Growth-fragmentation equations in biology

Marie Doumic

UNIVERSITE PARIS VI - PIERRE ET MARIE CURIE

Habilitation à diriger des recherches

Spécialité : Mathématiques appliquées et applications des mathématiques

présentée par

Marie Doumic Jauffret

Etudes de modèles de croissance et fragmentation et applications en biologie

Growth-fragmentation equations in biology

devant le jury composé de

Monsieur Vincent BERINGUE	Invité
Monsieur Jean-Michel CORON	Invité
Monsieur Odo DIEKMANN	Rapporteur
Madame Irene M. GAMBA	Rapporteur
Monsieur Christophe GIRAUD	Invité
Monsieur François GOLSE	Invité
Monsieur Stéphane MISCHLER	Rapporteur
Monsieur Benoît PERTHAME	Invité

A la mémoire de Julien Geitner

Contents

Part I
Résumé étendu (short French version)

Dans ce mémoire d'habilitation, je présente les travaux de recherche effectués depuis mars 2007 à Inria Paris-Rocquencourt, au sein de l'équipe-projet BANG dirigée par Benoît Perthame. Avant d'entrer dans le vif du sujet, le paragraphe ci-dessous résume mes travaux de recherche antérieurs, effectués au cours de ma thèse, et explique mon parcours professionnel. J'aborde ensuite mon sujet de recherche actuelle: les phénomènes de croissance et de division (ou fragmentation) en biologie. J'en détaille de façon plus approfondie les divers aspects: analyse mathématique, problème inverse, modélisation et applications en biologie. Enfin, j'esquisse les perspectives de ma recherche en cours et à venir.

1 Travaux antérieurs et parcours professionnel

1.1 Travaux de thèse: l'interaction laser-matière

Ma thèse se déroula au Commissariat à l'Energie Atomique et à l'Ecole Normale Supérieure, sous la direction de Rémi Sentis et de François Golse. Le projet Laser Mégajoule du CEA fournit mon sujet d'étude: l'interaction laser-matière. Ces travaux présentent une forte cohérence, et je n'ai pas abordé de nouveau ce thème de recherche depuis ma soutenance, en mai 2005: je laisse donc le lecteur intéressé se reporter à ce précédent manuscrit [3], et je me contente d'en faire ici un très bref résumé.

L'objectif était de modéliser puis de simuler la propagation laser dans un plasma lorsque la direction de la lumière, loin d'être orthogonale au domaine de simulation, présente un angle d'incidence prononcé. Les travaux antérieurs ne permettaient qu'un angle très faible [65]. Cela est problématique lorsqu'il s'agit, comme dans le projet Laser Mégajoule, de prendre en compte des variations dans la direction de propagation ou encore des croisements de faisceaux.

Partant de travaux de physiciens [65], et par un développement asymptotique de type Chapman-Enskog [44], nous bâtimes tout d'abord un modèle de propagation laser oblique en milieu inhomogène, modèle que nous justifiâmes par des estimations d'énergie (voir l'introduction de ma thèse [3], ainsi que [5]). Nous construisîmes alors un schéma numérique capable de rendre compte de cette propagation pour un angle d'incidence quelconque, allant jusqu'à presque 90 degrés [8]. Ce schéma est fondé sur la résolution analytique de l'équation dans un cas simple (coefficients constants) par transformée de Fourier [5]. La méthode est également efficace dans le cas de coefficients variables [6], et fut intégrée par F. Duboc à la plateforme de calcul HERA du Commissariat à l'Energie Atomique [26].

1.2 Parcours professionnel

Je commençai ma thèse en septembre 1999. Après un an, je décidai de me tourner vers une carrière opérationnelle et donc de quitter la filière scientifique – définitivement, croyais-je: je candidatai et je fus reçue au Corps des Ponts et Chaussées, devenu depuis le Corps des Ponts et Forêts. De 2000 à 2003, je suivis la scolarité normale des ingénieurs-élèves, alternant cours et stages professionnels.

En août 2003, je pris le poste de chef du service "Techniques de la voie d'eau", au service Navigation de la Seine. J'y étais responsable d'un service d'environ quarante personnes (ingénieurs, techniciens, personnel administratif et surveillants de travaux), service qui avait pour fonction de planifier et d'organiser les appels d'offres puis de suivre la réalisation des grands travaux fluviaux sur le bassin de la Seine (barrages de navigation, passes à poissons et écluses principalement). J'y passai trois années passionnantes et formatrices. Mes publications de l'époque, en collaboration avec un grand nombre d'auteurs, s'appellent la reconstruction des barrages de l'Oise, la reconstruction du barrage de Chatou, la passe à poissons d'Andrésy, la réorganisation de la filière ingénierie.

En parallèle, j'achevai ma thèse et la soutins en mai 2005. Je pensais mettre ainsi un point final à mes travaux scientifiques. Pourtant, lorsque deux ans plus tard, comme prévu dans par ma carrière d'ingénieur, je commençai à chercher un deuxième poste, je me tournai de nouveau vers la recherche.

Trois facteurs jouèrent dans ma décision. Le premier fut la maturité que j'avais acquise dans l' "opérationnel": je n'envisageais pas cela comme un retour en arrière ni même comme un retour à de premières amours, suite à un égarement passagé, mais comme une façon différente de faire de la science que lorsque j'étais étudiante. Le deuxième fut que mes travaux de thèse aient été fructueux, repris par F. Duboc et mis en oeuvre dans la plateforme HERA: mon travail avait donc été plus utile que ce que j'en avais perçu sur le coup. Le troisième, et l'élément déclencheur, fut une petite phrase – presque anachronique dans le contexte de ma soutenance de thèse – que Claude Bardos tint à ajouter à mon compte-rendu de soutenance: "la candidate pourrait faire une très bonne carrière dans la recherche scientifique". Je voulus tenter ce pari, aussi étrange qu'il semblât, et malgré un certain scepticisme ambiant - à commencer par le mien propre. Si aujourd'hui je crois pouvoir dire qu'il est tenu, je le dois à de nombreux soutiens et aux collaborations développées depuis cinq ans, et plus que tout autre à Benoît Perthame.

En mars 2007, je rejoignis en tant que chercheur (mon statut exact est celui d'ingénieur en détachement du Corps des Ponts, des Eaux et des Forêts) l'équipe-projet BANG dirigée par Benoît Perthame. Mon activité de recherche porta, et continue de porter, sur la compréhension de phénomènes de croissance et de fragmentation, avec un souci constant que les problèmes mathématiques envisagés soient liés à une question de biologie contemporaine. De juin 2008 à juin 2011, j'encadrai avec B. Perthame la thèse de Pierre Gabriel sur les équations de transport-fragmentation et applications aux maladies à prion. Il est maintenant maître de conférences à l'université de Versailles-St Quentin en Yvelines et fut

lauréat du prix solennel de la chancellerie des Universités de Paris "Thiessé de Rosemont / Demassieux" en 2012. Je co-encadre actuellement avec B. Perthame les thèses de H.W. Haffaf depuis septembre 2011 et T. Bourgeron depuis septembre 2012.

De septembre 2009 à décembre 2012, je coordonnai un projet ANR du programme blanc, TOPPAZ (pour *Theory and Observation of protein Polymerization in Amyloid Diseases (Prion, Alzheimer's)*), dans le cadre duquel la thèse de P. Gabriel fut effectuée. Dans la ligne de ce projet ANR, de décembre 2012 à décembre 2017, je dirige un projet européen du programme ERC Starting Grant, le projet SKIPPER^{AD} acronyme de *Simulation of the Kinetics and Inverse Problems for protein Polymerization in Amyloid Diseases*. L'idée de cet axe de recherche est la confrontation de nos modèles de croissance et fragmentation avec les données expérimentales les plus récentes, tant par des techniques de problèmes inverses que par l'analyse mathématique.

2 Phénomènes de croissance et de fragmentation

2.1 Quelques équations de population structurée [4, 12, 14]

Soit une population dont les individus grandissent et se reproduisent, et dont on souhaite suivre - et comprendre - le comportement au cours du temps. Les équations dites "de population structurée" ont été écrites dans ce but: elles rendent compte de l'évolution de la population au cours du temps, en fonction d'une variable qu'on appelle "structurante", c'est-à-dire qu'elle est choisie pour être caractéristique de la croissance et de la division des individus.

A ce stade, avant même l'écriture des équations, une question se pose: comment bien choisir la ou les variables structurantes ? Quelle grandeur *intrinsèque* permet de comprendre le développement de la population ? La réponse n'est bien sûr pas la même selon la population que l'on considère. Dans mes travaux, j'ai étudié dans [4] un modèle écrit pour des populations de cellules en division [34] où les variables structurantes étaient l'âge et le contenu en protéine (ce qui est mathématiquement le cas structuré en taille). Dans [12, 14], nous avons analysé des modèles non linéaires structurés en *maturité*, ce qui est utile, par exemple, pour modéliser l'hématopoièse où les cellules se différencient au fur et à mesure, selon un arbre qui va des cellules souches aux cellules sanguines. Dans [31, 1], l'équipe de H.T. Banks, avec laquelle j'ai collaboré, a étudié un modèle structuré en *label fluorescent*, qui présente l'avantage de pouvoir - contrairement à l'âge ou à la maturité - être directement mesuré, mais qui présente l'inconvénient concomitant de ne pas être à proprement parler *structurant*, c'est-à-dire qu'il s'agit d'un marqueur qui est rarement le moteur de la croissance ou de la division. Dans [7, 2, 15, 11, 13, 16] enfin, nous avons étudié, mes co-auteurs et moi-même, divers aspects de l'équation suivante:

$$\frac{\partial}{\partial t}n(t,x) + \frac{\partial}{\partial x}(g(x)n(t,x)) + B(x)n(t,x) = 2\int_x^\infty k(x,y)B(y)n(t,y)dy, \quad x > 0, \qquad (2.1)$$

que l'on complète par une condition initiale $n(t=0,x) = n^{in}(x)$ et une condition de flux au bord $g(0)n(t,0) = 0$.

Il n'y a pas de terminologie unifiée pour cette équation: selon le contexte, elle est appelée équation de division structurée en taille [128], équation de croissance-fragmentation [42], de transport-fragmentation [67], d'agrégation-fragmentation [7], de division cellulaire lorsque les cellules se divisent par fission [79]. Dans la suite, j'y ferai référence sous le nom d'équation de *croissance-fragmentation*.

Ici, la densité de population est décrite par $u(t,x)$, où t est le temps et x la variable structurante - typiquement, la taille d'un individu. Les individus grossissent avec un taux de croissance $g(x)$, et se divisent avec un taux de division $B(x)$. La probabilité de donner naissance à deux individus de taille respective x et $y-x$, pour un individu parent de taille y, est donnée par le noyau de fragmentation $k(x,y)$.

2.2 Etudes de l'équation de croissance-fragmentation [7, 2]

Croissance et fragmentation sont deux phénomènes de dynamique opposée: la première entraîne la population vers les grandes tailles, tandis que la seconde la ramène vers les plus petites. Lorsque leurs actions sont de force disons, d'une façon heuristique, *comparable* (cf. [7] pour une définition plus précise), il a été démontré qu'asymptotiquement la solution converge vers le produit d'un profil stationnaire $N(x)$ et d'une exponentielle en temps $e^{\lambda t}$. Ce phénomène est une réalité non seulement mathématique mais biologique, illustré par exemple par les phénomènes de *désynchronisation* des cellules [45]: au bout d'un certain temps, les cellules oublient leur répartition initiale (synchronisée) et se répartissent selon un profil stationnaire.

Les premiers résultats sur la convergence asymptotique remontent à ma connaissance 1967, dans des articles de biophysique de Bell et Anderson [37, 36]. Des démonstrations mathématiques fondées sur l'analyse des systèmes dynamiques, avec des méthodes de semi-groupe et l'application du théorème de Krein-Rutman [88], suivirent dans les années 80 [79, 80, 52]. Plus récemment, le comportement asymptotique a été généralisé à l'aide de deux outils fondamentaux: d'une part et comme précédemment, l'existence et l'unicité d'un premier vecteur propre positif associé à une valeur propre positive; d'autre part, la gamme d'inégalités de type entropique baptisée *General Relative Entropy* par B. Perthame et ses co-auteurs [108].

La première question qui se pose est celle des conditions d'existence et d'unicité de ce premier vecteur propre. Il s'agit de démontrer l'existence et l'unicité du couple (λ, N) avec $\lambda > 0$, $N \in L^1(\mathbb{R}_+)$, solutions de

$$\begin{aligned} &\lambda N(x) + \tfrac{\partial}{\partial x}\big(g(x)N(x)\big) + B(x)N(x) = 2\int_x^\infty k(x,y)B(y)N(y)dy, \quad x>0, \\ &\textstyle\int_0^\infty N(x)dx = 1, \qquad N(x>0)>0, \qquad g(0)N(0) = 0. \end{aligned} \tag{2.2}$$

A la suite des travaux de Diekmann, Heijmans et Thieme d'une part [79, 80, 52], de P. Michel

[105], B. Perthame et L. Ryzhik [128] d'autre part, nous voulûmes chercher les hypothèses les plus générales possibles sous lesquelles obtenir une telle propriété, autorisant en particulier le taux de croissance $g(x)$ à s'annuler en 0. C'est l'objet de l'article [7].

Les preuves reposent sur le théorème de Krein-Rutman ([88] pour l'article fondateur, ou [51] ch.VIII) vérifié par une équation régularisée, et par des estimations fines permettant d'obtenir de la compacité sur la suite des solutions régularisées, et donc sa convergence vers une solution de l'équation originelle. Ces estimations font jouer tour à tour la domination du taux de division par le taux de croissance, au voisinage de 0, et la domination réciproque au voisinage de l'infini. Pour le vecteur propre adjoint, on utilise aussi un principe du maximum pour construire une sur-solution de la suite des solutions régularisées. Enfin, la preuve de l'unicité repose sur une méthode entropique (*l'entropie relative généralisée* introduite par P. Michel, S. Mischler et B. Perthame [107, 108, 127]).

Les estimations récursives de [7] se sont révélées caractéristiques de l'équation, et utiles à d'autres études. Dans [42, 24], elles sont à la base d'estimations plus précises, nécessitant des hypothèses plus restrictives, qui servent à prouver l'existence d'un *trou spectral*, et donc une convergence exponentiellement rapide vers le comportement asymptotique. Dans [16], nous nous en servîmes pour étudier la régularité de l'opérateur $\Gamma : (g, B) \to (\lambda, N)$.

Dans [2] enfin, avec un changement de variables auto-similaire bien choisi, ces estimations nous conduisirent à mettre en évidence un comportement inattendu de l'équation. Nous souhaitions étudier le comportement du vecteur propre et de la valeur propre dominante lorsque la fragmentation (ou, de façon symétrique, la croissance) est modulée par un paramètre $\mathfrak{a} > 0$, ce qui peut se modéliser ainsi:

$$\lambda_{\mathfrak{a}} N_{\mathfrak{a}}(x) + \frac{\partial}{\partial x}(g(x)N_{\mathfrak{a}}(x)) + \mathfrak{a}B(x)N_{\mathfrak{a}}(x) = 2\mathfrak{a}\int_x^\infty k(x,y)B(y)N_{\mathfrak{a}}(y)dy, \quad x > 0. \tag{2.3}$$

Une telle étude peut servir de modèle-jouet pour comprendre l'effet d'une action influençant la fragmentation, comme dans le protocole appelé *PMCA* où la fragmentation des polymères de la protéine Prion est amplifiée par des ondes sonores (processus de *sonication*). Elle peut aussi être un pas pour mieux comprendre des modèles non linéaires, comme celui du Prion (cf. ci-dessous) ou encore pour optimiser l'action de médicaments anti-cancéreux [46].

Nous avons montré que le comportement de la valeur propre dominante $\lambda_{\mathfrak{a}}$ ne dépendait pas nécessairement de façon monotone d'un tel paramètre, car lorsque $\mathfrak{a} \to 0$ (respectivement $\mathfrak{a} \to \infty$) il est déterminé par le comportement du taux de croissance au voisinage de l'infini (respectivement de 0).

2.3 Problème inverse [15, 13, 16]

Un autre regard sur l'équation de croissance-fragmentation est donné par les études de problèmes inverses [57], souvent riches de connaissances insoupçonnées sur un modèle donné.

En effet, le problème dit *direct* se pose en général d'une seule façon: étant données certaines valeurs de paramètres et un problème d'évolution, comment montrer que le problème est bien posé (et dans quel espace), et que dire du comportement asymptotique en temps (et éventuellement en espace). Il n'en est pas de même pour les problèmes *inverses*, qui répondent chacun à une question du type: étant donné tel observable sur ma solution, telles hypotèses *a priori* sur les paramètres, puis-je reconstruire tels autres paramètres et/ou l'état initial du système; et les estimations obtenues sur cette reconstruction sont-elles optimales? A chaque observable et à chaque question applicative correspond donc un problème inverse différent, bien que relié à un unique problème d'évolution.

Ainsi, l'équation "cousine", où la variable structurante est l'âge, a fait l'objet de nombreuses études portant sur différents problèmes inverses - cf. par exemple [75, 76, 141, 142, 131, 58].

Pour l'équation de transport-fragmentation, le problème inverse de principale importance ne consiste bien sûr pas à rechercher la condition initiale, étant donné l'oubli de cette condition liée à la convergence rapide vers le profil stationnaire. On peut aussi concevoir que dans bien des cas, il est possible de mesurer d'une façon directe le taux de croissance $g(x)$. En revanche, développer des méthodes pour estimer le taux de division $B(x)$ peut se révéler d'un grand intérêt.

La méthode introduite dans [130] consiste à utiliser le comportement asymptotique pour simplifier le problème: à partir de mesures bruitées du profil stationnaire $N(x)$, et du taux de croissance exponentiel de la population λ, comment estimer le taux de division B dans le cadre du modèle dit de la *mitose égale*, avec $g(x) = 1$ et $k(x, y) = \delta_{x=\frac{y}{2}}$? L'équation considérée est donc toujours (2.2), mais l'inconnue n'est plus le couple (λ, N) mais le taux de division $B(x)$. Notant $H = BN$, et $L(N) = \lambda N + \frac{\partial}{\partial x}(gN)$, on peut diviser ce problème en trois étapes.

- Résoudre, dans des espaces bien choisis, l'équation

$$4H(2x) - H(x) = L(x), \tag{2.4}$$

  et montrer la continuité de l'opérateur $\mathcal{L} : L \to H$ dans ces espaces bien choisis.

- Sachant qu'on ne mesure non pas N mais N_ε, proche de N en un certain sens, et que donc on ne connaît pas l'exacte valeur de L dans l'équation (2.4), comment en déduire un H_ε proche de la solution H de (2.4) ? La notion de régularisation intervient ici, et consiste à faire en sorte de se placer dans le cadre des espaces de régularité de l'opérateur $\mathcal{L}$.

- Diviser H_ε par N_ε de façon à estimer le taux de division B. Cette division pose deux problèmes: tout d'abord, N_ε n'est pas exactement N, ensuite, l'estimation est nécessairement de très mauvaise qualité lorsque la densité N tend vers 0.

L'article fondateur [130] étudie le problème inverse dans le cas ci-dessus dit de la *mitose égale*, dans l'espace $L^2(\mathbb{R}_+)$ - donc on suppose qu'on mesure N_ε avec $\|N - N_\varepsilon\|_{L^2(\mathbb{R}_+)} \leq \varepsilon$.

Dans L^2, on peut montrer que l'opérateur $\mathcal{L}$ est continu. Les auteurs proposent alors une méthode de régularisation dite de *quasi-réversibilité*, en écho à [90]: par l'ajout d'un petit terme de dérivée $\alpha\frac{\partial BN}{\partial x}$, les régularités de BN et de N deviennent comparables, et on obtient une estimation en $\frac{\varepsilon}{\alpha}+\alpha$, où régularisation et bruit s'équilibrent, de telle sorte que le choix optimal est $\alpha=\sqrt{\varepsilon}$ et que l'on arrive à estimer le taux de division à $\sqrt{\varepsilon}$ près, modulo la division par N_ε.

A la suite des travaux de B. Perthame et J. Zubelli [130] et en collaboration avec eux, dans [15], nous avons tout d'abord proposé une deuxième méthode de régularisation possible, consistant à convoler $L(N_\varepsilon)$ avec une suite régularisante. Nous avons également construit une méthode numérique, ce qui a fait apparaître un point délicat: l'équation (2.4) possède en réalité une infinité de solutions au sens des distributions, parmi lesquelles il s'agit de sélectionner la "bonne" - de notre point de vue.

Ce problème apparut de façon plus aigue encore lorsque, en collaboration avec L.M. Tine [16], nous avons généralisé l'étude théorique et numérique ci-dessus à un noyau de fragmentation $k(x,y)$ quelconque. En effet, l'équation (2.4) généralisée s'écrit

$$2\int_x^\infty k(y,x)H(y)dy - H(x) = L(x), \tag{2.5}$$

où le terme $4H(2x)$ est remplacé par un terme intégral non local. On ne peut donc plus considérer cette équation que comme écrite en x, et non plus en $y=2x$. Cela entraîne que le problème se résout bien dans un espace $L^2(x^p dx)$ avec $p>3$, comme précédemment (c'est le cas ou le terme identité, deuxième terme du membre de gauche, domine le terme intégral), mais pas lorsque $p<3$ (où l'on espérerait que le terme intégral domine). D'un point de vue numérique, le problème est le même: contraints d'employer des schémas d'implémentation qui partent de la connaissance de $H(y)$ pour $y\geq x$ pour en déduire la valeur de H en x, une amplification importante du bruit au voisinage de $x=0$ est observée. Dans le travail en cours [19], nous proposons une méthode pour résoudre ce problème dans le cas des fragmentations auto-similaires, *i.e.* si $k(y,x)=\frac{1}{y}k_0(\frac{x}{y})$ avec k_0 une mesure de probabilité sur $[0,1]$.

2.4 Point de vue statistique [11]

Comme vu ci-dessus, nous avons jusqu'à présent considéré un bruit exprimé de façon déterministe comme une erreur de mesure dans L^2, et avons résolu le problème également dans L^2. Nos travaux se généraliseraient d'ailleurs sans difficulté à un bruit dans un espace de Hilbert H^s et à une recherche de solution dans un espace $H^{s'}$; on peut prouver que dans ce cas, sous certaines hypothèses bien choisies de régularité des coefficients, l'estimation optimale est en $0(\varepsilon^{\frac{s}{s+1}})$. En effet, notre problème rentre "presque" dans le cadre général de la théorie des problèmes inverses linéaires, cf. par exemple [57], et les résultats obtenus dans ce cadre général s'adaptent sans difficulté majeure à notre cas (de même que d'autres méthodes de

régularisation, comme le montrent par exemple les travaux de A. Groh, J. Krebs et M. Wagner [73]).

Mais en quoi consiste ce bruit exactement, et comment le modéliser d'une façon qui rende compte au mieux des mesures expérimentales ? Dans [13], nous avions appliqué notre méthode à des données sur les volumes de la bactéries E. Coli, mais en prenant comme point de départ des données pré-traitées [89].

Pour modéliser les mesures d'une population d'individus, dans [11], en collaboration avec M. Hoffmann, P. Reynaud-Bouret et V. Rivoirard, nous avons supposé que nous considérions un échantillon de n cellules, dont les tailles étaient des variables aléatoires indépendantes de loi $N(x)$ la solution de l'équation (2.2). Il s'agit alors, pour proposer une estimation du second membre L de l'équation 2.4, d'estimer la densité $N(x)$ ainsi que le terme dérivé $\frac{\partial}{\partial x}(gN)$; ensuite, le reste de la méthode décrite ci-dessus dans le cadre déterministe reste inchangé.

Nous avons de plus adapté la méthode introduite par Goldenschluger et Lepski [70, 69] qui permet de choisir automatiquement le paramètre de régularisation de façon optimale en fonction de la régularité inconnue de la densité $N(x)$, sous-jacente à l'échantillon. L'équivalent déterministe serait des méthodes comme le *discrepancy principle* [57]. La preuve fait intervenir entre autres des inégalités fines de concentration [102], permettant d'estimer la distance entre la variable aléatoire définissant notre estimateur et son espérance en fonction de la taille n de l'échantillon; cette partie a été réalisée par mes collaborateurs statisticiens.

2.5 Modélisation de la polymérisation des protéines [18]

2.5.1 Modèle du Prion: équation de croissance-fragmentation non linéaire

Ce n'est que récemment que l'équation de croissance-fragmentation ci-dessus (2.1) a été introduite pour modéliser la polymérisation et plus particulièrement la polymérisation des protéines. A ma connaissance, les premiers travaux sont ceux de Greer et al. [72, 71]. Les équations originelles sont en effet des systèmes différentiels, obtenus à partir de la loi d'action de masse. Par exemple, pour modéliser la polymérisation des protéines prion, Masel et al. [101] écrit le système ci-dessous:

$$\frac{dn_i}{dt} = -V(t)\big(g_i n_i - g_{i-1} n_{i-1}\big) - B_i n_i + 2\sum_{j=i+1}^{\infty} B_j k_{i,j} n_j, \tag{2.6}$$

$$\frac{dV}{dt} = \lambda - \gamma V - V\sum_{i=i_0}^{\infty} g_i n_i + 2\sum_{j\geq i_0}\sum_{i<i_0} i k_{i,j} B_j n_j, \tag{2.7}$$

où V désigne la concentration de monomères et n_i la concentration de polymères de taille i, *i.e.* formés par l'agrégation de i monomères. Greer et al. proposent une version continue de

ce modèle dans [134], sous la forme

$$\frac{\partial n}{\partial t} + V(t)\frac{\partial}{\partial x}\big(g(x)n(t,x)\big) + B(x)n(t,x) = 2\int_x^\infty k(x,y)B(y)n(t,y)dy, \tag{2.8}$$

$$\frac{dV}{dt} = \lambda - \gamma V - V(t)\int_{x_0}^\infty g(x)n(t,x)dx, \tag{2.9}$$

$$g(x_0)n(t,x_0) = 0. \tag{2.10}$$

On reconnaît bien dans (2.8) l'équation de croissance-fragmentation, mais elle est ici couplée, via le terme de croissance, à l'équation pour les monomères (2.9). Ici, $n(t,x)$ représente la densité de polymères de taille x à l'instant t, qui grossit par addition de monomères avec un taux de croissance $g(x)V(t)$ proportionnel à la densité de monomères, par la loi d'action de masse. A cela s'ajoute, dans un modèle *in vivo*, le taux de production de monomères λ et les taux de mort γ et μ.

Dans les articles d'origine [101] et [72], dans leurs versions tant discrète que continue, les paramètres $g(x)$, $B(x)$ sont supposés constants, et le noyau de fragmentation est uniforme $k(x,y) = \frac{1}{y}\mathbb{1}_{x\leq y}$. Par sommation, cela permet de réduire le système infini ou l'EDP à un système fermé de trois équations portant sur V, sur le moment d'ordre 0 $P = \sum n_i$, et sur le moment d'ordre 1 $M = \sum i n_i$, qui représente aussi la masse totale polymérisée.

Outre le remplacement des sommes par des intégrales et des différences par des dérivées, il existe une différence majeure entre le système discret (2.6)–(2.7) et le système (2.8)–(2.9): la condition au bord (2.9) n'a pas d'équivalent, mais est indispensable pour que le problème soit bien posé, contrairement au cas discret. Cependant, rien ne justifie a priori - sauf l'intuition d'une condition de flux entrant nul - la condition au bord nulle.

Nous nous sommes donc penchés sur cette question dans l'article [9], en collaboration avec T. Goudon et T. Lepoutre, en nous appuyant sur des travaux portant sur des équations proches (Lifshitz-Slyozov d'une part dans [48], où seul le mécanisme de polymérisation/dépolymériss apparaît ; l'équation de fragmentation-coalescence d'autre part dans [91], où la polymérisation est remplacée par de la coalescence). Nous avons étudié comment déduire le système (2.8)–(2.9) du système (2.6), par un changement d'échelle approprié, par une définition précise de ce qu'est une solution admissible, ainsi que par des estimations sur les moments. Nous nous sommes placés dans un espace de mesures, de façon à inclure le cas de la formation de poussières (*i.e.*, l'apparition d'une masse de Dirac en $x = 0$). Nous n'avons pu obtenir la condition au bord que dans certains cas, celui où $x_0 > 0$ (ce qui est assez improbable d'un point de vue pratique, car cela signifie que la taille du plus petit polymère est déjà très grande par rapport à 1), ou encore celui où la polymérisation s'annule au voisinage de x_0 (ce qui là encore est assez improbable, car il s'agit de l'initiation de la réaction en chaîne), ou encore si le taux de polymérisation est constant près de 0 - cette dernière hypothèse étant raisonnable.

2.5.2 Généralisation: modèles de polymérisation de protéines

Pour intéressant qu'il soit, le modèle dit (assez improprement) "prion" ci-dessus, est un cas particulier par rapport à la complexité des réactions en chaîne possibles. Il suppose par exemple que la polymérisation a lieu par addition de monomères. En réalité, rien n'est moins sûr, H. Rezaei et al. ayant par exemple mis en évidence l'existence de plusieurs voies de polymérisation chez la protéine PrPc (protéine prion *recombinante*, obtenue à partir de bactéries [136]). Ainsi, la protéine PrPc forme tout d'abord des oligomères de taille 12, 24 et 36 monomères, seuls les polymères de taille 36 étant ensuite capable de polymériser en très longs polymères, de taille pouvant aller jusqu'à 10^5 monomères.

Le modèle prion apparaît ainsi comme le modèle-jouet d'une maladie infectieuse où la transmission se fait par le seul jeu de la polymérisation/fragmentation des protéines. En effet, il suffit d'une très faible quantité de polymères initiale pour que la réaction démarre, entraînant la formation de nombreux autres polymères capables à leur tour de transmettre la maladie [43, **?**]. C'est d'ailleurs le sens de l'acronyme PRION, pour PRoteinasceous Infectious ONly (cet acronyme a été inventé par Stanley Prusiner en 1982 [133]). En revanche, le modèle ne peut rendre compte de l'apparition spontanée de la maladie: si aucun polymère n'est présent initialement, rien ne se passe, contrairement à un modèle qui inclue une réaction de *nucléation*, où la rencontre de monomères entraîne la formation spontanée de petits oligomères.

Afin donc de généraliser la dérivation de [9] et de prendre en compte l'apparition spontanée de la maladie, nous avons écrit un modèle plus complet dans [18], modèle destiné non pas à être employé tel quel pour toute protéine s'agrégant en longs polymères mais plutôt exemple d'une méthode à adapter à chaque situation. Ce modèle, dans une version simplifiée, s'écrit:

$$\frac{dV}{dt} = -\frac{i_0\, k_{on}^N\, V^{i_0+1} g(x_0)}{k_{off}^N + g(x_0)V} - V \int_{x_0}^{\infty} g(x)\, n(t,x)dx + \int_{x_0}^{\infty} g^-(x) n(t,x)dx, \tag{2.11}$$

$$\begin{aligned}\frac{\partial n}{\partial t} &= -V\frac{\partial}{\partial x}\big(g(x)n(t,x)\big) + \frac{\partial}{\partial x}(g^-(x)n(t,x)\big) + 2\int_x^{\infty} B(y)k(x,y)\, n(t,y)dy \\ &-B(x)n(t,x) + \frac{1}{2}\int_{x_0}^{x} k_{col}(y,x-y)n(t,y)n(t,x-y)dy - \int_{x_0}^{\infty} k_{col}(x,y)n(t,x)\, n(t,y)dy,\end{aligned} \tag{2.12}$$

$$g(x_0)n(t,x_0) = g(x_0)\frac{k_{on}^N\, V^{i_0}}{k_{off}^N + g(x_0)V}. \tag{2.13}$$

La principale originalité réside dans l'introduction de la nucléation, qui apparaît désormais comme une condition au bord (2.13). Dans [18], nous avons aussi appliqué notre méthode à

la polymérisation de la protéine PolyQ, responsable de la maladie de Huntington. Une comparaison de notre modèle à des données expérimentales nous a permis de mettre en évidence le mécanisme réactionnel de nucléation, qui s'est révélé être un échange conformationnel entre plusieurs espèces de monomères.

2.6 Perspectives

2.6.1 Applications à la polymérisation des protéines

Le modèle (2.11)–(2.13), proposé dans [18], a servi de base à un projet ERC Starting Grant, $SKIPPER^{AD}$, qui se déroule jusqu'en novembre 2017 et structurera fortement mon activité de recherche sur cette période. Il s'agit à la fois d'élargir ma recherche actuelle à des modèles plus généraux, et de l'appliquer à la polymérisation des protéines caractéristique des maladies amyloïdes.

Voici différentes directions de recherche qui font partie de ce projet et sont spécifiques de l'application à la polymérisation des protéines. D'autres sont des problèmes proprement mathématiques, et en tant que tels peuvent ou non s'y rattacher, ayant un champ d'application plus large.

- Analyse de sensibilité du modèle par rapport à ses paramètres

  La question est de savoir quels mécanismes influencent davantage quelle phase de la réaction afin de déterminer quelles données sont les plus fondamentales et quels paramètres sont les plus importants. Les outils utilisés seront les fonctions de sensibilité généralisées [30]. Ce travail se fera en lien avec l'équipe de H. Thomas Banks aux Etats-Unis, et au cours de la thèse de H.W. Haffaf.

- Dérivation rigoureuse du modèle continu

  De la même façon que dans [9], il s'agit de justifier théoriquement le modèle continu obtenu de façon formelle dans [18]. Ce travail pourra être proposé à un étudiant.

- Adaptation du modèle-cadre de [18] à chaque cas

  Les mécanismes en jeu pour les différentes protéines ne sont pas les mêmes: par exemple, on a montré expérimentalement que pour PolyQ il n'y a ni fragmentation ni coalescence. Pour le Prion, ce sont non pas des monomères mais des oligomères qui s'agrègent les uns aux autres. Il s'agit donc d'adapter le modèle général à chacun de ces cas.

- Problèmes inverses

  Une première question consiste à améliorer les résultats de [16] pour l'estimation du taux de division dans l'équation de croissance-fragmentation générale. En effet, comme dit plus haut, les résultats théoriques sont dans $L^2(x^p dx)$ avec $p > 3$, ce qui donne des

résultats numériques de qualité moyenne. Dans le cas d'un noyau de fragmentation auto-similaire, une méthode fondée sur la transformée de Mellin [155] permet des estimations dans tous les $L^2(x^p dx)$ pour $p \geq 0$. Cette étude est l'objet du début de thèse de Thibault Bourgeron (en co-direction avec Benoît Perthame), en collaboration avec Miguel Escobedo [19].

La question de retrouver les paramètres du modèle de [18] pose une série de questions de type "problèmes inverses": comment utiliser au mieux les données? A partir de distribution de tailles de polymères, peut-on appliquer ou adapter les méthodes développées dans [13, 15, 16]? Comment mettre au point des logiciels de simulation utilisables par les biophysiciens?

Ces questions font l'objet du travail de thèse de Hadjer Wafaâ Haffaf, à partir de septembre 2012, en lien étroit avec l'équipe d'Human Rezaei, ainsi qu'avec l'équipe-projet MACS (Philippe Moireau et Marc Fragu).

2.6.2 Interactions statistiques et analyse

Cette partie de ma recherche a commencé par des discussions avec Marc Hoffmann, Patricia Reynaud, Vincent Rivoirard, au sein d'un groupe de travail informel que nous avions organisé au cours de l'année 2008-2009. Ces discussions avaient pour point de départ un certain parallélisme entre les travaux des statisticiens et les problèmes inverses déterministes: les théorèmes et méthodes présentent de nombreuses similarités, sans pour autant qu'une des approches soit réductible à l'autre.

Suite à notre premier article [11], cette collaboration prend maintenant deux directions.

D'une part, j'ai été associée à un projet ANR blanc (projet "CALIBRATION") 2011-2014 dirigé par Vincent Rivoirard, puis au projet ANR blanc "PIECE" de Florent Malrieu 2012-2015. Dans ce cadre nous pourrons donner plusieurs suites à ce travail: méthode de Lepski dans un cadre déterministe ou à l'inverse méthode du *discrepancy principle* dans un cadre statistique [1]; méthodes de reconstruction lorsque le niveau de bruit n'est pas connu ; mise en équation puis analyse de modèles statistiques de neurosciences [126] ou de génétique [135] et, de façon plus générale, des processus de Hawkes [78, 41].

D'autre part, nous travaillons maintenant en lien avec Lydia Robert, biologiste à l'ENS, qui dispose de données expérimentales sur la croissance des bactéries, est très intéressée par une interaction avec les mathématiques, et nous a soumis une série de problèmes fascinants en lien direct avec les méthodes mathématiques développées. Pour la partie la plus théorique, nous proposons dans [10] (non inclus dans ce mémoire) un modèle probabiliste complet au niveau microscopique, modèle qui converge macroscopiquement vers une dynamique décrite par les EDP de croissance-fragmentation. Diverses extensions de ce modèle sont possibles, ce qui permet une compréhension unifiée des modèles: c'est le travail de thèse d'Adélaïde

[1] Il s'agit de méthodes de reconstruction, l'une statistique, l'autre déterministe, qui s'adaptent automatiquement à la régularité de la fonction

Olivier, en co-direction avec Marc Hoffmann. Ensuite, dans un travail en cours nous testons nos méthodes de reconstruction sur les données expérimentales de Lydia Robert [157, 152] et interprétons nos résultats afin de sélectionner le modèle adéquat. Enfin, nous enrichissons au fur et à mesure nos méthodes pour répondre à des problématiques plus riches soulevées par Lydia Robert - prise en compte de la variabilité entre cellules, structure en ADN.

2.6.3 Dynamique des populations

A la suite des travaux de [4, 7, 2], de nombreux problèmes restent ouverts, certains particulièrement ardus. Je cite ici deux voies qui me semblent particulièrement importantes.

- Dynamique en temps long

  Le comportement en temps long du mécanisme de polymérisation régit les expériences in vitro. La difficulté mathématique de son étude réside dans le couplage non linéaire des équations (2.8)–(2.9) (ce couplage quadratique pourrait d'ailleurs devenir polynomial dans le cas où ce seraient des oligomères et non plus des monomères qui s'agrègeraient). Numériquement, on observe que la population tend toujours vers un état stationnaire infectieux. De premiers résultats théoriques ont été obtenus dans [43, **?**], concernant la stabilité ou l'instabilité de zéro, à l'aide de l'entropie relative généralisée, mais le problème de la convergence vers un état stationnaire non nul demeure. De manière générale, cela rejoint la question du comportement asymptotique de nombreux problèmes non linéaires en dynamique des populations - question encore largement ouverte (cf. [129, 106, 109, 67, 42, 24, 123, 124, 125]).

- Autres convergences asymptotiques

  Dans [7], nous avons donné des conditions, les plus générales possibles, pour que l'équation linéarisée (2.12) (*i.e.* découplée de l'équation pour V (2.11)) ait comme comportement asymptotique une croissance exponentielle avec un profil stable, *i.e* $u(t,x) \to U(x)e^{\lambda t}$. D'autres articles ([60, 61, 63] étudient au contraire des cas où des équations proches (de fragmentation ou de coagulation) soient tendent vers un Dirac en zéro (phase poussière), soient forment en temps fini des polymères infinis (gélation, possible uniquement si de la coagulation s'ajoute à l'équation). Il resterait à achever la théorie pour avoir une vision complète de tous les comportements possibles: quelles sont les hypothèses optimales pour chaque cas ? Quels sont les cas limites, et que s'y passe-t-il ?

Part II
Presentation of Research (English version)

1 Structured population equations

1.1 A brief history and some philosophy

1.1.1 Structured populations in biology

Let us consider a population, *i.e.*, a group of individuals that grow, change, live, reproduce and die. How can we follow and understand its evolution over time? The field of "structured population equations" has been developed for this purpose: describe adequately a population dynamics, not only by means of average quantities, but by its distribution along "structuring" variables. Such parameters are termed *structuring* because they are supposed to be well-chosen traits that characterize the individuals' behaviour.

The first "structured equations" go back to the beginning of the twentieth century, with the work of Sharpe and Lotka [144], Mc Kendrick [103], Kermack [83, 84] on an age structure. In 1911, in a discrete setting, Sharpe and Lotka had paved the way by establishing the major facts concerning such a population [144]: when birth and death are independent of time, the age distribution does not depend asymptotically on the initial distribution, and tends to a steady profile. The proof relies on the Perron-Frobenius theorem. In one of its variants, such an equation can be written as follows:

$$\begin{cases} \frac{\partial}{\partial t}n(t,a) + \frac{\partial}{\partial a}n(t,a) = -\mu(a)n(t,a), \\ n(t,a=0) = \int_0^\infty B(a)n(t,a)da, \end{cases} \tag{1.1}$$

where $n(t,a)$ denotes the density of cells of age a at time t, $B(a)$ the birth rate and $\mu(a)$ is the death rate. The case where a cell divides into two offspring can be considered as a particular case of Equation (1.1), where $\mu(a)$ is replaced by $\mu(a) + B(a)/2$. Size structure was introduced only in the second half of the twentieth century, with the work of Bell and Anderson [36, 37], Sinko and Streifer [148, 149]. The size-structured equation can be written in full generality as follows:

$$\begin{cases} \frac{\partial}{\partial t}n(t,x) + \frac{\partial}{\partial x}\big(g(x)n(t,x)\big) = -\mu(x)n(t,x) - B(x)n(t,x) + 2\int_x^\infty k(x,y)B(y)n(t,y)dy, \\ g(x=0)n(t,x=0) = 0, \qquad n(t=0,x) = n^0(x). \end{cases} \tag{1.2}$$

In this equation, $n(t,x)$ denotes the density of cells of size x at time t, which can die with a rate $\mu(x)$, or divide with a rate $B(x)$, giving rise to one individual of size y and one of size

$x - y$ with a probability kernel $k(y, x)$. There exist several variants of this equation, and it can also be generalized to non binary fragmentation. Also, what is called "size" may refer to any physical quantity of the individual such as volume, length, protein or parasite content.

The nomenclature has not yet been unified and remains a little confusing even today, due to the fact that refering to a "size" or "age" or "protein" structure refers to the biological application rather than to the mathematical structure of the model. Size-structured models are an illuminating example: in a wide range of articles, *e.g.* [20, 64, 29], what is called a "size structure" rather refers to what I would call an age structure [2], since the size at birth is the same for any cell, and thus can be set at zero without lack of generality - exactly as in an age model. In such a way, age-structure appears as a particular case of the growth-fragmentation equation, where $k(x, y) = \frac{1}{2}(\delta_{x=0} + \delta_{x=y})$.

To avoid such confusion, I usually refer to the size-structured model as the growth-fragmentation equation, as it is called when applied to physics - "transport-fragmentation" or "drift-fragmentation" would maybe appear even clearer, since growth could occur in different ways, for instance by coagulation of individuals.

Simultaneously to the size-structured model, other *physiologically* structured equations emerged, such as age-size [148], maturity [139], satiety [104], age and generation-time [140].

1.1.2 Coagulation-fragmentation equations in physics

As concerns applications in physics, the growth-fragmentation equation may be viewed as a particular case of coagulation-fragmentation models. These were initially written in a discrete way, as infinite systems of ordinary differential equations [146, 33, 150, 151]. Such systems consist of the equations satisfied by c_i, the concentration of the clusters containing i objects (this object's concentration, when unbounded to any other, is denoted c_1 and it may be a dust, a unitary particle or a monomer). In 1917, Smoluchowski proposed the first coagulation model for colloidal solutions. In 1935, R. Becker and W. Döring proposed a system which modelled the interaction between monomers and large clusters, in the theory of nucleation of liquid droplets in solids, and in 1941, R. Simha wrote a discrete system for the degradation (fragmentation) of long chain polymers.

When the average sizes of the clusters are large enough for them to be coarse-grained and viewed as a continuous variable, continuous PDE were derived from these models. In 1961, I.M. Lifshitz and V.V. Slyozov proposed such a model to approximate the Becker-Döring system [99]. It consists in a transport equation for the clusters (structured in size as a one-dimensional continuous variable), quadratically coupled with an ODE for the monomers, see [48][3]. Similar coarse-graining was also performed on coagulation equations [53, 21, 91].

To the best of my knowledge, the discrete growth-fragmentation equation, when applied

[2]The age-structured equation (1.1) could be written with a variable "aging speed" $g(a)$ not necessarily equal to 1.

