorionic Gonadotropin Display Biased Agonism at the LH and LH/CG Receptors. *Sci. Rep.*, 7(1) :940, April 2017.

[A17] R. Yvinec, M.A. Ayoub, F. De Pascali, Crépieux P., E. Reiter, and A. Poupon. Workflow description to dynamically model β-arrestin signaling networks. *Methods Mol. Biol.*, 13(1957) :195–215, 2019.

[A18] R. Yvinec, S. Bernard, E. Hingant, and L. Pujo-Menjouet. First passage times in homogeneous nucleation : Dependence on the total number of particles. *J. Chem. Phys.*, 144(34106), 2016.

[A19] R. Yvinec, P. Crépieux, E. Reiter, A. Poupon, and F. Clément. Advances in computational modeling approaches of pituitary gonadotropin signaling. *Expert Opin. Drug. Discov.*, 13(9) :799–813, 2018.

[A20] R. Yvinec, M. R. D'Orsogna, and T. Chou. First passage times in homogeneous nucleation and self-assembly. *J. Chem. Phys.*, 137(24) :244107, 2012.

Bibliography

[21] D. V. Alexandrov and A. P. Malygin. Transient nucleation kinetics of crystal growth at the intermediate stage of bulk phase transitions. *J. Phys. A*, 46(45) :455101, 2013.

[22] D. F. Anderson, G. Craciun, M. Gopalkrishnan, and C. Wiuf. Lyapunov Functions, Stationary Distributions, and Non-equilibrium Potential for Reaction Networks. *Bull. Math. Biol.*, 77(9) :1744–1767, 2015.

[23] D. F. Anderson, G. Craciun, and T. G. Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biol.*, 72(8) :1947–1970, 2010.

[24] J. M. Ball, J. Carr, and O. Penrose. The Becker-Döring cluster equations : basic properties and asymptotic behaviour of solutions. *Comm. Math. Phys.*, 104(4) :657–692, 1986.

[25] R. Becker and W. Döring. Kinetische Behandlung der Keimbildung in Übersättigten Dämpfen. *Ann. Phys.*, 416(8) :719–752, 1935.

[26] S. K. Bristol-Gould, P. K. Kreeger, C. G. Selkirk, S. M. Kilen, K. E. Mayo, L. D. Shea, and T. K. Woodruff. Fate of the initial follicle pool : empirical and mathematical evidence supporting its sufficiency for adult fertility. *Dev. Biol.*, 298(1) :149–154, 2006.

[27] J. J. Burton. Nucleation Theory. In B. J. Berne, editor, *Statistical Mechanics : Part A : Equilibrium Techniques*, volume 5 of *Modern Theoretical Chemistry*, pages 195–234. Springer US, Boston, MA, 1977.

[28] M. Bächler, D. Menshykau, Ch. De Geyter, and D. Iber. Species-specific differences in follicular antral sizes result from diffusion-based limitations on the thickness of the granulosa cell layer. *Mol. Hum. Reprod.*, 20(3) :208–221, 2014.

[29] V. Cadoret, C. Frapsauce, P. Jarrier, V. Maillard, A. Bonnet, Y. Locatelli, D. Royère, D. Monniaux, F. Guérif, and P. Monget. Molecular evidence that follicle development is accelerated in vitro compared to in vivo. *Reproduction*, 153(5) :493–508, 2017.

[30] L. P. Cahill and P. Mauleon. Influences of season, cycle and breed on follicular growth rates in sheep. *J. Reprod. Fert.*, 58(2) :321–328, 1980.

[31] J. A. Cañizo, A. Einav, and B. Lods. Trend to Equilibrium for the Becker-Döring Equations : An Analogue of Cercignani's Conjecture. *Analysis & PDE*, 10(7) :1663–1708, 2017.

[32] A. R. Clark and J. A. Kruger. Mathematical modeling of the female reproductive system : from oocyte to delivery. *Wires. Syst. Biol. Med.*, 9(1) :e1353, 2017.

[33] H. G. Clarke, S. A. Hope, S. Byers, and R. J. Rodgers. Formation of ovarian follicular fluid may be due to the osmotic potential of large glycosaminoglycans and proteoglycans. *Reproduction*, 132(1) :119–131, 2006.

[34] F. Clément, P. Michel, D. Monniaux, and T. Stiehl. Coupled somatic cell kinetics and germ cell growth : multiscale model-based insight on ovarian follicular development. *Multiscale Model. Simul.*, 11(3) :719–746, 2013.

[35] F. Clément and D. Monniaux. Multiscale modelling of follicular selection. *Prog. Biophys. Mol. Biol.*, 113 :398–408, 2013.