[3]Taking into account the second-order term in the asymptotic development leads to a diffusion term [49].

to physical phenomena, initially appeared in the context of protein polymerization [122]. Fragmentation was viewed as a secondary process, which balances and speeds-up the polymerization dynamics [39, 132]. More recently, it was applied to an *in vivo* prion model [101], for which a continuous version was proposed by Greer et al. [72].

1.1.3 Some philosophy

Before writing any equation, we need to answer the following question: how can we choose the appropriate structuring variables for the population under concern? Which quantity gives access to the effective behaviour of the individuals? The answer depends on each specific situation. I would like to focus on a specific difficulty when using abstract variables, for which no measure or easy relation to a physical quantity is available, to serve as a structure. A telling example concerns models for maturing/differentiating processes.

An initial model of maturity-structured population was proposed in [139] in 1968. Since then, many studies have been carried out, producing various types of models, from individual-based [138] to PDE through differential equations [55], each of these models considering one specific situation (haematopoiesis probably being the most prolific). Their aim is often to provide a qualitative explanation or even a prediction for experimental observations.

Modelling questions arise; for instance, are discrete modelling better, equivalent or less suitable than continuous ones? This was one of the questions we investigated in [14], where we compared the discrete model written in [100, 153] with its natural continuous counterpart. It was clear that the two kinds of models were far from being equivalent, even qualitatively, since they exhibit different asymptotic behaviours.

As a second example of the difficulties of modelling maturity (or "satiety" [104] or any abstract quantity), let us consider the individual-based model (IBM) built by Roeder et al. [138]. In [86], P. Kim, P. Lee and D. Levy derived a PDE model that perfectly fits the IBM. We investigated the PDE model theoretically and numerically in [12]. In [137], Roeder *et al.* also derived an approximate PDE model, that turned out not to be the exact equivalent of the one in [86]. This difference is explained by the following elementary considerations.

The derivation of conservation laws is well-known in physics, where the variables are time, denoted by t, the particles' speed v, and a space variable x. The equation for the space density $n(t, x, v)$ can be derived in two equivalent ways. The first one consists in considering a small fixed volume $V = dx^3$, and establishing the variations of a quantity $N(t, x)$ during a small time interval dt, then letting the small quantities dx and dt vanish. The second one consists in following a small fixed quantity $n(t, x(t)).V(t)$ on a small interval of time dt along the characteristic curves. I must stress the fact that we do not directly consider the density $n(t, x)$, but rather the physically meaningful object which is a *quantity* $n(t, x)V(t, x)$, V being a volume.

In biology, the above-mentioned structured equations are also derived as conservation laws, or, more accurately, as balance laws since there is not conservation [127], in the same

spirit as in physics, but giving other meanings to the variable x. As seen above, this variable can represent a somewhat physical quantity, which is easy to imagine if not to measure, like age (increasing over time), size, weight or position. But it can also represent an abstract quantity: an aggregated variable representing a protein content for instance, or maturity, affinity, satiety [104]. For such abstract quantities, the conservation law can be derived in the same way as in physics, but the concept of a "density per unit of satiety" for instance is quite abstract... For this reason, one has to bear in mind that what is experimentally observed is not the density $n(t,x)$ but the quantity $n(t,x).dV$.

In the example of Ingo Roeder's model (I'll skip all the details and only keep what is useful here: the maturity structure), we have an individual-based model, discrete in time, where the maturity variable is called the *affinity*. It can either decrease by a factor $1/d$ in a certain region (called *niche*), or increase by a factor r in another one. Let us denote a the affinity, $n_a(t,a)$ the density "per unit of affinity" of the number of cells. The general form of the mass conservation equation (in the case of no source term) is then:

$$\frac{\partial}{\partial t}n_a(t,x) + \frac{\partial}{\partial t}(v_a(t,a)n_a(t,a)) = 0, \tag{1.3}$$

where $v_a(t,a) = \frac{da}{dt}$ represents the instantaneous velocity of the density $n_a(t,a)$. If we make the change of variables $x = g(a)$ where g is monotonous, and if we look for the mass balance equation in this new abstract variable x (*a priori* no better or worse than a), we will obtain the same general equation:

$$\frac{\partial}{\partial t}n_x(t,x) + \frac{\partial}{\partial t}(v_x(t,x)n_x(t,x)) = 0, \tag{1.4}$$

where $v_x(t,x) = \frac{dx}{dt}$ represents the instantaneous velocity of the density $n_x(t,x)$. There is an easy link between v_x and v_a given by $v_x = \frac{dx}{dt} = \frac{dx}{da}\frac{da}{dt} = g'(a)v_a$. The sensitive point is not to forget the indices when writing n_x and n_a, since these two quantities do not represent the same thing. In [137], the equation was written in the variable a, whereas in [86] it was written for $x = Log(a)$ due to the exponential behaviour of the affinity.

Both equations are correct, but do not refer to the same object. The question is now: How can we compare the results obtained with each other, and how can they be compared with those obtained in the IBM? It is possible through comparison of *quantities* rather than *densities*, *i.e.* compare $\int_{a_1}^{a_2} n_a(t,a)da = \int_{x_1}^{x_2} n_x(t,x)dx$, if we choose $x_1 = g(a_1)$, $x_2 = g(a_2)$. In the case of the IBM of Ingo Roeder's model, it leads to comparing $N(t,x)$ with either $n_x(t,x)dx$ or $an_a(t,a)da$, due to the exponential behaviour of the affinity, leading to an infinitesimal volume ada. Here, the "physical" reference is given by the IBM; this illustrates the unavoidable role of arbitrary choices in such modelling.

1.2 Nonlinear maturity-structured equations [12, 14]

In this section, I will briefly review the results of the two articles [12, 14], concerning models of structured populations of the family of age-structured equations.

1.2.1 Modelling differentiation through nonlinear age-structured equations

Cell differentiation is a process by which dividing cells become specialized and equipped to perform specific functions such as nerve cell communication or muscle contraction. Differentiation occurs many times during the growth of a multicellular organism as the organism changes from a single zygote to a complex system with cells of different types. Differentiation is also a common process in adult tissues. During tissue repair and during normal cell turnover a steady supply of somatic cells is ensured by proliferation of corresponding adult stem cells, which retain the capability for self-renewal. The two following models aim at modelling two different aspects of this stem cell self-renewal capability.

Modelling stem cells and haematopoeiseis in chronic myelogeneous leukemia (CML) [12]

Various cancers are likely to originate from a population of cancer stem cells that have properties comparable to those of stem cells, as proposed in a new paradigm by Bonnet et al [40]. This hypothesis states that a variety of cancers originate from a self-replenishing, cancer population, now known as cancer stem cells. Using this idea, Roeder et al. built a mathematical model of chronic myelogeneous leukemia (CML) stem cells [138]. In their model, leukemia stem cells continually circulate between proliferating and quiescent states. This formulation contrasts with the alternative paradigm of Michor et al. in which leukemia cells differentiate progressively from stem cells to differentiated cells without circulating back to previous and more dormant states [111, 110]. Both the Roeder and Michor models are directed to studying the dynamics of imatinib treatment. However, each model also presents a general paradigm for haematopoiesis that describes blood cell development with or without chronic myelogeneous leukemia.

Starting from the original agent-based model formulated by Roeder et al. [138], a complete PDE formulation was made in [85], and proved to coincide very well with the simulation results obtained by Roeder et al. In order to make this PDE model amenable to analysis, we simplified it as much as possible, distinguishing between four successive approximation steps, and justifying our approximations by numerical comparisons. It lead us to the conclusion that a conceptually important feature of Roeder et al. IBM, namely that the maturing process of early stem cells may be reversible, can be simplified into a kind of renewal equation without any loss in the quantitative results: instead of a backward transport equation, this means that cells entering the quiescent *niche* are instantaneously renewed, *i.e* their maturity restarts from 0.

The theoretical analysis of the asymptotic behaviour of the simplified PDE model proved that it possessed the same properties than the original agent-based model: according to the parameters' region, either 0 is an attractive steady state, or there is a unique positive

steady state, either stable or unstable (in which case periodic solutions exist). The proofs for stability of the trivial steady state rely on a well-chosen entropy (Lyapunov) functional for the nonlinear system. As concerns the (linearised) stability of the non trivial equilibrium, the nonlinear system does not seem to possess any entropy, but numerical computations of the spectrum gave the stability regions.

The simplified model is thus of value in understanding the dynamics of haematopoiesis and of chronic myelogeneous leukemia, and it presents the advantage of having fewer parameters, which makes comparison with both experimental data and alternative models much easier. In its simplest form (Approximation 4 of [12]), it is stated as follows:

$$\frac{\partial \Omega}{\partial t} + \rho_d \frac{\partial \Omega}{\partial x} = (-\alpha(A) + b)\,\Omega(t,x), \tag{1.5}$$

$$\frac{dA}{dt} = -\omega\left(\int_0^1 \Omega(t,x)dx\right)A(t) + \alpha(A)\int_0^1 \Omega(t,x)dx, \tag{1.6}$$

with boundary condition

$$\Omega(0,t) = \frac{\omega\left(\int_0^1 \Omega(t,x)dx\right)}{\rho_d}A(t). \tag{1.7}$$

Here, $\Omega(t,x)$ represents the maturing cells density at time t of maturity x, and ρ_d is the maturing velocity. The cells in the Ω compartment reproduce with an average fixed rate b and exchange with the quiescent state A with a rate $\alpha(A)$ which decreases with the density of quiescent cells A. This dependence models the fact that the exchange is possible only if there is enough available room in the quiescent *niche*. In their turn, dormant cells in the A compartment can go back to the maturing compartment, starting from 0 again, with a rate ω depending on the total quantity of maturing cells $\int_0^1 \Omega(t,x)dx$, modelling here again that the rate is room-dependent. When reaching $x = 1$, the cells commit to differentiation, do not transfer back to the A compartment and quit the system.

This system can be thought of as a variant of an age-structured model, due to the nonlocal integral term in Equation (1.6), even if it is not directly in a Dirichlet boundary condition but in its time derivative.

From discrete to continuous models of differentiation [14]

P. Kim *et al.* model focused on the spontaneous self-replenishing of stem cells, with nonlinear loops only from stem cells to stem cells. I presented only the simplest case; the following step consisted in adding a cancer stem cell compartment and study the competition between normal and cancer stem cell, with or without a chemical treatment.

To model the case of feedback coming from mature cells (which is proved to occur for instance through the signalling process due to cytokines), what is useful for another type of applications, A. Marciniak *et al.* proposed the following model (here in a simplified version:

System (18)–(21) of [14]):

$$\begin{aligned}
\frac{d}{dt}w(t) &= \alpha(v(t))w(t), && (1.8)\\
\partial_t u(t,x) + \partial_x\big[g\big(x,v(t)\big)u(t,x)\big] &= p(x)u(t,x), && (1.9)\\
u(t,0) &= w(t), \quad t>0, && (1.10)\\
\frac{d}{dt}v(t) &= g(v(t),x^*)u(t,x^*) - \mu v(t), && (1.11)
\end{aligned}$$

together with initial data. Here, $u(t,x)$ (the counterpart of $\Omega(t,x)$ of the previous model) is the density of maturing cells at time t of maturity x, and the maturing speed $g(x,s(v))$ depends also on the quantity of mature cells $v(t)$. Mature cells v can only die, they don't reproduce anymore. Immature cells (stem cells, counterpart of both $\Omega(t,0)$ and A) proliferate or die with a rate α also depending on v, and is a source term for u.

First of all, we linked this continuous model with the discrete model proposed in [153], showing a weak convergence of the discrete model to the continuous when the number of compartments tends to infinity under appropriate assumptions on the orders of magnitude of the coefficients. This was done in the same spirit than [9], and I refer to Section 5 for more details. Under continuity and positivity assumptions on g, and mainly supposing that $\alpha(\cdot)$ is a decreasing function with $\alpha(+\infty) < 0$, we proved boundedness for the system, by the use of the intermediate functional $\partial_x(ln(u))$. By the use of a Lyapunov/entropy functional, we proved extinction if $\alpha(0) < 0$, persistence if $\alpha(0) > 0$. When there exists a positive steady state, we treat it in a similar manner than in [12], studying the linearized problem around this steady state; we showed that Hopf bifurcations occur, so that it can be either stable or unstable. We illustrated this by simulations.

Small discussion and comparison between the models of [12] and [14] Though the maturity structure is basically the same in both models, leading to simple transport equations (1.5) and (1.9), the focus is different. For P. Kim's model, originated in Ingo Roeder's IBM, the idea is to model the self-replenishment of early stem cells. When cells of the Ω compartment reach $x = 1$, they are not fully differentiated, but they begin their differentiation process and cannot go back to stem cells. The paradigm of stem cell *niches*, as places either more favourable to growth and preparation to differentiation (compartment Ω) or more favourable to quiescence (compartment A), and the fact that the exchange between these compartments leads to self-regeneration, is specific to this model, and to the best of my knowledge was first proposed by I. Roeder. Till now, no experimental evidence has proved or contradicted such an assumption, but the model turned out to fit properly data about CML cells (though only indirect measures, through differentiated cells, are possible).

A. Marciniak's model focus on what follows this early stage. In a sense, one could plug both models, in which case P. Kim's one would replace $w(t)$ in A. Marciniak's model, and we set $w(t) = \Omega(t, x = 1)$. A. Marciniak's model aims at modelling the feedback loop that

exists (through signalling molecules like cytokins) between the amount of fully differentiated cells $v(t)$ and stem cells $w(t)$ and $u(t,x)$, what is not present in Ingo Roeder's model.

1.3 Asymptotic analysis for a nonlinear age-size structured model [4]

My very first study when I restarted research in Inria in 2007 consisted in the analysis of the eigenvalue problem for the age-size structured equation, namely:

$$\begin{cases} \frac{\partial}{\partial t}n(t,a,x) + \frac{\partial}{\partial a}n(t,a,x) + \frac{\partial}{\partial x}\big(g(a,x)n(t,a,x)\big) + B(a,x)n(t,a,x) = 0, \\ n(t,a=0,x) = 2\iint_0^\infty k(y,x)B(a,y)n(t,a,y)dyda. \end{cases} \tag{1.12}$$

I proved existence and uniqueness of a positive eigentriplet (λ, N, ϕ) to this problem and its adjoint, under assumptions which include cases where the growth rate $g(a,x)$ vanish. As for the growth-fragmentation equation (see Section 2.1 below), the proof first relies on the Krein-Rutman theorem applied to a regularised problem. To formulate it in a convenient way as a fixed point problem, I used the method of characteristics, with the age variable playing the role of time. In fact, the recursive estimates that we used in [7] for the growth-fragmentation equation are no more valid in higher dimensions, so that other techniques have to be used.

Among its numerous applications, this study was motivated by a two-compartment system proposed by F. Bekkal Brikci, J. Clairambault and B. Perthame in [34, 35] to model the cell cycle. The above-mentioned equation is satisfied by proliferating cells, and the variable x models a protein content. These proliferating cells exchange with a quiescent compartment, where cells do not undergo any change (neither in their age nor in their protein content): they can only turn back to the proliferating compartment. This model presents a nonlinear feedback: the exchange rate G, denoting the rate at which proliferating cells become quiescent, is supposed to depend on the total (weighted) population N, and to decrease with it as $\alpha_2 + \frac{\alpha_1}{1+N^n}$ for some α_1, α_2, $n > 0$.

By the use of relative entropy inequalities [108, 34], using weight functions coming from the linearised adjoint eigenproblem around infinity (*i.e.* for $\tilde{G} = G(\infty)$), I showed that if the quiescent cells do not die and if $\tilde{G} = 0$ then the asymptotic regime for the total population is a robust polynomial growth. Such a growth is in accordance both with experimental results and with individual based simulations [54].

2 The growth-fragmentation equation

The core of my research since 2007 is the linear growth-fragmentation equation, namely:

$$\frac{\partial}{\partial t}n(t,x) + \frac{\partial}{\partial x}(g(x)n(t,x)) + B(x)n(t,x) = 2\int_x^\infty k(x,y)B(y)n(t,y)dy, \quad x > 0, \tag{2.1}$$

with an initial condition $n(t=0,x) = n^{in}(x)$ and a boundary condition $g(0)n(t,0) = 0$. Though very simple to write - this is a one-dimension linear integro-PDE, this is a very rich problem. With various collaborators I studied it from different points of view: eigenvalue problem, self-similarity, inverse problem, link with fragmentation processes in statistics and probability.

The density of the population is given by $n(t,x)$ at time t of size x. The growth rate g depends on x, as well as the division rate B. When division occurs at size y, the probability density to give birth to two children of respective sizes x and $y-x$ is $k(x,y)$.

A particular case, and historically the first to have been studied, is given by the cell division equation with equal mitosis:

$$\frac{\partial}{\partial t}n(t,x) + \frac{\partial}{\partial x}(n(t,x)) + B(x)n(t,x) = 4B(2)n(t,2x), \quad x > 0, \tag{2.2}$$

which corresponds to Equation (2.1) with $g(x) \equiv 1$ and $k(x,y) = \delta_{x=\frac{y}{2}}$.

2.1 Eigenvalue problem [7]

2.1.1 Asymptotic behaviour and the eigenproblem

Growth and fragmentation have two opposite dynamics: growth leads the population towards larger sizes, whereas fragmentation leads to smaller and smaller sizes. We can roughly distinguish three cases.

- Fragmentation dominates: ultimately, $g(x) \equiv 0$. It fits then the pure fragmentation equation, studied for instance in [63, 42]. In such a case, under proper assumptions on the fragmentation rate near zero, all the population concentrates in a Dirac mass δ_0.

- Growth dominates: ultimately, $B(x) \equiv 0$. The population sizes grow more and more, following the pure transport equation.

- In a certain way, detailed below, there is a balance between growth and division, so that the population is maintained around finite sizes.

For our applications to biological populations, the third case is the most interesting, since as it is easy to imagine cellular sizes can neither grow to infinity nor vanish. In this case, under suitable assumptions, the solution n to (2.1) tends asymptotically to a stationary profile $N(x)$ times a time exponential $e^{\lambda t}$. This convergence is not only a mathematical result but a biological fact [89].

To the best of my knowledge, the first results on the asymptotic behaviour lie in the biophysical articles of Bell and Anderson [37, 36] in the late sixties. Detailed mathematical proofs followed in the eighties [79, 80, 52]. They were based on dynamical systems analysis, with semi-group methods and with the help of the Krein-Rutman theorem [88].

More recently, the asymptotic behaviour has been proved by the use of two fundamental methods: first, the existence and uniqueness of a positive dominant eigenvalue linked to a positive eigenvector; second, to the range of inequalities called *General Relative Entropy* by B. Perthame and co-authors [107, 108, 127].

The first question is to find optimal assumptions for existence and uniqueness of this first eigenvector: we want to show under which conditions there exists a unique triplet (λ, N, ϕ) with $\lambda > 0$, $N \in L^1(\mathbb{R}_+)$, $N \geq 0$ solution of the following problem

$$\begin{aligned}
&\lambda N(x) + \tfrac{\partial}{\partial x}\big(g(x)N(x)\big) + B(x)N(x) = 2\int_x^\infty k(x,y)B(y)N(y)dy, \quad x > 0,\\
&-g(x)\tfrac{\partial}{\partial x}(\phi(x)) + \big(B(x)+\lambda\big)\phi(x) = 2B(x)\int_0^x k(y,x)\phi(y)dy, \quad x > 0, \qquad (2.3)\\
&g(0)N(0) = 0, \quad \phi(x) \geq 0,\ N(x) \geq 0, \quad \int_0^\infty N(x)dx = \int_0^\infty \phi(x)N(x)dx = 1.
\end{aligned}$$

The existence and positivity of the adjoint vector ϕ is fundamental to prove relative entropy inequalities, where ϕ appears as the suitable weight to counterbalance the terms coming from N (see [127] for a complete explanation, or [107, 108] for the original articles).

Following B. Perthame and L. Ryzhik [128], and P. Michel [105], with P. Gabriel, we investigated what could be the optimal assumptions for this eigenvalue problem [7]. In particular, these assumptions include cases where the growth rate $g(x)$ vanishes around 0. We proved the following result.

Theorem 1 (Theorem 1 of [7])**.** *Under assumptions (5)–(13) of [7], there exists a unique solution* (λ, N, ϕ) *to the eigenproblem* (2.3) *with* $\lambda > 0$*, and we have*

$$x^\alpha gN \in L^p(\mathbb{R}^+), \quad \forall \alpha \geq -\gamma, \quad \forall p \in [1,\infty], \quad x^\alpha gN \in W^{1,1}(\mathbb{R}^+), \quad \forall \alpha \geq 0$$

$$\exists k > 0 \ s.t. \ \frac{\phi}{1+x^k} \in L^\infty(\mathbb{R}^+), \quad g\frac{\partial}{\partial x}\phi \in L^\infty_{loc}(\mathbb{R}^+).$$

2.1.2 "optimal" assumptions for the eigenproblem (Theorem 1)

Without entering into too much detail, here are the main assumptions. Concerning the fragmentation kernel, for modelling reasons we assumes symetry (what could be relaxed), and we exclude the case of the age-structured equation [4] by assuming (Assumptions (5)–(7) of [7])

$$\int k(x,y)dx = 1, \ \int xk(x,y)dx = \frac{y}{2}, \ \int \frac{x^2}{y^2}\,k(x,y)dx \leq c < 1/2. \qquad (2.4)$$

[4] For studies of the age-structured equation, see [127, 22, 74] for instance.

Then, Assumptions (8)–(10) of [7] are positivity or regularity assumptions on B and g. A space $\mathbb{P}$ of at most polynomial growth or decay is defined by

$$\mathbb{P} := \Big\{ f \in L^1_{\rm loc}(\mathbb{R}^*_+),\ f \geq 0\ :\ \exists\, \mu, \nu \geq 0,\ \limsup_{x\to\infty} x^{-\mu} f(x) < \infty,\ \liminf_{x\to\infty} x^{\nu} f(x) > 0 \Big\}, \quad (2.5)$$

and Assumptions (8)–(10) are given by

$$B \in L^1_{\rm loc}(\mathbb{R}^*_+) \cap \mathbb{P}, \quad \exists\, \alpha_0 \geq 0,\ g \in L^\infty_{\rm loc}(\mathbb{R}_+, x^{\alpha_0} dx) \cap \mathbb{P}, \quad (2.6)$$

$$\forall\, K \text{ compact in }]0, +\infty[,\ \exists\, m_k > 0 : g(x) \geq m_k\ \forall x \in K, \quad (2.7)$$

$$\exists\, b \geq 0,\ \text{supp} B = [b, +\infty). \quad (2.8)$$

The three last assumptions (11)–(13) of [7] are the most important, and seem to characterize more profoundly the behaviour of the equation. They link the three coefficients g, B and k in an intricate way, so that under these assumptions, an extra hypothesis on one of these three parameters has an influence on the others. They can be considered as "optimal" in the sense that we can exhibit counterexamples as soon as one of them is not satisfied.

To avoid shattering (zero-size polymers formation, see [27, 93]), we assume (Assumption (11) of [7])

$$\exists\, C > 0, \gamma \geq 0 \quad s.t. \qquad \int_0^x k(z,y)\, dz \leq \min\Big(1, C\Big(\frac{x}{y}\Big)^\gamma\Big) \qquad \text{and} \qquad \frac{x^\gamma}{g(x)} \in L^1_0 \quad (2.9)$$

where we denote L^1_0 the space of locally L^1 functions near zero. This assumption links implicitely g to k, and does not exclude the case $\gamma = 0$ as soon as $\frac{1}{g} \in L^1_0$, *i.e.* g is not too small around zero. We also assume (Assumption (12) of [7])

$$\frac{B}{g} \in L^1_0, \quad (2.10)$$

another way to say "g is not too small around 0 compared to B".

On the other hand, to avoid forming infinitely long polymers (gelation phenomenon, see [60, 62]), we assume

$$\lim_{x\to+\infty} \frac{xB(x)}{g(x)} = +\infty. \quad (2.11)$$

Contrarily to [128] or [127], we do not obtain exponential decay for N when $x \to \infty$; this is due to Assumption (2.11). Such a decay requires a slightly stronger assumption than (2.11), namely that $\liminf_{x\to\infty} \frac{B}{g} > 0$: see [24] for more details.

2.1.3 Some ideas for the proof

The proofs first rely on the positivity properties of the equation, linked to the fact that they are mass balance equations. We can thus apply the Krein-Rutman theorem (see [88] for the original article, or [51] ch.VIII for a recent proof and applications) to a regularised and truncated equation, that easily converges to a truncated positive equation. We then prove recursively a series of estimates, each relying on the previous ones. The main tool is to multiply the (truncated) equation by weights x^α and integrate it. We begin with α large enough to uniformally bound $\int x^\alpha B(x)N(x)dx$ thanks to Assumption (2.11). We then take $\alpha = 0$ to bound $\int B(x)N(x)dx$ thanks to both the previous estimate and Assumption (2.10). It implies L^∞ bounds for $x^\alpha g(x)N(x)$ for $\alpha \geq 0$; finally, to ensure compactness of the regularised sequence of solutions in L^1, we prove a uniform bound for $\alpha = -\gamma$, using all the above estimates and Assumption (2.9).

For the adjoint eigenvector ϕ, we use a maximum principle to obtain boundedness in $W^{1,\infty}_{loc}(\mathbb{R}^*_+)$.

These recursive estimates make use of all the structure of the equation, and proved useful for other studies. In [42, 24], they have been used to obtain a spectral gap, under some extra assumptions concerning in particular the fragmentation kernel. In [16], we used them to show the regularity of the operator $\Gamma : (g, B) \to (\lambda, N)$.

2.2 Self-similar behaviour [2]

As seen above, the first eigenvalue solution of the eigenproblem is the asymptotic exponential growth rate of the population. As such, it governs the long-term behaviour of the population, and is often called the *Malthus* parameter or the *fitness* of the population. Hence it is of great interest to know how it depends on the coefficients: for given parameters, is it favourable or unfavourable to increase fragmentation? Is it more efficient to modify the transport rate g or to modify the fragmentation rate B? Such concerns may have a deep impact on therapeutic strategy (see [34, 35, 46, 4]) or on the design of experimental protocols such as PMCA[5] (see [94] and references therein). Moreover, when modelling polymerization processes (see Section 5), Equation (2.1) is quadratically coupled with the density of monomers $V(t)$, which appears as a multiplier for the growth rate $g(x)$. The dynamics of $V(t)$ is governed by one or more ODE - see for instance [43, 72, 94]). The asymptotic study of such nonlinear versions of Equation (2.1) closely depends on such a dependence: in [43, **?**], the asymptotic results are obtained under the assumption of a monotonic dependence of λ with respect to the growth rate $V(t)g(x)$.

Based on simple previous case studies [66, 72, 59, 134], where $g(x) = 1$, $B(x) = x$ and $k(x, y) = 1/y$ for instance, it might be intuitively assumed that the eigenvalue λ always

[5] *PMCA*, Protein Misfolded Cyclic Amplification, is a protocol designed to amplify the quantity of prion protein aggregates due to periodic sonication pulses. In this application, u represents the density of protein aggregates and x their size; the division rate B is modulated by ultrasound waves.

increases when growth or when fragmentation increases. Nevertheless, we proved in [2] that this is not the case.

To study the dependence of the eigenproblem on its parameters, we fixed coefficients $g(x)$ and $B(x)$, and studied how the problem is modified under the action of a multiplier of either the growth or the fragmentation rate. We thus considered two problems, that we treated simulaneously: first, we modulate the growth rate $g(x)$ by a parameter $\alpha > 0$,

$$\begin{cases} \alpha \dfrac{\partial}{\partial x}(g(x)N_\alpha(x)) + (B(x) + \lambda_\alpha)N_\alpha(x) = 2\displaystyle\int_x^\infty B(y)k(x,y)N_\alpha(y)dy, \qquad x \geqslant 0, \\ g(x)N_\alpha(x=0) = 0, \qquad N_\alpha(x) > 0 \ \text{ for } x > 0, \qquad \int_0^\infty N_\alpha(x)dx = 1, \end{cases} \tag{2.12}$$

where α measures the strength of the growth rate, as in the Prion problem (see [72]), and second, with the modulation in front of the fragmentation coefficients, replacing $B(x)$ by $\mathfrak{a}B(x)$. We treated these two problems simultaneously, by a proper rescaling shown below. We wanted to see how the solutions N_α behave in the limit cases $\alpha \to \infty$ and $\alpha \to 0$. As already said, it is clear that if $\alpha \to \infty$, the growth dominates and the mass tends to infinity, whereas it tends to $x = 0$ when $\alpha \to 0$. Hence, the behaviour of N_α and λ_α will depend on the behaviour of the coefficients g and B around L when $\alpha \to L$.

Our main assumption is power-like behaviours of the coefficients in the neighbourhood of $L = 0$ or $L = +\infty$, namely that

$$\exists\, \nu,\ \gamma \in \mathbb{R},\ \bar{B} > 0,\ \bar{g} > 0, \quad \text{such that} \quad g(x) \underset{x\to L}{\sim} \bar{g}x^\nu, \quad B(x) \underset{x\to L}{\sim} \bar{B}x^\gamma. \tag{2.13}$$

Theorem 2 (Theorem 1 of [2])**.** *Under Assumption* (2.13) *and the assumptions of Theorem 1, we have, for* $L = 0$ *or* $L = +\infty$,

$$\lim_{\alpha\to L} \lambda_\alpha = \lim_{x\to L} B(x).$$

Contrarily to what it seems, the limit does *not* depend only on B, because B and g are intertwined by the assumptions of Theorem 1; in particular, they imply that $1 + \gamma - \nu > 0$. This result shows that it is easy to obtain nonmonotonicity of the function $\alpha \to \lambda_\alpha$: for instance if B vanishes near zero and infinity.

We proved more than this asymptotic result, namely that when conveniently rescaled, the eigenvector N_α tends to a steady self-similar profile, which is the unique solution of the eigenproblem (2.3) with $g(x) = gx^\nu$ and $B(x) = Bx^\gamma$. The proof for this convergence relies on similar estimates than the ones of Theorem 1.

3 Inverse problems for growth models

Another point of view on growth-fragmentation problems is given by the inverse problem approach [57]. Speaking of *inverse problem* is quite vague and refers to a physical, intuitive

or historical vision of what can be considered as *direct* and what should be considered as *inverse*.

In all what is above, departing from assumptions on the initial condition and the parameters, we studied different properties of the structured population, viewed as the solution to our problem (2.1): its existence and uniqueness, its asymptotic behaviour, how it depends on its coefficients, etc. All these questions refer to what we call the *direct problem*. Let us take another viewpoint: we now consider the distribution of the population $n(t, x)$ as being a parameter of a new problem, which consists in finding one or several of the components of the problem - *e.g.*, the growth rate, and/or the fragmentation rate, and/or the initial condition - and think of this component as the sought solution of Problem (2.1). Such a problem is called *inverse* simply because it is not the usual or natural way it is handled. It becomes useful and appears naturally when parameters or initial condition are difficult to measure, contrarily to the solution or part of it. In such cases, inverse problems appear in order to calibrate the model and make it predictive; or even, to refine noisy measures of the solution.

This shows that there exist as many inverse problems as direct ones; even if general methods have been developed and turned out to be accurate in many situations [57, 113], each problem remains specific and requires a specific study, as shown below. Moreover, results on inverse problems often reveal interesting aspects of the related direct problem.

3.1 Historical context and setting of the problem

Let us briefly recall some examples of inverse problems for structured populations that can be found in the literature.

The initially-studied inverse problems concerned the age-structured equation. In several articles [141, 75, 76], the inverse problem consists in estimating the birth function or the birth and death functions from a measure of the evolution of the total population over time, plus some extra information like *a priori* knowledge of the death function [141] and initial distribution, or measure of the cumulative number of births [75, 76]. In [131], the initial condition is to be estimated from the total population and *a priori* knowledge of birth and death functions. In many of these articles, assumptions have to be made to ensure that the asymptotic behaviour is *not* reached yet: if it were, the evolution of the total population would be simply exponential, and would not provide substantial information on the birth function. Such assumptions appear to be practically very restrictive, due to the fact that convergence to the asymptotic behaviour is exponential under fairly general assumptions that ensure the existence of a spectral gap [127].

To the best of my knowledge, the first study of the inverse problem for the growth-fragmentation equation was done by [130]. This article is the basis of my research on the topic.

The point of the article [130] as well as the one of the following articles [15, 13, 16, 11] is to estimate the division rate $B(x)$ in Equation (2.3)**.** It appears

in many applications as being the quantity which is the most difficult to estimate: direct measures of the size distribution of individuals is often easy [89]; the growth and death rates may also be measured; the fragmentation kernel $k(x,y)$ can be guessed or even measured (for equal mitosis, relevant for E. Coli for instance, it is $k(x,y) = \delta_{x=\frac{y}{2}}$). On the contrary, there exists no direct way to measure or estimate the division rate B. In the biological litterature, this parameter B is almost unknown, whereas size distributions are classical (see [89] for E. Coli), and debates on the growth rate $g(x)$ took place (still for E. Coli, see [50] to know whether it is constant, which gives linear growth, or linear, which gives exponential growth).

The method introduced in [130] consists in making use of the asymptotic behaviour to simplify the problem. In this article, the authors focused on the equal mitosis case of Equation (2.2), and solved the following problem: How can we estimate the fragmentation rate $B(x)$ from (noisy) measurements of the steady distribution $N(x)$ and of the exponential growth rate λ of the population ?

3.2 The associated direct problem [16]

Before going further in the *inverse* problem defined right above, the associated *direct* problem needs to be studied and gives light to it.

In [130], The direct map (*i.e.* the operator we want to inverse) was $\Gamma : B \to (\lambda, N)$ solution to

$$\begin{cases} \frac{\partial}{\partial x}N + (\lambda + B(x))N = 4B(2x)N(2x), \quad x \geq 0, \\ N(x=0) = 0, \\ N(x) > 0 \,\text{for}\, x > 0, \quad \int_0^\infty N(x)dx = 1. \end{cases} \tag{3.1}$$

As already said, this problem corresponds to the case $g(x) = 1$ and $k(x,y) = \delta_{x=\frac{y}{2}}$ of Equation (2.3). The authors proved (Theorem 3.2. of [130]) that for two given constants $0 < B_m < B_M$, the map Γ is Lipshitz-continuous - and therefore class C^1 as a general fact on the Fréchet derivatives of mapping with quadratic nonlinearities - from the space $L^2(\mathbb{R}_+) \cap \{B(x), B_m \leq B \leq B_M\}$ into $L^2(\mathbb{R}_+)$. The boundedness condition for B is the same than in [128], where Problem (2.2) is studied and its exponential speed of convergence towards equilibrium is proved.

We generalized this result in [16] to general kernels k and growth rate g. Furthermore, we slightly modified Γ, to make use of the measure of the growth rate λ in order to identify a constant c in front of the growth rate $g(x)$. In [13], practical application to experimental data showed that even if the general growth behaviour of the individuals may be guessed, for instance by supposing $g(x)$ constant or linear, its scaling is linked to the time-scale of the equation and thus to the exponential growth rate λ of the population. We formulate the problem as follows: the direct map that we want to inverse is $\Gamma : (c, B) \to (\lambda, N)$ solution of

$$\begin{cases} c\frac{\partial}{\partial x}\big(g(x)N\big) + (\lambda + B(x))N = 2\int_x^\infty B(y)N(y)k(x,y)dy, \quad x \geq 0, \\ gN(x=0) = 0, \\ N(x) > 0 \,\text{for}\, x > 0, \quad \int_0^\infty N(x)dx = 1. \end{cases} \tag{3.2}$$

Similarly than for Theorem 3.2. of [130], which is based on the study of the eigenproblem made in [128] under the boundedness condition $B_m < B < B_M$, we built the domain of definition for Γ on the ground of the study made in [7]. The difficulty is to contour which kind of division rates B are acceptable for given g and k. This lead us to the following definition.

Definition 3.1 (Definition 1.1. of [16]). *Let g, k satisfying Assumptions* (2.4) *and* (2.6)–(2.8) *(which are Assumptions (2) and (8)–(10) of [16], or Assumptions (5)–(9) of [7]). For a constant $b \geq 0$ and functions $f_0 \in L^1_0$, $f_\infty \xrightarrow[x\to+\infty]{} \infty$, the set $\mathcal{D}(b, f_0, f_\infty)$ is defined by*

$$\mathcal{D}(b, f_0, f_\infty) := \left\{ B \in L^\infty_{\text{loc}}(\mathbb{R}^*_+) \cap \mathbb{P}, \quad \text{supp}(B) = [\tilde{b} \leq b, +\infty), \quad \frac{B}{g} \leq f_0, \quad \frac{xB}{g} \geq f_\infty \right\}.$$

The regularity of the map Γ was then stated as follows.

Theorem 3 (Theorem 1.1. of [16]). *Let parameters g and k satisfy the assumptions of Definition 3.1. Then*

i) The map $\Gamma : (c, B) \longmapsto (\lambda_0, N)$ is injective and continuous under the $L^\infty-$ weak-$*$topology for B from any set $\mathbb{R}^*_+ \times \mathcal{D}(b, f_0, f_\infty)$ to $\mathbb{R}^*_+ \times L^1(\mathbb{R}_+) \cap L^\infty(\mathbb{R}_+)$.

ii) Let moreover $\frac{1}{g} \in L^2_0$. Then the map Γ is locally Lipschitz-continuous (and therefore class C^1, see [130]) under the strong topology of $\mathbb{R}^*_+ \times L^2(\mathbb{R}_+) \cap \mathcal{D}(b, f_0, f_\infty)$.

The proof partly relies on similar estimates than those of [7], so that we could simplify the proof of [130] as concerns the most important point, *i.e.* the Lipshitz-continuity.

Having proved that Γ is of class C^1 highlights the regularizing effect of the direct problem, whose counterpart is the de-regularizing effect of an inverse mapping for Γ. If we seek an inverse from L^2 to L^2, it cannot be continuous as it is: Γ^{-1} will need to be regularised. This is discussed in the next section.

3.3 Deterministic setting [15, 16]

Let us now go back to the inverse problem, which consists in inverting the map $\Gamma : B \to (\lambda, N)$ or $\Gamma : (c, B) \to (\lambda, N)$ defined above.

As concerns the estimate of the constant c in Problem (3.2), using the first moment of the equation, we obtain the identity

$$c = \lambda \frac{\int xN(x)dx}{\int g(x)N(x)dx}, \tag{3.3}$$

which shows how we define an estimate c_ε of c from a noisy measure $(\lambda_\varepsilon, N_\varepsilon)$ of (λ, N). This part does not present any major difficulty, so that I refer to [16] for more details but will not

discuss it anymore, and I focus on the estimate of the division rate B, which is the heart of the problem.

We now view Equation (2.3) as a problem where the unknown is $B(x)$, and the couple (λ, N), a given parameter only known with some uncertainty, so that we only have access to $(\lambda_\varepsilon, N_\varepsilon)$ with $|\lambda - \lambda_\varepsilon| \leq \varepsilon$ and $\|N - N_\varepsilon\|_{\mathcal{Z}} \leq \varepsilon$ in a certain space $\mathcal{Z}$. In [130, 15, 13], we have $\mathcal{Z} = L^2(\mathbb{R}_+)$; in [16], $\mathcal{Z} = L^2(x^p dx)$. In [11], the noise is a gaussian white noise that could be heuristically compared to $\mathcal{Z} = H^{-\frac{1}{2}}(\mathbb{R}_+)$.

To end the definition of the inverse problem to solve, we define the space where an estimate for B has to be found. In [130], it is $L^2(N_\varepsilon^2 dx)$. Due to the multiplication of B by N in the equation, it is more convenient to define a space $\mathcal{X}$ where $H = BN$ has to be found. Once an estimate for H is found, it remains to divide it by N_ε or a regularised version in order to get the estimate of the division rate B. The major difficulty linked to this division is that the estimate necessarily becomes poor when the underlying distribution vanishes, which is the case where x tends to ∞.

However, this difficulty is mainly theoretical, since the value of the division rate for such large x has little influence on the distribution N, as shown by numerical simulations.

For the estimate of H, any space L^p or weighted L^p seems convenient from a modelling point of view, and similar choices were made in [15, 16] as well as in [11]. All these elements lead to the following complete definition.

Definition of the Inverse Problem: Given (λ, N) solution of Problem (3.2) for parameters g, k, B satisfying the assumptions of Theorem 1, given two spaces $\mathcal{X}$, $\mathcal{Z}$ and a measure $(\lambda_\varepsilon, N_\varepsilon) \in \mathbb{R}_+^* \times \mathcal{Z}$ such that $|\lambda - \lambda_\varepsilon| \leq \varepsilon$ and $\|N - N_\varepsilon\|_{\mathcal{Z}} \leq \varepsilon$, the inverse problem is: How can we get an estimate $(c_\varepsilon, H_\varepsilon^\alpha)$ of $H = BN$ such that $\|H - H_\varepsilon^\alpha\|_{\mathcal{X}} \to_{\varepsilon \to 0} 0$?

Denoting $L(\lambda, N) = \lambda N + \frac{\partial}{\partial x}(gN)$, we decompose our problem in two main steps.

1. Set and solve in appropriate spaces $(\mathcal{X}, \mathcal{Y})$ the following equation

$$\mathcal{L}(H)(x) := 2\int_x^\infty k(x,y)H(y)dy - H(x) = L(x), \tag{3.4}$$

and prove the continuity of the operator $\mathcal{L}^{-1} :\ L \to H$ in these appropriate spaces.

In the case of equal mitosis studied in [130, 15], Equation (3.4) becomes the following dilation equation:

$$\mathcal{L}(H)(x) := 4H(2x) - H(x) = L(x), \tag{3.5}$$

and continuity of $\mathcal{L}^{-1}$ was proved *e.g.* for $\mathcal{X} = \mathcal{Y} = L^2(\mathbb{R}_+)$. It shows that Equation (3.5) *a priori* requires the same regularity for H than for L.

One also notices that this first step is a linear problem, and is specific of our growth-fragmentation equation. On the contrary, the second step can be formulated as a generic problem of estimating a function from its antiderivative (see [130] and [11]).

2. Given that we do not have access to the exact value of (λ, N) but only to a noisy measure $(\lambda_\varepsilon, N_\varepsilon)$, close to (λ, N) in the sense that $\|N - N_\varepsilon\|_{\mathcal{Z}} \leq \varepsilon$ and $|\lambda - \lambda_\varepsilon| \leq \varepsilon$, we do not know the exact value of $L(\lambda, N)$ either, but only $L(\lambda_\varepsilon, N_\varepsilon)$. How can we thus find an estimate $H_\varepsilon \in \mathcal{X}$ of the exact solution H of (3.4) ?

   The operator L is not linear but quadratic, and becomes linear if we forget the dependency on λ_ε. Hence by easy manipulations, the same estimates as for a linear operator may be obtained, and we can apply the general theory for linear inverse problems ([57] for instance).

   If the space $\mathcal{Z}$ where the noise for N is measured is such that $L(\mathbb{R}_+^* \times \mathcal{Z}) \subset \mathcal{Y}$, due to the continuity of $\mathcal{L}^{-1}$, this step is empty: we can simply take $H_\varepsilon = \mathcal{L}\big(L(\lambda_\varepsilon, N_\varepsilon)\big)$, and we will obtain an estimate in the order of ε. For equal mitosis, this would be the case for instance if $\mathcal{Z} = H^1(\mathbb{R}_+)$: in such a case, $L(\mathbb{R}_+^* \times H^1) \subset L^2(\mathbb{R}_+) = \mathcal{Y}$ and $\|L(\lambda_\varepsilon, N_\varepsilon) - L(\lambda, N)\|_{\mathcal{Y}} = O(\varepsilon)$.

   But if the space $\mathcal{Z}$ does not have a sufficient regularity with respect to the desired space $\mathcal{X}$, a regularisation step is required, characterized by a second parameter α that needs to be conveniently scaled with respect to ε. The distance between the spaces $L(\mathbb{R}_+ \times \mathcal{Z})$ and $\mathcal{Y}$ can be evaluated in the framework of *Hilbert scales* (see [57]) and characterizes the *degree of ill-posedness* of the problem. In a heuristic way, taking the Hilbert spaces $H^s(\mathbb{R}_+)$ as a Hilbert scale and $\mathcal{X} = L^2$, we see that since $\mathcal{Y} = L^2$, if $\mathcal{Z} = L^2$ the distance between $L(\mathbb{R}_+ \times \mathcal{Z}) = H^{-1}$ and $L^2 = H^0$ corresponds to a *degree of ill-posedness* $a = 1$; if $\mathcal{Z} = H^{-\frac{1}{2}}$, the distance between $L((\mathbb{R}_+ \times \mathcal{Z}) = H^{-\frac{3}{2}}$ and $L^2 = H^0$ corresponds to a *degree of ill-posedness* $a = \frac{3}{2}$.

   Such a regularisation can be done in many different ways, among which I distinguish two main types (that may be combined, as the *approximate inverse method*, applied to our problem in [73]).

   - From the measure N_ε, define a function or distribution L_ε^α in the convenient regularity space $\mathcal{Y}$, at a distance from $L(\lambda, N)$ vanishing with ε in $\mathcal{Y}$ and with α in $L(\mathbb{R}_+^* \times \mathcal{Z})$, and apply the operator $\mathcal{L}$ to it. Such a regularisation is done in [15] by what we called the *filtering* method for instance, which simply consists in a convolution with a mollifier $\rho_\alpha(x) = \frac{1}{\alpha}\rho(\frac{x}{\alpha})$, $\rho \in C_c^\infty(\mathbb{R})$, $\int \rho(x)dx = 1$. The balance between α and ε appears clearly by noting that $\|\rho_\alpha * N_\varepsilon - N_\varepsilon\|_{L^2} \leq C\alpha\|N_\varepsilon\|_{L^2}$, and $\|\rho_\alpha * \frac{\varepsilon df}{dx}\|_{L^2} \leq \frac{C\varepsilon}{\alpha}\|f\|_{L^2}$. Since we want both terms to vanish, it appears that the optimal choice for α is $\alpha = O(\sqrt{\varepsilon})$.
   - Approximate the operator $\mathcal{L}^{-1}$ by a more regular one $\mathcal{L}_\alpha^{-1}$, continuous from the space $L(\mathbb{R}_+^* \times \mathcal{Z})$ to $\mathcal{X}$ and converging to $\mathcal{L}^{-1}$ when α vanishes. Then, apply this operator $\mathcal{L}_\alpha^{-1}$ instead of $\mathcal{L}$ to $L(\lambda_\varepsilon, N_\varepsilon)$, and choose α appropriately to ensure a sufficient regularity (so that α cannot be too small) together with a sufficiently good approximation of $\mathcal{L}^{-1}$ (so that it cannot be too large).

The original method of [130], called in [15] the *quasi-reversibility* method in reference to [90], is of this kind: we have $\mathcal{Z} = L^2$, so that $L(\mathbb{R}_+^* \times \mathcal{Z}) = H^{-1}$ and we need a regularised operator $\mathcal{L}_\alpha^{-1}$ such that $\mathcal{L}_\alpha^{-1}(\mathcal{Z}) \subset \mathcal{X} = L^2$. Hence $\mathcal{L}^{-1}$ is replaced by $\mathcal{L}_\alpha^{-1} : L \to H_\alpha$ solution of the following regularised problem:

$$\begin{cases} \alpha \frac{\partial}{\partial x}\big(H_\alpha(2x)\big) + 4H_\alpha(2x) - H_\alpha(x) = L(x), \\ H_\alpha(0) = 0. \end{cases} \tag{3.6}$$

Let us now detail how we solved each specific step, and which difficulties appeared.

3.3.1 Theoretical results - step 1: Solving equation (3.4)

In [130] and [15], the particular case of equal mitosis was studied, *i.e.* Equation (3.5). In [15], Equation (3.5) did not appear explicitely, since the quasi-reversibility method considers the regularizing equation (3.6).

Such a dilation equation also appears in other application contexts, like TCP-IP protocoles [23] and the construction of wavelets [154], and was studied by probabilists [112]. The Lax-Milgram lemma lead us to the following result.

Proposition 1 (Proposition A.1. and Lemma A.2. of [15])**.** *Let $L \in L^2(\mathbb{R}_+, x^p dx)$, with $p \neq 3$, then there exists a unique solution $H \in L^2(\mathbb{R}_+, x^p dx)$ to (3.5) and*
• *for $p < 3$, this solution is given explicitly by the formula*

$$H^{(1)}(x) = \sum_{n=1}^{+\infty} 2^{-2n} L(2^{-n}x). \tag{3.7}$$

Furthermore, for $1 \leq q \leq \infty$, if $L \in L^q(\mathbb{R}_+)$ then $H^{(1)} \in L^q(\mathbb{R}_+)$.
• *for $p > 3$, this solution is given explicitly by the formula*

$$H^{(2)}(x) = -\sum_{n=0}^{+\infty} 2^{2n} L(2^n x). \tag{3.8}$$

The solutions to (3.5) with $L = 0$ in $\mathcal{D}'(0, \infty)$ have the form $\frac{f(\log(x))}{x^2}$ with $f \in \mathcal{D}'(\mathbb{R})$ a $\log(2)-$ periodic distribution.

This result emphasized two important points. First, the solution depends on the chosen space, and in general is not the same in L^2 and in $L^2(x^4 dx)$ for instance, since there is no reason, for a general L, even belonging to $L^2((1 + x^4)dx)$, that $H^{(1)} = H^{(2)}$. Second, the infinite number of distribution solutions may perturb numerical simulations.

In [16], we partly generalized Proposition 1, as follows.

Proposition 2 (Proposition 2.1. of [16]). *Let k satisfy Assumption* (2.4). *For $r, q \geq 0$, we define the following quantities:*

$$C_r := \sup_x \int_x^\infty \frac{x^r}{y^r} k(x,y)dy, \qquad D_q := \sup_y \int_0^y \frac{x^q}{y^q} k(x,y)dx. \tag{3.9}$$

If $0 \leq r \leq p$ are such that

$$C_r D_{p-r} < \frac{1}{4}, \tag{3.10}$$

Then for all $L \in L^2(\mathbb{R}_+, x^p \,\mathrm{d}x)$ there exists a unique solution $H \in L^2(x^p \,\mathrm{d}x)$ to Problem (3.4), *and we have the following estimate*

$$||H||_{L^2(x^p \,\mathrm{d}x)} \leq \frac{1}{1 - 2\sqrt{C_r D_{p-r}}} ||L||_{L^2(x^p \,\mathrm{d}x)}.$$

As for Proposition 1, the proof relies on the Lax-Milgram lemma and on Young inequalities. It corresponds to the case $p > 3$ when the identity dominates the integral kernel (see Remark 1 of [16]).

3.3.2 Theoretical results - step 2: error estimates and convergence rates

As already said, the pioneering article [130] studied the inverse problem for the case of equal mitosis of Equation (3.5), and solved it with the *quasi-reversibility* method described by Equation (3.6). Theorem 4.1. of [130] proved that for $L \in L^2(\mathbb{R}_+)$, there exists a unique solution $H \in H^1$ which, for some constant $C > 0$, satisfied the following estimates

$$\|H\|_{L^2}^2 + \alpha^2 \|H\|_{H^1}^2 \leq C\|L\|_{L^2}^2, \qquad \|H\|_{H^1}^2 \leq C\|L\|_{H^1}^2. \tag{3.11}$$

Such estimates are the key of this second step. Defining $H_{\varepsilon,\alpha}$ as the solution of Equation (3.6) for $L = L(\lambda, N_\varepsilon)$, the error is decomposed as

$$\|H - H_{\varepsilon,\alpha}\|_{L^2} \leq \|H - H_\alpha\|_{L^2} + \|H_\alpha - H_{\varepsilon,\alpha}\|_{L^2}, \tag{3.12}$$

where H_α is the solution of Equation (3.6) for $L = L(\lambda, N)$. The first term of the right-hand side is (up to a constant) smaller than $\alpha\|L(\lambda, N)\|_{H^1}$, and thus by $\alpha\|N\|_{H^2}$ (Theorem 4.2. of [130]). The second term may be bounded by $\frac{C}{\alpha}\|N - N_\varepsilon\|_{L^2}$ (Proposition 5.2. of [130]). All this lead them to the following error estimate (Theorem 5.1. of [130])

$$\|B_{\varepsilon,\alpha} - B\|_{L^2(N_\varepsilon^2 dx)} \leq C\alpha\|N\|_{H^2} + \frac{C + \|B\|_{L^\infty}}{\alpha}\|N_\varepsilon - N\|_{L^2}, \tag{3.13}$$

which showed an optimal convergence rate of order $\sqrt{\varepsilon}$ as soon as $N \in H^2(\mathbb{R}_+)$, in accordance with the general theory of linear inverse problems [57].

In collaboration with these authors, in [15] we applied a second regularisation technique, the filtering method - refining only slightly the problem by assuming also that the measure of λ was noisy. Similarly, we defined $H_{\varepsilon,\alpha}$ as the solution of Equation (3.4) for $L = L(\lambda_\varepsilon, \rho_\alpha * N_\varepsilon)$, and an intermediate function H_α solution of Equation (3.5) for $L = L(\lambda, \rho_\alpha * N)$ may also be defined. It lead to the exact equivalent of Estimate (3.13) (see Theorem 2.1. and Proposition 2.2. of [15]).

As shown by the description of step 2 seen above, and as exemplified by the work of another group [73], we could apply successfully any technique applicable to the classical inverse problem of estimating a function from its antiderivative.

In the case of the filtering method, for a mollifier ρ such that $\int x^p \rho(x)dx = 0$ for $p = 1, 2, \cdots, m_0$ and $s \leq m_0$, we could also easily generalize Estimate (3.13) as follows: if $N \in H^{s+1}$, in Eq. (3.12) the first term (the *bias*) becomes of order $O(\alpha^s)\|N\|_{H^{s+1}}$, hence the optimal value for α in Eq. (3.13) becomes $\alpha = O(\varepsilon^{\frac{1}{s+1}})$, what leads to an estimate in the order of $\varepsilon^{\frac{s}{s+1}}$. This rejoins the general theory of linear inverse problems [57].

In [16], we generalized both methods (filtering and quasi-reversibility) to Equation (3.4). We obtained similar estimates, but in $L^2(x^p dx)$ with $p > 3$ instead of L^2 due to Proposition 2 (Theorems 2.1. and 2.2. of [16]).

3.3.3 Numerical solution ([13, 15, 16]

The key point for the numerical solution of our inverse problem consists in solving numerically Equation (3.4); the other details refer to classical numerical techniques, such as FFT or upwind schemes, and I let the interested reader refer to [15] and [16].

The case of equal mitosis described by Equation (3.5) was solved in [15]. When examining this equation, two strategies appeared.

Strategy 1. Compute $H(2x)$ from $H(x)$. This means that we re-write Equation (3.5) with the new variable $y = 2x$. The scheme departs from zero, and one deduces the values of H_i step by step, from the knowledge of H_j for $j \leqslant i - 1$.

Strategy 2. Compute $H(x)$ from $H(2x)$. The scheme departs from the largest point $x = T$ of the simulation domain. We suppose that for $x \geqslant T$ we have $H(x) = H(T) = 0$ (it is relevant since we suppose that N vanishes for x large: see Theorem 1), and then deduce the smaller values H_i step by step, from the knowledge of H_j for $j \geqslant i + 1$.

The two approaches do not necessarily lead to the same result because as shown in Proposition 1, the first strategy selects an approximation of the solution $H^{(1)}$ whereas the second selects an approximation of the solution $H^{(2)}$. In the case of a very regular data N, then $H^{(2)}$ will perform better around infinity, whereas $H^{(1)}$ will be better around zero. However, if N is a solution of Equation (2.3), when we increase the number of points, the two approaches converge to the same solution since $H^{(2)} = H^{(1)}$. Since our simulation domain $[0, L]$ is bounded and contains zero, we prefered the first strategy. This choice was confirmed by all the numerical tests we performed: the second approach leads to a solution exploding around zero.

We proposed the following scheme, based on two other important requirements: that the two fundamental relations of "quantity" balance and of mass balance, given by the moments of order respectively 0 and 1 of the solution, are asymptotically preserved by the discretization.

We defined a regular grid on $[0,T]$ with mesh $k^{-1}T$ by

$$0 = x_{0,k} < x_{1,k} < \cdots < x_{i,k} := \tfrac{i}{k}T < \ldots < x_{k,k} = T.$$

We set

$$\varphi_{i,k} := \frac{k}{T}\int_{x_i,k}^{x_{i+1,k}} \varphi(x)dx \ \text{ for } \ i = 0,\ldots,k-1,$$

and define by induction the sequence [6]

$$H_{i,k}(\varphi) := \frac{1}{4}(H_{i/2,k}(\varphi) + \varphi_{i/2,k}), \quad i = 0,\ldots,k-1,$$

what gives, for $i = 0$ and $i = 1$

$$H_0(\varphi) := \tfrac{1}{3}\varphi_{0,k}, \ \ H_1(\varphi) := \tfrac{4}{21}\varphi_{0,k} + \tfrac{1}{7}\varphi_{1,k}.$$

Finally, we define

$$\mathcal{L}_k^{-1}(\varphi)(x) := \sum_{i=0}^{k-1} H_{i,k}(\varphi)1_{[x_{i,k},x_{i+1,k})}(x). \tag{3.14}$$

The convergence of this scheme was proved in [11] by the following estimate.

Proposition 3 (Proposition 4 of [11])**.** *Let $T > 0$ and $L \in \mathcal{H}^1$. Let $H = \mathcal{L}^{-1}(L)$ denote the unique solution of* (3.5) *belonging to $L^2(\mathbb{R}_+)$. We have for $k \geq 1$:*

$$\|\mathcal{L}_k^{-1}(L) - \mathcal{L}^{-1}(L)\|_{2,T} \leq C\frac{T}{\sqrt{k}}\|L\|_{\mathcal{H}^1},$$

with $C < \frac{1}{\sqrt{6}}$.

Following this algorithm, numerical tests were performed in [15] and proved satisfactory, even though the estimates become poor when N vanishes, as it was expected.

In [16], we implemented numerical schemes for general kernels $k(x,y)$. The main problem is that for this general case, we do not have the choice between Strategy 1 and Strategy 2: we cannot avoid Strategy 2, defining $H(x)$ from larger values present in $\int_x^\infty k(x,y)H(y)dy$. It lead to correct approximations for large x, but not sufficiently good for small ones, as illustrated in [16].

[6] for any sequence $u_i, i = 1,2,\ldots$, we define

$$u_{i/2} := \begin{cases} u_{i/2} & \text{if } i \text{ is even} \\ \frac{1}{2}(u_{(i-1)/2} + u_{(i+1)/2}) & \text{otherwise.} \end{cases}$$

4 Statistical viewpoint on the inverse problem [11]

As shown above, till now we modeled the noise in a deterministic way, supposing that $\|N_\varepsilon - N\|_{\mathcal{Z}} \leq \varepsilon$, with $\mathcal{Z} = L^2$ or $\mathcal{Z} = L^2(x^p dx)$. In [13], we applied our method on experimental pre-processed data [89], without taking care of what the measure could be. However, take the most of non pre-processed data is a major question, in order to use our techniques not only to estimate B but also to validate or discriminate models.

In [11], we replaced the assumption $N_\varepsilon \in L^2$, which does not correspond to any realistic model for the noise, by supposing that we measured the sizes $x_1, \cdots, x_n$ of a sample of n individuals. Supposing that this sample was chosen uniformly randomly in a steady population lead us to assume that the sizes were the realizations of $X_1, \cdots X_n$ independent random variables whose law was N the solution of Equation 2.3. Mathematically, it may be written

$$\mathbb{P}(X_1 \in dx_1, \ldots, X_n \in dx_n) = \prod_{i=1}^{n} N(x_i) dx_i,$$

where the symbol $\mathbb{P}$ stands for probability. Asymptotics are taken as the number of observations n grows to infinity.

When we made this modelling assumption, we were guided by intuitive ideas and by existing work on branching processes [32]. Till then, recent studies confirmed this point [47, 10].

With such a statistical model, Step 2 now consists in estimating $L(\lambda, N)$ from the n-sample, *i.e.*, the density N and its weighted derivative. Once this is achieved, Step 1 is unchanged, so that we can combine the rest of the method with this statistical estimation.

Density estimation is a classical statistical problem, as well as estimation of the derivative of a density. Many different techniques are possible, as in the deterministic setting. We chose kernel regularisation partly because it is nothing else that what we called the filtering method in a deterministic setting. As in a deterministic setting, we regularised the empirical measure $\frac{1}{n}\sum_{i=1}^{n} \delta_{x=x_i}$, which stands for N_ε and represents an (ugly) estimate for the density N, by convolution with a mollifier. In order to keep standard statistical notations, we denoted this mollifier K_h, h replacing α and K replacing ρ in the notations of the deterministic filtering method. Estimates are thus denoted for instance $\hat{N}_h$, which stands for $N_{\varepsilon,\alpha}$.

In [11], we went a step further, by adapting the method of Goldenschluger and Lepski [70, 69]. This method selects automatically the regularisation parameter h (called *bandwidth* in statistics), without any *a priori* knowledge on the regularity of the underlying density N. Such methods are called *adaptive*, or in deterministic inverse problems *a posteriori* methods, in contrast with *a priori* methods where the bandwidth/regularisation parameter is chosen from an assumption on the regularity of the function, as for instance the above-mentioned estimates of Theorem 2.1. of [15] lead to an optimal choice $\alpha = O(\sqrt{\varepsilon})$ under the *a priori* assumption that $N \in H^2$. Among such methods, one can quote the *discrepancy principle* [57], recently adapted to a statistical setting.

4.1 The Goldenschluger and Lepski's method

Without going into technical details, let us briefly explain in what consists the Goldenschluger and Lepski's method. We take here the example of estimating the density N; to estimate $L(\lambda, N)$, we proceeded in a similar way (though rates of convergence are different due to the derivative). For $h > 0$ and $x \in \mathbb{R}$, we defined

$$\hat{N}_h(x) := \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i), \tag{4.1}$$

where $K_h(x) = h^{-1}K(h^{-1}x)$. Note in particular that $\mathbb{E}(\hat{N}_h) = K_h * N$, where $*$ denotes convolution: in expectancy, it is what we denoted above $N_\alpha = \rho_\alpha * N$. As above in the deterministic setting, we measured the performance of $\hat{N}_h$ via its squared integrated error, *i.e.* the average L^2 distance between N and $\hat{N}_h$. This error is decomposed as

$$\mathbb{E}[\|N - \hat{N}_h\|_2] \leq \big\|N - K_h * N\big\|_2 + \mathbb{E}[\|K_h * N - \hat{N}_h\|_2]. \tag{4.2}$$

We can compare this estimate with the above-mentioned Estimate (3.12): on the right-hand side, the first term is exactly the same. It corresponds to what is called a *bias* term in statistics: it biases the estimate in the sense that it does not depend on the measure but only on the regularisation parameter/bandwidth h.

The second term on the right-hand side is the same in both estimates, except that in Estimate (4.2) it is viewed in expectancy. It is easy to see that

$$\mathbb{E}[\|K_h * N - \hat{N}_h\|_2^2] \leq \frac{\big\|K\big\|_2^2}{nh}.$$

If one has to choose h in a family $\mathcal{H}$ of possible bandwidths, the best choice is $\bar{h}$ where

$$\bar{h} := \mathrm{argmin}_{h \in \mathcal{H}} \big\{ \big\|N - K_h * N\big\|_2 + \frac{1}{\sqrt{nh}} \big\|K\big\|_2 \big\}. \tag{4.3}$$

This ideal compromise $\bar{h}$ is called the "oracle" in statistics: it depends on N and then cannot be used in practice. In our deterministic setting, it would consists similarly in choosing

$$\bar{\alpha} := \mathrm{argmin}_{\alpha \in \mathcal{H}} \big\{ \big\|N - \rho_\alpha * N\big\|_2 + \frac{\varepsilon}{\sqrt{\alpha}} \big\|\rho\big\|_2 \big\}.$$

Hence one wants to find an automatic (data-driven) method for selecting a bandwidth able to achieve this minimum up to a constant. The Lepski method [95, 96, 97, 98] is one of the various theoretical adaptive methods available for selecting a density estimator. In particular it is the only known method able to select a bandwidth for kernel estimators. However the method does not usually provide a non asymptotic[7] oracle inequality. Recently, Goldenschluger and Lepski [70] developed powerful probabilistic tools that enable to

[7] *i.e.* valid for finite n and not only when $n \to \infty$.

overcome this weakness and that can provide with a fully data-driven bandwidth selection method. In [11], we gave a practical illustration of their work: how should one select the bandwidth for a given kernel in dimension 1?

The main idea is to estimate the bias term by looking at several estimators. The method consists in setting, for any x and any $h, h' > 0$,

$$\hat{N}_{h,h'}(x) := \frac{1}{n}\sum_{i=1}^{n}(K_h * K_{h'})(x - X_i) = (K_h * \hat{N}_{h'})(x).$$

Next, for any $h \in \mathcal{H}$, define

$$A(h) \;:=\; \sup_{h' \in \mathcal{H}} \big\{ \|\hat{N}_{h,h'} - \hat{N}_{h'}\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2 \big\}_+,$$

where χ is a well-chosen constant. The quantity $A(h)$ is a good estimator of $\big\|N - K_h * N\big\|_2$ (see (4.2) and (4.3) in [11]). The next step consists then in setting

$$\hat{h} := \arg\min_{h \in \mathcal{H}} \big\{ A(h) + \frac{\chi}{\sqrt{nh}}\|K\|_2 \big\}, \tag{4.4}$$

and our final estimator of N is obtained by defining $\hat{N} := \hat{N}_{\hat{h}}$. With this choice, the following optimal estimate is obtained.

Proposition 4 (Proposition 2 in [11])**.** *Assume $N \in L^\infty$ and let $K \in \mathcal{C}(\mathbb{R}) \cap L^2(\mathbb{R})$, $\int K(x)dx = 1$. If $\mathcal{H} \subset \{D^{-1}, D = 1, \ldots, D_{\max}\}$ with $D_{\max} = \delta n$ for $\delta > 0$, then,*

$$\mathbb{E}\left[\|\hat{N} - N\|_2^2\right] \le C \inf_{h \in \mathcal{H}} \big\{ \|K_h * N - N\|_2^2 + \frac{\big\|K\big\|_2^2}{(hn)} \big\} + C_1 n^{-1},$$

where C and C_1 are constants depending on χ, δ, $\|K\|_2$, $\|K\|_1$ and $\|N\|_\infty$.

This inequality is called an oracle inequality, for we have $\mathbb{E}[\|\hat{N} - N\|_2] \le (\mathbb{E}[\|\hat{N} - N\|_2^2])^{1/(2)}$ and $\hat{h}$ is performing as well as the oracle $\bar{h}$ up to some multiplicative constant. In that sense, we are able to select the *best bandwidth* within our family $\mathcal{H}$.

A similar bandwidth selection and oracle inequality was obtained for the estimate of $\frac{\partial}{\partial x}(g(x)N)$ in Proposition 3 of [11], the term $\frac{\big\|K\big\|_2^2}{(hn)}$ being replaced by $\frac{\big\|K'\big\|_2^2}{nh^3}$ due to the derivation.

The proofs used subtle concentration inequalities (see [102]), in order to estimate the distance between the random variable $\hat{N}$ and its expectancy. This was performed by my co-authors.

4.2 Rates of convergence

These oracle inequalities, together with the convergence of the numerical scheme of [15] expressed by Proposition 3, lead to an upper bound for the estimate of B. Theorem 1 of [11] gave a complete statement for this upper bound, which in its turn provided the following rate of convergence.

Proposition 5 (From Proposition 1 of [11])**.** *Under Assumptions of Theorem 1, if furthermore*

- N, $gN \in H^{s+1}(\mathbb{R}_+)$, *and* $g \in L^\infty(\mathbb{R}_+)$,
- $K \in \mathcal{C}^1(\mathbb{R}) \cap L^2(\mathbb{R}) \cap L^1((1+x^{m_0})dx)$ *is a kernel such that* $\int K(x)dx = 1$ *and* $\int x^p K(x)dx = 0$ *for* $p = 1, 2, \cdots m_0 - 1$,
- *the estimate* $\hat{\lambda}$ *for* λ *satisfies Assumption 2 of [11].*

Let us define $\hat{H} = \mathcal{L}_k^{-1}(\hat{L})$ *with* $k = n$ *and* $\hat{L}$ *selected as in [11] by the Goldenschluger and Lepski's method. Then* $\hat{H}$ *satisfies, for all* $s \in [1; m_0 - 1]$

$$\mathbb{E}\left[\left\|\hat{H} - H\right\|_{2,T}\right] = O\big(n^{-\frac{s}{2s+3}}\big). \tag{4.5}$$

Let us compare this rate of convergence with the one obtained in the deterministic setting. After renormalization, we obtained the rate $n^{-s/(2s+3)}$ for estimating H, and this corresponds to ill-posed inverse problems of order 1 in nonparametric statistics. We can make a parallel with additive deterministic noise following Nussbaum and Pereverzev [121].

In the deterministic setting, we supposed that we had an approximate knowledge of N up to a deterministic error that we can write $\varepsilon\zeta = N_\varepsilon - N \in L^2$ with $\|\zeta\|_2 \leq 1$. As discussed above in Section 3.3, the problem of estimating H is ill-posed of degree $a = 1$ in the terminology of Wahba [57, 156] for it involves the computation of the derivative of N. By classical inverse problem theory for linear cases, this means that if $N \in \mathcal{H}^{s+1}$, the optimal recovery rate in $\mathbb{L}^2$-error norm should be $\varepsilon^{s/(s+a)} = \varepsilon^{s/(s+1)}$.

Suppose now that we replace the deterministic noise ζ by a random Gaussian *white noise*: we observe

$$N_\varepsilon = N + \varepsilon\mathbb{B} \tag{4.6}$$

where $\mathbb{B}$ is a Gaussian white noise, *i.e.* a random distribution in $H^{-1/2}$, that operates on test functions $\varphi \in \mathbb{L}^2$ and such that $\mathbb{B}(\varphi)$ is a centered Gaussian variable with variance $\|\varphi\|_2^2$. Model (4.6) serves as a representative toy model for most stochastic error models such as density estimation or signal recovery in the presence of noise.

Since we can make a heuristic correspondence between $\mathbb{B}$ and a deterministic distribution in $H^{-\frac{1}{2}}$, this leads to a degree of ill-posedness $a = 3/2$, and so the rate of convergence becomes in this case $\varepsilon^{\frac{s}{s+\frac{3}{2}}}$ instead of $\varepsilon^{\frac{s}{s+1}}$.

Comparing now the level of noise ε with n, we have a formal correspondence $\varepsilon = n^{-1/2}$, which can be seen as due to the Berry-Esseen theorem (see also [120]).

We thus obtained the rate

$$\varepsilon^{s/(s+3/2)} = \varepsilon^{2s/(2s+3)} = n^{-s/(2s+3)}.$$

This is exactly the rate we found in Proposition 5: the deterministic error model and the statistical error model coincide to that extent. The statistician reader will note that the rate $n^{-s/(2s+3)}$ is also the minimax rate of convergence when estimating the derivative of a density, see [68], and the analyst that $\varepsilon^{\frac{s}{s+1}}$ is the rate for estimating a function from its antiderivative.

5 Modelling and application to protein polymerization

5.1 Prion model: the nonlinear growth-fragmentation equation [9]

Though fragmentation-coagulation equations are known and used in physical applications for a while, the growth-fragmentation equation (2.1) was introduced only recently to model polymerization, and, to be specific, protein polymerization. To the best of my knowledge, the first was the model by Greer et al. in [72, 134, 59].

The original equations are differential systems obtained by the law of mass action. To model protein polymerization for instance, Masel et al. [101] wrote the following system:

$$\frac{dn_i}{dt} = -V(t)\big(g_i n_i - g_{i-1} n_{i-1}\big) - B_i n_i + 2\sum_{j=i+1}^{\infty} B_j k_{i,j} n_j, \qquad i \geq i_0, \tag{5.1}$$

$$\frac{dV}{dt} = \lambda - \gamma V - V\sum_{i=i_0}^{\infty} g_i n_i + 2\sum_{j\geq i_0}\sum_{i<i_0} i k_{i,j} B_j n_j, \tag{5.2}$$

where n_i denotes the concentration of polymers containing i monomers, and $V = n_1$ is the concentration of monomers. Each term of the equations is obtained by the law of mass action for the following reactions.

- Polymers of size i can grow into polymers of size $i+1$ by the addition of monomers with a reaction rate g_i. It gives the first two terms on the right-hand side, with a quadratic coupling $V(t)n_i(t)$.
- Polymers of size i can break down into smaller ones with a rate B_i, and with a probability $k_{j,i}$ to give rise to two polymers of respective sizes j and $i-j$, $1 \leq j \leq i-1$. Hence we have the following discrete equivalent of Assumptions (2.4):

$$\sum_{j=1}^{i-1} k_{j,i} = 1, \quad k_{i,j} \geq 0, \quad k_{i,j} = k_{j-i,j}, \quad k_{i,j} = 0 \text{ for } j \leq i. \tag{5.3}$$

- Monomers are produced at a constant rate λ and die with a rate γ.
- Polymers of size smaller than i_0 are unstable, and immediately degraded into monomers, what gives the last double summation term in the equation for V.

A fundamental equation is the mass balance: summing Equation (5.2) with the sum of Equations (5.1) multiplied by i, we obtain

$$\frac{d}{dt}\left(V(t) + \sum_{i=i_0}^{\infty} i n_i(t)\right) = \lambda - \gamma V. \tag{5.4}$$

Such a system is far from being the first of this kind: see [39, 122] among many seminal references. Greer et al. wrote a continuous version, under the form

$$\frac{\partial n}{\partial t} + V\frac{\partial}{\partial x}\big(g(x)n(t,x)\big) + B(x)n = 2\int_x^{\infty} k(x,y)B(y)n(t,y)dy, \quad x \geq x_0, \tag{5.5}$$

$$\frac{dV(t)}{dt} = \lambda - \gamma V - V\int_{x_0}^{\infty} g(x)n(t,x)dx + \int_0^{x_0}\int_{x_0}^{\infty} xk(x,y)B(y)n(t,y)dydx, \tag{5.6}$$

$$g(x_0)n(t,x_0) = 0. \tag{5.7}$$

As for the discrete system, the mass balance is obtained by summation:

$$\frac{d}{dt}\left(V(t) + \int_{x_0}^{\infty} xn(t,x)dx\right) = \lambda - \gamma V. \tag{5.8}$$

The link between the discrete system (5.1)–(5.2) and the continuous one (5.5)–(5.6) is easy to see, simply by replacing i by a continuous index x, sums by integrals and differences by derivatives.

We recognize the growth-fragmentation equation in Equation (2.8), but here coupled with the monomers equation (5.6) *via* the growth term. The unknown $n(t,x)$ represents the concentration of polymers of size x at time t, growing by monomer addition with a growth rate $g(x)V(t)$ proportionnal to the concentration of monomers by the law of mass action.

In the original articles [101] and [59, 134], as in most of the articles of the biological literature [122, 87, 158], in their discrete version as in their continuous ones, parameters $g(x)$, $B(x)$ are assumed to be constant, and the fragmentation kernel k is uniform $k(x,y) = \frac{1}{y}\mathbb{1}_{x\leq y}$. Such assumptions make it possible to reduce the system to a closed system of 3 ODE whose unknowns are V, the total number of polymers $P = \sum n_i$, and the total mass $M = \sum i n_i$. However, such assumptions are difficult if not impossible to justify biologically. Some experimental results even tend to prove the contrary [145].

Another major difference between the discrete system (5.2)–(5.1) and the continuous version (5.5)–(5.7) lies in the boundary condition (5.7). This condition needs to be done for the continuous problem to be well-posed. However, nothing justifies it *a priori*, except the intuition of no influx. In [9], we gave it a weak definition and proved it under some assumptions on the coefficients (Theorem 3 of [9]).

We settled our study on the basis of existing work on close equations: Lifshitz-Slyozov on the one hand, in [48], where only the polymerization/depolymerization appears, coagulation-fraagmentation equation on the other hand, in [91], where polymerization is replaced by coagulation. In collaboration with T. Goudon and T. Lepoutre, we studied in [9] in which sense and under which assumptions System (5.5)–(5.6) is the asymptotic limit of System (5.1)–(5.2).

The first step to understand formally the link between the two settings is to adimension the discrete equations. We pass from the discrete model to the continuous model by associating to the u_i's a stepwise constant function, constant on each interval $(\varepsilon i, \varepsilon(i+1))$. Then sums over the index i will be interpreted as Riemann sums which are expected to tend to integrals in the continuum limit while finite differences will give rise to derivatives. This makes appear an appropriate scaling and a small parameter ε. We define

$$\chi_i^\varepsilon(x) = \chi_{[i\varepsilon,(i+1)\varepsilon)(x)}, \qquad u^\varepsilon(t,x) := \sum_{i=n_0(\varepsilon)}^{\infty} u_i^\varepsilon(t)\chi_i^\varepsilon(x),$$

and we study the limit $\varepsilon \to 0$, $i \to \infty$ whereas εi remains finite. To understand the physical interpretation of ε, the best way is to look at the mass balance (5.4). In an adimensionned version, it becomes

$$\frac{d}{dt}\left(V^\varepsilon(t) + \frac{\mathcal{U}^\varepsilon}{\mathcal{V}^\varepsilon}\sum_{i=i_0}^{\infty} i n_i^\varepsilon\right) = 0, \tag{5.9}$$

where $\mathcal{U}^\varepsilon$ is the characteristic value for the concentration of polymers u_i, whereas $\mathcal{V}^\varepsilon$ is the one for the monomers V. Therefore, to admit Equation (5.8) as a formal limit, we are led to set

$$\frac{\mathcal{U}^\varepsilon}{\mathcal{V}^\varepsilon} = \varepsilon^2.$$

Let i_M be the average length of a polymer: the polymerized mass $\sum i u_i^\varepsilon$ is thus in the order of $\sum i_M \mathcal{U}^\varepsilon = i_M^2 \mathcal{U}^\varepsilon$. Supposing that the polymerized mass and the monomers' mass V are in the same order of magnitude, it leads to $i_M^2 \mathcal{U}^\varepsilon \approx \mathcal{V}^\varepsilon$. Hence, we get

$$\varepsilon = \frac{1}{i_M},$$

which corresponds to εi_M remaining finite whereas $\varepsilon \to 0$ and $i_M \to \infty$. Our limit corresponds to the case of very large polymers, also called *fibrils* or *amyloids* in our applications.

The average size of polymers go from some hundreds to some hundreds of thousands. Under the following assumption (which is Assumption (8) in [9])

$$\exists\, K,\ \alpha \geq 0,\ 0 \leq \theta \leq 1, \quad 0 \leq B_i \leq K i^{\alpha}, \quad 0 \leq g_i \leq K i^{\theta}, \tag{5.10}$$

the orders of magnitude of each characteristic value should lead to the following rescaled equations

$$\frac{dn_i^{\varepsilon}}{dt} = -\varepsilon^{\theta-1} V^{\varepsilon}(t)\big(g_i n_i^{\varepsilon} - g_{i-1} n_{i-1}^{\varepsilon}\big) - \varepsilon^{\alpha} B_i n_i^{\varepsilon} + 2\varepsilon^{\alpha} \sum_{j=i+1}^{\infty} B_j k_{i,j} n_j^{\varepsilon}, \qquad i \geq i_0, \tag{5.11}$$

$$\frac{dV^{\varepsilon}}{dt} = \lambda - \gamma V^{\varepsilon} - \varepsilon^{\theta+1} V^{\varepsilon} \sum_{i=i_0}^{\infty} g_i n_i^{\epsilon} + 2\varepsilon^{2+\alpha} \sum_{j \geq i_0} \sum_{i<i_0} i k_{i,j} B_j n_j. \tag{5.12}$$

To obtain the convergence, we now need continuity assumptions on the coefficients, namely (Assumptions (22) and (23) of [9])

$$|B_{i+1} - B_i| \leq K i^{\alpha-1}, \quad |g_{i+1} - g_i| \leq K i^{\theta-1}, \quad |\sum_{p=0}^{i-1} \sum_{r=0}^{p-1} k_{r,j+1} - \sum_{p=0}^{i-1} \sum_{r=0}^{p-1} k_{r,j}| \leq K. \tag{5.13}$$

Under such a compactness assumption, up to a subsequence, the coefficients converge towards continuous functions, as expressed by Lemma 1 and 2 of [9]. The proper definition of the convergence for $k(x,y)$ is the most sensitive point. Denoting $\mathcal{M}^1([0,\infty))$ the space of Radon measures on $\mathbb{R}_+$, the limit k is such that $k : y \to k(dx,y) \in \mathcal{M}^1_+([0,+\infty))$ is in $\mathcal{C}([0,\infty); \mathcal{M}^1_+([0,\infty) - weak - \star)$.

Our proof relies on weak compactness arguments in $\mathcal{M}^1$, making use of moments conservation through time. Hence the last important assumption is the following initial condition (Assumption (21) in [9]):

$$\exists\, \sigma > max(\theta, \alpha - 1), \quad V^{\varepsilon}(0) + \sum_{i=n_0}^{\infty} \big(\varepsilon + \varepsilon^2 i + \varepsilon^{2+\sigma} i^{1+\sigma}\big) n_i^{\varepsilon}(0) < +\infty. \tag{5.14}$$

Finally, we needed to exchange the equation for the monomers (5.12) by the mass balance (5.9), for convergence reasons: weak convergence as we obtained is not sufficient for the convergence towards $\int g(x) n(t,x) dx$ in Equation (5.12). In Definition 1 of [9], we defined a *monomer preserving solution* as satisfying both the mass balance and the equation for V in a weak way; this is equivalent to setting the boundary condition (5.7).

We obtained the following convergence results.

Theorem 4 (Theorems 2 and 3 in [9])**.** *Under Assumptions* (5.3)*,* (5.10)*,* (5.13) *and* (5.14)*, there exists a subsequence* (ε_n) *such that*

$$n^{\varepsilon_n} \to n \text{ in } \mathcal{C}([0,T]; \mathcal{M}^1([0,\infty)) - weak - \star, \quad V^{\varepsilon_n} \to V \text{ uniformly on } [0,T].$$

The limit (n,V) *is a weak solution of Equations* (5.5) *and* (5.8)*. Moreover,* (n,V) *is a monomer preserving solution in the following situations:*

- $x_0 = 0$ *and either* $\theta > 0$ *or* τ_i *is constant,*
- $x_0 > 0$ *and the discrete fragmentation coefficients fulfill a strengthened assumption (Assumption (25) in [9]).*

The assumption $x_0 > 0$, if helpful from the point of view of analysis, is not biologically relevant, since this means that the minimal stable size of a polymer (this polymer is called the *nucleus*) is very large compared to 1. In practice, this is not satisfied, typical sizes going from 2 to 5 or 6.

We also noticed that if the fragmentation kernel contains a Dirac mass in x_0^+, then this should appear in the boundary condition (5.7) instead of 0.

All these observations helped us to generalize the prion model to other types of polymerization, including spontaneous formation of polymers from monomers alone.

5.2 General protein polymerization models [18]

The above-mentioned "Prion" model may be viewed as a toy model, which as such remains amenable to qualitative analysis and helps us to understand typical behaviour of fibrils population. In this spirit, its nonlinear asymptotic behaviour was investigated and discussed in [43, **?**].

However, it remains oversimplified in view of the complexity of the possible chain reactions. For instance, there is no biological evidence that polymerization occurs through monomer addition, rather than oligomer addition; on the contrary, H. Rezaei et al. [56] showed the existence of several polymerization pathways for recombining PrPc

Moreover, the prion model is a toy model for a prion-like infectious disease: the disease, characterized by the presence of the fibrils, is able to propagate through the polymerization-fragmentation mechanism only. As shown in [43], a small amount of polymers at initial time is sufficient for the process to initiate, forming a large number of other fibrils. But the equations do not model the spontaneous appearance of the disease: if no polymer is initially present, nothing happens, contrarily to models including *nucleation*, which is the spontaneous formation of small oligomers by the encounter of monomers.

In order to generalize the derivation of [9] and to take into account such reactions, in [18] we wrote a more complete model. It is not designed for direct use and study, but to give a method to adapt to each situation.

In its discrete version, the model is the transcription in differential equations of the law of mass action for the following reactions.

- **Conformational exchange:** inert monomers of concentration V^* may become active conformers of concentration V, able to polymerize

$$V^* \underset{k_I^-}{\overset{k_I^+}{\rightleftharpoons}} V, \qquad \frac{dV^*}{dt} = -k_I^+ V^* + k_I^- V + \cdots, \quad \frac{dV}{dt} = k_I^+ V^* - k_I^- V + \cdots$$

- **Nucleation:** the smallest stable polymer of concentration n_{i_0} is of size i_0, and may be formed from i_0 monomers according the following reactions of association and dissociation and their translation in ODE

$$\underbrace{V+\cdots+V}_{i_0} \underset{k_{off}^N}{\overset{k_{on}^N}{\rightleftharpoons}} n_{i_0}, \qquad \frac{dV}{dt} = -i_0 k_{on}^N V^{i_0} + i_0 k_{off}^N n_{i_0}, \quad \frac{dn_{i_0}}{dt} = k_{on}^N V^{i_0} - k_{off}^N n_{i_0} \tag{5.15}$$

- polymerization and depolymerization occur by monomer addition, at rates g_i and g_i^-,
- fragmentation of $j-$mers into $i-$mers occurs like in the prion model, with a fragmentation rate $B_j k_{i,j}$,
- coalescence of i-mers and j-mers giving rise to $(i+j)-$ mers occurs at a rate $k_{col}^{i,j}$.
- we can add degradation (as in the prion model), but this reaction being very simple, for the sake of simplicity we don't consider it here and let the reader refer to [9, 18].

From the discrete transcription of these reactions, we formally derived the following PDE model (keeping the same notations)

$$\frac{dV^*}{dt} = -k_I^+ V^* + k_I^- V, \tag{5.16}$$

$$\frac{dV}{dt} = k_I^+ V^* - k_I^- V - \frac{i_0\, k_{on}^N\, V^{i_0+1} g(x_0)}{k_{off}^N + g(x_0)V} - V \int_{x_0}^{\infty} g(x)\, n(t,x)dx + \int_{x_0}^{\infty} g^-(x) n(t,x)dx, \tag{5.17}$$

$$\begin{aligned}\frac{\partial n}{\partial t} &= -V\frac{\partial}{\partial x}\big(g(x)n(t,x)\big) + \frac{\partial}{\partial x}\big(g^-(x)n(t,x)\big) + 2\int_x^{\infty} B(y)k(x,y)\, n(t,y)dy \\ &\quad -B(x)n(t,x) + \frac{1}{2}\int_{x_0}^{x} k_{col}(y,x-y)n(t,y)n(t,x-y)dy \; - \int_{x_0}^{\infty} k_{col}(x,y)n(t,x)\, n(t,y)dy,\end{aligned} \tag{5.18}$$

$$g(x_0)n(t,x_0) = g(x_0)\frac{k_{on}^N\, V^{i_0}}{k_{off}^N + g(x_0)V}. \tag{5.19}$$

The main novelty is the nucleation term, which appears now as a source term in the boundary condition (5.19) replacing (5.7). By this model, we proposed a new framework that can be adapted to most protein polymerization reactions. While keeping all the richness of the reactions, it may reduce the computational cost and is much more amenable to analysis; this is particularly convenient for parameter estimation in which the model has to be evaluated a large number of times ([158, 28]. Moreover, it proved to give at least as good results as

previously used methods in the biophysical literature [158, 87]. Recent analytical tools as the ones described above can also be applied.

As a proof of concept, we applied our method to new experimental data on Polyglutamine aggregation, which is involved in Huntington's disease. Our biologists collaborators carried out biophysical analyses to investigate the aggregation kinetics of PolyQ41, which are peptides containing a repetition of 41 glutamine residues per monomer. Such a length of PolyQ repetition per molecule was proved to be sufficient to induce aggregation *in vitro* and in transfected cells [143]. The experiments consist in letting an initial bath of monomers aggregate spontaneously, till all the monomers become polymerized.

Due to its simplicity, PolyQ provides an excellent model system to test our mathematical model. According to the experimental observations, fragmentation can be ignored as well as coalescence. In order to determine whether coalescence occurs, we monitored simultaneously two types of measurements. The first one, by static light scattering, can be viewed as a measurement of $I_2(t) = \sum\limits_{i \geq i_0} i^2 c_i = \int x^2 c(t,x)dx$. The second one, by thioflavine T (ThT) fluorescence, is mathematically expressed by $M(t) = \sum_{i \geq i_0} i c_i = \int x c(t,x)dx$. If there were coalescence, the weighted average polymer size would continue to grow even when the total polymerized mass $M(t)$ reached a plateau, so the second moment $I_2(t)$ would continue to grow after the plateau has been reached by $M(t)$. In our experiments, however, both curves reached the plateau roughly simultaneously; therefore we concluded that coalescence is negligible.