[36] J. F. Collet, T. Goudon, F. Poupaud, and A. Vasseur. The Beker-Döring system and its Lifshitz-Slyozov limit. *SIAM J. Appl. Math.*, 62(5) :1488–1500, 2002.

[37] J.E. Coxworth and K. Hawkes. Ovarian follicle loss in humans and mice : lessons from statistical model comparison. *Hum. Reprod.*, 25(7) :1796–1805, 2010.

[38] J. A. Cãnizo, A. Einav, and B. Lods. Uniform moment propagation for the Becker–Döring equations. *P Roy Soc Edinb A*, 149(04) :995–1015, 2019.

[39] M. A. Driancourt. Regulation of ovarian follicular dynamics in farm animals. Implications for manipulation of reproduction. *Theriogenology*, 55(6) :1211–1239, April 2001.

[40] D. B. Duncan and R. M. Dunwell. Metastability in the classical truncated Becker-Döring equations. *Proc. Edinb. Math. Soc.*, 45(3) :701–716, 2002.

[41] D. B. Duncan and A. R. Soheili. Approximating the Becker-Döring cluster equations. *Appl. Numer. Math.*, 37(1-2) :1–29, 2001.

[42] M. J. Faddy, Esther C. Jones, and R. G. Edwards. An analytical model for ovarian follicle dynamics. *J. Exp. Zool.*, 197(2) :173–185, 1976.

[43] M. J. Faddy and M. C. Jones. Fitting time-dependent multicompartment models : a case study. *Biometrics*, 44(2) :587–593, 1988.

[44] M. J. Faddy, E. Telfer, and R. G. Gosden. The kinetics of pre-antral follicle development in ovaries of CBA/Ca mice during the first 14 weeks of life. *Cell Proliferation*, 20(6) :551–560, 1987.

[45] Y. Feng, P. Cui, X. Lu, B. Hsueh, Fredrik Möller B., Livia Zarnescu Y., R. Tomer, D. Boerboom, P. Carmeliet, K. Deisseroth, and A. J. W. Hsueh. CLARITY reveals dynamics of ovarian follicular architecture and vasculature in three-dimensions. *Sci. Rep.*, 7 :44810, 2017.

[46] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, Cambridge, 2009.

[47] J. E. Fortune. The early stages of follicular development : activation of primordial follicles and growth of preantral follicles. *Anim. Reprod. Sci.*, 78(3-4) :135–163, 2003.

[48] N. Fournier and P. Laurençot. Marcus-Lushnikov processes, Smoluchowski's and Flory's models. *Stochastic Process. Appl.*, 119(1) :167–189, 2009.

[49] G. Freiman and B. Granovsky. Clustering in coagulation-fragmentation processes, random combinatorial structures and additive number systems : Asymptotic formulae and limiting laws. *Trans. Amer. Math. Soc.*, 357(6) :2483–2507, 2005.

[50] T. Goudon and L. Monasse. Fokker-Planck approach of Ostwald Ripening : simulation of a modified Lifshitz-Slyozov-Wagner system with a diffusive correction. *SIAM J. Sci. Comput.*, 42(1) :B157–B184, 2020.

[51] A. Gougeon. Dynamics of human follicular growth : a morphologic perspective. In E.Y. Adashi and P.C.K. Leung, editors, *The ovary*. Raven Press, New York, 1993.

[52] A. Gougeon. Regulation of Ovarian Follicular Development in Primates : Facts and Hypotheses. *Endocr. Rev.*, 17(2) :121–155, 1996.

[53] A. Gougeon, R. Ecochard, and J. C. Thalabard. Age-related changes of the population of human ovarian follicles : increase in the disappearance rate of non-growing and early-growing follicles in aging women. *Biol. Reprod.*, 50(3) :653–663, 1994.

[54] J. Griffin, B. R. Emery, I. Huang, C. M. Peterson, and D. T. Carrell. Comparative analysis of follicle morphology and oocyte diameter in four mammalian species (mouse, hamster, pig, and human). *J Exp Clin Assist Reprod*, 3 :2, 2006.

[55] T. E. Harris. *The theory of branching processes*. Springer-Verlag, 1963.

[56] P. Jagers and F. C. Klebaner. Population size dependent and age dependent branching processes. *Stochastic Process. Appl.*, 87(2) :235–254, 2000.

[57] I. Jeon. Existence of gelling solutions for coagulation- fragmentation equations. *Commun. Math. Phys.*, 567 :541–567, 1998.

[58] S. Karlin and J. McGregor. The classification of birth and death processes. *Trans. Amer. Math. Soc.*, 86(2) :366–400, 1957.

[59] F. P. Kelly. *Reversibility and Stochastic Networks*. Cambridge University Press, Cambridge, 1979.