We also ignored depolymerization in the course of the experiment, so that the model became quite simple: a pure transport equation, with a nonlinear source term for the nucleation, and quadratically coupled with two ODEs standing for the monomer's and conformer's concentrations

$$\frac{dV^*}{dt} = -k_I^+ V^* + k_I^- V, \tag{5.20}$$

$$\frac{dV}{dt} = k_I^+ V^* - k_I^- V - \frac{i_0\, k_{on}^N\, g(0) V^{i_0+1}}{k_{off}^N + g(0)V} - V \int\limits_{x_0}^{\infty} g(x)\, n(t,x)dx, \tag{5.21}$$

$$\frac{\partial n}{\partial t} = -V \frac{\partial}{\partial x}(g(x)n(t,x)), \tag{5.22}$$

$$n(t,0) = \frac{k_{on}^N V^{i_0}}{k_{off}^N + g(0)V}. \tag{5.23}$$

As an initial approach, we tested piecewise linear polymerization rates.

The parameters of this model were then estimated by fitting experimental data on PolyQ41 protein polymerization. We performed this in two successive ways. The first consisted in fitting separately each experimental curve, corresponding to a given experiment, at a given concentration. The result is that whatever i_0 is, the fit is excellent for any curve, with a measurement error from 0.5 to 2% in L^2 adimensioned norms. It gives almost undistinguish-

able curves. However, the variability among the optimal coefficients was large, which led us to the second step. This consisted in fitting *simultaneously* all the curves of experiments carried out in identical experimental conditions, but for different concentrations. The total adimensioned error (in L^2-norm) diminished with i_0, and reached its lowest level for $i_0 = 1$. For larger values of the nucleus, the error is moreover too large for the model to be acceptable. It gave solid ground to the assumption, already suggested in the literature [82], that the nucleus is of size 1, but with a specific and unconventional nucleation-elongation reaction scheme, where the elongating species V and the nucleus $n_{i_0} = \tilde{V}$ are distinct conformers.

Another result of our simulations is that k_I^- is negligible, thus we can suppose that $c_1 = c_0 e^{-k_I^+ t}$. In the same way, we can compare V to the solution of the following differential equation

$$\frac{dV_{test}}{dt} = k_I^+ V(t=0) e^{-k_I^+ t} - i_0 k_{on}^N V_{test}^{i_0}, \quad V_{test}(0) = 0,$$

i.e., ignore the contribution of polymers in the equation for V : it fits perfectly for the total duration of the *lag phase*, which is the period during which the reaction initiates - an easy definition for the lag phase could be, for instance, before the polymerized mass had reached a level of 5% or 10% of the total mass.

6 Perspectives

6.1 Application to protein polymerization

The framework model (5.16)–(5.19) is the basis of a project funded by the European Research Council as a Starting Grant of the program called IDEAS. It is called $SKIPPER^{AD}$, which stands for *Simulation of the Kinetics and Inverse Problem for the Protein polymERization in Amyloid Diseases (Prion, Alzheimer's).*

Amyloid diseases are a group of diseases which involve the aggregation and the deposition of misfolded proteins, called *amyloid*, which are specific for each disease (PrP for Prion, Aβ for Alzheimer's). In a healthy state, they remain monomeric, but when misfolded they propagate the abnormal configuration and aggregate to others, forming very long polymers also called *fibrils*. Elucidating the intrinsic mechanisms of these chain reactions, most probably specific for each disease, is a major challenge of molecular biology: do polymers break or do they coalesce? Do some specific sizes polymerize faster? Does polymerization occur by monomer, dimer, or i-mer addition? On which part of the reactions should a treatment focus to arrest the disease? Up to now, only very partial and partially justified answers have been provided.

The project aims at developing new mathematical methods in order to model fibrillization reactions, analyse experimental data, help the biologists to discover the key mechanisms of polymerization in these diseases and predict the effects of new therapies. Through $SKIPPER^{AD}$, I intend to enlarge my current research to more general models (built from

the framework method described in [18]), combine analytical with statistical and probabilistic viewpoints on growth-fragmentation processes, and permanently adjust my mathematical research to the application to protein polymerization.

The project takes place from December 2012 to December 2017, and will be the guideline for my research activity during these years. Human Rezaei, research director in biophysics at INRA Jouy-en-Josas and specialist of Prion diseases, is fully involved in the project to which he will devote 30% of his working time. 2 post-doctoral students and 2 PhD students will be recruited, and the project will also facilitate the ongoing collaborations.

Here are the research directions which are specific of this project. However, many problems detailed in the other sections also belong to the $SKIPPER^{AD}$ project, but as mathematical problems, they have a wider field of application so that they can be considered *per se*.

6.1.1 Adapting the framework model (5.16)–(5.19) to specific situations

This model can be adapted, simplified or made more complex to fit particular cases, for instance to take into account a conformer/monomer exchange, or to ignore coalescence or depolymerization, etc., with much freedom and many possible variants. We depart from existing experimental data and biological questions. Through careful discussions with H. Rezaei, we shall define how much we can rely on these data, what are the origins of the noises and measurement errors, and we will formulate a model and one (or better several) suitable numerical scheme(s). To build such a model, we will adapt our general problem (5.16)-(5.19) to the case corresponding to the experiment under consideration. Even for the same disease and the same protein, it may not be the same part of the model that will play the main role according to the experimental conditions (sonication or not, high or low temperature, monomeric or fibril initial conditions etc.)

This modelling part will be the basis of mathematical studies, even if these ones could then be considered as problems with potentially wider applications and interesting in their own right.

6.1.2 Sensitivity analysis (in collaboration with H.T. Banks)

Once such a model is designed, to gain a first insight into which part of the reaction which type of process influences, an interesting research direction is given by the generalized sensitivity functions derived by H.T. Banks and co-authors [28, 30]. This method could also help us to design better experiments, in order to obtain experimental data at the right moment where it depends most on the parameters being sought.

6.1.3 Fibrils Depolymerization: Lifshitz-Slyozov revisited and data assimilation methods

As an example of such an adaptation of the general model, here is an ongoing work in collaboration with H. Rezaei, S. Prigent, J. Torrent (for the biological part), P. Moireau, H.W. Haffaf and M. Fragu (for the mathematical part).

Departing only from fibrils (of $PrPsc$, the Prion proteins) as an initial condition at time $t = 0$ and with no monomers, H. Rezaei's team perform measurements of the total population of polymers, under experimental conditions that favour depolymerization. Surprisingly, they observe oscillations.

Our first model to describe the behaviour of the population follows either the Becker-Döring system, in a discrete way [25], or the Lifshitz-Slyozov equations [117], namely:

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x}\Big(\big(V(t)g(x) - g^-(x)\big)\, n\Big) = 0 \tag{6.1}$$

$$\frac{dV}{dt} = \int_0^\infty \big(g^-(x) - Vg(x)\big)n(t,x)dx, \tag{6.2}$$

$$n(t, x = 0) = \frac{k_{on}^N V^{i_0}}{k_{off}^N + g(0)V}. \tag{6.3}$$

Although this system (except the boundary condition (6.3)) and its asymptotic limits have been widely studied, our biological focus sheds new light on this problem since our assumptions regarding the coefficients are far from the original ones as stated in the seminal paper of Lifshitz and Slyozov [99, 119, 118] or Becker and Döring [33, 25, 81]. We will investigate the following questions.

1. Is it possible to obtain permanent oscillations in the measured quantity $I_2 = \int x^2 n(t,x)dx$, as experimental observations seem to show? Our initial studies seem to answer in a negative way, and also to show that the fact that we measure I_2 rather than the total polymerised mass $I = \int xn(t,x)dx$ has a major influence. If so, which kind of reactions could explain this unusual phenomenon ?

   We are also studying under which conditions on the polymerization and depolymerization coefficients we can observe multiple steady states for a given initial condition. Hopf bifurcation and the studies carried out for related equations such as Prion provide us with different research directions.

2. Can we deduce the polymerization and depolymerization size-structured rates from the observation of I_2 or I over time - with several experiments corresponding to different initial concentrations?

This second question belongs to the general problem of identifiability or observability of our system through the data provided. Ideally, an observability inequality involving the

parameters to identify and the given data can be obtained, justifying the well-posedness of the inverse problem.

To go further, a direction well-adapted is provided by data assimilation methods.

Data assimilation strategies [116], initially defined for evolution equations in weather forecasting and geophysics, are now widely used for evolution problems in chemical systems, biomechanical systems etc. We plan to investigate how such strategies can be applied to our problems.

In essence, data assimilation consists in merging model and data in order to circumvent the initial model uncertainties and improve the prediction computed. In the data assimilation context, we distinguish two different approaches: first, the variational approach which estimates the uncertainties by minimizing a least square criterion involving the discrepancy between the data and the corresponding outputs from a model simulation [116]; secondly, the sequential approach which filters the uncertainties over time to stabilize the computed numerical system on the actual system partially observed – see [147] and references therein. The classic examples of this approach are Kalman based filtering methods. These two strategies are different in their practical use but both rely on the same fundamental *observability condition* which expresses the fact that observing the system even partially – through a time sequence of mean quantities observed in our case – is sufficient to compensate the lack of initial knowledge about the system [147]. Moreover, the two approaches can be proved to be equivalent in various cases. Despite their easier implementation in comparison to variational strategies, most sequential approaches are known to suffer from a "curse of dimensionality" as recalled by Bellman [38], and therefore are not widely used for systems derived from the discretization of partial differential equations.

However, some recent studies [115, 113] have paved the way to adapt filtering strategies to the joint state-parameter estimation of large dimensional problems. P. Moireau and co-authors have demonstrated the efficiency of some adapted filtering methods from a theoretical point of view [114] and also in a clinical context. In collaboration with this group, we expect to generalize these results to our polymerization-fragmentation problems.

It seems very likely that our inverse problems, in their most general formulation, fall within the scope of data assimilation methods which will give a well identified methodological framework for solving them in practice. We can then rely on a generic data assimilation library called Verdandi (http://verdandi.gforge.inria.fr/) already available and developed by P. Moireau and co-authors, where their methods are available, alongside state-of-the-art classical data assimilation algorithms. The numerical results obtained will help us identify observability issues, for example by analyzing uncertainty covariances resulting from the application of well-adapted methods.

6.2 Combining Statistical and Analytical Approaches

This part or my research, originated by a very informal workshop with Marc Hoffmann, Patricia Reynaud and Vincent Rivoirard in 2008, progressively grows in importance.

Following our common work [11], this collaboration now takes two directions

On the one hand, I have been associated to the project CALIBRATION funded by the ANR, and supervised by Vincent Rivoirard, and then to the project "PIECE" of Florent Malrieu. In this context, we imagine several possible directions: comparison between the Goldenschluger and Lepski's method and of the deterministic *discrepancy principle*; reconstruction methods with unknown level of noise ; analytical viewpoint on the Hawkes processes, which seem to have much in common with nonlinear renewal equations [135].

On the other hand, we now collaborate with Lydia Robert, biologist at INRA and ENS, on the cell division cycle of the bacteria E. Coli. For the most theoretical part, we build a probabilistic modelling of the division, through a branching process whose empirical distribution satisfies the growth-fragmentation equation in expectancy. We also propose a statistical method to estimate the division rate B from measurements of the distribution of the sizes of dividing cells (and not of all the sizes, as in [11]). Finally, we progressively enrich our methods to give answers to biological issues studied by Lydia Robert. Among these issues, we faced again the fundamental question raised in the introduction: what is the really intrinsic structuring variable ? Is it size, age, or still some other hidden parameter?

6.3 Population Dynamics

Following [4, 7, 2], there remain many open problems in population dynamics, some known to be very hard. I only quote here three possible directions.

- Nonlinear dynamics

  Partial results have been proved for the nonlinear asymptotic behaviour of of the growth-fragmentation equation in [67, 43, **?**, 123, 124, 125], but the convergence towards a nontrivial steady state has not yet been proved in general cases.

  This rejoins the general question of nonlinear asymptotic behaviour of structured equations, which remains largely open [129, 106, 109].

- Optimal assumptions for a spectral gap

  Several important advances to prove an exponential rate of decay towards the steady distribution have been made; on the one hand, for perturbations of constant or decaying division rate [128, 92, 124, 125], and on the other hand for perturbations of a uniform fragmentation kernel [42, 24]. However, more general assumptions, maybe as general as for the well-posedness of the eigenvalue problem [7], could probably yet be found.

- Non standard asymptotic behaviours

  In [7], we found (probably optimal) assumptions under which the solution to Equation (2.1) behaves asymptotically as an exponential growth decoupled of a stable steady profile, *i.e.* $u(t,x)e^{-\lambda t} \to U(x)$. Other articles [60, 61, 63] focus on cases where fragmentation or coagulation equations tend either to Dirac masses or to infinitely-long

polymers. To complete the theory, it would be necessary to have optimal assumptions for each case, as well as a study of limit cases - where oscillations may appear [77].

Part III
Selection of articles [7, 15, 11, 9, 18]

1 Eigenvalue Problem [7]

Eigenelements of a General Aggregation-Fragmentation Model.

Marie Doumic * Pierre Gabriel †

June 12, 2009

Abstract

We consider a linear integro-differential equation which arises to describe both aggregation-fragmentation processes and cell division. We prove the existence of a solution $(\lambda, \mathcal{U}, \phi)$ to the related eigenproblem, where λ denotes its first eigenvalue. Our study concerns a non-constant transport term that can vanish at $x = 0$. Non lower-bounded transport terms bring difficulties to find *a priori* estimates. We use weighted-norms to solve this problem.

Keywords Aggregation-fragmentation equations, eigenproblem, size repartition, polymerization process, cell division, long-time asymptotic.

AMS Class. No. 35A05, 35B40, 45C05, 45K05, 82D60, 92D25

1 Introduction

Competition between growth and fragmentation is a common phenomenon for a structured population. It arises for instance in a context of cell division (see, among many others, [22, 1, 4, 5, 6, 14, 30, 18, 20]), polymerization (see [7, 13]), telecommunication (see [2]) or neurosciences (see [27]). It is also a mechanism which rules the proliferation of prion's proteins (see [10, 19, 21]). These proteins are responsible of spongiform encephalopaties and appear in the form of aggregates in infected cells. Such polymers grow attaching non infectious monomers and converting them into infectious ones. On the other hand they increase their number by splitting. To describe such phenomenons, we write the following integro-differential equation, used in [10, 19, 21] to study the proliferation of prion,

$$\begin{cases} \dfrac{\partial}{\partial t}u(x,t) + \dfrac{\partial}{\partial x}\big(\tau(x)u(x,t)\big) + \beta(x)u(x,t) = 2\displaystyle\int_x^\infty \beta(y)\kappa(x,y)\,u(y,t)\,dy, \qquad x \geqslant 0, \\ u(x,0) = u_0(x), \\ u(0,t) = 0. \end{cases} \tag{1}$$

*INRIA Rocquencourt, projet BANG, Domaine de Voluceau, BP 105, F-78153 Rocquencourt, France.

†Université Pierre et Marie Curie-Paris 6, UMR 7598 LJLL, BC187, 4, place de Jussieu, F-75252 Paris cedex 5; corresponding author, email: gabriel@ann.jussieu.fr

The aggregates density function $u(x,t)$ represents the quantity of polymers of size x at time t. They aggregate monomers with the rate $\tau(x)$. Equation (1) also takes into account the fragmentation of a polymer of size y into two smaller polymers of size x and $y-x$. This fragmentation occurs with a rate $\beta(y)$ and produce an aggregate of size x with the rate $\kappa(x,y)$. Equation (1) is a particular case of the more general one

$$\frac{\partial}{\partial t}u(x,t)+\frac{\partial}{\partial x}\big(\tau(x)u(x,t)\big)+[\beta(x)+\mu(x)]u(x,t)=n\int_x^\infty \beta(y)\kappa(x,y)\,u(y,t)\,dy, \qquad x\geqslant x_0, \quad (2)$$

with the bound condition $u(x_0,t)=0$ (see [3, 10, 21]). In this model, polymers are broken in an average of $n>1$ smaller ones by the fragmentation process, there is a death term $\mu(x)\geq 0$ representing degradation, and a minimal size of polymers x_0 which can be positive. Our results remain true for this generalization.

To study the asymptotic growth of the quantity of polymers, we are looking at the eigenelements $(\lambda,\mathcal{U},\phi)$ of (1), *i.e.*, the solution of the equation

$$\begin{cases} \dfrac{\partial}{\partial x}(\tau(x)\mathcal{U}(x))+(\beta(x)+\lambda)\mathcal{U}(x)=2\displaystyle\int_x^\infty \beta(y)\kappa(x,y)\mathcal{U}(y)dy, \qquad x\geqslant 0,\\ \\ \tau\mathcal{U}(x=0)=0, \qquad \mathcal{U}(x)\geq 0, \qquad \int_0^\infty \mathcal{U}(x)dx=1,\\ \\ -\tau(x)\dfrac{\partial}{\partial x}(\phi(x))+(\beta(x)+\lambda)\phi(x)=2\beta(x)\displaystyle\int_0^x \kappa(y,x)\phi(y)dy, \qquad x\geqslant 0,\\ \\ \phi(x)\geq 0, \qquad \int_0^\infty \phi(x)\mathcal{U}(x)dx=1. \end{cases} \quad (3)$$

Here λ is the first eigenvalue, $\mathcal{U}$ the corresponding eigenfunction and ϕ the dual eigenfunction. Our parameters $\beta,\ \tau$ and κ are not necessarily regular. For the first equation (equation on $\mathcal{U}$) we are looking for $\mathcal{D}'$ solutions defined as follows : $\mathcal{U}\in L^1(\mathbb{R}^+)$ is a $\mathcal{D}'$ solution if $\forall\varphi\in\mathcal{C}_c^\infty(\mathbb{R}^+)$,

$$-\int_0^\infty \tau(x)\mathcal{U}(x)\partial_x\varphi(x)\,dx+\lambda\int_0^\infty \mathcal{U}(x)\varphi(x)\,dx=\int_0^\infty \beta(x)\mathcal{U}(x)\Big(2\int_0^\infty \varphi(y)\kappa(y,x)\,dy-\varphi(x)\Big)\,dx. \quad (4)$$

Concerning the dual equation, we are looking for a solution $\phi\in W^{1,\infty}_{loc}(0,\infty)$ such that the equality holds in $L^1_{loc}(0,\infty)$, *i.e.* almost everywhere.

Existence and uniqueness of such elements has already been proved for general fragmentation kernels $\kappa(x,y)$ and fragmentation rates $\beta(x)$, but with very particular polymerization rates $\tau(x)$, namely constant ($\tau\equiv 1$ in [28]), homogeneous ($\tau(x)=x^\mu$ in [23]) or with a compact support ($Supp\,\tau=[0,x_M]$ in [14]). It seems however of deep interest to consider more general τ as [10, 32] suggest. The proof of [28] can be adapted for non constant rates but still positive and bounded ($0<m<\tau(x)<M$). The paper [23] gives results for $\tau(0)=0$, but for a very restricted class of shape for τ. The paper [14] gives results for τ with general shape in the case where there is also an age variable (integration in age then allows to recover Problem (1)), but requires a compact support and regular parameters. Here we consider polymerization rates that can vanish at $x=0$, positive for x positive, with general shape and few regularity for the all parameters ($\tau,\ \beta$ and κ), and we have to reformulate the space for solutions.

Theorem 1 (Existence and Uniqueness) *Under assumptions* (5)-(13)*, there exists a unique solution* $(\lambda,\mathcal{U},\phi)$ *(in the sense we have defined before) to the eigenproblem* (3) *and we have*

$$\lambda > 0,$$

$$x^\alpha \tau \mathcal{U} \in L^p(\mathbb{R}^+), \quad \forall \alpha \geq -\gamma, \quad \forall p \in [1,\infty],$$

$$x^\alpha \tau \mathcal{U} \in W^{1,1}(\mathbb{R}^+), \quad \forall \alpha \geq 0$$

$$\exists k > 0 \ s.t. \ \frac{\phi}{1+x^k} \in L^\infty(\mathbb{R}^+),$$

$$\tau \frac{\partial}{\partial x}\phi \in L^\infty_{loc}(\mathbb{R}^+).$$

The end of this paper is devoted to define precisely the assumptions and prove this theorem. It is organized as follows : in Section 2 we describe the assumptions and give some examples of interesting parameters. In Section 3 we prove Theorem 1 using *a priori* bounds on weighted norms and then we give some consequences and perspectives in Section 4. The proof of technical lemmas and theorem can be found in the Appendix.

2 Coefficients

2.1 Assumptions

For all $y \geq 0$, $\kappa(.,y)$ is a nonnegative measure with a support included in $[0,y]$. We define κ on $(\mathbb{R}_+)^2$ as follows : $\kappa(x,y) = 0$ for $x > y$. We assume that for all continuous function ψ, the application $f_\psi : y \mapsto \int \psi(x)\kappa(x,y)\,dx$ is Lebesgue measurable.
The natural assumptions on κ (see [19] for the motivations) are that polymers can split only in two pieces which is taken into account by

$$\int \kappa(x,y)dx = 1. \tag{5}$$

So $\kappa(y,.)$ is a probability measure and $f_\psi \in L^\infty_{loc}(\mathbb{R}^+)$. The conservation of mass imposes

$$\int x\kappa(x,y)dx = \frac{y}{2}, \tag{6}$$

a property that is automatically satisfied for a symetric fragmentation (*i.e.* $\kappa(x,y) = \kappa(y-x,y)$) thanks to (5). For the more general model (2), assumption (6) becomes $\int x\kappa(x,y)dx = \frac{y}{n}$ to preserv the mass conservation.
We also assume that the second moment of κ is less than the first one

$$\int \frac{x^2}{y^2}\,\kappa(x,y)dx \leq c < 1/2 \tag{7}$$

(it becomes $c < 1/n$ for model (2)). We refer to the Examples for an explanation of the physical meaning.

For the polymerization and fragmentation rates τ and β, we introduce the set

$$\mathcal{P} := \big\{ f \geq 0 \,:\, \exists \mu, \nu \geq 0,\ \limsup_{x\to\infty} x^{-\mu} f(x) < \infty \text{ and } \liminf_{x\to\infty} x^{\nu} f(x) > 0 \big\}$$

and the space

$$L^1_0 := \big\{ f,\ \exists a > 0,\ f \in L^1(0,a) \big\}.$$

We consider

$$\tau,\, \beta \in L^1_{loc}(\mathbb{R}^+) \cap \mathcal{P} \tag{8}$$

satisfying

$$\forall K \text{ compact of } (0,\infty),\ \exists m_K > 0 \quad s.t. \quad \tau(x) \geq m_K \text{ for } a.e.\ x \in K \tag{9}$$

and

$$\exists b \geq 0, \quad Supp\beta = [b,\infty). \tag{10}$$

Assumption (10) is necessary to prove uniqueness and existence for the adjoint problem.

To avoid shattering (zero-size polymers formation, see [3, 21]), we assume

$$\exists\, C > 0, \gamma \geq 0 \quad s.t. \qquad \int_0^x \kappa(z,y)\, dz \leq \min\Big(1, C\Big(\frac{x}{y}\Big)^{\gamma}\Big) \qquad \text{and} \qquad \frac{x^\gamma}{\tau(x)} \in L^1_0 \tag{11}$$

$$\frac{\beta}{\tau} \in L^1_0. \tag{12}$$

On the other hand, to avoid forming infinitely long polymers (gelation phenomenon, see [17, 16]), we assume

$$\lim_{x\to+\infty} \frac{x\beta(x)}{\tau(x)} = +\infty. \tag{13}$$

Remark 1 *In case when* (11) *is satisfied for* $\gamma > 0$, *then* (7) *is automatically fulfilled (see Lemma 3 in the Appendix).*

2.2 Examples

First we give some examples of coefficients which satisfy or not our previous assumptions.
For the fragmentation kernel, we first check the assumptions (5) and (6). They are satisfied for autosimilar measures, namely $\kappa(x,y) = \frac{1}{y}\kappa_0(\frac{x}{y})$, with κ_0 a probability measure on $[0,1]$, symmetric in $1/2$. Now we exhibit some κ_0.

General mitosis :

$$\kappa_0^r = \frac{1}{2}(\delta_r + \delta_{1-r}) \qquad \text{for} \qquad r \in [0, 1/2]. \tag{14}$$

Assumption (11) is satisfied for any $\gamma > 0$ in the cases when $r \in (0, 1/2]$. So (7) is also fulfilled thanks to Remark 1. The particular value $r = 1/2$ leads to equal mitosis ($\kappa(x,y) = \delta_{x=\frac{y}{2}}$).
The case $r = 0$ corresponds to the renewal equation ($\kappa(x,y) = \frac{1}{2}(\delta_{x=0} + \delta_{x=y})$). In this case, the size of the polymers does not decrease during the fragmentation process. It is typically what we want to avoid with assumption (7). For such a fragmentation kernel, assumption (11) is satified only for $\gamma = 0$,

and the moments $\int z^k \kappa_0(z)dz$ are equal to $1/2$ for all $k > 0$, so (7) does not hold true. However, if we consider a convex combination of κ_0^0 with an other kernel such as κ_0^r with $r \in (0, 1/2]$, then (11) remains false for any $\gamma > 0$ but (7) is fulfilled. Indeed we have for $\rho \in (0, 1)$

$$\int z^2(\rho\kappa_0^0(z) + (1-\rho)\kappa_0^r(z))\, dz = \frac{\rho}{2} + \frac{1-\rho}{2}(r^2 + (1-r)^2) = \frac{1}{2}(1 - 2r(1-r)(1-\rho)) < \frac{1}{2}.$$

Homogeneous fragmentation :

$$\kappa_0^\alpha(z) = \frac{\alpha+1}{2}(z^\alpha + (1-z)^\alpha) \qquad \text{for} \qquad \alpha > -1. \tag{15}$$

The parameter $\gamma = 1 + \alpha > 0$ suits for (11) and so (7) is fulfilled. Uniform repartition ($\kappa(x, y) = \frac{1}{y}\mathbb{1}_{0\le x\le y}$) corresponds to $\alpha = 0$ and is also included.

This last case of uniform repartition is useful because it provides us with explicit formulas for the eigenelements. For instance, we can consider the two following examples.

First example : $\tau(x) = \tau_0,\ \beta(x) = \beta_0 x$.
In this case, widely used by [19], the eigenelements exist and we have

$$\lambda = \sqrt{\beta_0\tau_0},$$

$$\mathcal{U}(x) = 2\sqrt{\frac{\beta_0}{\tau_0}}\Big(X + \frac{X^2}{2}\Big)e^{-X-\frac{X^2}{2}}, \quad \text{with } X = \sqrt{\frac{\beta_0}{\tau_0}}x,$$

$$\phi(x) = \frac{1}{2}(1+X).$$

Second example : $\tau(x) = \tau_0 x$.
For such β for which there exists eigenelements, we have

$$\lambda = \tau_0 \qquad \text{and} \qquad \phi(x) = \frac{x}{\int y\mathcal{U}(y)}.$$

For instance when $\beta(x) = \beta_0 x^n$ with $n \in \mathbb{N}^*$, then the eigenelements exist and we can compute $\mathcal{U}$ and ϕ and we have the formulas in Table 1. In this table we can notice that $\mathcal{U}(0) > 0$ but the boundary condition $\tau\mathcal{U}(0) = 0$ is fulfilled.

$n = 1$	$\lambda = \tau_0$	$\mathcal{U}(x) = \frac{\beta_0}{\tau_0}e^{-\frac{\beta_0}{\tau_0}x}$	$\phi(x) = \frac{\beta_0}{\tau_0}x$
$n = 2$	$\lambda = \tau_0$	$\mathcal{U}(x) = \sqrt{\frac{2\beta_0}{\pi\tau_0}}e^{-\frac{1}{2}\frac{\beta_0}{\tau_0}x^2}$	$\phi(x) = \sqrt{\frac{\pi\beta_0}{2\tau_0}}x$
n	$\lambda = \tau_0$	$\mathcal{U}(x) = \Big(\frac{n\beta_0}{\tau_0}\Big)^{\frac{1}{n}}\frac{n}{\Gamma(\frac{1}{n})}e^{-\frac{1}{n}\frac{\beta_0}{\tau_0}x^n}$	$\phi(x) = \Big(\frac{\tau_0}{n\beta_0}\Big)^{\frac{1}{n}}\frac{\Gamma(\frac{1}{n})}{\Gamma(\frac{2}{n})}x$

Table 1: The example $\tau(x) = \tau_0 x,\ \beta(x) = \beta_0 x^n$ and uniform repartition $\kappa(x, y) = \frac{1}{y}\mathbb{1}_{0\le x\le y}$. The table gives the eigenelements solution to (3).

Now we turn to non-existence cases. Let us consider constant fragmentation $\beta(x) = \beta_0$ with an affine polymerization $\tau(x) = \tau_0 + \tau_1 x$, and any fragmentation kernel κ which satisfies to assumptions (5)-(6). We notice that (13) is not satisfied and look at two instructive cases.

First case : $\tau_0 = 0$.
In this case assumption (12) does not hold true. Assume that there exists $\mathcal{U} \in L^1(\mathbb{R}^+)$ solution of (3) with the estimates of Theorem 1. Integrating the equation on $\mathcal{U}$ we obtain that $\lambda = \beta_0$, but multiplying the equation by x before integration we have that $\lambda = \tau_1$. We conclude that eigenelements cannot exist if $\tau_1 \neq \beta_0$.
Moreover, if we take $\kappa(x,y) = \frac{1}{y}\mathbb{1}_{0\leq x\leq y}$, then a formal computation shows that any solution to the first equation of (3) belongs to the plan $Vect\{x^{-1}, x^{-\frac{2\beta_0}{\tau_1}}\}$. So, even if $\beta_0 = \tau_1$, there does not exist an eigenvector in L^1.

Second case : $\tau_0 > 0$.
In this case (12) holds true but the same integrations than before lead to

$$\int x\mathcal{U}(x)\,dx = \frac{\tau_0}{\beta_0 - \tau_1}.$$

So there cannot exist any eigenvector $\mathcal{U} \in L^1(x\,dx)$ for $\tau_1 \geq \beta_0$.

3 Proof of the main theorem

3.1 A preliminary lemma

Before proving Theorem 1, we give a preliminary lemma, useful to prove uniqueness of the eigenfunctions.

Lemma 1 (Positivity) *Consider $\mathcal{U}$ and ϕ solutions to the eigenproblem* (3)*.*
We define $m := \inf_{x,y}\{x \,:\, (x,y) \in Supp\,\beta(y)\kappa(x,y)\}$. Then we have, under assumptions (5)*,* (6)*,* (9) *and* (10)

$$Supp\,\mathcal{U} = [m,\infty) \qquad and \qquad \tau\mathcal{U}(x) > 0 \quad \forall x > m,$$
$$\phi(x) > 0 \quad \forall x > 0.$$

If additionaly $\frac{1}{\tau} \in L^1_0$, then $\phi(0) > 0$.

Remark 2 *In case $Supp\,\kappa = \{(x,y)/x \leq y\}$, then $m = 0$ and Lemma 1 and Theorem 1 can be proved without the connexity condition* (10) *on the support of β.*

Proof. Let $x_0 > 0$, we define $F : x \mapsto \tau(x)\mathcal{U}(x)e^{\int_{x_0}^x \frac{\lambda+\beta(s)}{\tau(s)}ds}$. We have that

$$F'(x) = 2e^{\int_{x_0}^x \frac{\lambda+\beta(s)}{\tau(s)}ds} \int \beta(y)\kappa(x,y)\mathcal{U}(y)\,dy \geq 0. \tag{16}$$

So, as soon as $\tau\mathcal{U}(x)$ once becomes positive, it remains positive for larger x.

We define $a := \inf\{x \,:\, \tau(x)\mathcal{U}(x) > 0\}$. We first prove that $a \leq \frac{b}{2}$. For this we integrate the equation on $[0,a]$ to obtain

$$\int_0^a \int_a^\infty \beta(y)\kappa(x,y)\mathcal{U}(y)\,dydx = 0,$$

$$\int_a^\infty \beta(y)\mathcal{U}(y)\int_0^a \kappa(x,y)\,dxdy = 0.$$

Thus for almost every $y \geq \max(a,b)$, $\int_0^a \kappa(x,y)\,dx = 0$. As a consequence we have

$$1 = \int \kappa(x,y)\,dx = \int_a^y \kappa(x,y)\,dx \leq \frac{1}{a}\int x\kappa(x,y)\,dx = \frac{y}{2a}$$

thanks to (5) and (6), and this is possible only if $b \geq 2a$.

Assume by contradiction that $m < a$, integrating (3) multiplied by φ, we have for all $\varphi \in \mathcal{C}_c^\infty$ such that $Supp\,\varphi \subset [0,a]$

$$\int\int \varphi(x)\beta(y)\kappa(x,y)\mathcal{U}(y)\,dydx = 0. \tag{17}$$

By definition of m and using the fact that $m < a$, there exists $(p,q) \in (m,a)\times(b,\infty)$ such that $(p,q) \in Supp\,\beta(y)\kappa(x,y)$. But we can choose φ positive such that $\varphi(p)\mathcal{U}(q) > 0$ and this is a contradiction with (17). So we have $m \geq a$.
To conclude we notice that on $[0,m]$, $\mathcal{U}$ satisfies

$$\partial_x(\tau(x)\mathcal{U}(x)) + \lambda\mathcal{U}(x) = 0.$$

So, thanks to the condition $\tau(0)\mathcal{U}(0) = 0$ and the assumption (9), we have $\mathcal{U} \equiv 0$ on $[0,m]$, so $m = a$ and the first statement is proved.

For ϕ, we define $G(x) := \phi(x)e^{-\int_{x_0}^x \frac{\lambda+\beta(s)}{\tau(s)}ds}$. We have that

$$G'(x) = -2e^{-\int_{x_0}^x \frac{\lambda+\beta(s)}{\tau(s)}ds}\beta(x)\int_0^x \kappa(y,x)\phi(y)\,dy \leq 0, \tag{18}$$

so, as soon as ϕ vanishes, it remains null. Therefore ϕ is positive on an interval $(0,x_1)$ with $x_1 \in \mathbb{R}_+^* \cup \{+\infty\}$. Assuming that $x_1 < +\infty$ and using that $x_1 > a = m$ because $\int \phi(x)\mathcal{U}(x)dx = 1$, we can find $X \geq x_1$ such that

$$\int_{x_1}^X G'(x)\,dx = -2\int_{x_1}^X\int_0^{x_1} e^{\int_{x_0}^x \frac{\lambda+\beta(s)}{\tau(s)}ds}\phi(y)\beta(x)\kappa(y,x)\,dy\,dx < 0.$$

This contradicts that $\phi(x) = 0$ for $x \geq x_1$, and we have proved that $\phi(x) > 0$ for $x > 0$.
If $\frac{1}{\tau} \in L_0^1$, we can take $x_0 = 0$ in the definition of G and so $\phi(0) > 0$ or $\phi \equiv 0$. The fact that ϕ is positive ends the proof of the lemma.

□

3.2 Truncated problem

The proof of the theorem is based on uniform estimates on the solution to a truncated equation. Let $\eta,\ \delta,\ R$ positive numbers and define

$$\tau_\eta(x) = \begin{cases} \eta & 0 \leq x \leq \eta \\ \tau(x) & x \geq \eta. \end{cases}$$

Then τ_η is lower bounded on $[0,R]$ thanks to (9) and we denote by $\mu = \mu(\eta,R) := \inf_{[0,R]} \tau_\eta$. The existence of eigenelements $(\lambda_\eta^\delta, \mathcal{U}_\eta^\delta, \phi_\eta^\delta)$ for the following truncated problem when $\delta R < \mu$ is standard (see Theorem 3 in the Appendix).

$$\left\{\begin{array}{l} \dfrac{\partial}{\partial x}(\tau_\eta(x)\mathcal{U}_\eta^\delta(x)) + (\beta(x)+\lambda_\eta^\delta)\mathcal{U}_\eta^\delta(x) = 2\displaystyle\int_x^R \beta(y)\kappa(x,y)\mathcal{U}_\eta^\delta(y)\,dy, \qquad 0<x<R, \\ \\ \tau_\eta\mathcal{U}_\eta^\delta(x=0) = \delta, \qquad \mathcal{U}_\eta^\delta(x) > 0, \qquad \int \mathcal{U}_\eta^\delta(x)dx = 1, \\ \\ -\tau_\eta(x)\dfrac{\partial}{\partial x}\phi_\eta^\delta(x) + (\beta(x)+\lambda_\eta^\delta)\phi_\eta^\delta(x) - 2\beta(x)\displaystyle\int_0^x \kappa(y,x)\phi_\eta^\delta(y)\,dy = \delta\phi_\eta^\delta(0), \qquad 0<x<R, \\ \\ \phi_\eta^\delta(R) = 0, \qquad \phi_\eta^\delta(x) > 0, \qquad \int \phi_\eta^\delta(x)\mathcal{U}_\eta^\delta(x)dx = 1. \end{array}\right. \tag{19}$$

The proof of the theorem 1 requires $\lambda_\eta^\delta > 0$. To enforce it, we take $\delta R = \frac{\mu}{2}$ and we consider R large enough to satisfy the following lemma.

Lemma 2 *Under assumptions* (5), (8) *and* (13), *there exists a* $R_0 > 0$ *such that for all* $R > R_0$, *if we choose* $\delta = \frac{\mu}{2R}$, *then we have* $\lambda_\eta^\delta > 0$.

Proof. Assume by contradiction that $R > 0$ and $\lambda_\eta^\delta \le 0$ with $\delta = \frac{\mu}{2R}$. Then, integrating between 0 and $x > 0$, we obtain

$$\begin{aligned} 0 &\ge \lambda\int_0^x \mathcal{U}(y)\,dy \\ &= \delta - \tau(x)\mathcal{U}(x) - \int_0^x \beta(y)\mathcal{U}(y)\,dy + 2\int_0^x\int_z^R \beta(y)\kappa(z,y)\mathcal{U}(y)\,dy\,dz \\ &= \delta - \tau(x)\mathcal{U}(x) + \int_0^x \beta(y)\mathcal{U}(y)\,dy + 2\int_x^R\Big(\int_0^x \kappa(z,y)\,dz\Big)\beta(y)\mathcal{U}(y)\,dy \\ &\ge \delta - \tau(x)\mathcal{U}(x) + \int_0^x \beta(y)\mathcal{U}(y)\,dy. \end{aligned}$$

Consequently

$$\tau(x)\mathcal{U}(x) \ge \delta + \int_0^x \frac{\beta(y)}{\tau(y)}\tau(y)\mathcal{U}(y)\,dy$$

and, thanks to Grönwall's lemma,

$$\tau(x)\mathcal{U}(x) \ge \delta e^{\int_0^x \frac{\beta(y)}{\tau(y)}dy}.$$

But assumption (13) ensures that for all $n \ge 0$, there is a $A > 0$ such that

$$\frac{\beta(x)}{\tau(x)} \ge \frac{n}{x}, \qquad \forall x \ge A$$

and thus we have

$$\tau(x)\mathcal{U}(x) \ge \delta x^n, \quad \forall x \ge A.$$

Thanks to (8) we can choose n and A such that $x^{-n}\tau(x) \le \frac{\mu}{4}$ for $x \ge A$ and then we have

$$1 = \int_0^R \mathcal{U}(x)\,dx \ge \int_A^R \mathcal{U}(x)\,dx \ge \delta \int_A^R \frac{x^n}{\tau(x)}\,dx \ge \frac{2}{R}(R-A)$$

what is a contradiction as soon as $R > 2A$; so Lemma 2 holds for $R_0 = 2A$.

□

3.3 Limit as $\delta \to 0$ for $\mathcal{U}_\eta^\delta$ and λ_η^δ

Fix η and let $\delta \to 0$ (then $R \to \infty$ since $\delta R = \frac{\mu}{2}$).

First estimate: λ_η^δ upper bound. Integrating equation (19) between 0 and R, we find

$$\lambda_\eta^\delta \le \delta + \int \beta(x)\mathcal{U}_\eta^\delta(x)\,dx,$$

then the idea is to prove a uniform estimate on $\int \beta\mathcal{U}_\eta^\delta$. For this we begin with bounding the higher moments $\int x^\alpha \beta\mathcal{U}_\eta^\delta$ for $\alpha \ge 2$.
Let $\alpha \ge 2$, according to (7) we have

$$\int \frac{x^\alpha}{y^\alpha}\kappa(x,y)\,dx \le \int \frac{x^2}{y^2}\kappa(x,y)\,dx \le c < \frac{1}{2}.$$

Multiplying the equation on $\mathcal{U}_\eta^\delta$ by x^α and then integrating on $[0,R]$, we obtain for all $A \ge \eta$

$$\begin{aligned}
\int x^\alpha\big((1-2c)\beta(x)\big)\mathcal{U}_\eta^\delta(x)\,dx &\le \alpha\int x^{\alpha-1}\tau_\eta(x)\mathcal{U}_\eta^\delta(x)\,dx \\
&= \alpha\int_{x\le A} x^{\alpha-1}\tau_\eta(x)\mathcal{U}_\eta^\delta(x)\,dx + \alpha\int_{x\ge A} x^{\alpha-1}\tau(x)\mathcal{U}_\eta^\delta(x)\,dx \\
&\le \alpha A^{\alpha-1}\sup_{(0,A)}\tau + \omega_{A,\alpha}\int x^\alpha\beta(x)\mathcal{U}_\eta^\delta(x)\,dx,
\end{aligned}$$

where $\omega_{A,\alpha}$ is a positive number chosen to have $\alpha\tau(x) \le \omega_{A,\alpha}x\beta(x),\ \forall x \ge A$. Thanks to (7) and (13), we can choose A_α large enough to have $\omega_{A_\alpha,\alpha} < 1-2c$. Thus we find

$$\forall\alpha \ge 2,\ \exists A_\alpha :\ \forall\eta,\delta > 0,\quad \int x^\alpha\beta(x)\mathcal{U}_\eta^\delta(x)\,dx \le \frac{\alpha {A_\alpha}^{\alpha-1}\sup_{(0,A)}\tau}{1-2c-\omega_{A_\alpha,\alpha}} := B_\alpha. \tag{20}$$

The next step is to prove the same estimates for $0 \le \alpha < 2$ and for this we first give a bound on $\tau_\eta\mathcal{U}_\eta^\delta$. We fix $\rho \in (0,1/2)$ and define $x_\eta > 0$ as the unique point such that $\int_0^{x_\eta}\frac{\beta(y)}{\tau_\eta(y)}dy = \rho$. It exists because β is nonnegative and locally integrable, and τ_η is positive. Thanks to assumption (12), we know that $x_\eta \underset{\eta\to 0}{\longrightarrow} x_0$ where $x_0 > 0$ satisfies $\int_0^{x_0}\frac{\beta(y)}{\tau(y)}dy = \rho$, so x_η is bounded by $0 < \underline{x} \le x_\eta \le \overline{x}$.

Then, integrating (19) between 0 and $x \le x_\eta$, we find

$$\begin{aligned}
\tau_\eta(x)\mathcal{U}_\eta^\delta(x) &\le \delta + 2\int_0^x \int \beta(y)\mathcal{U}_\eta^\delta(y)\kappa(z,y)\,dy\,dz \\
&\le \delta + 2\int \beta(y)\mathcal{U}_\eta^\delta(y)\,dy \\
&= \delta + 2\int_0^{x_\eta} \beta(y)\mathcal{U}_\eta^\delta(y)\,dy + 2\int_{x_\eta}^\infty \beta(y)\mathcal{U}_\eta^\delta(y)\,dy \\
&\le \delta + 2\sup_{(0,x_\eta)}\{\tau_\eta\mathcal{U}_\eta^\delta\}\int_0^{x_\eta}\frac{\beta(y)}{\tau_\eta(y)}\,dy + \frac{2}{x_\eta^2}\int_0^\infty y^2\beta(y)\mathcal{U}_\eta^\delta(y)\,dy \\
&\le \delta + 2\rho\sup_{(0,x_\eta)}\{\tau_\eta\mathcal{U}_\eta^\delta\} + \frac{2}{x_\eta^2}B_2.
\end{aligned}$$

Consequently, if we consider $\delta \le 1$ for instance, we obtain

$$\sup_{x\in(0,\underline{x})}\tau_\eta(x)\mathcal{U}_\eta^\delta(x) \le \frac{1+2B_2/\underline{x}^2}{1-2\rho} := C \tag{21}$$

so $\tau_\eta\mathcal{U}_\eta^\delta$ is uniformly bounded in a neighborhood of zero.
Now we can prove a bound B_α for $x^\alpha\beta\mathcal{U}_\eta^\delta$ in the case $0 \le \alpha < 2$. Thanks to the estimates (20) and (21) we have

$$\begin{aligned}
\int x^\alpha\beta(x)\mathcal{U}_\eta^\delta(x)\,dx &= \int_0^{\overline{x}} x^\alpha\beta(x)\mathcal{U}_\eta^\delta(x)\,dx + \int_{\overline{x}}^R x^\alpha\beta(x)\mathcal{U}_\eta^\delta(x)\,dx \\
&\le \overline{x}^\alpha\sup_{(0,\overline{x})}\{\tau_\eta\mathcal{U}_\eta^\delta\}\int_0^{\overline{x}}\frac{\beta(y)}{\tau_\eta(y)}\,dy + \overline{x}^{\alpha-2}\int_{\overline{x}}^R x^2\beta(x)\mathcal{U}_\eta^\delta(x)\,dx \\
&\le C\rho\overline{x}^\alpha + B_2\overline{x}^{\alpha-2} := B_\alpha.
\end{aligned} \tag{22}$$

Combining (20) and (22) we obtain

$$\forall\alpha \ge 0,\ \exists B_\alpha :\ \forall\eta,\delta>0,\quad \int x^\alpha\beta(x)\mathcal{U}_\eta^\delta(x)\,dx \le B_\alpha, \tag{23}$$

and finally we bound λ_η^δ

$$\lambda_\eta^\delta = \delta + \int \beta\mathcal{U}_\eta^\delta \le \delta + B_0. \tag{24}$$

So the family $\{\lambda_\eta^\delta\}_\delta$ belong to a compact interval and we can extract a converging subsequence $\lambda_\eta^\delta \underset{\delta\to 0}{\longrightarrow} \lambda_\eta$.

Second estimate : $W^{1,1}$ bound for $x^\alpha\tau_\eta\mathcal{U}_\eta^\delta,\ \alpha \ge 0$. We use the estimate (23). First we give a L^∞bound for $\tau_\eta\mathcal{U}_\eta^\delta$ by integrating (19) between 0 and x

$$\tau_\eta(x)\mathcal{U}_\eta^\delta(x) \le \delta + 2\int_0^R \beta(y)\mathcal{U}_\eta^\delta(y)\,dy \le \delta + 2B_0 := D_0. \tag{25}$$

Then we bound $x^\alpha \tau_\eta \mathcal{U}_\eta^\delta$ in L^1 for $\alpha > -1$. Assumption (13) ensures that there exists $X > 0$ such that $\tau(x) \le x\beta(x),\ \forall x \ge X$, so we have for $R > X$

$$\begin{aligned}\int x^\alpha \tau_\eta(x)\mathcal{U}_\eta^\delta(x)\,dx &\le \sup_{(0,X)}\{\tau_\eta \mathcal{U}_\eta^\delta\}\int_0^X x^\alpha\,dx + \int_X^R x^{\alpha+1}\beta(x)\mathcal{U}_\eta^\delta(x)\,dx \\ &\le \sup_{(0,X)}\{\tau_\eta \mathcal{U}_\eta^\delta\}\frac{X^{\alpha+1}}{\alpha+1} + B_{\alpha+1} := C_\alpha.\end{aligned}$$

Finally

$$\forall \alpha > -1,\ \exists C_\alpha :\ \forall \eta, \delta > 0, \quad \int x^\alpha \tau_\eta(x)\mathcal{U}_\eta^\delta(x)\,dx \le C_\alpha \tag{26}$$

and we also have that $x^\alpha \mathcal{U}_\eta^\delta$ is bounded in L^1 because $\tau \in \mathcal{P}$ (see assumption (8)).
A consequence of (23) and (26) is that $x^\alpha \tau_\eta \mathcal{U}_\eta^\delta$ is bound in L^∞ for all $\alpha \ge 0$. We already have (25) and for $\alpha > 0$, we multiply (19) by x^α, integrate on $[0, x]$ and obtain

$$x^\alpha \tau_\eta(x)\mathcal{U}_\eta^\delta(x) \le \alpha\int_0^R y^{\alpha-1}\tau_\eta(y)\mathcal{U}_\eta^\delta(y)\,dy + 2\int_0^R y^\alpha\beta(y)\mathcal{U}_\eta^\delta(y)\,dy \le \alpha C_\alpha + 2B_\alpha := D_\alpha,$$

that give immediately

$$\forall \alpha \ge 0,\ \exists D_\alpha :\ \forall \eta, \delta > 0, \quad \sup_{x>0} x^\alpha \tau_\eta(x)\mathcal{U}_\eta^\delta(x) \le D_\alpha. \tag{27}$$

To conclude we use the fact that neither the parameters nor $\mathcal{U}_\eta^\delta$ are negative and we find by the chain rule, for $\alpha \ge 0$

$$\begin{aligned}\int \Big|\frac{\partial}{\partial x}(x^\alpha \tau_\eta(x)\mathcal{U}_\eta^\delta(x))\Big|dx &\le \alpha\int x^{\alpha-1}\tau_\eta(x)\mathcal{U}_\eta^\delta(x)\,dx + \int x^\alpha\big|\partial_x(\tau_\eta(x)\mathcal{U}_\eta^\delta(x))\big|\,dx \\ &\le \alpha\int x^{\alpha-1}\tau_\eta(x)\mathcal{U}_\eta^\delta(x)\,dx + \lambda_\eta^\delta\int x^\alpha\mathcal{U}_\eta^\delta(x)\,dx + 3\int x^\alpha\beta(x)\mathcal{U}_\eta^\delta(x)\,dx\end{aligned} \tag{28}$$

and all the terms in the right hand side are uniformly bounded thanks to the previous estimates.

Since we have proved that the family $\{x^\alpha \tau_\eta \mathcal{U}_\eta^\delta\}_\delta$ is bounded in $W^{1,1}(\mathbb{R}^+)$ for all $\alpha \ge 0$, then, because τ_η is positive and belongs to $\mathcal{P}$, we can extract from $\{\mathcal{U}_\eta^\delta\}_\delta$ a subsequence which converges in $L^1(\mathbb{R}^+)$ when $\delta \to 0$. Passing to the limit in equation (19) we find that

$$\begin{cases} \dfrac{\partial}{\partial x}(\tau_\eta(x)\mathcal{U}_\eta(x)) + (\beta(x) + \lambda_\eta)\mathcal{U}_\eta(x) = 2\displaystyle\int_x^\infty \beta(y)\kappa(x,y)\mathcal{U}_\eta(y)\,dy, \\ \mathcal{U}_\eta(0) = 0, \quad \mathcal{U}_\eta(x) \ge 0, \quad \int \mathcal{U}_\eta = 1, \end{cases} \tag{29}$$

with $\lambda_\eta \ge 0$.

3.4 Limit as $\eta \to 0$ for $\mathcal{U}_\eta$ and λ_η

All the estimates (20)-(28) remain true for $\delta = 0$. So we still know that the family $\{x^\alpha \tau_\eta \mathcal{U}_\eta\}_\eta$ belongs to a compact set of L^1, but not necessarily $\{\mathcal{U}_\eta\}_\eta$ because in the limit τ can vanish at zero. We need one more estimate to study the limit $\eta \to 0$.

Third estimate: L^∞ bound for $x^\alpha \tau_\eta \mathcal{U}_\eta$, $\alpha \geq -\gamma$. We already know that $x^\alpha \tau_\eta \mathcal{U}_\eta$ is bounded for $\alpha \geq 0$. So, to prove the bound, it only remains to prove that $x^{-\gamma}\tau_\eta \mathcal{U}_\eta$ is bounded in a neighborhood of zero. Let define $f_\eta : x \mapsto \sup_{(0,x)} \tau_\eta \mathcal{U}_\eta$. If we integrate (29) between 0 and $x' < x$, we find

$$\tau_\eta(x')\mathcal{U}_\eta(x') \leq 2\int_0^{x'} \int \beta(y)\mathcal{U}_\eta(y)\kappa(z,y)\,dy\,dz \leq 2\int_0^{x} \int \beta(y)\mathcal{U}_\eta(y)\kappa(z,y)\,dy\,dz$$

and so for all x

$$f_\eta(x) \leq 2\int_0^{x} \int \beta(y)\mathcal{U}_\eta(y)\kappa(z,y)\,dy\,dz.$$

We consider x_η and $\underline{x}$ defined in the first estimate and, using (11) and (12), we have for all $x < x_\eta$

$$\begin{aligned}
f_\eta(x) &\leq 2\int_0^{x} \int \beta(y)\mathcal{U}_\eta(y)\kappa(z,y)\,dy\,dz \\
&= 2\int \beta(y)\mathcal{U}_\eta(y)\int_0^{x} \kappa(z,y)\,dz\,dy \\
&\leq 2\int_0^{\infty} \beta(y)\mathcal{U}_\eta(y)\min\Big(1, C\Big(\frac{x}{y}\Big)^\gamma\Big)\,dy \\
&= 2\int_0^{x} \beta(y)\mathcal{U}_\eta(y)\,dy + 2C\int_x^{x_\eta} \beta(y)\mathcal{U}_\eta(y)\Big(\frac{x}{y}\Big)^\gamma\,dy + 2C\int_{x_\eta}^{\infty} \beta(y)\mathcal{U}_\eta(y)\Big(\frac{x}{y}\Big)^\gamma\,dy \\
&= 2\int_0^{x} \frac{\beta(y)}{\tau_\eta(y)}\tau_\eta(y)\mathcal{U}_\eta(y)\,dy + 2Cx^\gamma\int_x^{x_\eta} \frac{\beta(y)}{\tau_\eta(y)}\frac{\tau_\eta(y)\mathcal{U}_\eta(y)}{y^\gamma}\,dy + 2C\int_{x_\eta}^{\infty} \beta(y)\mathcal{U}_\eta(y)\Big(\frac{x}{y}\Big)^\gamma\,dy \\
&\leq 2f_\eta(x)\int_0^{x_\eta} \frac{\beta(y)}{\tau_\eta(y)}\,dy + 2Cx^\gamma\int_x^{x_\eta} \frac{\beta(y)}{\tau_\eta(y)}\frac{f_\eta(y)}{y^\gamma}\,dy + 2C\|\beta\mathcal{U}_\eta\|_{L^1}x^\gamma.
\end{aligned}$$

We set $\mathcal{V}_\eta(x) = x^{-\gamma}f_\eta(x)$ and we obtain

$$(1-2\rho)\mathcal{V}_\eta(x) \leq K + 2C\int_x^{x_\eta} \frac{\beta(y)}{\tau_\eta(y)}\mathcal{V}_\alpha(y)\,dy.$$

Hence, using Grönwall's lemma, we find that $\mathcal{V}_\eta(x) \leq \dfrac{Ke^{\frac{2C\rho}{1-2\rho}}}{1-2\rho}$ and consequently

$$x^{-\gamma}\tau_\eta(x)\mathcal{U}_\eta(x) \leq \frac{Ke^{\frac{2C\rho}{1-2\rho}}}{1-2\rho} := \widetilde{C}, \quad \forall x \in [0,\underline{x}]. \tag{30}$$

This last estimate allows us to bound $\mathcal{U}_\eta$ by $\frac{x^\gamma}{\tau}$ which is in L^1_0 by the assumption (11). Thanks to the second estimate, we also have that $\int x^\alpha \mathcal{U}_\eta$ is bounded in L^1 and so, thanks to the Dunford-Pettis theorem (see [8] for instance), $\{\mathcal{U}_\eta\}_\eta$ belong to a L^1-weak compact set. Thus we can extract a subsequence which converges L^1-weak toward $\mathcal{U}$. But for all $\varepsilon > 0$, $\{x^\alpha \mathcal{U}_\eta\}_\eta$ is bounded in $W^{1,1}([\varepsilon,\infty))$ for all $\alpha \geq 1$ thanks to (28) and so the convergence is strong on $[\varepsilon,\infty)$. Then we write

$$\begin{aligned}
\int |\mathcal{U}_\eta - \mathcal{U}| &= \int_0^{\varepsilon} |\mathcal{U}_\eta - \mathcal{U}| + \int_\varepsilon^{\infty} |\mathcal{U}_\eta - \mathcal{U}| \\
&\leq 2\widetilde{C}\int_0^{\varepsilon} \frac{x^\gamma}{\tau(x)} + \int_\varepsilon^{\infty} |\mathcal{U}_\eta - \mathcal{U}|.
\end{aligned}$$

The first term on the right hand side is small for ε small because $\frac{x^\gamma}{\tau} \in L^1_0$ and then the second term is small for η small because of the strong convergence. Finally $\mathcal{U}_\eta \underset{\eta\to 0}{\longrightarrow} \mathcal{U}$ strongly in $L^1(\mathbb{R}^+)$ and $\mathcal{U}$ solution of the eigenproblem (3).

3.5 Limit as $\delta, \eta \to 0$ for ϕ^δ_η

We prove uniform estimates on ϕ^δ_η which are enough to pass to the limit and prove the result.

Fourth estimate : uniform ϕ^δ_η-bound on $[0, A]$. Let $A > 0$, our first goal is to prove the existence of a constant $C_0(A)$ such that

$$\forall \eta, \delta, \qquad \sup_{(0,A)} \phi^\delta_\eta \le C_0(A).$$

We divide the equation on ϕ^δ_η by τ_η and we integrate between x and x_η with $0 < x < x_\eta$, where x_η, bounded by $\underline{x}$ and $\overline{x}$, is defined in the first estimate. Considering $\delta < \frac{\mu(1-2\rho)}{\overline{x}}$ (fulfilled for $R > \frac{\overline{x}}{2(1-2\rho)}$ since $\delta = \frac{\mu}{2R}$), we find

$$\begin{aligned}
\phi^\delta_\eta(x) &\le \phi^\delta_\eta(x_\eta) + 2\int_x^{x_\eta} \frac{\beta(y)}{\tau_\eta(y)} \int_0^y \kappa(z,y)\phi^\delta_\eta(z)\,dz + x_\eta \frac{\delta}{\mu}\phi^\delta_\eta(0) \\
&\le \phi^\delta_\eta(x_\eta) + \sup_{(0,x_\eta)} \{\phi^\delta_\eta\}\Big(2\int_0^{x_\eta} \frac{\beta(y)}{\tau_\eta(y)} \int_0^y \kappa(z,y)\,dz + x_\eta \frac{\delta}{\mu}\Big)
\end{aligned}$$

and we obtain

$$\sup_{x\in(0,\underline{x})} \phi^\delta_\eta(x) \le \frac{1}{1-2\rho-\delta\overline{x}/\mu}\phi^\delta_\eta(x_\eta).$$

Using the decay of $\phi^\delta_\eta(x) e^{-\int_{\underline{x}}^x \frac{\beta+\lambda^\delta_\eta}{\tau_\eta}}$, there exists $C(A)$ such that

$$\sup_{x\in(0,A)} \phi^\delta_\eta(x) \le C(A)\phi^\delta_\eta(x_\eta).$$

Noticing that $\int \phi^\delta_\eta(x)\mathcal{U}^\delta_\eta(x)dx = 1$, we conclude

$$1 \ge \int_0^{x_\eta} \phi^\delta_\eta(x)\mathcal{U}^\delta_\eta(x)dx \ge \phi^\delta_\eta(x_\eta)\int_0^{x_\eta} e^{-\int_x^{x_\eta}\frac{\beta+\lambda^\delta_\eta}{\tau_\eta}}\mathcal{U}^\delta_\eta(x)\,dx,$$

so, as $x_\eta \to x_0$ and $\int_0^{x_0}\mathcal{U}(x)dx > 0$ (thanks to Lemma 1 and because $x_0 > b \ge a$), we have

$$\sup_{(0,A)} \phi^\delta_\eta \le C_0(A). \tag{31}$$

Fifth estimate : uniform ϕ_η^δ-bound on $[A,\infty)$. Following an idea introduced in [29] we notice that the equation in (19) satisfied by ϕ_η^δ is a transport equation and therefore satisfies the maximum principle (see Lemma 4 in the Appendix). Therefore it remains to build a supersolution $\overline{\phi}$ that is positive at $x=R$, to conclude $\phi_\eta^\delta(x)\le\overline{\phi}(x)$ on $[0,R]$.

This we cannot do on $[0,R]$, but on a subinterval $[A_0,R]$ only. So we begin with an auxiliary function $\overline{\varphi}(x)=x^k+\theta$ with k and θ positive numbers to be determined. We have to check that on $[A_0,R]$

$$-\tau(x)\frac{\partial}{\partial x}\overline{\varphi}(x)+(\lambda_\eta^\delta+\beta(x))\overline{\varphi}(x)\ge 2\beta(x)\int\kappa(y,x)\overline{\varphi}(y)\,dy+\delta\phi_\eta^\delta(0),$$

i.e.

$$-k\tau(x)x^{k-1}+(\lambda_\eta^\delta+\beta(x))\overline{\varphi}(x)\ge\Big(2\theta+2\int\kappa(y,x)y^k\,dy\Big)\beta(x)+\delta\phi_\eta^\delta(0).$$

For $k\ge 2$, we know that $\int\kappa(y,x)\frac{y^k}{x^k}\,dy\le c<1/2$ so it is sufficient to prove that there exists $A_0>0$ such that we have

$$-k\tau(x)x^{k-1}+(\lambda_\eta^\delta+\beta(x))(x^k+\theta)\ge(2\theta+2cx^k)\beta(x)+\delta C_0(1)\tag{32}$$

for all $x>A_0$, where C_0 is defined in (31). For this, dividing (32) by $x^{k-1}\tau(x)$, we say that if we have

$$(1-2c)\frac{x\beta(x)}{\tau(x)}\ge k+\frac{2\theta\beta(x)+\delta C_0(1)}{x^{k-1}\tau(x)},\tag{33}$$

then (32) holds true. Thanks to assumptions (8) and (13) we know that there exists $k>0$ such that for any $\theta>0$, there exists $A_0>0$ for which (33) is true on $[A_0,+\infty)$.

Then we conclude by choosing the supersolution $\overline{\phi}(x)=\frac{C_0(A_0)}{\theta}\overline{\varphi}(x)$ so that

$$\overline{\phi}(x)\ge\phi_\eta^\delta(x)\quad\text{on }[0,A_0],$$

and on $[A_0,R]$, we have

$$\begin{cases}-\tau(x)\frac{\partial}{\partial x}\overline{\phi}(x)+(\lambda_\eta^\delta+\beta(x))\overline{\phi}(x)\ge 2\beta(x)\int_0^x\kappa(y,x)\overline{\phi}(y)\,dy+\delta\phi_\eta^\delta(0),\\ \overline{\phi}(R)>0,\end{cases}\tag{34}$$

which is a supersolution to the equation satisfied by ϕ_η^δ. Therefore $\phi_\eta^\delta\le\overline{\phi}$ uniformly in η and δ and we get

$$\exists k,\theta,C\ s.t.\ \forall\eta,\delta,\quad\phi_\eta^\delta(x)\le(Cx^k+\theta).\tag{35}$$

Equation (19) and the fact that ϕ_η^δ is uniformly bounded in $L^\infty_{loc}(\mathbb{R}^+)$ give immediately that $\partial_x\phi_\eta^\delta$ is uniformly bounded in $L^\infty_{loc}(\mathbb{R}^+,\tau(x)dx)$, so in $L^\infty_{loc}(0,\infty)$ thanks to (9).

Then we can extract a subsequence of $\{\phi_\eta^\delta\}$ which converges $\mathcal{C}^0(0,\infty)$ toward ϕ. Now we check that ϕ satisfied the adjoint equation of (3). We consider the terms of (19) one after another.
First $(\lambda_\eta^\delta+\beta(x))\phi_\eta^\delta(x)$ converges to $(\lambda+\beta(x))\phi(x)$ in L^∞_{loc}.
For $\partial_x\phi_\eta^\delta$, we have an L^∞ bound on each compact of $(0,\infty)$. So it converges $L^\infty-*weak$ toward $\partial_x\phi$.
It remains the last term which we write, for all $x>0$,

$$\int_0^x\kappa(y,x)(\phi_\eta^\delta(y)-\phi(y))\,dy\le\|\phi_\eta^\delta-\phi\|_{L^\infty(0,x)}\underset{\eta,\delta\to 0}{\longrightarrow}0.$$

The fact that $\int \phi\mathcal{U} = 1$ comes from the convergence $L^\infty - L^1$ when written as

$$1 = \int \phi_\eta^\delta(x)\mathcal{U}_\eta^\delta(x)\,dx = \int \frac{\phi_\eta^\delta(x)}{1+x^k}(1+x^k)\mathcal{U}_\eta^\delta(x)\,dx \longrightarrow \int \frac{\phi(x)}{1+x^k}(1+x^k)\mathcal{U}(x)\,dx = \int \phi\mathcal{U}.$$

At this stage we have found $(\lambda, \mathcal{U}, \phi) \in \mathbb{R}^+ \times L^1(\mathbb{R}^+) \times \mathcal{C}(\mathbb{R}^+)$ solution of (3). The estimates announced in Theorem 1 also follow from those uniform estimates. It remains to prove that $\lambda > 0$ and the uniqueness.

3.6 Proof of $\lambda > 0$

We prove a little bit more, namely that

$$\lambda \geq \frac{1}{2}\sup_{x\geq 0}\{\tau(x)\mathcal{U}(x)\}. \tag{36}$$

We integrate the first equation of (3) between 0 and x and find

$$\begin{aligned}
0 \leq \lambda\int_0^x \mathcal{U}(y)\,dy &= -\tau(x)\mathcal{U}(x) - \int_0^x \beta(y)\mathcal{U}(y)\,dy + 2\int_0^x\int_z^\infty \beta(y)\kappa(z,y)\mathcal{U}(y)\,dy\,dz \\
&\leq -\tau(x)\mathcal{U}(x) + 2\int_0^\infty\int_z^\infty \beta(y)\kappa(z,y)\mathcal{U}(y)\,dy\,dz \\
&= -\tau(x)\mathcal{U}(x) + 2\int_0^\infty \beta(y)\mathcal{U}(y)\,dy \\
&= -\tau(x)\mathcal{U}(x) + 2\lambda,
\end{aligned}$$

Hence $2\lambda \geq \tau(x)\mathcal{U}(x)$ and (36) is proved.

3.7 Uniqueness

We follow the idea of [23]. Let $(\lambda_1, \mathcal{U}_1, \phi_1)$ and $(\lambda_2, \mathcal{U}_2, \phi_2)$ two solutions to the eigenproblem (3). First we have

$$\begin{aligned}
\lambda_1\int \mathcal{U}_1(x)\phi_2(x)\,dx &= \int\Big(-\partial_x(\tau(x)\mathcal{U}_1(x)) - \beta(x)\mathcal{U}_1(x) + 2\int_x^\infty \beta(y)\kappa(x,y)\mathcal{U}_1(y)\,dy\Big)\phi_2(x)\,dx \\
&= \int\Big(\tau(x)\partial_x\phi_2(x) - \beta(x)\phi_2(x) + 2\beta(x)\int_0^x \kappa(y,x)\phi_2(y)\,dy\Big)\mathcal{U}_1(x)\,dx \\
&= \lambda_2\int \mathcal{U}_1(x)\phi_2(x)\,dx
\end{aligned}$$

and then $\lambda_1 = \lambda_2 = \lambda$ because $\int \mathcal{U}_1\phi_2 > 0$ thanks to Lemma 1.
For the eigenvectors we use the General Relative Entropy method introduced in [25, 26]. For $C > 0$, we test the equation on $\mathcal{U}_1$ against $\operatorname{sgn}\big(\frac{\mathcal{U}_1}{\mathcal{U}_2} - C\big)\phi_1$,

$$0 = \int\Big[\partial_x(\tau(x)\mathcal{U}_1(x)) + (\lambda+\beta(x))\mathcal{U}_1(x) - 2\int_x^\infty \beta(y)\kappa(x,y)\mathcal{U}_1(y)\,dy\Big]\operatorname{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x) - C\Big)\phi_1(x)\,dx.$$

Deriving $\Big|\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big|\tau(x)\mathcal{U}_2(x)\phi_1(x)$ we find

$$\int \partial_x(\tau(x)\mathcal{U}_1(x))\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\phi_1(x)\,dx=\int \partial_x\Big(\Big|\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big|\tau(x)\mathcal{U}_2(x)\phi_1(x)\Big)\,dx$$

$$+\int \partial_x(\tau(x)\mathcal{U}_2(x))\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\phi_1(x)\,dx-\int\Big|\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big|\partial_x(\tau(x)\mathcal{U}_2(x)\phi_1(x))\,dx$$

and then

$$\int \partial_x(\tau(x)\mathcal{U}_1(x))\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\phi_1(x)\,dx=$$

$$2\int\Big|\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big|\Big[\int_0^x\beta(x)\kappa(y,x)\mathcal{U}_2(x)\phi_1(y)\,dy-\int_x^\infty\beta(y)\kappa(x,y)\mathcal{U}_2(y)\phi_1(x)\,dy\Big]\,dx$$

$$+2\int\int_x^\infty\beta(y)\kappa(x,y)\mathcal{U}_2(y)\,dy\,\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\phi_1(x)\,dx$$

$$-\int(\lambda+\beta(x))\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\mathcal{U}_2(x)\phi_1(x)\,dx,$$

$$\int \partial_x(\tau(x)\mathcal{U}_1(x))\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\phi_1(x)\,dx=$$

$$2\int\int\beta(y)\kappa(x,y)\Big[\Big|\frac{\mathcal{U}_1}{\mathcal{U}_2}(y)-C\Big|-\Big|\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big|\Big]\mathcal{U}_2(y)\phi_1(x)\,dxdy$$

$$+2\int\int_x^\infty\beta(y)\kappa(x,y)\mathcal{U}_2(y)\,dy\,\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\phi_1(x)\,dx$$

$$-\int(\lambda+\beta(x))\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\mathcal{U}_2(x)\phi_1(x)\,dx.$$

So

$$0=2\int\int\beta(y)\kappa(x,y)\Big[\Big|\frac{\mathcal{U}_1}{\mathcal{U}_2}(y)-C\Big|-\Big|\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big|\Big]\mathcal{U}_2(y)\phi_1(x)\,dxdy$$

$$+2\int\int_x^\infty\beta(y)\kappa(x,y)\mathcal{U}_2(y)\,dy\,\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\phi_1(x)\,dx$$

$$-2\int\int_x^\infty\beta(y)\kappa(x,y)\mathcal{U}_1(y)\,dy\,\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\phi_1(x)\,dx$$

$$0=\int\int\beta(y)\kappa(x,y)\mathcal{U}_2(y)\Big|\frac{\mathcal{U}_1}{\mathcal{U}_2}(y)-C\Big|\Big[1-\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(y)-C\Big)\Big]\phi_1(x)\,dxdy.$$

Hence $\Big[1-\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(x)-C\Big)\mathrm{sgn}\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(y)-C\Big)\Big] = 0$ on the support of $\kappa(x,y)$ for all C thus $\frac{\mathcal{U}_1}{\mathcal{U}_2}(x) = \frac{\mathcal{U}_1}{\mathcal{U}_2}(y)$ on the support of $\kappa(x,y)$ and

$$\partial_x \frac{\mathcal{U}_1}{\mathcal{U}_2}(x) = \int \beta(y)\kappa(x,y)\Big(\frac{\mathcal{U}_1}{\mathcal{U}_2}(y) - \frac{\mathcal{U}_1}{\mathcal{U}_2}(x)\Big)\frac{\mathcal{U}_2(y)}{\mathcal{U}_2(x)}\,dy = 0 \tag{37}$$

so $\dfrac{\mathcal{U}_1}{\mathcal{U}_2} \equiv cst = 1$.

We can prove in the same way that $\phi_1 = \phi_2$ even if we can have $\mathcal{U} \equiv 0$ on $[0,m]$ with $m > 0$. Indeed in this case we know that $\beta \equiv 0$ on $[0,m]$ and so

$$\phi_i(x) = \phi_i(0)e^{\int_0^x \frac{\lambda}{\tau(s)}ds} \quad \forall x \in [0,m],\ i \in \{1,2\}.$$

4 Consequences, Perspectives

Existence of eigenelements is the basic stone for asymptotic studies when $t \to \infty$. The so-called "General Relative Entropy" method, introduced in [28, 25] and widely used (see [10, 9, 5] for instance) takes advantage of the eigenvalue problem and its adjoint to build entropy functionals. We recall for instance the following theorem (see [25, 26, 29]) for the linear case of Equation (1).

Theorem 2 *Let u_0 a function defined on $\mathbb{R}^+$ satisfying*

$$\exists C_0 > 0,\ s.t.\ \forall x \geq 0\ |u_0(x)| \leq C_0\mathcal{U}(x).$$

Then there exists a unique solution $u(x,t)$ to the equation (1) and we have for all $t > 0$ the a priori bounds

$$|u(x,t)|e^{-\lambda t} \leq C_0\mathcal{U}(x),$$

$$\int u(y,t)e^{-\lambda t}\phi(y)dy = \int u_0(y)\phi(y)dy := \langle u_0, \phi\rangle,$$

and the asymptotic behaviour

$$\int_0^\infty \big|u(y,t)e^{-\lambda t} - \langle u_0,\phi\rangle\mathcal{U}(y)\big|\phi(y)\,dy \underset{t\to\infty}{\longrightarrow} 0.$$

The eigenvalue problem can also be used in nonlinear cases, such as prion proliferation equations, where there is a quadratic coupling of Equation (1) with a differential equation. In [10, 9] for instance, the stability of steady states is investigated. The use of entropy methods in the case of nonlinear problems remains however a challenging and widely open field (see [30] for a recent review).

A following work is to study the dependency of the eigenvalue λ on parameters τ and β (see [24]). For instance, our assumptions allow τ to vanish at zero, what is a necessary condition to ensure that λ tends to zero when the fragmentation tends to infinity. Such results give precious information on the qualitative behaviour of the solution.

Another possible extension of the present work is to prove existence of eigenelements in the case of time-periodic parameters, using the Floquet's theory, and then compare the new λ_F with the time-independent one λ (see [11]). Such studies can help to choose a right strategy in order to optimize, for instance, the total mass $\int xu(t,x)dx$ in the case of prion proliferation (see [10]) or on the contrary minimize the total population $\int u(t,x)dx$ in the case of cancer therapy (see [12, 11]).

Finally, this eigenvalue problem could be used to recover some of the equation parameters like τ and β from the knowledge of the asymptotic profile of the solution, as introduced in [31, 15] in the case of symmetric division ($\tau = 1$ and $\kappa = \delta_{x=\frac{y}{2}}$), by the use of inverse problems techniques. The method of [31] has to be adapted to our general case, in order to model prion proliferation for instance, or yet to recover the aggregation rate τ ; this is another direction of future research.

Aknowledgment
The authors thank a lot Benoît Perthame for his precious help and his corrections.

Appendix

A Assumption on κ.

Lemma 3 *Assumptions* (5),(6) *and* (11) *with* $\gamma > 0$ *imply that*

$$\inf_y \lim_{\eta\to 0} \int_{\eta y}^{(1-\eta)y} \kappa(x,y)\,dx > 0,$$

which means that polymers undergo a decrease in the size during fragmentation process. As a consequence, assumption (7) *holds true.*

Proof. With the first assumption (5) we have

$$1 = \int_0^y \kappa(x,y)\,dx = \int_0^{\eta y} \kappa(x,y)\,dx + \int_{\eta y}^{(1-\eta)y} \kappa(x,y)\,dx + \int_{(1-\eta)y}^{y} \kappa(x,y)\,dx.$$

The two other assumptions (6) and (11) allow to control the mass of κ at the ends :

$$\begin{aligned}
\int_{\eta y}^{(1-\eta)y} \kappa(x,y)\,dx &= 1 - \int_0^{\eta y} \kappa(x,y)\,dx - \int_{(1-\eta)y}^{y} \kappa(x,y)\,dx \\
&\geq 1 - C\eta^\gamma - \frac{1}{1-\eta}\int_{(1-\eta)y}^{y} \frac{x}{y}\kappa(x,y)\,dx \\
&\geq 1 - C\eta^\gamma - \frac{1}{2(1-\eta)} \quad \underset{\eta\to 0}{\longrightarrow} \quad \frac{1}{2},
\end{aligned}$$

which gives the first assertion of the lemma.
Now we can prove (7) :

$$\begin{aligned}
\int_0^y \frac{x^2}{y^2}\kappa(x,y)\,dx &\leq \Big[\int_0^{\eta y} \frac{x}{y}\kappa(x,y)\,dx + \int_{(1-\eta)y}^{y} \frac{x}{y}\kappa(x,y)\,dx\Big] + \int_{\eta y}^{(1-\eta)y} \frac{x^2}{y^2}\kappa(x,y)\,dx \\
&\leq \Big[\frac{1}{2} - \int_{\eta y}^{(1-\eta)y} \frac{x}{y}\kappa(x,y)\,dx\Big] + (1-\eta)\int_{\eta y}^{(1-\eta)y} \frac{x}{y}\kappa(x,y)\,dx \\
&= \frac{1}{2} - \eta\int_{\eta y}^{(1-\eta)y} \frac{x}{y}\kappa(x,y)\,dx \\
&\leq \frac{1}{2} - \eta^2\int_{\eta y}^{(1-\eta)y} \kappa(x,y)\,dx.
\end{aligned}$$

We use the first part of the proof to conclude. Taking $\eta = \min\Big(\frac{1}{4}, \frac{1}{(4C)^{1/\gamma}}\Big)$ for instance, we obtain

$$\int_{\eta y}^{(1-\eta)y} \kappa(x,y)\,dx \geq \frac{1}{3},$$

and the lemma is proved for $c = \frac{1}{2} - \frac{1}{48}$. $\square$

B Krein-Rutman

We prove existence of solution for the truncated equation (19). In this part η and δ are fixed (with $\delta R < \mu$), so we will omit these indices for τ, λ, $\mathcal{U}$ and ϕ but we keep in mind that $\tau(x) \geq \mu > 0$. We use the Krein-Rutman theorem which requires working in the space of continuous functions (see [28] for instance). First we define regularized parameters as follows :

$$\tau_\varepsilon = \rho_\varepsilon * \tau, \quad \beta_\varepsilon = \rho_\varepsilon * \beta, \quad \text{and} \quad \forall y \geq 0,\ \kappa_\varepsilon(.,y) = \rho_\varepsilon * \kappa(.,y),$$

where $\rho_\varepsilon(x) = \frac{1}{\varepsilon}\rho(\frac{x}{\varepsilon})$ with $\rho \in \mathcal{C}_c^\infty((0,\infty))$, positive and such that $\int_0^\infty \rho = 1$. Then we have the theorem

Theorem 3 *Under assumptions* (5)-(13) *on the parameters and for all* $\varepsilon > 0$, *there is a unique solution* $\lambda_\varepsilon \in \mathbb{R}$ *and* $\mathcal{U}_\varepsilon, \phi_\varepsilon \in \mathcal{C}^1([0,R])$ *to the regularized eigenproblem*

$$\begin{cases} \dfrac{\partial}{\partial x}(\tau_\varepsilon(x)\mathcal{U}_\varepsilon(x)) + (\beta_\varepsilon(x) + \lambda_\varepsilon)\mathcal{U}_\varepsilon(x) = 2\displaystyle\int_0^R \beta_\varepsilon(y)\kappa_\varepsilon(x,y)\mathcal{U}_\varepsilon(y)\,dy, \qquad 0 < x < R, \\ \tau_\varepsilon\mathcal{U}_\varepsilon(x=0) = \delta \int_0^R \mathcal{U}_\varepsilon(y)\,dy, \qquad \mathcal{U}_\varepsilon(x) > 0, \qquad \int_0^R \mathcal{U}_\varepsilon(x)dx = 1, \\ -\tau_\varepsilon(x)\dfrac{\partial}{\partial x}\phi_\varepsilon(x) + (\beta_\varepsilon(x) + \lambda_\varepsilon)\phi_\varepsilon(x) - 2\beta_\varepsilon(x)\displaystyle\int_0^R \kappa_\varepsilon(y,x)\phi_\varepsilon(y)\,dy = \tau_\varepsilon(0)\delta\phi_\varepsilon(0), \quad 0 < x < R, \\ \phi_\varepsilon(R) = 0, \qquad \phi_\varepsilon(x) > 0, \qquad \int_0^R \phi_\varepsilon(x)\mathcal{U}_\varepsilon(x)dx = 1. \end{cases} \tag{38}$$

Proof. We follow the proof of [28]. We define linear operators on $E := \mathcal{C}^0([0,R])$ to apply the Krein-Rutman theorem.

Direct equation. For $\nu > 0$ we consider the following equation on E

$$\begin{cases} \dfrac{\partial}{\partial x}(n(x)) + \dfrac{\nu + \beta_\varepsilon(x)}{\tau_\varepsilon(x)}n(x) - 2\displaystyle\int_0^R \frac{\beta_\varepsilon(y)}{\tau_\varepsilon(y)}\kappa_\varepsilon(x,y)n(y)\,dy = \frac{f(x)}{\tau_\varepsilon(x)}, \qquad 0 \leq x \leq R, \\ n(x=0) = \delta \int_0^R \frac{n(y)}{\tau_\varepsilon(y)}dy, \end{cases} \tag{39}$$

and we prove that the linear operator $A : f \mapsto n$ (solution of (39)) satisfies to the assumptions of the Krein-Rutman theorem.

First step: construction of A. Fix $f \in E$ and for $m \in E$, we define $n = T(m) \in E$ as the (explicit) solution to

$$\begin{cases} \dfrac{\partial}{\partial x}(n(x)) + \dfrac{\nu + \beta_\varepsilon(x)}{\tau_\varepsilon(x)}n(x) = 2\displaystyle\int_0^R \frac{\beta_\varepsilon(y)}{\tau_\varepsilon(y)}\kappa_\varepsilon(x,y)m(y)\,dy + \frac{f(x)}{\tau_\varepsilon(x)}, \qquad 0 \leq x \leq R, \\ n(x=0) = \delta \int_0^R \frac{m(y)}{\tau_\varepsilon(y)}dy, \end{cases}$$

We prove that T is a strict contraction. Therefore it has a unique fixed point thanks to the Banach-Picard theorem. This fixed point is a solution to (39).

In order to prove that T is a strict contraction, we consider m_1 and m_2 two functions in E, we compute for $n = n_1 - n_2,\ m = m_1 - m_2$,

$$\begin{cases} \dfrac{\partial}{\partial x}(n(x)) + \dfrac{\nu + \beta_\varepsilon(x)}{\tau_\varepsilon(x)} n(x) = 2 \displaystyle\int_0^R \frac{\beta_\varepsilon(y)}{\tau_\varepsilon(y)} \kappa_\varepsilon(x,y) m(y)\, dy, \qquad 0 \le x \le R, \\ \\ n(x=0) = \delta \int_0^R \frac{m(y)}{\tau_\varepsilon(y)} dy, \end{cases}$$

therefore

$$\begin{cases} \dfrac{\partial}{\partial x}|n(x)| + \dfrac{\nu + \beta_\varepsilon(x)}{\tau_\varepsilon(x)} |n(x)| \le 2 \displaystyle\int_0^R \frac{\beta_\varepsilon(y)}{\tau_\varepsilon(y)} \kappa_\varepsilon(x,y) |m(y)|\, dy, \qquad 0 \le x \le R, \\ \\ n(x=0) \le \delta \int_0^R \frac{|m(y)|}{\tau_\varepsilon(y)} dy. \end{cases}$$

After integration, we obtain

$$|n(x)| e^{\int_0^x \frac{\mu+\beta_\varepsilon}{\tau_\varepsilon}} \le \delta \int_0^R \frac{|m(y)|}{\tau_\varepsilon(y)} dy + \int_0^x e^{\int_0^{x'} \frac{\nu+\beta_\varepsilon}{\tau_\varepsilon}} \int_0^R \frac{\beta_\varepsilon(y)}{\tau_\varepsilon(y)} \kappa_\varepsilon(x',y) |m(y)|\, dy dx'$$

and thus

$$\begin{aligned} |n(x)| &\le \delta \int_0^R \frac{|m(y)|}{\tau_\varepsilon(y)} dy + \int_0^x e^{-\int_{x'}^x \frac{\nu+\beta_\varepsilon}{\tau_\varepsilon}} \int_0^R \frac{\beta_\varepsilon(y)}{\tau_\varepsilon(y)} \kappa_\varepsilon(x',y) |m(y)|\, dy dx' \\ &\le \|m\|_E \frac{1}{\mu} \Big[\delta R + \int_0^x e^{-\int_{x'}^x \frac{\nu+\beta_\varepsilon}{\tau_\varepsilon}} \int_0^R \beta_\varepsilon(y) \kappa_\varepsilon(x',y)\, dy dx' \Big] \\ &\le \|m\|_E \frac{1}{\mu} \Big[\delta R + \Big\| \int_0^R \beta_\varepsilon(y) \kappa_\varepsilon(.,y) dy \Big\|_{L^\infty} \int_0^x e^{-\frac{\nu}{\|\tau_\varepsilon\|_{L^\infty}}(x-x')}\, dx' \Big] \\ &\le \|m\|_E \underbrace{\frac{1}{\mu} \Big[\delta R + \nu^{-1} \|\tau_\varepsilon\|_{L^\infty} \Big\| \int_0^R \beta_\varepsilon(y) \kappa_\varepsilon(.,y) dy \Big\|_{L^\infty} \Big]}_{:=k}. \end{aligned}$$

Because $\delta R < \mu$ by assumption, we can choose ν large so that $k < 1$ and we obtain

$$\|n\|_E \le k \|m\|_E.$$

Thus T is a strict contraction and we have proved the existence of a solution to (39).

Second step: A is continuous. This relies on a general argument which in fact shows that the linear mapping A is Lipschitz continuous. Indeed, arguing as above

$$|n(x)| e^{\int_0^x \frac{\nu+\beta_\varepsilon}{\tau_\varepsilon}} \le \delta \int_0^R \frac{|n(y)|}{\tau_\varepsilon(y)} dy + \int_0^x e^{\int_0^{x'} \frac{\nu+\beta_\varepsilon}{\tau_\varepsilon}} \int_0^R \frac{\beta_\varepsilon(y)}{\tau_\varepsilon(y)} \kappa_\varepsilon(x',y) |n(y)|\, dy dx' + \int_0^x e^{\int_0^{x'} \frac{\nu+\beta_\varepsilon}{\tau_\varepsilon}} \frac{|f(x')|}{\tau_\varepsilon(x')}\, dx',$$

and thus

$$|n(x)| \le k \|n\|_E + \int_0^R \frac{|f(x')|}{\tau_\varepsilon(x')}\, dx' \le k \|n\|_E + \frac{R}{\mu} \|f\|_E.$$

This indeed proves that

$$\|n\|_E \le \frac{R}{\mu(1-k)} \|f\|_E.$$

Third step: A is strongly positive. For $f \geq 0$, the operator T of the first step maps $m \geq 0$ to $n \geq 0$. Therefore the fixed point n is nonnegative. In other words $n = A(f) \geq 0$. If additionally f does not vanish, then n does not vanish either. Therefore $n(0) = \delta \int_0^R \frac{n(y)}{\tau_\varepsilon(y)} dy > 0$ and thus

$$n(x) \geq n(0) + e^{-\int_0^x \frac{\nu+\beta_\varepsilon}{\tau_\varepsilon}} \int_0^x e^{-\int_0^{x'} \frac{\nu+\beta_\varepsilon}{\tau_\varepsilon}} \frac{f(x')}{\tau_\varepsilon(x')}\, dx' > 0.$$

Fourth step: A is compact. For $\|f\|_E \leq 1$, the third step proves that n is bounded in E and thus

$$\frac{\partial}{\partial_x} n = -\frac{\nu + \beta_\varepsilon}{\tau_\varepsilon} n + \int \frac{\beta_\varepsilon(y)}{\tau_\varepsilon(y)} \kappa(x,y) n(y) dy + \frac{f}{\tau_\varepsilon}$$

is also bounded in E. Therefore by the Ascoli-Arzela theorem the family n is relatively compact in E.