[60] M. Kreer. Classical Becker-Döring cluster equations : rigorous results on metastability and long-time behaviour. *Ann. Physik (8)*, 2(4) :398–417, 1993.

[61] C. Kuehn. *Multiple Time Scale Dynamics*. Springer, 2015.

[62] J. Kuntz, P. Thomas, G.-B. Stan, and M. Barahona. The exit time finite state projection scheme : bounding exit distributions and occupation measures of continuous-time Markov chains. *SIAM J. Sci. Comput.*, 41(2) :A748–A769, 2019.

[63] T. G. Kurtz. Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes. *J. Appl. Probab.*, 7(1) :49–58, 1970.

[64] T. G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Process. Appl.*, 6 :223–240, 1978.

[65] T. G. Kurtz. Averaging for martingale problems and stochastic approximation. In I. Karatzas and D. Ocone, editors, *Applied Stochastic Analysis*, volume 177, pages 186–209. Springer-Verlag, 1992.

[66] T. G. Kurtz and H.-W. Kang. Separation of Time-scales and model reduction for stochastic reaction networks. *Ann. Appl. Probab.*, 23(2) :529–583, 2013.

[67] P. Laurençot. Weak solutions to the Lifshitz-Slyozov-Wagner equation. *Indiana Univ. Math. J.*, 50(3) :1319–1346, 2001.

[68] P. Laurençot and S. Mischler. From the Becker-Döring to the Lifshitz-Slyozov-Wagner equations. *J. Statist. Phys.*, 106(5-6) :957–991, 2002.

[69] P. Laurençot. Weak Compactness Techniques and Coagulation Equations. In *Evolutionary Equations with Applications in Natural Sciences*, Lecture Notes in Mathematics, pages 199–253. Springer, Cham, 2015.

[70] L. Lei and A. C. Spradling. Female mice lack adult germ-line stem cells but sustain oogenesis using stable primordial follicles. *Proc. Natl. Acad. Sci. USA*, 110(21) :8585–8590, 2013.

[71] I. M. Lifshitz and V. V. Slyozov. The kinetics of precipitation from supersaturated solid solutions. *J. Phys. Chem. Solids*, 19(1-2) :35–50, 1961.

[72] T. Lundy, P. Smith, A. O'Connell, N. L. Hudson, and K. P. McNatty. Populations of granulosa cells in small follicles of the sheep ovary. *J. Reprod. Fert.*, 115(2) :251–262, 1999.

[73] K. P. McNatty, P. Smith, N. L. Hudson, D. A. Heath, D. J. Tisdall, W. S. O, and R. Braw-Tal. Development of the sheep ovary during fetal and early neonatal life and the effect of fecundity genes. *J. Reprod. Fert. Suppl.*, 49 :123–135, 1995.

[74] P. Michel. Large population and size scale limit of a stochastic particle model describing an age and space-structured population. *Math. Models Methods Appl. Sci.*, 27(3) :581–615, 2017.

[75] P. Monget, J. Bobe, A. Gougeon, S. Fabre, D. Monniaux, and R. Dalbies-Tran. The ovarian reserve in mammals : A functional and evolutionary perspective. *Mol. Cell. Endocrinol.*, 356(1-2) :2–12, 2012.

[76] D. Monniaux. Driving folliculogenesis by the oocyte-somatic cell dialog : Lessons from genetic models. *Theriogenology*, 86(1) :41–53, 2016.

[77] D. Monniaux. Factors influencing establishment of the ovarian reserve and their effects on fertility. *Anim. Reprod.*, 15(Suppl.1) :635–647, 2018.

[78] D. Monniaux, V. Cadoret, F. Clément, R. Dalbies-Tran, S. Elis, S. Fabre, V. Maillard, P. Monget, and S. Uzbekova. Folliculogenesis. In *Encyclopedia of Endocrine Diseases*, volume 2, pages 377–398. Elsevier, 2019.

[79] D. Monniaux, F. Clement, R. Dalbies-Tran, A. Estienne, S. Fabre, C. Mansanet, and P. Monget. The Ovarian Reserve of Primordial Follicles and the Dynamic Reserve of Antral Growing Follicles : What Is the Link ? *Biol. Reprod.*, 90(4) :85–85, 2014.

[80] D Monniaux, B Mandon-Pépin, and P Monget. L'atrésie folliculaire, un gaspillage programmé. *Médecine / sciences*, 15(2) :157, 1999.

[81] D. Monniaux, P. Michel, M. Postel, and F. Clément. Multi-scale modelling of ovarian follicular development : From follicular morphogenesis to selection for ovulation. *Biol. Cell*, 108(6) :149–160, 2016.