Adjoint equation. A function ϕ is a solution to the adjoint equation of (38) if and only if $\tilde{\phi}(x) := \phi(R - x)$ satisfies

$$\begin{cases} \tilde{\tau}_\varepsilon(x) \dfrac{\partial}{\partial x} \tilde{\phi}(x) + (\tilde{\beta}_\varepsilon(x) + \lambda_\varepsilon) \tilde{\phi}(x) - 2\tilde{\beta}_\varepsilon(x) \displaystyle\int_0^R \kappa_\varepsilon(y, R-x) \tilde{\phi}_\varepsilon(y)\, dy = \delta \tilde{\phi}_\varepsilon(R), \quad 0 < x < R, \\ \\ \tilde{\phi}_\varepsilon(0) = 0, \end{cases} \tag{40}$$

where $\tilde{\tau}_\varepsilon(x) = \tau_\varepsilon(R - x)$ and $\tilde{\beta}_\varepsilon(x) = \beta_\varepsilon(R - x)$. Then the same method than for the direct equation give the result, namely the existence of λ and $\tilde{\phi}$ solution to (40).

Finally we have proved existence of $(\lambda_\mathcal{U}, \mathcal{U}_\varepsilon)$ and $(\lambda_\phi, \phi_\varepsilon)$ solution to the direct and adjoint equations of (38). It remains to prove that $\lambda_\mathcal{U} = \lambda_\phi$ but it is nothing but integrating the direct equation against the ajoint eigenvector, what gives

$$\lambda_\mathcal{U} \int \mathcal{U}_\varepsilon \phi_\varepsilon = \lambda_\phi \int \mathcal{U}_\varepsilon \phi_\varepsilon.$$

□

To have existence of solution for (19), it remains to do $\varepsilon \to 0$. For this we can prove uniform bounds in L^∞ for $\mathcal{U}_\varepsilon$ and ϕ_ε because we are on the fixed compact $[0, R]$. Then we can extract subsequences which converge $L^\infty *-$weak toward $\mathcal{U}$ and ϕ, solutions to (19) because τ_ε and β_ε converge in L^1 toward τ and β. Concerning κ_ε, we have that for all $\varphi \in \mathcal{C}_c^\infty$, $\int \varphi(x) \kappa_\varepsilon(x,y) dx \longrightarrow \int \varphi(x) \kappa(x,y) dx \quad \forall y$, and it is sufficient to pass to the limit in the equations.

C Maximum principle

Lemma 4 *If there exists $A_0 > 0$ such that $\overline{\phi} \geq \phi$ on $[0, A_0]$ and $\overline{\phi}$ a supersolution of (3) on $[A_0, R]$ with $\overline{\phi}(R) \geq \phi(R)$, then $\overline{\phi} \geq \phi$ on $[0, R]$.*

Proof. The proof is based on the same tools than to prove uniqueness (see above) or to establish GRE principles (see [25, 26] for instance).
We know that $\overline{\phi} \geq \phi$ on $[0, A_0]$ and that $\overline{\phi}$ is a supersolution to the equation satisfied by ϕ on $[A_0, R]$, *i.e.* there exists a function $f \geq \delta\phi(0)$ such that

$$-\tau(x)\frac{\partial}{\partial x}\overline{\phi}(x) + (\lambda + \beta(x))\overline{\phi}(x) = 2\beta(x)\int_0^x \kappa(y,x)\overline{\phi}(y)\,dy + f(x), \quad \forall x \in [A_0, R].$$

So we have for all $x \in [A_0, R]$

$$-\tau(x)\frac{\partial}{\partial x}(\phi(x) - \overline{\phi}(x)) + (\lambda + \beta(x))(\phi(x) - \overline{\phi}(x)) = 2\beta(x)\int_0^x \kappa(y,x)(\phi(y) - \overline{\phi}(y))\,dy - f(x).$$

Then, multiplying by $\mathbb{1}_{\phi \geq \overline{\phi}}$, we obtain (see [28] for a justification)

$$-\tau(x)\frac{\partial}{\partial x}(\phi - \overline{\phi})_+(x) + (\lambda + \beta(x))(\phi - \overline{\phi})_+(x) \leq 2\beta(x)\int_0^x \kappa(y,x)(\phi - \overline{\phi})_+(y)\,dy - f(x)\mathbb{1}_{\phi \geq \overline{\phi}}(x),$$

and this inequality is satisfied on $[0, R]$ since $(\phi - \overline{\phi})_+ \equiv 0$ on $[0, A_0]$.
If we test against $\mathcal{U}$ we have, using the fact that $\phi(R) = 0 < \overline{\phi}(R)$,

$$\begin{aligned}\int_0^R (\phi - \overline{\phi})_+(x)\frac{\partial}{\partial x}(\tau(x)\mathcal{U}(x))\,dx + \int_0^R (\lambda + \beta(x))(\phi - \overline{\phi})_+(x)\mathcal{U}(x)\,dx \\ \leq \quad 2\int_0^R (\phi - \overline{\phi})_+(y)\int_y^R \beta(x)\kappa(y,x)\mathcal{U}(x)\,dxdy - \int_0^R f(x)\mathbb{1}_{\phi \geq \overline{\phi}}(x)\mathcal{U}(x)\,dx.\end{aligned}$$

But if we test the equation (3) satisfied by $\mathcal{U}$ against $(\phi - \overline{\phi})_+$, we find

$$\begin{aligned}\int_0^R (\phi - \overline{\phi})_+(x)\frac{\partial}{\partial x}(\tau(x)\mathcal{U}(x))\,dx + \int_0^R (\lambda + \beta(x))(\phi - \overline{\phi})_+(x)\mathcal{U}(x)\,dx \\ = \quad 2\int_0^R (\phi - \overline{\phi})_+(y)\int_y^R \beta(x)\kappa(y,x)\mathcal{U}(x)\,dxdy,\end{aligned}$$

and finally, substracting,

$$0 \leq -\int_0^R f(x)\mathbb{1}_{\phi \geq \overline{\phi}}(x)\mathcal{U}(x)\,dx,$$

so

$$\delta\phi(0)\int_0^R \mathbb{1}_{\phi \geq \overline{\phi}}(x)\mathcal{U}(x)\,dx \leq 0$$

and this can hold only if $\mathbb{1}_{\phi \geq \overline{\phi}} \equiv 0$ or $\phi(0) = 0$. But we deal with the truncated problem with $\tau(x) \geq \eta > 0$, so $\frac{1}{\tau} \in L^1_0$ and $\phi(0) > 0$ thanks to the lemma 1. Thus $\mathbb{1}_{\phi \geq \overline{\phi}} \equiv 0$ and the lemma 4 is proved. $\square$

References

[1] M. Adimy and L. Pujo-Menjouet. Asymptotic behaviour of a singular transport equation modelling cell division. *Dis. Cont. Dyn. Sys. Ser. B*, 3(3):439–456, 2003.

[2] F. Baccelli, D. R. McDonald, and J. Reynier. A mean-field model for multiple tcp connections through a buffer implementing red. *Performance Evaluation*, 49(1-4):77 – 97, 2002.

[3] J. Banasiak and W. Lamb. On a coagulation and fragmentation equation with mass loss. *Proc. Roy. Soc. Edinburgh Sect. A*, 136(6):1157–1173, 2006.

[4] B. Basse, B.C. Baguley, E.S. Marshall, W.R. Joseph, B. van Brunt, G. Wake, and D. J. N. Wall. A mathematical model for analysis of the cell cycle in cell lines derived from human tumors. *J. Math. Biol.*, 47(4):295–312, 2003.

[5] F. Bekkal Brikci, J. Clairambault, and B. Perthame. Analysis of a molecular structured population model with possible polynomial growth for the cell division cycle. *Math. Comput. Modelling*, 47(7-8):699–713, 2008.

[6] F. Bekkal Brikci, J. Clairambault, B. Ribba, and B. Perthame. An age-and-cyclin-structured cell population model for healthy and tumoral tissues. *J. Math. Biol.*, 57(1):91–110, 2008.

[7] T. Biben, J.-C. Geminard, and F. Melo. Dynamics of Bio-Polymeric Brushes Growing from a Cellular Membrane: Tentative Modelling of the Actin Turnover within an Adhesion Unit; the Podosome. *J. Biol. Phys.*, 31:87–120, 2005.

[8] H. Brezis. *Analyse fonctionnelle*. Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree]. Masson, Paris, 1983. Théorie et applications. [Theory and applications].

[9] V. Calvez, N. Lenuzza, M. Doumic, J.-P. Deslys, F. Mouthon, and B. Perthame. Prion dynamic with size dependency - strain phenomena. *J. of Biol. Dyn.*, in press, 2008.

[10] V. Calvez, V. Lenuzza, D. Oelz, J.-P. Deslys, P. Laurent, F. Mouthon, and B. Perthame. Size distribution dependence of prion aggregates infectivity. *Math. Biosci.*, 1:88–99, 2009.

[11] J. Clairambault, S. Gaubert, and T. Lepoutre. Comparison of perron and floquet eigenvalues in age structured cell division cycle models. *Math. Model. Nat. Phenom.*, page in press, 2009.

[12] J. Clairambault, S. Gaubert, and B. Perthame. An inequality for the Perron and Floquet eigenvalues of monotone differential systems and age structured equations. *C. R. Math. Acad. Sci. Paris*, 345(10):549–554, 2007.

[13] O. Destaing, F. Saltel, J.-C. Geminard, P. Jurdic, and F. Bard. Podosomes Display Actin Turnover and Dynamic Self- Organization in Osteoclasts Expressing Actin-Green Fluorescent Protein. *Mol. Biol. of the Cell*, 14, 2003.

[14] M. Doumic. Analysis of a population model structured by the cells molecular content. *Math. Model. Nat. Phenom.*, 2(3):121–152, 2007.

[15] Marie Doumic, Benoît Perthame, and Jorge P. Zubelli. Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems*, 25, 2009.

[16] M. Escobedo, Ph. Laurençot, S. Mischler, and B. Perthame. Gelation and mass conservation in coagulation-fragmentation models. *J. Differential Equations*, 195(1):143–174, 2003.

[17] M. Escobedo, S. Mischler, and B. Perthame. Gelation in coagulation and fragmentation models. *Comm. Math. Phys.*, 231(1):157–188, 2002.

[18] J.Z. Farkas. Stability conditions for a nonlinear size-structured model. *Nonlin. Anal. Real World App.*, 6:962–969, 2005.

[19] M.L. Greer, L. Pujo-Menjouet, and G.F. Webb. A mathematical analysis of the dynamics of prion proliferation. *J. Theoret. Biol.*, 242(3):598–606, 2006.

[20] M. Gyllenberg and G.F. Webb. A Nonlinear Structured Population Model of Tumor Growth With Quiescence. *J. Math. Biol*, 28:671–694, 1990.

[21] P. Laurençot and C. Walker. Well-posedness for a model of prion proliferation dynamics. *J. Evol. Equ.*, 7(2):241–264, 2007.

[22] J. A. J. Metz and O. Diekmann, editors. *The dynamics of physiologically structured populations*, volume 68 of *Lecture Notes in Biomathematics*. Springer-Verlag, Berlin, 1986. Papers from the colloquium held in Amsterdam, 1983.

[23] P. Michel. Existence of a solution to the cell division eigenproblem. *Math. Models Methods Appl. Sci.*, 16(7, suppl.):1125–1153, 2006.

[24] P. Michel. Optimal proliferation rate in a cell division model. *Math. Model. Nat. Phenom.*, 1(2):23–44, 2006.

[25] P. Michel, S. Mischler, and B. Perthame. General entropy equations for structured population models and scattering. *C. R. Math. Acad. Sci. Paris*, 338(9):697–702, 2004.

[26] P. Michel, S. Mischler, and B. Perthame. General relative entropy inequality: an illustration on growth models. *J. Math. Pures Appl. (9)*, 84(9):1235–1260, 2005.

[27] K. Pakdaman, B. Perthame, and D. Salort. Dynamics of a structured neuron population. (submitted), 2009.

[28] B. Perthame. *Transport equations in biology*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007.

[29] B. Perthame and L. Ryzhik. Exponential decay for the fragmentation or cell-division equation. *J. Differential Equations*, 210(1):155–177, 2005.

[30] B. Perthame and T.M. Touaoula. Analysis of a cell system with finite divisions. *Boletín SEMA*, (44), 2008.

[31] Benoît Perthame and Jorge P. Zubelli. On the inverse problem for a size-structured population model. *Inverse Problems*, 23(3):1037–1052, 2007.

[32] J.R. Silveira, G.J. Raymond, A.G. Hughson, R.E. Race, V.L. Sim, S.F. Hayes, and B. Caughey. The most infectious prion protein particles. *Nature*, 437(7056):257–261, September 2005.

2 Haematopoiesis Modelling [14]

A Structured Population Model of Cell Differentiation

Marie Doumic [*‡] Anna Marciniak-Czochra[†] Benoît Perthame[‡*]
Jorge P. Zubelli[§]

December 3, 2010

ABSTRACT

We introduce and analyze several aspects of a new model for cell differentiation. It assumes that differentiation of progenitor cells is a continuous process. From the mathematical point of view, it is based on partial differential equations of transport type. Specifically, it consists of a structured population equation with a nonlinear feedback loop. This models the signaling process due to cytokines, which regulate the differentiation and proliferation process. We compare the continuous model to its discrete counterpart, a multi-compartmental model of a discrete collection of cell subpopulations recently proposed by Marciniak-Czochra *et al.* [17] to investigate the dynamics of the hematopoietic system. We obtain uniform bounds for the solutions, characterize steady state solutions, and analyze their linearized stability. We show how persistence or extinction might occur according to values of parameters that characterize the stem cells self-renewal. We also perform numerical simulations and discuss the qualitative behavior of the continuous model *vis a vis* the discrete one.

Key-words. Structured population dynamics; transport equation; stem cells; cell differentiation;

Contents

[*]INRIA Paris-Rocquencourt, EPI BANG, Domaine de Voluceau, F 78153 Le Chesnay cedex; email: marie.doumic@inria.fr

[†]University of Heidelberg Center for Modelling and Simulation in the Biosciences (BIOMS) Interdisciplinary Center for Scientific Computing (IWR) Institute of Applied Mathematics and BIOQUANT

[‡] Université Pierre et Marie Curie, CNRS UMR 7598 LJLL, BC187, 4, place Jussieu, F-75252 Paris cedex 5; email: benoit.perthame@upmc.fr

[§]IMPA, Est. D. Castorina 110, Rio de Janeiro, RJ 22460-320 Brazil; email: zubelli@impa.br

Introduction

Cell differentiation is a process by which dividing cells become specialized and equipped to perform specific functions such as nerve cell communication or muscle contraction. Differentiation occurs many times during the development of a multicellular organism as the organism changes from a single zygote to a complex system with cells of different types. Differentiation is also a common process in adult tissues. During tissue repair and during normal cell turnover a steady supply of somatic cells is ensured by proliferation of corresponding adult stem cells, which retain the capability for self-renewal. Also various cancers are likely to originate from a population of cancer stem cells that have properties comparable to those of stem cells [3].

Stem cell state and fate depends on the environment, which ensures that the critical stem cell character and activity in homeostasis is conserved, and that repair and development are accomplished [19]. Cell differentiation and the maintenance of self-renewal are intrinsically complex processes requiring the coordinated dynamic expression of hundreds of genes and proteins in response to external signaling. During differentiation, certain genes become activated and other genes inactivated in an intricately regulated fashion. As a result, differentiated cells develop specific structures and performs specific functions. There exists evidence that disorder in self-renewal behavior may lead to neoplasia [3, 4]. For example, it has been shown that acute myeloid leukemia originates from a hierarchy of cells that differ with respect to self-renewal capacities [13, 23]. Although much progress has been made in identifying the specific factors and genes responsible for stem cells decisions [20], the mechanisms involved in these processes remain largely unknown.

While different genetic and epigenetic processes are involved in formation and maintenance of different tissues, the dynamics of population depends on the relative importance of symmetric and asymmetric cell divisions, cell differentiation and death. The same genes

and proteins are observed to be essential for regulation of different tissues [21]. This unity and conservation of basic processes implies that their mathematical models can apply across the spectrum of normal and pathological (cancer stem cells) development.

One established method of modeling such systems is to use a discrete collection of ordinary differential equations describing dynamics of cells at different maturation stages and transition between the stages. These so called multi-compartmental models are based on the assumption that in each lineage of cell precursors there exists a discrete chain of maturation stages, which are sequentially traversed, e.g., [15, 27]. However, it is also becoming progressively clear that the differentiated precursors form such sequence only under homeostatic (steady-state) conditions. Committed cells generally form a continuous sequence, which may involve incremental stages, part of which may be reversible. As an example, cell differentiation without cell divisions is observed during neurogenesis. Moreover, in some tissues such as the mammary gland, different stages of differentiation are not well identified [9].

These observations invoke not only the fundamental biological question of whether cell differentiation is a discrete or a continuous process and what is the measure of cell differentiation, but also how to choose an appropriate modeling approach. Is the pace of maturation (commitment) dictated by successive divisions, or is maturation a continuous process decoupled from proliferation? In leukemias, it seems to be decoupled. The classical view in normal hematopoiesis seems to be opposite.

To address these questions and to investigate the impact of possible continuous transformations on the differentiation process, we introduce a new model based on partial differential equations of transport type and compare this model to its discrete counterpart. The point of departure is a multi-compartmental model of a discrete collection of cell subpopulations, which was recently proposed in [17] to investigate dynamics of the hematopoietic system with cell proliferation and differentiation regulated by a nonlinear feedback loop. Furthermore, since self-renewal is an important parameter in our models, the proposed models seem to be a right departure point to investigate cancer development, for example in leukemias [23].

In the present paper we extend the discrete model to a structured population model accounting for a continuous process of differentiation of progenitor cells. Models of the latter type have been already applied to the description of some aspects of hematopoiesis [6, 5, 1, 2, 11]. These models are based on the assumption that differentiation of progenitor cells is a continuous process, which progresses with a constant velocity. Mathematical description involves so called age-structured population equations. The model presented here is novel due to the nonlinearities in the coupling of the model equations, in particular the nonlinear coupling in the maturity rate function.

The paper is organized as follows. In Section 1, we formulate the new model. In Section 1.2 the link between our model and the discrete model of [17] is accomplished. Sections 2, 3 and 4 are devoted to the analysis of the model. In Section 2 the existence and uniform boundedness of the solutions are shown. In Section 3, it is shown that depending on the value of a parameter characterizing stem cells self-renewal model solutions tend to zero or they stay separated from zero. Section 4 provides the structure of steady states and

conditions for existence of a positive stationary solution, while Section 4.2 is devoted to a linearized problem around the positive steady state when this state exists to investigate its stability. Using the characteristic equation we study some special cases, for which we show stability or instability of the positive stationary solution. In Section 5, a numerical approach and some results on stability and instability are presented. We conclude in Section 6 with some final comments and suggestions for further investigation.

1 Model of Cell Differentiation

1.1 Continuous Model

In the following we assume that the dynamics of differentiated precursors can be approximated by a continuous maturation model. Under this assumption we extend the multi-compartmental system from [17]. Let $w(t)$ denote the number of stem cells, $v(t)$ the number of mature cells and $u(x,\, t)$ the distribution density of progenitor cells structured with respect to the maturity level x, so that $\int_{x_1}^{x_2} u(x,t)dx$ is equal to the number of progenitors with maturity between x_1 and x_2. This includes maturity stages between stem cells and differentiated cells. Thus, $u(0,t)$ describes a population of stem cells and $u(x,\, t)$, for $x > 0$, corresponds to progenitor cells. We assume that $x = x^*$ denotes the last maturity level of immature cells, and therefore, $u(x^*,t)$ describes the concentration of cells which differentiate into mature cells.

The model takes the form

$$\begin{aligned}
\frac{d}{dt}w(t) &= [2a_w(s)-1]p_w(s)w(t) - d_w w(t), && (1)\\
\partial_t u(x,t) + \partial_x[g(x,s)u(x,t)] &= p(x,s)u(x,t) - d(x)u(x,t), && (2)\\
g(0,s)u(0,t) &= 2[1-a_w(s)]p_w(s)w(t), \quad t>0, && (3)\\
\frac{d}{dt}v(t) &= g(x^*,s)u(x^*,t) - \mu v(t), && (4)
\end{aligned}$$

together with initial data

$$w(0) = w_0 \geq 0, \qquad u(0,x) = u_0(x) \geq 0, \qquad v(0) = v_0 \geq 0.$$

Integrating formally Equation (2) and adding it to Equations (1) and (4) yields the following cell number balance equation

$$\frac{d}{dt}\big[w(t) + \int_0^{x^*} u(x,t)dx + v(t)\big] = (p_w(s) - d_w)w(t) + \int_0^{x^*} \big(p(x,s) - d(x)\big)u(x,t)dx - \mu v(t). \tag{5}$$

System (1)-(4) describes the following scenario: After division a stem cell gives rise to two progeny cells. Cell divisions can be symmetric or asymmetric. We assume that on the average the fraction a_w of progeny cells remains at the same stage of differentiation as the parent cell, while the $1 - a_w$ fraction of the progeny cells differentiates, i.e. transfers

to the higher differentiation stage. This covers the symmetric and asymmetric scenarios. Parameters p_w and d_w denote the proliferation rate of stem cells and their death rate, respectively. Progenitor cells differentiate at the rate g, which depends on their maturity stage and is also regulated by the feedback from mature cells given by a signaling factor s. Parameters $p(x)$ and $d(x)$ denote the proliferation and death rates of precursor cells and depend on the level of cell maturation. Mature cells do not divide and die at the rate μ. The whole process is regulated by a single feedback mechanism based on the assumption that there exist signaling molecules (cytokines) which regulate the differentiation or proliferation process. The intensity of the signal depends on the level of mature cells, and is modeled using the dependence

$$s = s[v(t)] = \frac{1}{1 + kv(t)},$$

which can be justified using a quasi-steady state approximation of the plausible dynamics of the cytokine molecules, see [17]. This expression reflects the heuristic assumption that signal intensity achieves its maximum under absence of mature cells and decreases asymptotically to zero if level of mature cells increases.

The concentration of signaling molecules $s(v)$ influences the length of the cell cycle (proliferation rate p) and/or the fraction of stem cells self-renewal (a_w) as well as the rate of cell maturation (g).

In this model, differentiation of stem cells takes place during mitosis. The differentiation of progenitor cells occurs independently of proliferation. In other words, cells undergo continuous transformations between divisions. We call this process maturation. In the terms of the model this means that in an infinitesimal time interval (t, $t+dt$), the following events occur to a cell of maturity x,

1. either the cell matures to level $x + dx$, which happens with probability $g[x, v(t)]dt$,
2. or the cell divides into 2 daughters, which happens with probability $p(x)dt$, with other events occurring with probabilities of the order $o(dt)$.

If we stick to the discrete model proposed in [17], we obtain the following relations linking proliferation and maturation

$$\begin{cases} g(x, v) = 2[1 - \frac{a(x)}{1+kv(t)}]p(x), \\ a_w = a(0), \qquad p_w = p(0), \qquad 0 < a_w = a(0) \leq 1, \end{cases} \tag{6}$$

which does not necessarily mean that differentiation can only occur by division; see the discussion at the end of Section 1.2.

This leads to a simplification of the boundary condition (3) that becomes

$$u(0, t) = w(t).$$

1.2 Discrete *versus* Continuous Models

General Setting

In this section we consider the relationship between the structured population model (1)-(4) and the multicompartmental model introduced in [17]. Following reference [18], the multicompartmental model can be formulated in a general way,

$$\frac{d}{dt}u_1 = p_1(s)u_1 - g_1(u_1, s) - d_1u_1, \tag{7}$$

$$\frac{d}{dt}u_i = p_i(s)u_i + g_{i-1}(u_{i-1}, s) - g_i(u_i, s) - d_iu_i, \quad \text{for } i = 1, ..., n-1 \tag{8}$$

$$\frac{d}{dt}u_n = g_{n-1}(u_{n-1}, s) - d_nu_n, \tag{9}$$

where $g_i(u_i, s)$ denotes a flux of cells from the subpopulation i differentiating to the subpopulation $i + 1$. The terms $p_i(s)u_i$ and d_iu_i describe cell fluxes due to proliferation and death, respectively. In the general case proliferation or differentiation may depend on signal intensity.

In reference [17], differentiation was linked to proliferation and the following expressions were proposed:

$$g_i(u_i, s) = 2(1 - \frac{a_i}{1 + ku_n})p_iu_i. \tag{10}$$

This is similar to what was defined for stem cells with p_w and a_w; here $\frac{a_i}{1+ku_n}$ represents the fraction of cells remaining at the same stage of differentiation i as the parent cell while the $1 - \frac{a_i}{1+ku_n}$ fraction of cells differentiates to the higher stage $i + 1$. Formulation (7)-(9) describes the differentiation process independently of cell proliferation in the sense that cells either multiply at stage i or differentiate from compartment i to $i + 1$ and so forth. Assuming that cell differentiation occurs at a properly-chosen time scale compared to the time scale of the cell division process, we show in the next paragraph how to obtain, after a suitable renormalization, the structured population model of the next paragraph.

Continuous Limit

Let us write System (7)–(9) in a dimensionless way. We define $\mathcal{P}$, $\mathcal{D}$, $\mathcal{G}_1$, $\mathcal{G}$, $\mathcal{U}_1$, $\mathcal{U}$ and $\mathcal{U}_n$ as characteristic values for the quantities p_i, d_i, g_1, g_i for $i \geq 2$, u_1, u_i for $2 \leq i \leq n-1$, and u_n, respectively. Then, we define dimensionless quantities by $\bar{p}_i = \frac{p_i}{\mathcal{P}}$ etc. We make the following hypothesis

$$g_i(u_i, s) = g_i(s)u_i.$$

Then, system (7)–(9) becomes

$$\frac{d}{dt}\bar{u}_1 = \mathcal{P}\bar{p}_1(s)\bar{u}_1 - \mathcal{G}_1\bar{g}_1\bar{u}_1 - \mathcal{D}\bar{d}_1\bar{u}_1, \tag{11}$$

$$\frac{d}{dt}\bar{u}_2 = \mathcal{P}\bar{p}_2(s)\bar{u}_2 + \mathcal{G}_1\frac{\mathcal{U}_1}{\mathcal{U}}\bar{g}_1\bar{u}_1 - \mathcal{G}\bar{g}_2\bar{u}_2 - \mathcal{D}\bar{d}_2\bar{u}_2, \tag{12}$$

$$\frac{d}{dt}\bar{u}_i = \mathcal{P}\bar{p}_i(s)\bar{u}_i + \mathcal{G}(\bar{g}_{i-1}\bar{u}_{i-1} - \bar{g}_i\bar{u}_i) - \mathcal{D}\bar{d}_i\bar{u}_i, \quad \text{for } i = 3, ..., n-1, \tag{13}$$

$$\frac{d}{dt}\bar{u}_n = \frac{\mathcal{G}\mathcal{U}}{\mathcal{U}_n}\bar{g}_{n-1}\bar{u}_{n-1} - \mathcal{D}\bar{d}_n\bar{u}_n. \tag{14}$$

Letting the number of compartments tend to infinity, we pass from the discrete model to the continuous model by associating to the u_i's a function, constant on intervals of type $(\varepsilon i, \varepsilon(i+1))$, with $\varepsilon \to 0$, $i \to \infty$ and the product εi remaining positive and finite, say $n = n_\varepsilon$, with $\varepsilon n \to x^* \in \mathbb{R}_+^* = (0, +\infty)$. Compartment dependent constants tend to continuous functions, sums over the index i are interpreted as Riemann sums tending to integrals while finite differences give rise to derivatives. A precise discussion of the limiting process is outside the scope of this presentation. We refer, for instance, to [7, 10] for recent examples of how to obtain such limits based on moments estimates.

In order to interpret the terms $\mathcal{G}(\bar{g}_{i-1}\bar{u}_{i-1} - \bar{g}_i\bar{u}_i)$ in Equation (13) and $\mathcal{G}_1\frac{\mathcal{U}_1}{\mathcal{U}}\bar{g}_1\bar{u}_1 - \mathcal{G}\bar{g}_2\bar{u}_2$ in Equation (12), as a finite differences tending to a derivative, we take

$$\mathcal{G} = \frac{1}{\varepsilon}, \qquad \mathcal{G}_1\frac{\mathcal{U}_1}{\mathcal{U}} = \frac{1}{\varepsilon}.$$

Assuming that $\mathcal{P}\bar{p}_i(s)\bar{u}_i$ and $\mathcal{D}\bar{d}_i\bar{u}_i$ tend toward limits $p(x,s)u(x,t)$ and $d(x)u(x,t)$ leads to

$$\mathcal{P} = 1, \qquad \mathcal{D} = 1.$$

In order to obtain that $\frac{\mathcal{G}\mathcal{U}}{\mathcal{U}_n}\bar{g}_{n-1}\bar{u}_{n-1}$ converges to the limit $g(x^*)u(t,x^*)$ in Equation (14), we require

$$1 = \frac{\mathcal{G}\mathcal{U}}{\mathcal{U}_n} = \frac{1}{\varepsilon}\frac{\mathcal{U}}{\mathcal{U}_n} \qquad \mathcal{U}_n = \frac{\mathcal{U}}{\varepsilon}.$$

This means that the order of magnitude of the number of mature cells is much larger than the one of the maturing cells and stem cells. This can be interpreted by the fact that u_i tends to $u(t,x)$ a *density* of cells per unit of maturity, whereas u_1 and u_n are *numbers* of cells. This scaling follows also from mass balance considerations: System (7)–(9) leads to the following mass balance

$$\frac{d}{dt}u_1 + \frac{d}{dt}\sum_{i=2}^{n-1}u_i + \frac{d}{dt}u_n = p_1u_1 + \sum_{i=2}^{n-1}p_iu_i - d_1u_1 - \sum_{i=2}^{n-1}d_iu_i - d_nu_n, \tag{15}$$

meaning that the exchange among compartments at maturation rate g_i does not influence the total growth of the population. In this equation, we keep the specific values u_1 and u_n

and interpret the sum for $2 \leq i \leq n-1$ as an integral. We obtain

$$\frac{d}{dt}u_1 + \frac{d}{dt}\sum_{i=2}^{n-1} u_i + \frac{d}{dt}u_n = \frac{d}{dt}\mathcal{U}_1\bar{u}_1 + \mathcal{U}\frac{d}{dt}\sum_{i=2}^{n-1}\bar{u}_i + \mathcal{U}_n\frac{d}{dt}\bar{u}_n,$$

which leads to the following choice

$$\frac{\mathcal{U}}{\mathcal{U}_n} = \frac{\mathcal{U}}{\mathcal{U}_1} = \varepsilon.$$

With this choice and the previous relations, we are led to choose $\mathcal{G}_1 = 1$, which allows a limit for Equation (11). We note however that this implies a different order of magnitude for $\mathcal{G}_1$ and for $\mathcal{G}$; the interpretation could be that we have divided the previous discrete compartments into smaller ones, of size ε, where division does not occur but where maturation occurs. In this framework, $\mathcal{G}_1$ is not homogeneous to $\mathcal{G}$ but rather to the integral of $\mathcal{G}$ over a small compartment of size ε.

Under these assumptions, let us set

$$\chi_i^\varepsilon(x) = \chi_{[i\varepsilon,(i+1)\varepsilon)}(x),$$

with χ_A being the indicator function of a set A. We introduce the piecewise constant function

$$u^\varepsilon(x,t) := \sum_{i=1}^{n_\varepsilon} u_i(t)\chi_i^\varepsilon(x).$$

By the same token, we associate the following functions to the coefficients

$$d^\varepsilon(x) := \sum_{i=1}^{n_\varepsilon} d_i\chi_i^\varepsilon(x), \qquad p^\varepsilon(x,s) := \sum_{i=1}^{n_\varepsilon} p_i(s)\chi_i^\varepsilon(x), \qquad g^\varepsilon(x,u(x),s) := \sum_{i=1}^{n_\varepsilon} g_i(u_i,s)\chi_i^\varepsilon(x).$$

We make the following continuity assumptions on the dimensionless system:

$$\begin{aligned} &\exists K > 0 \quad \text{s.t.} \quad |g_i| + |d_i| + |p_i| \leq K, \quad |g_{i+1} - g_i| + |d_{i+1} - d_i| + |p_{i+1} - p_i| \leq \tfrac{K}{i} \\ &p_i, g_i \text{ are uniformly continuous with respect to the variable } s. \end{aligned} \tag{16}$$

We define the piecewise constant functions g^ε, d^ε and p^ε on the respective basis of the discrete coefficients g_i, d_i and p_i, similarly as u^ε was defined for u_i. Assumption (16) leads to their convergence (up to subsequences) to continuous functions g, d and p of both variables x and s (see Lemma 1 of [10] for instance). We can prove (based e.g. on [7, 10]) the following result.

Proposition 1.1 *Suppose that u_i^ε is a solution of Equation* (13) *verifying $(u_i(t=0)) \in l^1$. Under Assumption* (16)*, for all $T > 0$, there exists a subsequence of (u_i^ε) converging towards a limit $u \in \mathcal{C}(0,T;\mathcal{M}^1([0,x^*])-weak-*)$ solution of Equation* (2)*, u_1^ε to a limit $u_1 \in \mathcal{C}(0,T)$ solution of Equation* (1) *where $s = s(v)$ with v a limit of a subsequence of u_n^ε. The boundary condition* (3) *is satisfied in a distributional sense and Equation* (5) *is satisfied, what is equivalent to a weak formulation of the boundary condition* (4)*.*

We note that we have used the fact that maturation and proliferation are decorrelated. If this were not the case, it would be impossible to make a $1/\varepsilon$ factor appear in $\mathcal{G}$, since $\mathcal{P} = 1$. In such case in the limit equation the transport appears as a first order corrective term and Equation (2) is replaced by

$$\partial_t u(x,t) + \varepsilon \partial_x [g(x,s)u(x,t)] = p(x,s)u(x,t) - d(x)u(x,t). \tag{17}$$

Figure 1 in Section 5.2 depicts related numerical simulations.

2 Uniform Bounds for the Continuous Model

In the remainder of this work we will consider a special version of the above model assuming time independent proliferation rates $p(x)$, and zero death rates of undifferentiated cells $d_w = 0$ and $d(x) = 0$. Indeed, neglecting death rates of immature cells does not change the analysis. Concerning the feedback loops it was shown in [17] for the discrete model that the feedback on the stem cells self-renewal fraction and on the maturation speed g is much more important for the efficiency of the process than the feedback on the proliferation rate $p(x)$. Therefore, in the reminder of this work we focus on the model with regulated self-renewal and maturation. We also introduce simpler notation which makes it easier for analysis. This yields the following system of differential equations for $t > 0$, $x > 0$.

$$\begin{aligned}
\frac{d}{dt} w(t) &= \alpha(v(t))w(t), && (18)\\
\partial_t u(x,t) + \partial_x [g(x,v(t))u(x,t)] &= p(x)u(x,t), && (19)\\
u(0,t) &= w(t), && (20)\\
\frac{d}{dt} v(t) &= g(x^*, v(t))u(x^*,t) - \mu v(t), && (21)
\end{aligned}$$

together with initial data

$$w(0) = w_0 \geq 0, \qquad u(0,x) = u_0(x) \geq 0, \qquad v(0) = v_0 \geq 0. \tag{22}$$

We obtain the cell number balance law

$$\frac{d}{dt}\big[w + \int u(x,t)dx + v\big] = [\alpha(v) + g(0,v)]w + \int p(x)u(x,t)dx - \mu v(t), \tag{23}$$

corresponding to the fact that the total population can only change by proliferation or death. Indeed, one can interpret $\alpha + g(0,v)$ as the stem cells proliferation rate, see above sections and Equation (5).
In the sequel we will study model (18)–(21) under the following assumptions

$$g_x,\ g_{xx} \in L^\infty([0,x^*] \times \mathbb{R}^+),\ \alpha(x) \in C([0,\infty)),\ p(x) \in C^1([0,x^*]), \tag{24}$$

$$\alpha(v) \in [\alpha_\infty, \alpha_0], \quad \alpha \text{ is decreasing }, \quad \alpha(+\infty) := \alpha_\infty < 0, \tag{25}$$

$$0 < g_- \leq g(x,v) \leq g_+ < \infty,\ \forall (x,v) \in [0,x^*] \times \mathbb{R}^+, \tag{26}$$

First we show that the model solutions are uniformly bounded

Theorem 2.1 *Under assumptions* (24)–(26) *and that* $u_0(x) \in C^1([0,x^*])$*, the solution to System* (18)–(22) *is uniformly bounded. More precisely, all the components* $w(t)$*,* $u(x,t)$*,* $v(t)$ *are uniformly bounded.*

The remainder of the section is devoted to the proof of this result, which uses some technical lemmas. We first prove the following estimate

Lemma 2.2 *Under the assumptions of Theorem 2.1, the function* $z(x,t) = \partial_x(\ln u)$ *is uniformly bounded on* $[0,x^*] \times \mathbb{R}^+$*.*

Proof. The equation for z reads

$$\left\{ \begin{array}{rl} \partial_t z + \partial_x(gz) & = -g_{xx} + p_x, \\ z(0,t) & = -\frac{\alpha(v)-p(0)}{g(0,v)} - \frac{g_x(0,v)}{g(0,v)} \in L^\infty(0,+\infty). \end{array} \right. \tag{27}$$

Indeed, we have $z(0,t) = \frac{u_x(0,t)}{u(0,t)}$ and thus we can compute

$$\begin{aligned} z(0,t) &= -\frac{\partial_t u(0,t) - p(0)u(0,t) + g_x(0,v)u(0,t)}{g(0,v)u(0,t)} \\ &= -\frac{\alpha(v)-p(0)}{g(0,v)} - \frac{g_x(0,v)}{g(0,v)}. \end{aligned}$$

And we conclude that $z(0,t)$ is uniformly bounded by assumptions (24)–(26).
Next, we rewrite the equation for z as

$$\partial_t z + g\partial_x z = -g_x z + Q(x,v), \tag{28}$$

where $Q = -g_{xx} + p_x$ is a bounded function of v and x.

In the following, we show that the solution to (28) satisfies the estimate

$$||z(x,t)||_{L^\infty} \leq M := \left(\sup_t |z(0,t)| + \sup_x |z(x,0)| + x^*\ ||\frac{Q}{g}||_{L^\infty} \right) e^{x^* ||\frac{g_x}{g}||_{L^\infty}}. \tag{29}$$

Indeed, since $g \geq g_- > 0$, we can rewrite Equation (28) as

$$\partial_x z + \frac{1}{g}\partial_t z = -\frac{g_x}{g} z + \frac{Q(x,v)}{g},$$

and apply the method of characteristics by defining as usual (except that x plays the role of time),

$$\frac{dT}{dx}(x,t) = \frac{1}{g}(x, v(T(x,t))), \quad T(x=0,t) = t,$$

$$\frac{dT^{-1}}{dx}(x,t') = -\frac{1}{g}(x, v(T^{-1}(x,t'))), \quad T^{-1}(x=0,t') = t'.$$

We look for solutions of the form $Z(x,t) = z(x,T(x,t))$, which satisfy the following equation

$$\partial_x Z + \frac{g_x}{g}(x,T(x,t))Z = \partial_x(Ze^{\int_0^x \frac{g_x}{g}(\zeta,T(\zeta,t))d\zeta})e^{-\int_0^x \frac{g_x}{g}(\zeta,T(\zeta,t))d\zeta} = \frac{Q}{g}(x, v(T(x,t))).$$

Integrating the above equation yields

$$\begin{aligned}
Z(x,t)e^{\int_0^x \frac{g_x}{g}(\zeta,T(\zeta,t))d\zeta} &= Z(0,t) + \int_0^x \frac{Q}{g}(\xi, v(T(\xi,t)))e^{\int_0^\xi \frac{g_x}{g}(\zeta,T(\zeta,t))d\zeta}d\xi, \\
Z(x,t) &= Z(0,t)e^{-\int_0^x \frac{g_x}{g}(\zeta,T(\zeta,t))d\zeta} + \int_0^x \frac{Q}{g}(\xi, v(T(\xi,t)))e^{-\int_\xi^x \frac{g_x}{g}(\zeta,T(\zeta,t))d\zeta}d\xi, \\
z(x,T(x,t)) &= z(0,t)e^{-\int_0^x \frac{g_x}{g}(\zeta,T(\zeta,t))d\zeta} + \int_0^x \frac{Q}{g}(\xi, v(T(\xi,t)))e^{-\int_\xi^x \frac{g_x}{g}(\zeta,T(\zeta,t))d\zeta}d\xi.
\end{aligned}$$

Defining $\bar{t} = T(x,t)$, or yet $t = T^{-1}(x,\bar{t})$, yields $T^{-1} \geq 0$ for $\bar{t} \geq \frac{x^*}{g_{min}}$, and we obtain

$$z(x,\bar{t}) = z(0,T^{-1}(x,\bar{t}))e^{-\int_0^x \frac{g_x}{g}(\zeta,T(\zeta,T^{-1}(x,\bar{t})))d\zeta} + \int_0^x \frac{Q}{g}(\xi, v(T(\xi,T^{-1}(x,\bar{t}))))e^{-\int_\xi^x \frac{g_x}{g}(\zeta,T(\zeta,T^{-1}(x,\bar{t})))d\zeta}d\xi.$$

Therefore, for $\bar{t} \geq \frac{x^*}{g_{min}}$ it holds

$$||z||_{L^\infty} \leq (|z(0,\cdot)| + x^*||\frac{Q}{g}||_{L^\infty})e^{x^*||\frac{g_x}{g}||_{L^\infty}}.$$

□

From Lemma 2.2, we deduce several useful estimates

Lemma 2.3 *There exist positive constants M_1, M_2, M_3 such that the solutions to system (18)–(22) satisfy*

(i) $w(t) \leq M_1 u(x,t)$,

(ii) $w(t) \leq M_2 v(t)$,

(iii) $u(x,t) \leq M_3 w(t)$.

Proof (i) Boundedness of $-z = -\frac{\partial}{\partial x}\ln u$ results in the following inequality

$$\ln\frac{1}{u} \leq \ln\frac{1}{w} + Mx,$$

which in turn yields assertion (i) with $M_1 = e^{Mx^*}$.

(ii) To bound w by v, we calculate

$$\frac{d}{dt}\frac{w}{v} = \frac{w}{v}\left(\alpha(v(t)) - g(x^*, v(t))\frac{u(x^*,t)}{v} + \mu\right).$$

Since $\alpha(v) \leq \alpha(0)$, $g(x^*, v(t)) \geq g_-$ and $u(x^*, t) \geq w(t)/M_1$ we obtain

$$\frac{d}{dt}\frac{w}{v} \leq \frac{w}{v}\left(\alpha(0) + \mu - \frac{g_-}{M_1}\frac{w}{v}\right).$$

This yields the estimate

$$w(t) \leq v(t)\max\left(\frac{w(0)}{v(0)}, M_1\frac{\alpha(0)+\mu}{g_-}\right) := M_2 v(t),$$

and the assertion (ii) is proved.

(iii) The proof follows as in (i), departing from $\ln u(x,t) \leq \ln w(x,t) + Mx$. □

As a consequence of Lemma 2.3, we derive

Corollary 2.4 *Under the assumptions of Theorem 2.1, the components $w(t)$, $u(x,t)$ and $v(t)$ of the solutions to System* (18)–(22) *are uniformly bounded.*

Proof. Applying Lemma 2.3 (ii) to equation (18), we obtain

$$\frac{dw}{dt} \leq \alpha\left(\frac{w}{M_2}\right)w.$$

This yields boundedness of w by Assumption (25).

Boundedness of w yields also boundedness of u using Lemma 2.3 (iii). Finally, boundedness of v results from Equation (21) due to boundedness of $u(x^*, t)$ because $g \leq g_+$. □

The proof of Theorem 2.1 is now complete. □

We also state another result, in the spirit of Lemma 2.3, that is used later on

Lemma 2.5 *There exists a constant $M_4 > 0$ and $0 < \gamma < 1$ such that $v(t) \leq M_4 w^\gamma(t)$.*

Proof. We calculate

$$\frac{d}{dt}\frac{v}{w^\gamma} \leq M_3 g(x^*, v(t))w^{1-\gamma} - \frac{v}{w^\gamma}(\mu + \gamma\alpha(v)).$$

We choose $\gamma > 0$ small enough such that $\mu + \gamma\alpha_\infty := \mu_1 > 0$ and $\gamma < 1$. Since w is uniformly bounded, we find

$$\frac{d}{dt}\frac{v}{w^\gamma} \leq C - \frac{v}{w^\gamma}\mu_1$$

which yields boundedness of $\frac{v}{w^\gamma}$. □

Finally, we conclude this section with a consequence of Theorem 2.1.

Corollary 2.6 *Under the Assumptions* (24)–(26) *and that* $u_0(x) \in C^1([0, x^*])$, *System* (18)–(22) *has a unique global solution. Furthermore, such solution is uniformly bounded.*

Proof. Local in time existence of the unique solution follows from the Cauchy-Lipschitz theorem. Theorem 2.1 provides uniform boundedness of solutions and hence the global existence. □

3 Extinction and Persistence

In this section we provide conditions for extinction and persistence of positive solutions.

First, we consider a case when $\alpha(0) < 0$. In this case there exists only a trivial steady state of the model and

Theorem 3.1 *Assume* (24)–(26). *If* $\alpha(0) < 0$, *then all solutions of system* (18)–(22) *converge to zero at an exponential rate.*

Proof First of all, notice that, since $\alpha(v) \leq \alpha(0) < 0$, it is obvious from equation (18) that w converges to 0 exponentially.
For the other components, we consider a functional $\gamma w(t) + \int_0^{x^*} e^{-\beta x}u(x,t)dx + e^{-\beta x^*}v$, with positive constants γ and β to be determined. We compute its time derivative,

$$\begin{aligned}
&\frac{d}{dt}\big(\gamma w(t) + \int_0^{x^*} e^{-\beta x}u(x,t)dx + e^{-\beta x^*}v\big)\\
=\ &\gamma\alpha(v)w(t) - \beta\int_0^{x^*} e^{-\beta x}g(x,v)u(x,t)dx - e^{-\beta x^*}g(x^*,v)u(x^*,t)\\
&+g(0,v)u(0,t) + \int_0^{x^*} p(x)u(x,t)e^{-\beta x}dx + g(x^*,v)u(x^*,t)e^{-\beta x^*} - \mu e^{-\beta x^*}v\\
=\ &[\gamma\alpha(v) + g(0,v)]w(t) + \int_0^{x^*} e^{-\beta x}u(x,t)\big(p(x) - \beta g(x,v)\big)dx - \mu e^{-\beta x^*}v.
\end{aligned}$$

Since $\alpha(v) \leq \alpha(0) < 0$ we may choose γ such that $\gamma\alpha(0)+\sup_v g(0,v) \leq -\Gamma < 0$. Moreover, choosing β such that $p(x) - \beta g(x,v) < -\Gamma$ implies that

$$\frac{d}{dt}\big(\gamma w(t) + \int_0^{x^*} e^{-\beta x} u(x,t)dx + e^{-\beta x^*} v\big) \leq -\Gamma w(t) - \Gamma \int_0^{x^*} e^{-\beta x} u(x,t)dx - \mu e^{-\beta x^*} v.$$

We conclude that the solutions converge to zero at an exponential rate for $t \to \infty$. □

Secondly, if it is the case that $\alpha(0) > 0$, then we conclude that the solutions to the system cannot become extinct.

Theorem 3.2 *Assume* (24)–(26)*,* $w(0) > 0$ *and* $u_0(x) \in C^1([0, x^*])$*. If* $\alpha(0) > 0$*, the solution* u*,* v*,* w *of system* (18)-(21) *with positive initial conditions remain bounded away from zero.*

Proof Applying Lemma 2.5 to equation (18), we obtain

$$\frac{dw}{dt} \geq \alpha\left(M_4 w^\gamma\right) w,$$

and the assumption $\alpha(0) > 0$ allows us to conclude. Then, the estimates of Lemma 2.3 conclude for u and v. □

4 Stationary Solutions and Their Stability

4.1 Stationary Solutions

As usual in dynamical systems, a natural question concerns the existence of steady states (stationary solutions). We shall now investigate this issue.

In our case, the steady states are given by the solutions $(\bar{w}, \bar{u}, \bar{v})$ to the system

$$\begin{aligned} \alpha(\bar{v})\bar{w} &= 0, && (30)\\ \frac{d}{dx}[\bar{g}(x)\bar{u}(x)] &= p(x)\bar{u}(x), && (31)\\ \bar{u}(0) &= \bar{w}, && (32)\\ \bar{g}(x^*)\bar{u}(x^*) - \mu\bar{v} &= 0, && (33) \end{aligned}$$

where $\bar{g}(x) := g(x, \bar{v})$.

System (18)–(21) always admits the trivial steady state $w = 0$, $u = 0$, $v = 0$, which we do not consider. Depending upon the value $\alpha(0)$ it may also have exactly one positive steady state $(\bar{w}, \bar{u}, \bar{v})$, as we state it in the

Lemma 4.1 *Under the Assumptions* (24)–(26)*, the System* (18)-(21) *has a strictly positive steady state if and only if* $\alpha(0) > 0$*. Furthermore, the steady state is unique.*

This condition is in agreement with biological observations concerning self-renewal of stem cell subpopulation [12] and an analogous condition for the compartmental model was discussed in [18].

Proof Since we discard the trivial steady state, from equation (30) we obtain the condition $\alpha(\bar{v}) = 0$. As we know that α decreases and tends to $\alpha_\infty < 0$ at infinity, there exists a unique solution $\bar{v}$ to

$$\alpha(\bar{v}) = 0, \tag{34}$$

if and only if the condition $\alpha(0) > 0$ holds. Thus, we may compute

$$\bar{u}(x^*) = \frac{\mu \bar{v}}{\bar{g}(x^*)} \ .$$

Solving differential equation (31) with this boundary condition at $x = x^*$ yields,

$$\bar{u}(x) = \frac{\bar{g}(x^*)}{\bar{g}(x)} \bar{u}(x^*) \exp\Big\{ - \int_x^{x^*} \frac{p(\xi)}{\bar{g}(\xi)} d\xi \Big\}. \tag{35}$$

We finally identify $\bar{w}$ using the boundary condition at $x = 0$, $\bar{w} = \bar{u}(0)$, which leads to

$$\bar{w} = \frac{\bar{g}(x^*)}{\bar{g}(0)} \bar{u}(x^*) \exp\Big\{ - \int_0^{x^*} \frac{p(\xi)}{\bar{g}(\xi)} d\xi \Big\} = \frac{\mu \bar{v}}{\bar{g}(0)} \exp\Big\{ - \int_0^{x^*} \frac{p(\xi)}{\bar{g}(\xi)} d\xi \Big\}. \tag{36}$$

This gives explicit values of the model and completes the proof. □

For α given explicitly by (6), we may compute

$$\left\{ \begin{array}{rl} \bar{v} & = \frac{2a_w - 1}{k}, \\ \bar{u}(x^*) & = \frac{\mu}{kp(x^*)} \ \frac{a_w(2a_w-1)}{2a_w - a(x^*)}. \end{array} \right. \tag{37}$$

4.2 The Linearized Problem around the Steady State

In order to investigate local linear stability, we consider in this section the linearization around the positive steady state. We first derive a characteristic equation for the eigenvalue problem. The signs of the real parts of these eigenvalues give stability (if they all are negative) or instability (if there exists one with positive real part). To emphasize our main point, which is that stability as well as instability of the positive steady state can take place for a suitable choice of model parameters, we shall focus on some simpler cases where stability analysis is more transparent.

The Characteristic Equation in the General Case

We denote by $(\bar{w}, \bar{u}, \bar{v})$ the steady state solution to Equations (30)–(33). Positivity of the considered steady state yields $\alpha(\bar{v}) = 0$. The linearized problem reads

$$\begin{aligned}
\frac{d}{dt}w(t) &= \frac{d\alpha}{dv}(\bar{v})\bar{w}v(t), && (38)\\
\partial_t u(x,t) + \partial_x[g(x,\bar{v})u(x,t)] + \partial_x[\frac{\partial g}{\partial v}(x,\bar{v})\bar{u}(x)]v(t) &= p(x)u(x,t), && (39)\\
u(0,t) &= w(t), && (40)\\
\frac{d}{dt}v(t) = g(x^*,\bar{v})u(x^*,t) + \frac{\partial g}{\partial v}(x^*,\bar{v})\bar{u}(x^*)v(t) &- \mu v(t), && (41)
\end{aligned}$$

where w, u, and v denote now the deviation of the solution from the steady state. Setting $w(t) = We^{\lambda t}$, $u(x,t) = U(x)e^{\lambda t}$ and $v(t) = Ve^{\lambda t}$ we obtain the eigenvalue problem of the form

$$\begin{aligned}
\lambda W &= \frac{d\alpha}{dv}(\bar{v})\bar{w}V, && (42)\\
\lambda U(x) + \partial_x[g(x,\bar{v})U(x)] + \partial_x[\frac{\partial g}{\partial v}(x,\bar{v})\bar{u}(x)]V &= p(x)U(x), && (43)\\
U(0) &= W, && (44)\\
\lambda V = g(x^*,\bar{v})U(x^*) + \frac{\partial g}{\partial v}(x^*,\bar{v})\bar{u}(x^*)V &- \mu V. && (45)
\end{aligned}$$

Defining an auxiliary function $f(x)$ such that

$$f(x)e^{\int_0^x \frac{-p(s)}{g(s,\bar{v})}ds} = -\partial_x[\frac{\partial g}{\partial v}(x,\bar{v})\bar{u}(x)],$$

we obtain

$$\partial_x[g(x,\bar{v})U(x)e^{\int_0^x \frac{\lambda-p(s)}{g(s,\bar{v})}ds}] = f(x)Ve^{\int_0^x \frac{\lambda}{g(s,\bar{v})}ds},$$

$$g(x,\bar{v})U(x) = g(0,\bar{v})U(0)e^{-\int_0^x \frac{\lambda-p(s)}{g(s,\bar{v})}ds} + Ve^{-\int_0^x \frac{\lambda-p(s)}{g(s,\bar{v})}ds}\int_0^x f(s)e^{\int_0^s \frac{\lambda}{g(\sigma,\bar{v})}d\sigma}ds.$$

Hence, using (42) and (44) leads to

$$g(x^*,\bar{v})U(x^*) = \left(g(0,\bar{v})\frac{d\alpha}{dv}(\bar{v})\frac{\bar{w}}{\lambda} + \int_0^{x^*} f(s)e^{\int_0^s \frac{\lambda}{g(\sigma,\bar{v})}d\sigma}ds\right)Ve^{-\int_0^{x^*} \frac{\lambda-p(s)}{g(s,\bar{v})}ds}.$$

We insert this expression in Equation (45) and obtain the characteristic equation

$$\lambda + \mu - \frac{dg}{dv}(x^*,\bar{v})\bar{u}(x^*) = \left(g(0,\bar{v})\frac{d\alpha}{dv}(\bar{v})\frac{\bar{w}}{\lambda} + \int_0^{x^*} f(s)e^{\int_0^s \frac{\lambda}{g(\sigma,\bar{v})}d\sigma}ds\right)e^{-\int_0^{x^*} \frac{\lambda-p(s)}{g(s,\bar{v})}ds}. \quad (46)$$

The Simplest Case: g independent of v

We first focus on the simplest case when the maturation rate $g(x,v) = g(x)$ does not depend on v. In other words, the feedback loop only affects stem cells. Although this case is very restrictive compared to the original discrete model, since it does not include relation (6), it is an illustrative example of a possible general behavior. Instability and appearance of the oscillations in this model suggest that regulation of the processes solely by the stem cell level is not enough to stabilize the system. Moreover, regulatory feedback between mature cells and progenitor cells has a stabilizing effect and is essential for efficient regulation of the process.
Since we have $f = 0$, combining Equation (36) with Equation (46), we arrive at

$$\lambda^2 + \mu\lambda = \mu\bar{v}\frac{d\alpha}{dv}(\bar{v})e^{-\tau\lambda}, \qquad \tau = \int_0^{x^*} \frac{1}{g(s)}ds > 0. \tag{47}$$

The relationship is identical with the characteristic equation of a delay differential system. Indeed, problem (18)–(21) can be reformulated as a delay differential system. We obtain the following result.

Proposition 4.2 *Assume that Equations* (24)–(26) *hold,* $\alpha(0) > 0$*, and* g *is independent of* v*. Consider the steady state* $(\bar{u}, \bar{v}, \bar{w})$ *given in Lemma 4.1. Then,*
(i) for $1 < \tau\bar{v}\,|\frac{d\alpha}{dv}(\bar{v})| < \frac{\pi}{2}$*, the system undergoes a Hopf bifurcation for a single value* $\mu_0 > 0$ *of the parameter* μ*. Therefore the steady state can be either locally stable or unstable.*
(ii) Further bifurcations also occur for $\tau\bar{v}\,|\frac{d\alpha}{dv}(\bar{v})| > 2k\pi + \frac{\pi}{2}$ *and* $k \geq 1$ *for at least one value* $\mu_k > 0$*.*

Because in the special case at hand, the system can be reduced to a delay differential equation, the linearised stability implies the stability of the nonlinear system, which then undergoes a Hopf bifurcation for certain values of the parameters (see [8, 16] and the references therein).

Proof. (i) In order to identify the parameter values for which the bifurcation occurs, we look for purely imaginary solutions $\lambda = i\omega$ with $\omega \in \mathbb{R}$. We obtain the two following relations

$$\omega^2 = \mu\bar{v}\,|\frac{d\alpha}{dv}(\bar{v})|\,\cos(\tau\omega), \qquad \tau\omega = \tau\bar{v}\,|\frac{d\alpha}{dv}(\bar{v})|\,\sin(\tau\omega).$$

By symmetry, we only consider $\omega > 0$. The second relation gives a single value $\tau\omega_0 \in (0, \frac{\pi}{2})$ as soon as $1 < \tau\bar{v}\,|\frac{d\alpha}{dv}(\bar{v})| < \frac{\pi}{2}$. We can enforce the first relation for a single μ, because $\cos(\tau\omega_0) > 0$. This proves statement (i). (ii) For $\tau\bar{v}\,|\frac{d\alpha}{dv}(\bar{v})| > 2k\pi + \frac{\pi}{2}$ and $k \geq 1$, the equation $\tau\omega = \tau\bar{v}\,|\frac{d\alpha}{dv}(\bar{v})|\sin(\tau\omega)$ also has a root $\tau\omega_k \in (2k\pi, 2k\pi + \frac{\pi}{2}$, for which $\cos(\tau\omega_k) > 0$ and thus we can find again a μ_k for which the first equation is satisfied. But there might be multiple compatible crossings and several bifurcations are possible. □

We now proceed numerically using, for instance Matlab's device DDE BIFTOOL. We

check that for the values $\mu = \tau = 1$, we get stability for $\mu\bar{v}\frac{d\alpha}{dv}(\bar{v}) = -1$ and instability for $\mu\bar{v}\frac{d\alpha}{dv}(\bar{v}) = -2$.
This proposition as well as numerical simulations (see Figures 4 and 5) show that instability occurs through a Hopf bifurcation, and that regular oscillations appear.

A Case Motivated by the Discrete Model

In this section we will study more closely the case given by the relations (6). We can use the values of the steady state computed in Equations (34)–(36), keeping $g(x, v)$ and $p(x)$ fully general. It implies, denoting $\bar{g}(x) = g(x, \bar{v})$

$$\frac{d\alpha}{dv}(\bar{v}) = -\frac{2ka_w p_w}{(1+k\bar{v})^2} = -\frac{kp_w}{2a_w}, \qquad \bar{w} = \frac{\bar{g}(x^*)}{p_w}\bar{u}(x^*)exp\Big\{-\int_0^{x^*}\frac{p(\xi)}{\bar{g}(\xi)}d\xi\Big\}.$$

We can show (see the Appendix for detailed calculations) that Equation (46) can be written as

$$\lambda + \mu = \frac{\mu}{k}(2a_w - 1)\left(\frac{k}{2a_w}\Big(-\frac{p_w}{\lambda} + 1\Big) + \int_0^{x^*}\frac{\partial g}{\partial v}(x, \bar{v})\frac{\lambda - p(x)}{\bar{g}(x)^2}e^{\int_0^x \frac{\lambda}{\bar{g}(s)}ds}dx\right)e^{-\int_0^{x^*}\frac{\lambda}{\bar{g}(s)}ds}. \quad (48)$$

Case of $\alpha(v) = p_w(\frac{2a_w}{1+kv} - 1)$ and g independent of x

Since $g(v)$ is now independent of x, we have that for all x

$$g(x, v) = g(0, v) = p_w - \alpha(v) = 2p_w(1 - \frac{a_w}{1+kv}) \ .$$

In particular $g(\bar{v}) = p_w$ so $\frac{dg}{dv}(\bar{v}) = p_w\frac{k}{2a_w}$. We now substitute $g(x, \bar{v}) = g(\bar{v})$ in Equation (48) and obtain

$$\lambda + \mu = \mu\frac{2a_w - 1}{2a_w}\left(-\frac{p_w}{\lambda} + 1 + \int_0^{x^*}\frac{\lambda - p(x)}{p_w}e^{\int_0^x \frac{\lambda}{p_w}ds}dx\right)e^{-\int_0^{x^*}\frac{\lambda}{p_w}ds}.$$

We compute the first part of the integral term: $\int_0^{x^*}\frac{\lambda}{p_w}e^{\int_0^x \frac{\lambda}{p_w}ds}dx = e^{\frac{\lambda}{p_w}x^*} - 1$, and so

$$\lambda + \frac{\mu}{2a_w} = -\frac{\mu}{2a_w}(2a_w - 1)\left(\frac{p_w}{\lambda} + \int_0^{x^*}\frac{p(x)}{p_w}e^{\frac{\lambda}{p_w}x}dx\right)e^{-\frac{\lambda}{p_w}x^*}. \quad (49)$$

Proposition 4.3 *Let $\alpha(v)$ be defined by (6) with $a_w > \frac{1}{2}$, and $(\bar{u}, \bar{v}, \bar{w})$ be defined by Equations (34)–(36) the unique steady state solution of System (18)–(21). If the maturation rate $g(x, v)$ is independent of the maturity of the cell x and if the proliferation rate p is constant, then the steady state $(\bar{u}, \bar{v}, \bar{w})$ is locally linearly stable. For a non-decreasing proliferation rate, instability may appear.*

We treat a case of non-decreasing proliferation rate because it is the most biologically relevant; however instability may appear for other cases, and is even easier to exhibit, as the proof (postponed to the Appendix) shows. Figures 2 and 3 below illustrate a case of instability with a nondecreasing proliferation rate.

5 Numerical Simulations

In this section we illustrate our theoretical results with a number of numerical simulations. We start with a description of our numerical methods.

5.1 The Numerical Scheme

We build a simple numerical scheme for System (18)–(21). We discretize the problem on a grid regular in space and adaptive in time. We denote by $\Delta t^k = t^{k+1} - t^k$ the time step between time t^{k+1} and time t^k, by $\Delta x = x^*/I$ the spatial step, where I denotes the number of points: $x_i = i\Delta x$, $0 \leqslant i \leqslant I$.

We use an explicit upwind finite volume method for u

$$u_i^k = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(t^k, y)dy, \qquad \frac{1}{\Delta t^k} \int_0^{\Delta t^k} u(t^k + s, x_{i+\frac{1}{2}})ds \approx u_i^k.$$

For time discretization, we use a marching technique. At each time t^k, we choose the time step Δt^k so as to satisfy the largest possible CFL stability criterion

$$\theta := g\frac{\Delta t^k}{\Delta x} \leq 1,$$

so that

$$\Delta t^k = \frac{\Delta x}{Max_x g(x, v^k)}.$$

In order to avoid a vanishing time step, it is necessary here to suppose $g \in L^\infty$. Also, more efficient schemes (of WENO type for instance, see [24, 22]) could be used to capture discontinuities of g.

The algorithm is the following:

- **Initialization** We use the initial data

$$w^0 = w_0, \qquad u_j^0 = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u_0(y)dy, \qquad v^0 = v_0.$$

- **From t^k to t^{k+1} :**

- We calculate $\alpha^k = \alpha(v^k)$ and define $w^{k+1} = (1 + \Delta t^k \alpha^k) w^k$.
- We calculate $\Delta t^k = \frac{\Delta x}{Max_i g(x_i, v^k)}$ and define $t^{k+1} = t^k + \Delta t^k$.
- For a boundary condition at $i = 0$, we define $u_0^{k+1} = w^{k+1}$.
- We define u_i^{k+1} by the following scheme

$$\frac{u_j^{k+1} - u_j^k}{\Delta t^k} + \frac{g(x_j, v^k) u_j^k - g(x_{j-1}, v^k) u_{j-1}^k}{\Delta x} = p_j u_j^k.$$

- We define v^{k+1} by

$$\frac{v^{k+1} - v^k}{\Delta t^k} = g(x_I, v^k) u_I^k - \mu v^{k+1},$$

and the term μv^{k+1}in the right hand side is discretized implicitely for stability. The reason for the choice of u_I^k instead of u_I^{k+1} in the right-hand side of this last scheme is due to cell number balance considerations as shown below.

- **Cell number balance.** From Equation (18)–(21) we have obtained the cell number balance (23). We check the equivalent discrete mass balance:

$$\frac{w^{k+1} - w^k}{\Delta t^k} + \sum_{j=0}^{I} \frac{u_j^{k+1} - u_j^k}{\Delta t^k} + \frac{v^{k+1} - v^k}{\Delta t^k} =$$

$$\big(\alpha^k + g(x_0, v^k)\big) w^k + \sum_{i=0}^{I} p_j u_j^k \Delta x - \mu v^{k+1}.$$

5.2 Numerical Simulations

First we compare results of the numerical simulations of the discrete and the continuous models. To do so, we depart from the discrete values of parameters given in [25]. The notations are those of System (7)-(9), with $g_i(s, u_i) = 2[1 - a_i(s)]p_i u_i$, p_i independent of s, $d_i = 0$ for $i < n$ and $a_i(s) = \frac{a_i}{1 + k u_n}$. It corresponds to the model 1 studied in [17, 18].

Parameter	Value	Parameter	Value	Parameter	Value
a_1	0.77	p_1	$2.15\dot{1}0^{-3}$ day^{-1}	d_8	0.6925 day^{-1}
a_2	0.7689	p_2	$11.2\dot{1}10^{-3}$ day^{-1}	k	$12.8.10^{-10}$
a_3	0.7359	p_3	$5.66\dot{1}0^{-2}$ day^{-1}		
a_4	0.7678	p_4	0.1586 day^{-1}		
a_5	0.154	p_5	0.32 day^{-1}		
a_6	0.11	p_6	0.7 day^{-1}		
a_7	0.605	p_7	1 day^{-1}		

To make comparison easier, for the continuous maturation model, we replace the interval $[0, x^*]$ by the interval $[1, 7]$ (7 being the number of maturing steps in the discrete model) and we define $a(x)$ and $p(x)$ based on parameters in Table 5.2 by piecewise linear continuous functions with values a_i and p_i at $x = i$. We take them along a regular grid to obtain approximations of $a(x)$ and $p(x)$.

In Figure 1, the results of the discrete model are identical with the ones of the continuous model if the grid is equal to $X = [1, 2, ...7]$ (case $I = 6$). If the grid becomes finer, we observe a slower convergence toward the steady state together with an increase of the relative importance of the stem cell population. Though unrealistic from a biological viewpoint, it was expected by the derivation of the continuous model from the discrete one. It shows that the analogy between the two models is limited. They exhibit different quantitative properties (see [18] for a study of the discrete model properties), as well as conditions for nontrivial steady state. Moreover, we see that the typical parameter sizes have to be adapted. Indeed, the time evolution is much too slow compared to experimental data.

Let us now focus on the stability and instability properties, in order to illustrate the theoretical results of Propositions 4.2 and 4.3.

Figures 2 and 3 are an illustration of the instability case stated in Proposition 4.3. Here, we took a maturity interval $[0, X^*]$ with $X^* = 50$, and a proliferation rate $p(x) = p_w + B\chi_{x \geq Y^*}$ with $p_w = 30$, $Y^* = 20$, $B = 50$. We keep $a(x)$ constant equal to $a_w = 0.75$ and $k = 1.28x10^{-9}$ as in the discrete case. The maturation speed $g(x)$ is given by Equation (6). We see that the destabilization is very slow, and our example is very unrealistic, since the stem cell population level is tiny.

In Figures 4 and 5, we illustrate instability in the case of Proposition 4.2. We have taken here $X^* = 1$, constant proliferation rate $p(x) = p_w = 6$ and maturation rate $g(x) = 1$ and $\alpha(v) = (\frac{2a_w}{1+kv} - 1)p_w$ with $a_w = 0.75$ and $k = 1.28x10^{-9}$.

6 Final Remarks

In this paper we have developed a structured population model of cell differentiation and self-renewal with a nonlinear regulatory feedback between the level of mature cells and the rate of the maturation process. We showed that perturbations in the regulatory mechanism may lead to the destabilization of the positive steady state, which corresponds to the healthy state of the tissue. In particular, we showed that the regulation of stem cells self-renewal is not sufficient for stability of the system and the lack of the regulation on the level of progenitor cells may lead to the persistent oscillations. This and other stability results suggest how imbalanced regulation of cell self-renewal and differentiation may lead to the destabilization of the system, which is observed during development of some cancers, such as leukemias. The model developed in this paper is rather general and, after adjusting it to specific biological assumptions, may serve as a tool to explore the role of different regulatory mechanisms in the normal and pathological development.

Comparing the model to its discrete counterpart we addressed the question of the choice

of the right class of models, for example discrete compartments versus continuous maturation, punctuated by division events. We showed that the models may exhibit different dynamics. Interestingly, the structure of steady states varies and the discrete compartmental model admits semi-trivial steady states of the form $(0, .., 0, \bar{u}_i, .., \bar{u}_n)$, which do not exist in the continuous differentiation model.

To understand the difference between the two models, we derived a limit equation for the discrete model assuming that a continuum of differentiation stages can be defined. The rationale for such assumption is provided by the fact that differentiation is controlled by intracellular biochemical processes, which are indeed continuous in time, at least when averaged over a large number of cells. Consequently, for the proper time scaling we have to assume that commitment and maturation of cell progenitors do not proceed by the division clock (one division = one step in the maturation process) but is a continuous process and can take place between the divisions. This observation explains the fundamental difference between the two models. The structured population model is indeed not a limit of the discrete model with the transitions between compartments correlated to the division of the cells. However, the models can exhibit exactly the same dynamics for a suitable choice of the maturation rate function g.

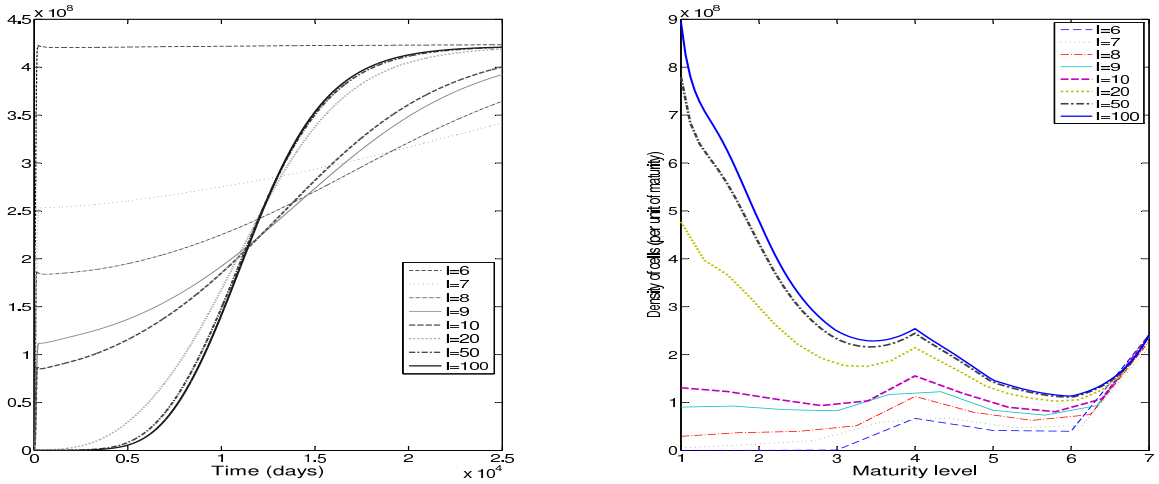

Figure 1: Comparison of numerical simulations using different grids on the interval $[1, 7]$, from $I = 6$ (7 points, maturity step $dx = 1$, discrete model) to $I = 100$. Left: mature cells evolution with time. Right: distribution of cell density along the maturation level, at steady state. One can see that the model is extremely sensitive to the number of steps (even 7 to 10): small numbers seem to be unstable, whereas for large numbers the numerical scheme converges.

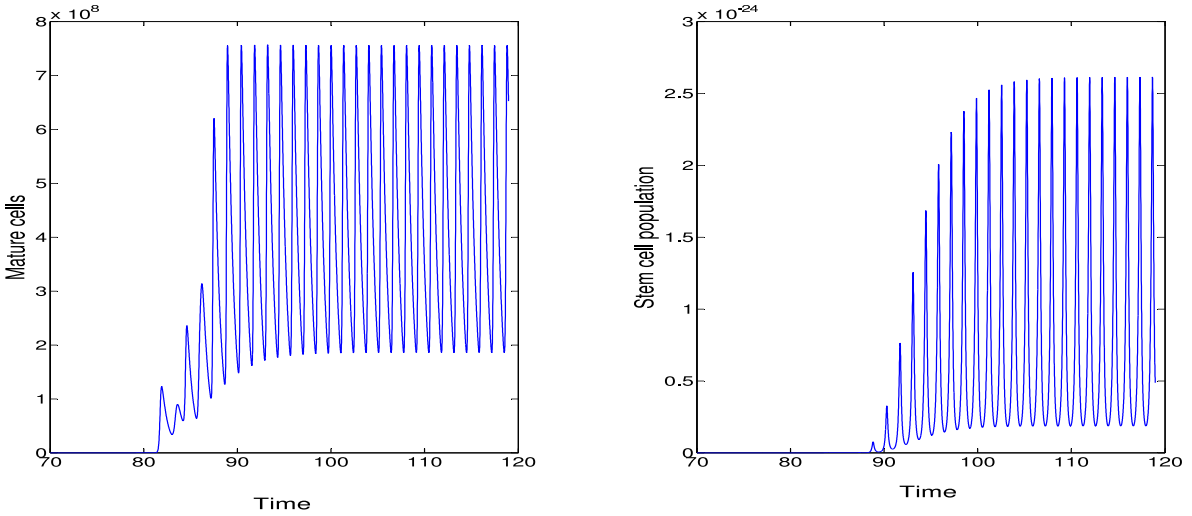

Figure 2: Example of instability, in illustration of Proposition 4.3. Left: evolution of mature cells. Right: evolution of stem cells.

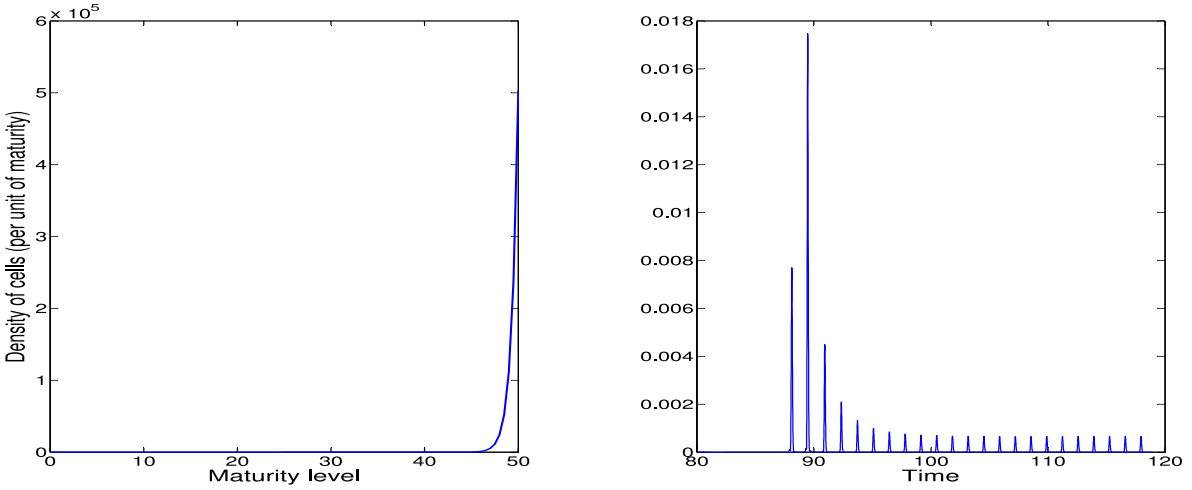

Figure 3: Same case as in Figure 2. Left: final distribution of cells according to their maturity level. Right: time evolution of $\sqrt{\frac{\int |\frac{\partial}{\partial t} u(x,t)|^2 dx}{\int |u(x,t)|^2 dx}}$, to measure the trend to a stable maturity level distribution.

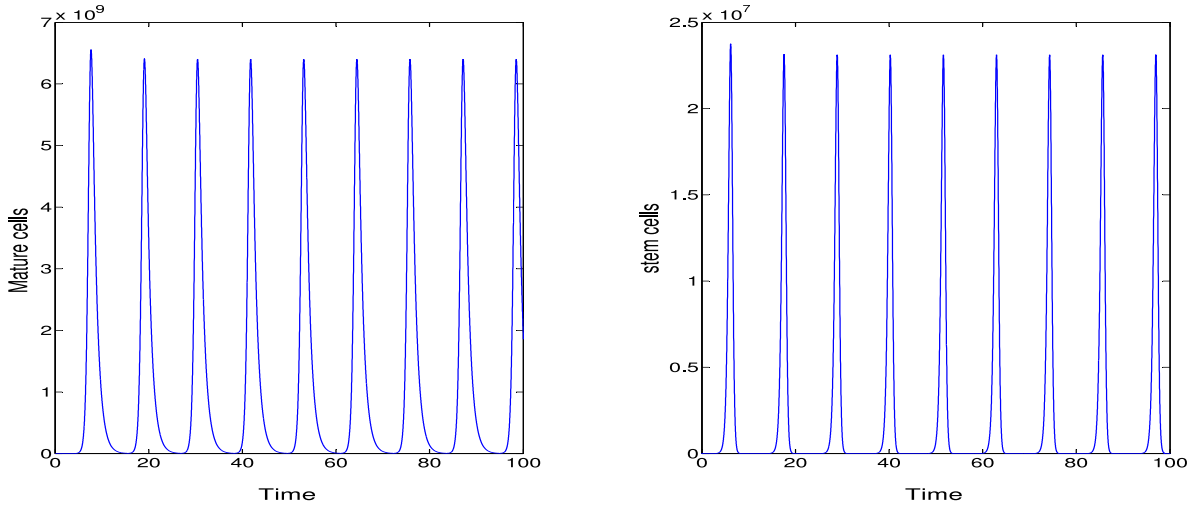

Figure 4: Example of instability, in illustration of Proposition 4.2. Left: evolution of mature cells. Right: evolution of stem cells.

Acknowledgments

The authors were supported by the CNPq-INRIA agreement INVEBIO. JPZ was supported by CNPq under grants 302161/2003-1 and 474085/2003-1. AM-C was supported by ERC Starting Grant "Biostruct", Emmy Noether Programme of German Research Council (DFG) and WIN Kolleg of Heidelberg Academy of Sciences and Humanities. JPZ and BP are thankful to the RICAM Special Semester on Quantitative Biology Analyzed by Mathematical Methods, October 1st, 2007 -January 27th, 2008, organized by RICAM, Austrian Academy of Sciences, to the MathAmSud Programm NAPDE and to the International Cooperation Agreement Brazil-France.

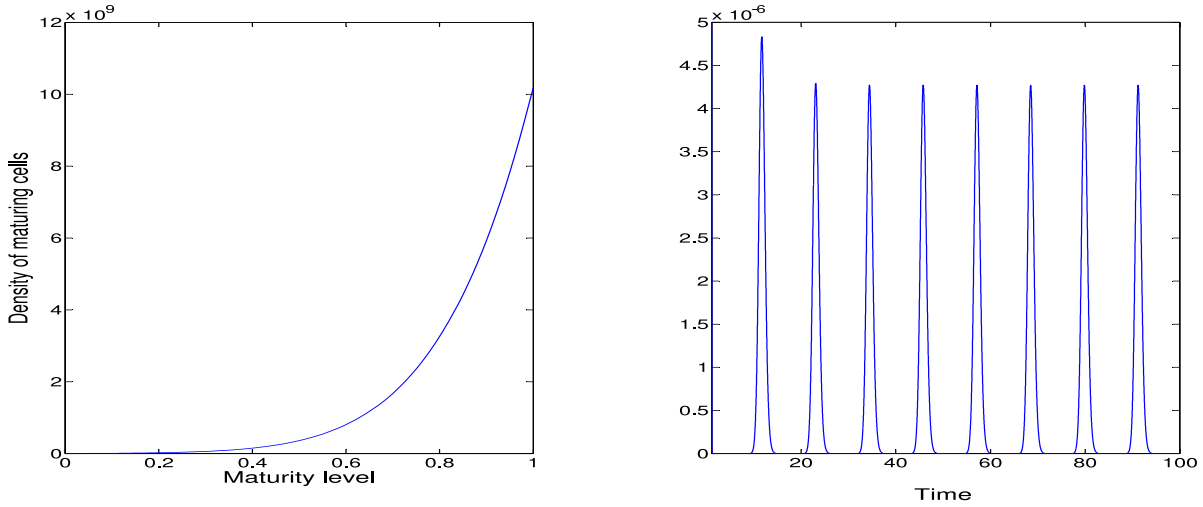

Figure 5: Same case as in Figure 4. Left: final distribution of cells according to their maturity level. Right: time evolution of $\sqrt{\frac{\int |\frac{\partial}{\partial t}u(x,t)|^2 dx}{\int |u(x,t)|^2 dx}}$, to measure the trend to a stable maturity level distribution.

A Appendix: Proofs of the Results in Section 4.2

A.1 The Characteristic Equation in the Case Derived from the Discrete Model

Using definitions given in (6), Equation (46) can be rewritten as

$$\lambda+\mu-\frac{dg}{dv}(x^*,\bar{v})\bar{u}(x^*)=\left(-\frac{k}{2a_w}p_w\frac{\bar{g}(x^*)}{\lambda}\bar{u}(x^*)+e^{\int_0^{x^*}\frac{p(\xi)}{\bar{g}(\xi)}d\xi}\int_0^{x^*}f(s)e^{\int_0^s\frac{\lambda}{g(\sigma,\bar{v})}d\sigma}ds\right)e^{-\int_0^{x^*}\frac{\lambda}{g(s,\bar{v})}ds}. \tag{50}$$

Calculating the term with f we obtain

$$\int_0^{x^*}f(x)e^{\int_0^x\frac{\lambda}{g(s,\bar{v})}ds}dx=\int_0^{x^*}-\partial_x[\frac{\partial g}{\partial v}(x,\bar{v})\bar{u}(x)]e^{\int_0^x\frac{\lambda-p(s)}{g(s,\bar{v})}ds}dx$$

$$=+\int_0^{x^*}\frac{\partial g}{\partial v}(x,\bar{v})\bar{u}(x)\partial_x[e^{\int_0^x\frac{\lambda-p(s)}{g(s,\bar{v})}ds}]dx-\frac{\partial g}{\partial v}(x^*,\bar{v})\bar{u}(x^*)e^{\int_0^{x^*}\frac{\lambda-p(s)}{g(\sigma,\bar{v})}ds}+\frac{\partial g}{\partial v}(0,\bar{v})\bar{u}(0)$$

$$=+\int_0^{x^*}\frac{\partial g}{\partial v}(x,\bar{v})\bar{u}(x)\frac{\lambda-p(x)}{g(x,\bar{v})}e^{\int_0^x\frac{\lambda-p(s)}{g(s,\bar{v})}ds}dx-\frac{\partial g}{\partial v}(x^*,\bar{v})\bar{u}(x^*)e^{\int_0^{x^*}\frac{\lambda-p(s)}{g(\sigma,\bar{v})}ds}+\frac{k}{2a_w}p_w\bar{w}.$$

Inserting it into Equation (50) leads to

$$\lambda+\mu = \left(-\frac{k}{2a_w}p_w\frac{\bar{g}(x^*)}{\lambda}\bar{u}(x^*)+e^{\int_0^{x^*}\frac{p(\xi)}{\bar{g}(\xi)}d\xi}\Big(\int_0^{x^*}\frac{\partial g}{\partial v}(x,\bar{v})\bar{u}(x)\frac{\lambda-p(x)}{\bar{g}(x)}e^{\int_0^x\frac{\lambda-p(s)}{\bar{g}(s)}ds}dx+\frac{k}{2a_w}p_w\bar{w}\Big)\right)e^{-\int_0^{x^*}\frac{\lambda}{\bar{g}(s)}ds}.$$

Due to the definition of $\bar{w}$ given by Equation (36), the first and the last term of the right-hand side can be written together as

$$\lambda+\mu = \left(\frac{k}{2a_w}\bar{g}(x^*)\bar{u}(x^*)\Big(-\frac{p_w}{\lambda}+1\Big)+e^{\int_0^{x^*}\frac{p(\xi)}{\bar{g}(\xi)}d\xi}\Big(\int_0^{x^*}\frac{\partial g}{\partial v}(x,\bar{v})\bar{u}(x)\frac{\lambda-p(x)}{\bar{g}(x)}e^{\int_0^x\frac{\lambda-p(s)}{\bar{g}(s)}ds}dx\Big)\right)e^{-\int_0^{x^*}\frac{\lambda}{\bar{g}(s)}ds}.$$

Using Equation (35) we obtain

$$\bar{u}(x)\frac{\lambda-p(x)}{\bar{g}(x)}e^{\int_0^x\frac{\lambda-p(s)}{\bar{g}(s)}ds} = \bar{g}(x^*)\bar{u}(x^*)e^{-\int_0^{x^*}\frac{p(s)}{\bar{g}(s)}ds}\frac{\lambda-p(x)}{\bar{g}(x)^2}e^{\int_0^x\frac{\lambda}{\bar{g}(s)}ds}.$$

So finally, using Equations (33) and (34) and taking $\bar{g}(x^*)\bar{u}(x^*) = \frac{\mu}{k}(2a_w-1)$, we obtain the expression given by Equation (48).

Proof of Proposition 4.3

The local linear stability is equivalent to the fact that all eigenvalues $\lambda \in \mathbb{C}$, given by solutions of Equation (49), have negative real parts.
First step: λ is a solution of the following equation

$$\lambda^2 + C\lambda + D = -\frac{\mu}{2a_w}(2a_w-1)\lambda\int_0^{x^*}b(x)e^{\frac{\lambda}{p_w}(x-x^*)}dx \tag{51}$$

with $C = \frac{\mu}{2a_w} > 0$, $D = p_w\mu\frac{2a_w-1}{2a_w} > 0$, and $b(x) = \frac{p(x)-p_w}{p_w} \geq 0$ a non-decreasing function. Indeed, with this definition of $b(x)$ we rewrite Equation (49) in the form

$$\lambda + \frac{\mu}{2a_w} = -\frac{\mu}{2a_w}(2a_w-1)\left(\frac{p_w}{\lambda}+\int_0^{x^*}(1+b(x))e^{\frac{\lambda}{p_w}x}dx\right)e^{-\frac{\lambda}{p_w}x^*}.$$

Integrating by parts we obtain $\frac{p_w}{\lambda}+\int_0^{x^*}e^{\frac{\lambda}{p_w}x}dx = \frac{p_w}{\lambda}e^{\frac{\lambda}{p_w}x^*}$.