[82] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124(4) :044104, 2006.

[83] B. Niethammer. On the Evolution of Large Clusters in the Becker-Döring Model. *J. Nonlinear Sci.*, 13(1) :115–155, 2003.

[84] B. Niethammer. A scaling limit of the Becker-Döring equations in the regime of small excess density. *J. Nonlinear Sci.*, 14(5) :453–468, 2004.

[85] B. Niethammer and R. L. Pego. On the Initial–Value Problem in the Lifshitz-Slyozov-Wagner Theory of Ostwald Ripening. *SIAM J. Math. Anal.*, 31(3) :467, 2000.

[86] T. Pedersen and N. R. Hartmann. The kinetics of granulosa cells in developing follicles in the mouse ovary. *Cell Tissue Kinet*, 4(2) :171–184, 1971.

[87] Torben Pedersen. Determination of follicle growth rate in the ovary of the immature mouse. *J. Reprod. Fertil.*, 21(1) :81–93, 1970.

[88] O. Penrose. Metastable states for the becker-döring cluster equations. *Comm. Math. Phys.*, 124(4) :515–541, 1989.

[89] O. Penrose. The Becker–Döring equations at large times and their connection with the LSW theory of coarsening. *J. Statist. Phys.*, 89(1-2) :305–320, 1997.

[90] O. Penrose, J. L. Lebowitz, J. Marro, M. H. Kalos, and A. Sur. Growth of clusters in a first-order phase transition. *J. Statist. Phys.*, 19(3) :243–267, 1978.

[91] H. M. Picton. Activation of follicle development : the primordial follicle. *Theriogenology*, 55(6) :1193–1210, April 2001.

[92] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15) :1923–1929, August 2009.

[93] A. Raue, B. Steiert, M. Schelker, C. Kreutz, T. Maiwald, H. Hass, J. Vanlier, C. Tönsing, L. Adlung, R. Engesser, W. Mader, T. Heinemann, J. Hasenauer, M. Schilling, T. Höfer, E. Klipp, F. Theis, U. Klingmüller, B. Schöberl, and J. Timmer. Data2Dynamics : a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*, 31(21) :3558–3560, 07 2015.

[94] F. Robin. *Modeling and analysis of cell population dynamics : application to the early development of ovarian follicles*. PhD thesis, University Paris-Saclay, 2019.

[95] C. Ruoss, A. Tadros, T. O'Shea, J. McFarlane, and G. Almahbobi. Ovarian follicle development in Booroola sheep exhibiting impaired bone morphogenetic protein signalling pathway. *Reproduction*, 138(4) :689–696, 2009.

[96] Prigent S., Ballesta A., Charles F., Lenuzza N., Gabriel P., Tine L. M., RezaeiH., and Doumic M. An efficient kinetic model for assemblies of amyloid fibrils and its application to polyglutamine aggregation. *PLoS One*, 7(11) :e43273, 2012.

[97] R. J. Scaramuzzi, D. T. Baird, B. K. Campbell, M.-A. Driancourt, J. Dupont, J. E. Fortune, R. B. Gilchrist, G. B. Martin, K. P. McNatty, A. S. McNeilly, P. Monget, D. Monniaux, C. Viñoles, and R. Webb. Regulation of folliculogenesis and the determination of ovulation rate in ruminants. *Reprod. Fertil. Dev.*, 23(3) :444–467, 2011.

[98] J. W. P. Schmelzer. *Nucleation Theory and Applications*. Wiley-VCH Verlag GmbH & Co. KGaA, 2005.

[99] M. Slemrod. The Becker-Döring equations. In *Modeling in applied sciences*, pages 149–171. Springer, 2000.

[100] Wen Sun. A Functional Central Limit Theorem for the Becker-Döring model. *J. Stat. Phys.*, 171 :145–165, 2018.

[101] C. Thibault and M.-C. Levasseur. *La reproduction chez les mammifères et l'homme*. Editions Quae, 2001.

[102] J. J. L. Velázquez. The Becker-Döring equations and the Lifshitz-Slyozov theory of coarsening. *J. Stat. Phys.*, 92(1-2) :195–236, 1998.

[103] H. M. Wear, M. J. McPike, and K. H. Watanabe. From primordial germ cells to primordial follicles : a review and visual representation of early ovarian development in mice. *J. Ovarian Res.*, 9(1) :36, 2016.

[104] T. Wilson, X. Y. Wu, J. L. Juengel, I. K. Ross, J. M. Lumsden, E. A. Lord, K. G. Dodds, G. A. Walling, J. C. McEwan, A. R. O'Connell, K. P. McNatty, and G. W. Montgomery. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. *Biol. Reprod.*, 64(4) :1225–1235, 2001.