Second step: the limiting case is for $b(x) = 0$, *i.e.*, p independent of x. In this case, the eigenvalues are given by

$$\lambda_\pm = \frac{-C \pm \sqrt{C^2-4D}}{2}.$$

For $C^2-4D>0$ these two eigenvalues are negative. If $C^2-4D<0$, they are complex conjugated with negative real parts. In any case, the steady state is locally linearly stable and the first part of the proposition is proved.
Third step: In the general case, in order to study the sign of the real part of the eigenvalues λ, we look for values of the parameters such that $\lambda = i\omega$ with $\omega \in \mathbb{R}$. It corresponds to a Hopf bifurcation and it leads to:

$$-\omega^2+iC\omega+D = -\frac{\mu}{2a_w}(2a_w-1)i\omega\int_0^{x^*} b(x)\cos(\frac{\omega}{p_w}(x-x^*))dx+\frac{\mu}{2a_w}(2a_w-1)\omega\int_0^{x^*} b(x)\sin(\frac{\omega}{p_w}(x-x^*))dx.$$

Taking the imaginary part of this equation yields, since $\omega \neq 0$:

$$C = -\frac{\mu}{2a_w}(2a_w-1)\int_0^{x^*} b(x)\cos(\frac{\omega}{p_w}(x-x^*))dx = -\frac{\mu}{2a_w}(2a_w-1)\int_0^{x^*} b(x^*-y)\cos(\frac{\omega}{p_w}y)dy$$

Since b is increasing, $b(x^*-\cdot)$ is decreasing. This leads to

$$\int_0^{Min(x^*,\frac{p_w}{\omega}\pi)} b(x^*-y)\cos(\frac{\omega}{p_w}y)dy \geq 0.$$

Indeed, either $x^* \leq \frac{\pi}{2}\frac{p_w}{\omega}$, in which case it is evident because $b(x^*-y)\cos(\frac{\omega}{p_w}y)\geq 0$ for all $0\leq x\leq x^*$, or $x^*\geq \frac{\pi}{2}\frac{p_w}{\omega}$ and we can write

$$\int_0^{Min(x^*,\frac{p_w}{\omega}\pi)} b(x^*-y)\cos(\frac{\omega}{p_w}y)dy \geq b(x^*-\frac{\pi}{2}\frac{p_w}{\omega})\int_0^{\frac{\pi}{2}\frac{p_w}{\omega}}\cos(\frac{\omega}{p_w}y)dy - b(x^*-\frac{\pi}{2}\frac{p_w}{\omega})|\int_{\frac{\pi}{2}\frac{p_w}{\omega}}^{\frac{p_w}{\omega}\pi}\cos(\frac{\omega}{p_w}y)dy = 0.$$

To end the proof, let us simply exhibit an example where instability can occur: Let χ denote the Heaviside function. Defining

$$b(x) = B\chi_{y^*\leq x\leq x^*},$$

we compute explicitely

$$-\omega^2+i\frac{\mu\omega}{2a_w}+p_w\mu\frac{2a_w-1}{2a_w} = -p_w\mu\frac{2a_w-1}{2a_w}iB\sin(\frac{\omega}{p_w}(x^*-y^*))+p_w\mu\frac{2a_w-1}{2a_w}B\big(\cos(\frac{\omega}{p_w}(x^*-y^*))-1\big).$$

It provides two relations

$$\frac{\omega}{p_w} = -(2a_w-1)B\sin(\frac{\omega}{p_w}(x^*-y^*)),\quad \frac{\omega^2}{p_w^2}\frac{2a_w}{2a_w-1} = \frac{\mu}{p_w}\left(1+B\big(1-\cos(\frac{\omega}{p_w}(x^*-y^*))\big)\right).$$

We can see that there exist many sets of parameters such that both relations are satisfied. Given a_w, B, p_w, and ω such that

$$\frac{\omega}{p_w(2a_w - 1)B} \leq 1,$$

we can always find $x^* - y^*$ such that the first relation is satisfied, and then fix μ using the second relation. □

References

[1] M. Adimy, F. Crauste and L. Pujo-Menjouet, *On the stability of a maturity structured model of cellular proliferation,* Dis. Cont. Dyn. Sys. Ser. A, 12 (3): 501–522, 2005.

[2] M. Adimy, F. Crauste and S. Ruan, *A mathematical study of the hematopoiesis process with applications to chronic myelogenous leukemia*, SIAM J. Appl. Math., 65 (4): 1328–1352, 2005.

[3] M. Al-Hajj and M.F. Clarke, *Self-renewal and solid tumor stem cells,* Oncogene 23: 7274-7282, 2004.

[4] P.A. Beachy, S.S. Karhadkar and D.M. Berman, *Tissue repair and stem cell renewal in carcinogenesis,* Nature 432: 324-331, 2004.

[5] J. Belair, M.C. Mackey and J. Mahaffy, *Age structured and two delay models for erythropoiesis*, Math. Biosci. 128: 317–346, 1995.

[6] C. Colijn and M.C. Mackey, *A mathematical model of hematopoiesis–I. Periodic chronic myelogenous leukemia*, J. Theor. Biol. 237: 117–132, 2005.

[7] J.-F. Collet, T. Goudon, F. Poupaud and A. Vasseur, *The Becker–Döring System and Its Lifshitz–Slyozov Limit*, SIAM J. on Appl. Math. 62 (5): 1488–1500, 2002.

[8] O. Diekmann, S.A van Gils, S.M. Verduyn-Lunel and H.C. Walther, *Delay equations, Functional-, Complex and Nonlinear Analysis*, ApMS, 110, NY, 1995.

[9] G. Dontu, M. Al-Hajj, W.M. Abdallah, M.F. Clarke and M.S. Wicha, *Stem cells in normal breast development and breast cancer*, Cell Prolif. 36: 59-72, 2003.

[10] M. Doumic, L. Goudon and T. Lepoutre, *Scaling limit of a discrete prion dynamics model*, Comm. in Math. Sc. 7(4): 839–865, 2009.

[11] M. Doumic, P. Kim, B. Perthame *Stability Analysis of a Simplified Yet Complete Model for Chronic Myelogenous Leukemia*, Bull. of Math. Biol. (doi: 10.1007/s11538-009-9500-0).

[12] S. He, D. Nakada, S.J. Morrison, *Mechanisms of stem cell self-renewal*, Annu. Rev. Cell. Dev. Biol. 25: 377-406, 2009.

[13] K.J. Hope, L. Jin and J.E. Dick, *Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity.* Nat Immunol. 5: 738-743, 2004.

[14] J. Klaus, D. Herrmann, I. Breitkreutz, U. Hegenbart, U. Mazitschek, G. Egerer, F.W. Cremer, R.M. Lowenthal, J. Huesing, S. Fruehauf, T. Moehler, A.D. Ho,and H. Goldschmidt, *Effect of cd34 cell dose on hematopoietic reconstitution and outcome in 508 patients with multiple myeloma undergoing autologous peripheral blood stem cell transplantation,* Eur J Haematol. 78(1): 21–28, 2007.

[15] B.I. Lord, *Biology of the haemopoietic stem cell*, In C. S. Potten, editor, Stem cells: 401–422. Academic Press, Cambridge, 1997.

[16] P. Magal and S. Ruan, *Center Manifolds for Semilinear Equations with Non-dense Domain and Applications on Hopf Bifurcation in Age Structured Models*, Memoirs of the AMS. 202(951), 2009.

[17] A. Marciniak-Czochra, T. Stiehl, W. Jäger, A. Ho and W. Wagner, *Modeling of asymmetric cell division in hematopoietic stem cells - regulation of self-renewal is essential for efficient repopulation,* Stem Cells Dev. 18(3): 377-385, 2009.

[18] A. Marciniak-Czochra, T. Stiehl, *Characterization of stem cells using mathematical models of multistage cell lineages,* Math. Comp. Models. (doi: 10.1016/ j.mcm.2010.03.057).

[19] K.A. Moore and I.R. Lemischka *Stem cells and their niches,* Science 311: 1880-1885, 2006.

[20] S.J. Morrison and J. Kimble, *Asymmetric and symmetric stem cell divisions in development and cancer*, Nature 441: 1068-1074, 2006.

[21] T. Reya, S.J. Morrison, M.F. Clarke and I.L. Weissman, *Stem cells, cancer, cancer stem cells*, Nature 414: 105-11 2001.

[22] G. Shan Jiang and D. Peng, *Weighted eno schemes for Hamilton-Jacobi equations*, SIAM J. Sci. Comput, 21: 2126-2143, 1997.

[23] T. Schroeder *Asymmetric Cell Division in Normal and Malignant Hematopoietic Precursor Cells* Cell Stem Cell 1: 479-481, 2007.

[24] C.-W. Shu, *Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws,* Springer: 325-432, 1998.

[25] T. Stiehl. Diploma thesis, University of Heidelberg, 2009.

[26] I. Thornley, D.R. Sutherland, R. Nayar, L. Sung, M.H. Freedman, and H.A. Messner, *Replicative stress after allogeneic bone marrow transplantation: changes in cycling of cd34+cd90+ and cd34+cd90- hematopoietic progenitors,* Blood, 97(6): 18768, 2001.

[27] N. Uchida, W.H. Fleming, E.J. Alpern, and I.L. Weissman, *Heterogeneity of hematopoietic stem cells,* Curr. Opin. Immunol., 5(2): 177184, 1993.

[28] I.L. Weissman, *Stem cells: units of development, units of regeneration, and units in evolution*, Cell 100(1): 157-68, 2000. PMID: 10647940.

3 Inverse Problem [15]

Numerical Solution of an Inverse Problem in Size-Structured Population Dynamics

Marie Doumic [*‡] Benoît Perthame[†‡] Jorge Zubelli[§]

June 26, 2008

Abstract

We consider a size-structured model for cell division and address the question of determining the division (birth) rate from the measured stable size distribution of the population. We propose a new regularization technique based on a filtering approach. We prove convergence of the algorithm and validate the theoretical results by implementing numerical simulations, based on classical techniques. We compare the results for direct and inverse problems, for the filtering method and for the quasi-reversibility method proposed in [1].

1 Introduction

The use of size-structured models to describe biological systems has attracted the interest of many authors and has a long standing tradition. In particular, the use of size structures was very well documented and compared to experiments in the 70's. This led to the survey book [2] and subsequent mathematical analysis (see also the references in [3]). Needless to say, in such models it is crucial for the analysis, computer simulation and prediction to calibrate the corresponding model parameters so as to obtain good quantitative results. Indeed, in the inverse problem literature, a number of authors have addressed the calibration of certain structured population models. See for example [4, 5, 6, 7] and references therein.

In this article, we consider theoretical and numerical aspects of the inverse problem of determining the division rate coefficient $B = B(x)$ in the following specific size-structured model for cell division:

$$\begin{cases} \frac{\partial}{\partial t} n(t,x) + \frac{\partial}{\partial x} n(t,x) + B(x) n(t,x) = 4B(2x) n(t,2x), \qquad x \geqslant 0,\ t \geqslant 0, \\ \\ n(t, x=0) = 0, t > 0, \\ \\ n(0,x) = n^0(x) \geq 0. \end{cases} \tag{1}$$

[*] Département de Mathématiques et Applications, École Normale Supérieure, INRIA projet BANG, 45 rue d'Ulm, F 75230 Paris cedex 05, France; email: doumic@dma.ens.fr

[†] Université Pierre et Marie Curie-Paris 6, UMR 7598 LJLL, BC187, 4, place Jussieu, F-75252 Paris cedex 5, and Institut Universitaire de France; email: perthame@ann.jussieu.fr

[‡]INRIA Rocquencourt, projet BANG, Domaine de Voluceau, BP 105, 781153 Rocquencourt, France; emails: marie.doumic@inria.fr, benoit.perthame@inria.fr

[§]IMPA, Est. D. Castorina 110, Rio de Janeiro, RJ 22460-320 Brazil; email: zubelli@impa.br

Here, the cell density is represented by $n(t,x)$ at time t and size x. The division rate B expresses the division of cells of size $2x$ into two cells of size x.

By making use of flux cytometry technologies for instance, it is possible to determine cell populations with certain properties as protein content on a large scale of tenths of thousands of cells. In other applications, like coagulation fragmentation equation [8, 9, 10, 11, 12], or prion aggregation and fragmentation [13, 14, 15], similar equations arise, and much less is known on aggregate size repartition. The division rate $B(x)$, on the contrary, is not directly measurable.

The long time behavior of solutions is well known. Indeed, it was proved in [16, 17] that under fairly general conditions on the coefficients, there is a unique solution (N, λ_0) to the following eigenvalue problem

$$\begin{cases} \frac{\partial}{\partial x}N + (\lambda_0 + B(x))N = 4B(2x)N(2x), \qquad x \geqslant 0, \\ \\ N(x=0) = 0, \\ \\ N(x) > 0 \ \text{ for } x > 0, \qquad \int_0^\infty N(x)dx = 1, \end{cases} \tag{2}$$

where $\lambda_0 > 0$ and $Ne^{\mu x} \in L^\infty \cap L^1$ for all $\mu < \lambda_0$.

It was shown in [18, 16]

$$n(t,x)e^{-\lambda_0 t} \xrightarrow[t\to\infty]{} m_0 N(x), \quad \text{ in } \ L^1(\mathbb{R}_+, \phi(x)dx),$$

where the weight ϕ is the unique solution to the adjoint problem

$$\begin{cases} -\frac{\partial}{\partial x}\phi + (\lambda_0 + B(x))\phi = 2B(x)\phi(\frac{x}{2}), \qquad x \geqslant 0, \\ \\ \phi(x) > 0, \qquad \int_0^\infty \phi(x)N(x)dx = 1. \end{cases} \tag{3}$$

In other words, λ_0 is the growth rate of such a system and is usually called "Malthus parameter" in population biology. From [18, 16, 3] we also know that λ_0 is related to N by the relation

$$\lambda_0 = \frac{\int_0^\infty N dx}{\int_0^\infty xN dx}. \tag{4}$$

The question we address here is the following: How can we estimate the division rate B from the knowledge of the steady dynamics N and λ_0 ? The inverse problem thus consists of finding B a solution to

$$4B(2x)N(2x) - B(x)N(x) = L(x) := \frac{\partial}{\partial x}N(x) + \lambda_0 N(x), \quad x \geqslant 0, \tag{5}$$

assuming that (N, λ_0) is known, or thanks to (4) that N is known. As seen in [1], this problem is well-posed if N satisfies strong regularity properties such as $\frac{\partial}{\partial x}N(x) \in L^p(\mathbb{R}_+)$ for some $p \geqslant 1$.

However, in practical applications we have only an *approximate* knowledge of (N, λ_0), given by noisy data $(N_\varepsilon, \lambda_\varepsilon)$, with $N_\varepsilon \in L^2_+(\mathbb{R}_+)$ for instance. [1] This means that we have no way of controlling $\frac{\partial}{\partial x}N_\varepsilon$, so we cannot control the precision of a solution B_ε to problem (5) when a perturbed N_ε replaces N. Furthermore, it is not even clear whether such a B_ε exists.

[1] Actually, our knowledge of λ_0 is presumably an order of precision higher than that of N, since the rate λ_0 can be estimated independently by means of time information.

The question we focus on is then: How to approximate the problem (5) in order to get a solution B_ε as close as possible to the exact division rate B?

We remark that, in the context of noisy data, the inverse problem under consideration is ill-posed [1] and thus regularization would be required. A natural tool to be invoked from the inverse problem literature would be some kind of Tikhonov regularization method [19, 20]. However, this would lead to computationally intensive problems. Indeed, for each forward problem evaluation a dilation-differential equation of the form (2) would have to be solved.

In [1], two of the present authors proposed a method of regularization consisting in the solution of the following approximate problem:

$$\begin{cases} \alpha\frac{\partial}{\partial y}(B_{\varepsilon,\alpha}N_\varepsilon) + 4B_{\varepsilon,\alpha}(y)N_\varepsilon(y) = B_{\varepsilon,\alpha}\left(\frac{y}{2}\right)N_\varepsilon\left(\frac{y}{2}\right) + \lambda_0 N_\varepsilon\left(\frac{y}{2}\right) + 2\frac{\partial}{\partial y}\left(N_\varepsilon\left(\frac{y}{2}\right)\right), \qquad y>0, \\ \\ (B_{\varepsilon,\alpha}N_\varepsilon)(0) = 0, \end{cases}$$

where α is a regularizing parameter. It was shown that a convergence rate of order $\sqrt{\varepsilon}$ could be obtained, for $\alpha = O(\sqrt{\varepsilon})$, where ε is the error on the data N in an appropriate norm.

The above method of the solution to the inverse problem will be called *quasi reversibility* in accordance with the general spirit of the terminology of [21, 22]. The main goal of this work is to investigate the numerics of such approach, to consider an alternative technique based on filtering ideas and to compare the performance of the different methods. The alternative technique is also analyzed from the theoretical point of view and estimates are presented.

In this work, we have modified slightly the original regularization equation by writing $\lambda_{\varepsilon,\alpha}$ instead of λ_0 for the reasons we shall explain in the sequel. Thus, we work with

$$\begin{cases} \alpha\frac{\partial}{\partial y}(B_{\varepsilon,\alpha}N_\varepsilon) + 4B_{\varepsilon,\alpha}(y)N_\varepsilon(y) = B_{\varepsilon,\alpha}\left(\frac{y}{2}\right)N_\varepsilon\left(\frac{y}{2}\right) + \lambda_{\varepsilon,\alpha} N_\varepsilon\left(\frac{y}{2}\right) + 2\frac{\partial}{\partial y}\left(N_\varepsilon\left(\frac{y}{2}\right)\right), \qquad y>0, \\ \\ (B_{\varepsilon,\alpha}N_\varepsilon)(0) = 0. \end{cases} \tag{6}$$

Indeed, in order to conserve regularity properties of the solution $H = BN$ to the inverse problem, we want it to be both in $L^1(\mathbb{R}_+)$ and in $L^1(\mathbb{R}_+, xdx)$ in order to express that both the total number of cells and the total biomass are finite. Hence, formal integration of Equation (6) gives

$$\lambda_{\varepsilon,\alpha}\int_0^\infty N_\varepsilon dx = \int_0^\infty B_{\varepsilon,\alpha}N_\varepsilon dx, \tag{7}$$

and integration against the weight x gives

$$-\alpha\int_0^\infty B_{\varepsilon,\alpha}N_\varepsilon dx = 4\lambda_{\varepsilon,\alpha}\int_0^\infty xN_\varepsilon dx - 4\int_0^\infty N_\varepsilon dx. \tag{8}$$

Hence, we have to choose, according to the eigenvalue theory:

$$\lambda_{\varepsilon,\alpha} = \frac{\int_0^\infty N_\varepsilon dx}{\int_0^\infty xN_\varepsilon dx + \frac{\alpha}{4}\int_0^\infty N_\varepsilon dx}. \tag{9}$$

The choice of $\lambda_{\varepsilon,\alpha}$ can be understood as a compatibility condition when $\alpha > 0$ and for $\alpha = 0$ it tells us that (N, λ_0) is overdetermined data for the inverse problem. Therefore, if we have *a priori* knowledge on λ_0, we could verify its distance to $\lambda_{\epsilon,\alpha}$ as a way of checking the error of the inverse problem solution.

The plan of this work is the following: In Section 2, we propose yet another method to regularize the inverse problem, and obtain a convergence rate. The convergence rate turns out to be as good as the one in [1]. In Section 3 we give a numerical method to solve it, and in Section 4 we show some numerical simulations so as to compare the accuracy of the different methods.

2 Regularization by Filtering

2.1 Filtering approach

Taking a closer look at Equation (5), we see that all the difficulties come from the differential term $\frac{\partial}{\partial x}N$. In [1], the choice was to add an equivalent derivative $\alpha\frac{\partial}{\partial x}(B\,N)$ to the equation; here on the contrary, we choose to regularize it by a convolution method.

For $\alpha > 0$, we use the notation

$$\rho_\alpha(x) = \frac{1}{\alpha}\rho(\frac{x}{\alpha}), \qquad \rho \in \mathcal{C}_c^\infty(\mathbb{R}), \quad \int_0^\infty \rho(x)\,dx = 1, \quad \rho \geqslant 0, \quad \text{Supp}(\rho) \subset [0,1], \tag{10}$$

and we replace in (5) the term $\frac{\partial}{\partial x}N_\varepsilon + \lambda_0 N_\varepsilon$ by

$$\big(\frac{\partial}{\partial x}N_\varepsilon + \lambda_{\varepsilon,\alpha}N_\varepsilon\big) * \rho_\alpha(x) = N_\varepsilon * \big(\frac{\partial}{\partial x}\rho_\alpha + \lambda_{\varepsilon,\alpha}\rho_\alpha\big)(x) = \int_0^\infty N_\varepsilon(x')\big(\frac{\partial}{\partial x}\rho_\alpha + \lambda_{\varepsilon,\alpha}\rho_\alpha\big)(x-x')dx'.$$

We now use the notation

$$N_{\varepsilon,\alpha} = N_\varepsilon * \rho_\alpha.$$

In this way, we obtain a smooth term in $L^2(\mathbb{R}_+)$. Furthermore, $N_{\varepsilon,\alpha}$ converges to N_ε in $L^2(\mathbb{R}_+)$ when α tends to zero. We now have to consider the following problem:
Find $B_{\varepsilon,\alpha}$ solution of

$$4B_{\varepsilon,\alpha}(2x)N_{\varepsilon,\alpha}(2x) + B_{\varepsilon,\alpha}(x)N_{\varepsilon,\alpha}(x) = \frac{\partial}{\partial x}N_{\varepsilon,\alpha} + \lambda_{\varepsilon,\alpha}N_{\varepsilon,\alpha}(x), \quad x \geqslant 0\ . \tag{11}$$

As in Equation (6), for the quasi-reversibility method, we need to choose $\lambda_{\varepsilon,\alpha}$ appropriately. Indeed, we perform the same manipulations leading to Equation (9) to get

$$\lambda_{\varepsilon,\alpha} = \frac{\int_0^\infty N_{\varepsilon,\alpha}(x)dx}{\int_0^\infty xN_{\varepsilon,\alpha}(x)dx}. \tag{12}$$

By Theorem A.3 (see the Appendix), we know that the problem in Equation (11) has a unique solution $B_{\varepsilon,\alpha} \in L^2(\mathbb{R}_+, N_{\varepsilon,\alpha}^2 dx)$.

2.2 Estimates for the filtering approach

The main result of this section establishes an estimate for the regularization of the inverse problem by means of the filtering method described above.

Theorem 2.1 *Suppose that $N \in H^2(\mathbb{R}_+)$ and $B \in L^\infty(\mathbb{R}_+)$, $B \geqslant 0$ verify (2). Let $\varepsilon > 0$ and $N_\varepsilon \in L^2(\mathbb{R}_+)$, $N_\varepsilon(x) > 0$ for $x > 0$, such that*

$$||N_\varepsilon - N||_{L^2(\mathbb{R}_+)} \leqslant \varepsilon||N||_{L^2(\mathbb{R}+)}.$$

Let $B_{\varepsilon,\alpha} \in L^2(\mathbb{R}_+, N_{\varepsilon,\alpha}^2 dx)$ be the unique solution of (10) and (11). We have the following estimate:

$$||B_{\varepsilon,\alpha} - B||_{L^2(N_\varepsilon^2 dx)} \leqslant C(\alpha + |\lambda_{\varepsilon,\alpha} - \lambda_0|)||N||_{H^2(\mathbb{R}_+)} + \frac{C}{\alpha}||N_{\varepsilon,\alpha} - N||_{L^2(\mathbb{R}_+)}, \tag{13}$$

where C is a constant depending only on $||B||_{L^\infty}$, $||B_{\varepsilon,\alpha}||_{L^\infty}$ and the regularizing function ρ.

This theorem relies on a first estimate.

Proposition 2.2 *Using the same notations as in Theorem 2.1, we have*

$$||B_{\varepsilon,\alpha}N_{\varepsilon,\alpha} - B\,N||^2_{L^2(dx)} \leqslant C\left(1+\lambda_0^2\right)\left(1+\frac{1}{\alpha^2}\right)||N_\varepsilon - N||^2_{L^2(dx)} + C(\alpha^2 + |\lambda_{\varepsilon,\alpha} - \lambda_0|^2)||N||^2_{H^2(\mathbb{R}_+)}\ , \tag{14}$$

where C depends only on the regularizing function ρ.

Proof of Prop. 2.2: Denote by $Q = B_{\varepsilon,\alpha}N_{\varepsilon,\alpha} - BN$, $R = N_{\varepsilon,\alpha} - N$ and $\delta = \lambda_{\varepsilon,\alpha} - \lambda_0$. From Equations (2) and (11), Q verifies:

$$\begin{cases} \frac{\partial}{\partial x}R(x) + \lambda_0 R(x) + \delta N_{\varepsilon,\alpha}(x) + Q(x) = 4Q(2x), \qquad x \geqslant 0, \\ \\ Q(x=0) = 0. \end{cases} \tag{15}$$

(Since $N_{\varepsilon,\alpha} \in H^1(\mathbb{R}_+)$, the definition of $Q(x=0)$ is not ambiguous.) Multiplying this equation by $Q(2x)$ and integrating on the interval $(0, y)$ yields

$$4\int_0^y Q(2x)^2 dx = \int_0^y Q(2x)\frac{\partial}{\partial x}R(x)dx + \lambda_0\int_0^y Q(2x)R(x)dx$$

$$+\delta\int_0^y Q(2x)N_{\varepsilon,\alpha}(x)dx + \int_0^y Q(2x)Q(x)dx.$$

From the Cauchy-Schwarz inequality, after the change of variables $x \to 2x$, we have

$$4\int_0^y Q(2x)^2 dx \leqslant \frac{1}{2}\int_0^y \left(\frac{\partial}{\partial x}R\right)^2(x)dx + \frac{1}{2}\int_0^y Q(2x)^2 dx + \frac{\lambda_0}{2}\int_0^y CR(x)^2 dx + \frac{\lambda_0}{2}\int_0^y \frac{Q(2x)^2}{C}dx$$

$$+\frac{|\delta|^2}{2}\int_0^y N_{\varepsilon,\alpha}(x)^2 dx + \frac{1}{2}\int_0^y Q(2x)^2 dx + \frac{1}{2}\int_0^y Q(2x)^2 dx + \int_0^{\frac{y}{2}} Q(2x)^2 dx.$$

We take, for instance, $C = \lambda_0$. We obtain

$$||B_{\varepsilon,\alpha}N_{\varepsilon,\alpha} - BN||^2_{L^2} \leqslant ||N_\varepsilon * \frac{\partial}{\partial x}\rho_\alpha - \frac{\partial}{\partial x}N||^2_{L^2} + \lambda_0^2||N_\varepsilon * \rho_\alpha - N||^2_{L^2} + |\lambda_{\varepsilon,\alpha} - \lambda_0|^2||N_\varepsilon * \rho_\alpha||^2_{L^2}. \tag{16}$$

The last two terms of this inequality are easy to estimate, writing

$$||N_\varepsilon * \rho_\alpha - N||_{L^2} \leqslant ||N_\varepsilon * \rho_\alpha - N * \rho_\alpha||_{L^2} + ||N * \rho_\alpha - N||_{L^2} \leqslant C\left(||N_\varepsilon - N||_{L^2} + \alpha||N||_{H^1}\right),$$

and

$$||N_\varepsilon * \rho_\alpha||_{L^2} \leqslant C||N||_{L^2}.$$

It remains to evaluate the first term on the right-hand side of inequality (16). We write

$$||N_\varepsilon * \frac{\partial}{\partial x}\rho_\alpha - \frac{\partial}{\partial x}N||^2_{L^2} \leqslant 2||N_\varepsilon * \frac{\partial}{\partial x}\rho_\alpha - N * \frac{\partial}{\partial x}\rho_\alpha||^2_{L^2} + 2||N * \frac{\partial}{\partial x}\rho_\alpha - \frac{\partial}{\partial x}N||^2_{L^2}.$$

By a convolution estimate we evaluate the first term as

$$||N_\varepsilon * \frac{\partial}{\partial x}\rho_\alpha - N * \frac{\partial}{\partial x}\rho_\alpha||^2_{L^2(\mathbb{R}_+,dx)} \leqslant ||N_\varepsilon - N||^2_{L^2(\mathbb{R}_+,dx)}||\frac{\partial}{\partial x}\rho_\alpha||^2_{L^1} \ .$$

Since $\int_0^\infty |\frac{\partial}{\partial x}\rho_\alpha(x)|dx = \frac{1}{\alpha}\int_0^\infty |\frac{\partial}{\partial x}\rho(y)|dy$, we have

$$||N_\varepsilon * \frac{\partial}{\partial x}\rho_\alpha - \frac{\partial}{\partial x}N||^2_{L^2(\mathbb{R}_+,dx)} \leqslant \frac{C(\rho)}{\alpha^2}||N_\varepsilon - N||^2_{L^2(\mathbb{R}_+,dx)} + 2||N * \frac{\partial}{\partial x}\rho_\alpha - \frac{\partial}{\partial x}N||^2_{L^2(\mathbb{R}_+,dx)} \ .$$

To evaluate the last term $||N * \frac{\partial}{\partial x}\rho_\alpha - \frac{\partial}{\partial x}N||^2_{L^2}$, we extend to $\mathbb{R}$ the functions N and $\frac{\partial}{\partial x}N$ by zero and consider their Fourier transforms. We denote $\hat{f}(\xi)$ the Fourier transform of $f \in L^2(\mathbb{R}_+)$ at ξ, where f is extended as zero on $\mathbb{R}_-$. We obtain by Fourier analysis

$$||N * \frac{\partial}{\partial x}\rho_\alpha - \frac{\partial}{\partial x}N||^2_{L^2(\mathbb{R}_+,dx)} = ||i\xi\hat{N}\hat{\rho}_\alpha - i\xi\hat{N}||^2_{L^2(\mathbb{R}_+,dx)} \leqslant \int_{-\infty}^{\infty} |\hat{N}(\xi)|^2|\xi|^4\frac{|\hat{\rho}_\alpha(\xi)-1|^2}{|\xi|^2}d\xi.$$

Using that

$$|\frac{|\hat{\rho}_\alpha(\xi)-1|^2}{\xi^2}| \leqslant C(\rho)\alpha^2, \tag{17}$$

where $C(\rho)$ only depends on the regularization function ρ, we have that

$$||N * \frac{\partial}{\partial x}\rho_\alpha - \frac{\partial}{\partial x}N||^2_{L^2(\mathbb{R}_+,dx)} \leqslant C(\rho)\alpha^2||N||^2_{H^2(\mathbb{R}_+)} \ .$$

Going back to (16), this concludes the proof of Proposition 2.2. □

We can now deduce the proof of Theorem 2.1. We write:

$$||B_{\varepsilon,\alpha} - B||_{L^2(N_\varepsilon^2 dx)} \leqslant ||B_{\varepsilon,\alpha}N_\varepsilon - B_{\varepsilon,\alpha}\,N_{\varepsilon,\alpha}||_{L^2(\mathbb{R}_+)} + ||B_{\varepsilon,\alpha}\,N_{\varepsilon,\alpha} - B\,N||_{L^2(\mathbb{R}_+)} + ||B\,N - B\,N_\varepsilon||_{L^2(\mathbb{R}_+)}.$$

Using Proposition 2.2, and the fact that

$$||N_\varepsilon - N_{\varepsilon,\alpha}||_{L^2} \leqslant 2||N_\varepsilon - N||_{L^2} + \alpha||N||_{H^1},$$

this inequality gives the result. □

3 Numerical Solution of the Inverse Problem

This section is concerned with the numerical aspects of the solution of the inverse problem. In order to do that we start with a description of the solution to the direct one in Subsection 3.1.

3.1 Direct Problem

In the direct problem, we assume we know the proliferation rate B, we look for N and $\lambda_0 > 0$ solutions of (2). For this purpose, we solve the time-dependent problem (1) and look for a steady dynamics. As already said, this problem is well-posed (see for instance [3]) and it was proved in [18] that solutions grow at an exponential rate towards $\rho N(x)e^{\lambda_0 t}$ with $\rho = \int_0^\infty n(0,x)\phi(x)dx$, recalling the notation in (3). Furthermore, under more restrictive conditions it was shown in [16] that there exists constants $\mu > 0$ and $C(n^0) > 0$, such that

$$||n(t,x)e^{-\lambda_0 t} - \rho N(x)||_{L^1(\mathbb{R}_+,\phi(x)dx)} \leqslant Ce^{-\mu t}.$$

To solve it numerically, we discretize the problem (1) along a regular grid, denote by Δt the time step and by $\Delta x = L/I$ the spatial step, where I denotes the number of points and L the computational domain length: $x_i = i\Delta x$, $0 \leqslant i \leqslant I$.

We use an upwind finite volume method (cf. [23, 24, 25])

$$n_i^k = \frac{1}{\Delta x}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} n(k\Delta t, y)dy, \qquad \frac{1}{\Delta t}\int_0^{\Delta t} n(k\Delta t + s, x_{i+\frac{1}{2}})ds \approx n_i^k.$$

For the time discretization, we use a marching technique. We choose the time step Δt so as to satisfy the largest possible CFL stability criteria $\theta := \frac{\Delta t}{\Delta x} = 1$.

The numerical scheme is given, for $i = 1, ..., I$, by $n_0^k = 0$ and

$$\frac{n_i^{k+1} - n_i^k}{\Delta t} + \frac{n_i^k - n_{i-1}^k}{\Delta x} + B_i n_i^{k+1} = B_{2i-1}n_{2i-1}^k + 2B_{2i}n_{2i}^k + B_{2i+1}n_{2i+1}^k \ , \tag{18}$$

with the convention that $n_j = 0$ for $j > I$. For stability reasons, we have used an implicit method for the division term in the left hand side and explicit for the right hand side of the equation. The specific form for the right hand side is simply motivated by the need of also dividing cells of odd labels.

According to the power algorithm, we do not keep n^{k+1} from (18) but rather renormalize it as

$$\widetilde{n}^{k+1} = \frac{n^k}{\Delta x \sum_{j=1}^{I} n_j^k}.$$

It is standard, for these positive matrices arising in (18), that

$$\widetilde{n}^{k+1} \underset{k\to\infty}{\longrightarrow} N, \qquad \sum_{i=1}^{I} N_i = 1, \qquad N_i > 0,$$

where N is the dominant eigenvector for the problem

$$\frac{N_i - N_{i-1}}{\Delta x} + (\lambda_0 + B_i)N_i = B_{2i-1}N_{2i-1} + 2B_{2i}N_{2i} + B_{2i+1}N_{2i+1}.$$

One can also find the dominant eigenvalue as

$$\lambda_0 = \lim_{k\to\infty} \frac{1}{\Delta t}\log\left(\frac{\sum_{i=1}^{I} n_i^{k+1}}{\sum_{i=1}^{I} n_i^k}\right).$$

For matrices with one dominant eigenvalue and a corresponding one-dimensional eigenspace, it is known that the power algorithm is fast and in fact converges with exponential rate [26]. In practice we can stop the iterations when the relative error on the normalized quantity

$$\frac{1}{\Delta t}\left(\sum_{i=1}^{I}\widetilde{n}_i^{k+1}-\sum_{i=1}^{I}\widetilde{n}_i^{k}\right)$$

is small enough, say of the order of 10^{-10}.

3.2 Inverse Problem: General Strategy

In the sequel, we denote by H the product $B.N$ and its approximations. Indeed, from Equations (6) or (11), we have to search for the product $H = B_{\varepsilon,\alpha}N_\varepsilon$ or $H = B_{\varepsilon,\alpha}N_{\varepsilon,\alpha}$ before computing $B_{\varepsilon,\alpha}$. In particular, we cannot avoid a loss of information where N_ε is small, i.e., for $x \approx 0$ or $x \gg 1$.

The inverse problem (5), as well as (11), can be written as

$$4H(2x) - H(x) = L(x), \tag{19}$$

with different expressions for H and L. We may think of two possible numerical approaches.

Strategy 1. Compute $H(2x)$ from $H(x)$: This means that we re-write Equation (19) with the new variable $y = 2x$, and arrive at

$$4H(y) - H(\frac{y}{2}) = L(\frac{y}{2}). \tag{20}$$

The scheme departs from zero, and one deduces the values of H_i step by step, from the knowledge of H_j for $j \leqslant i-1$.

Strategy 2. Compute $H(x)$ from $H(2x)$: The scheme departs from the largest point $x = L$ of our simulation domain. We suppose that for $x \geqslant L$ we have $H(x) = H(L) = 0$ (it is relevant since we suppose that N vanishes for x large: see below), and then deduce the smaller values H_i step by step, from the knowledge of H_j for $j \geqslant i+1$.

The two approaches do not necessarily lead to the same result because the continuous equation

$$4H(2x) - H(x) = 0 \tag{21}$$

has infinitely many solutions. This issue is interesting on its own and is related to the construction of wavelets, see [27]. It is discussed in Proposition A.1 of the Appendix.

By imposing $H \in L^2(\mathbb{R}_+)$, we select a unique solution, as shown in Theorem A.3. The question is then: Which numerical strategy should we use to select the *correct* solution, *i.e.* the one in $L^2(\mathbb{R}_+)$?

Among the solutions of Equation (19), we single out two, defined by the power series:

$$H^{(1)}(x) = \sum_{n=1}^{+\infty} 2^{-2n}L(2^{-n}x) \quad \text{and } H^{(2)}(x) = -\sum_{n=0}^{+\infty} 2^{2n}L(2^n x) \quad , \quad \forall\, x > 0.$$

Proposition A.1 shows that for $L \in L^2(\mathbb{R}_+, x^p dx)$, there is a unique solution in $L^2(\mathbb{R}_+, x^p dx)$, given by $H^{(1)}$ if $p < 3$ and by $H^{(2)}$ if $p > 3$ (and the power series converge in the corresponding spaces).

For $B > 0$ smooth and bounded from above and from below, we know that N is smooth and vanishes at $x \approx 0$ and $x \approx \infty$, and BN inherits these properties. For instance, we know that

$H \in L^2(dx) \cap L^2(x^4 dx)$. By uniqueness of a solution in each space, Proposition A.1 implies that $H^{(2)} = H^{(1)}$, or equivalently:

$$\sum_{n=-\infty}^{+\infty} 2^{2n} L(2^n x) = 0, \quad \forall\, x \geqslant 0.$$

This very particular property cannot be verified at the discrete level. Hence, the two strategies generally give two different approximations of the same solution of (19). The first strategy selects an approximation of the solution $H^{(1)}$ whereas the second selects an approximation of the solution $H^{(2)}$. In the case of a very regular data N, then $H^{(2)}$ will perform better around infinity, whereas $H^{(1)}$ will be better around zero. However, if N is a solution of Equation (2), when we increase the number of points, the two approaches converge to the same solution since $H^{(2)} = H^{(1)}$.

Since our simulation domain $[0, L]$ is bounded and contains zero, we prefer the first strategy. This choice is confirmed by all the numerical tests we have performed: the second approach has always lead to a solution exploding around zero. However, for the sake of completeness, we also describe the scheme we used for the second approach.

3.3 Inverse Problem: Filtering Approach

According to strategies 1 and 2, we now present two approaches to handle the numerical solution of the inverse problem regularized with the *filtering approach*. Both need to first compute the convolution terms arising in (11). To do so we first take the Fast Fourier Transform F of N_ε, multiply it by $i\xi\hat{\rho}_\alpha(\xi)$, and then take the inverse Fast Fourier Transform F^*. We choose and define the regularization function ρ_α by its Fourier transform:

$$\hat{\rho}_\alpha(\xi) = \frac{1}{\sqrt{1+\alpha^2\xi^2}}.$$

This leads us to the numerical approximation

$$\frac{\partial}{\partial x} N_{\varepsilon,\alpha} \approx dN_\alpha = \mathsf{F}^*\Big(i\xi\hat{\rho}_\alpha(\xi)\mathsf{F}(N_\varepsilon)(\xi) \Big). \tag{22}$$

We also impose $dN_{\alpha,0} = 0$ for compatibility with the continuous equation and further use.

As mentioned earlier, there are two alternatives, either starting from zero or coming from infinity.

The Filtering Approach Starting from Zero (strategy 1). We solve Equation (11) considered as an equation in the variable $y = 2x$, that is to say (20), in order to compute its solution $H^{(1)}(x)$. At the discrete level, we use the notations

$$H_i^f \approx B_i N_i, \qquad L_i^f = dN_{\alpha,i} + \lambda_{\varepsilon,\alpha} N_i, \quad L_0^f = 0.$$

The discrete version of (20) reads

$$4H_i^f = H_{\frac{i}{2}}^f + L_{\frac{i}{2}}^f, \qquad \forall\, 0 \leqslant i \leqslant I, \tag{23}$$

and we need to define the quantities $G_{\frac{i}{2}}$. We choose

$$G_{\frac{i}{2}} = \begin{cases} G_{\frac{i}{2}} & \text{when } i \text{ is even,} \\ \frac{1}{2}\big(G_{\frac{i-1}{2}} + G_{\frac{i+1}{2}}\big) & \text{when } i \text{ is odd.} \end{cases} \tag{24}$$

In particular, we have $H_0^f = 0$.

Summing up all the terms in (23) for $1 \leqslant i \leqslant I$, we find (with I even to simplify):

$$4\sum_{i=0}^{I} H_i^f = 2\sum_{i=0}^{\frac{I}{2}}(H_i^f + L_i^f) - \frac{1}{2}(H_{I/2}^f + L_{I/2}^f).$$

Since we have assumed that N has exponential decay for $x \gg 1$, it follows that

$$\sum_{i=0}^{I} H_i^f = \sum_{i=1}^{I} L_i^f + E_I, \qquad \text{with} \qquad |E_I| \leqslant 2\sum_{i=\frac{I}{2}}^{I} |H_i|. \tag{25}$$

Multiplying (23) by x_i and summing up again, we find

$$\sum_{i=1}^{\frac{I}{2}} x_i L_i^f = F_I, \qquad \text{with} \qquad |F_I| \leqslant \sum_{i=\frac{I}{2}}^{I} x_i |H_i|. \tag{26}$$

As a consequence, we can choose:

$$\lambda_{\varepsilon,\alpha} = -\frac{\sum x_i dN_{\alpha,i}}{\sum x_i N_i}, \tag{27}$$

as the discrete version of the relations (4) or (12).

The Filtering Approach Starting from Infinity (strategy 2). Another method is to discretize the formulation (19) in order to compute its solution $H^{(2)}(x)$. We define the extension $H_i^f = 0$ for $i \geqslant I+1$, and for $2 \leq i \leq I$, we define by backward iterations

$$H_i^{f\infty} = 2H_{2i}^{f\infty} + H_{2i+1}^{f\infty} + H_{2i-1}^{f\infty} - L_i^f. \tag{28}$$

This however does not apply to the indices $i = 0,\ 1$ and we set $H_0^f = \frac{L_0^f}{3} = 0$ and $H_1^f = 4H_2^f - L_1^f$. By summing up all the terms in (28), we find balance properties equivalent to (25)–(26), but with remainders E_I and F_I depending on H_1 and H_2 instead of $H_{i\geqslant\frac{I}{2}}$. One has to check *a posteriori* that these last quantities are very small ; it is not the case in a standard calculation, but becomes true when the precision of the direct problem scheme increases.

3.4 Inverse problem: Quasi-Reversibility Approach

In this section, we present a numerical scheme for the regularized inverse problem proposed in [1]. This problem leads to solving (6) taken at $y = 2x$, that is

$$\begin{cases} \alpha\frac{\partial}{\partial y}(B_{\varepsilon,\alpha}N_\varepsilon) + 4B_{\varepsilon,\alpha}(y)N_\varepsilon(y) = B_{\varepsilon,\alpha}\left(\frac{y}{2}\right)N_\varepsilon\left(\frac{y}{2}\right) + \lambda_{\varepsilon,\alpha}N_\varepsilon\left(\frac{y}{2}\right) + 2\frac{\partial}{\partial y}\left(N_\varepsilon\left(\frac{y}{2}\right)\right), \qquad y > 0, \\ \\ (B_{\varepsilon,\alpha}N_\varepsilon)(0) = 0, \end{cases}$$

where $\alpha > 0$ is the regularizing parameter and $\lambda_{\varepsilon,\alpha}$ is defined by (9). This gives, in a discretized version, after dropping the index ε,

$$\lambda_{\varepsilon,\alpha} = \frac{\sum N_i}{\sum x_i N_i + \frac{\alpha}{4}\sum N_i}. \tag{29}$$

For the numerical discretization we set $H^Q_{-1} = 0$ and also recall that $N_0 = 0$ and assume that the data satisfies $N_{I+1} = 0$. We use a standard upwind scheme for the differential term:

$$\frac{\alpha}{\Delta x}(H^Q_i - H^Q_{i-1}) + 4H^Q_i = H^Q_{\frac{i}{2}} + L^Q_{\frac{i}{2}}, \tag{30}$$

where we have defined the fractional indices as in the filtering approach by (24), and here

$$L^Q_i = \lambda_\varepsilon N_i + \frac{N_{i+1} - N_i}{\Delta x}.$$

If we neglect the terms $H^Q_{i \geqslant \frac{I}{2}+1}$, we can easily verify a discrete version of the balance laws (7) and (9), equivalent to (25)–(26).

4 Numerical Tests

As input data, we take the values of the function N obtained by the numerical solution of the direct problem in Section 3.1, we add a random noise uniformly distributed in $[-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]$, and we enforce nonnegativity of the data

$$N_\varepsilon = \max(N + \varepsilon\, r, 0).$$

We solve the direct problem on a regular grid of $I + 1$ points, on an interval $[0, 2L]$. We need L large enough, such that it is possible to assume that $N(x \geqslant L) \approx 0$ and we have checked it *a posteriori.* Indeed, we have seen that this property is essential when we use the inverse schemes on a domain $[0, L]$ in order to verify the balance laws (7)–(9). In other words, we solve the direct problem on a domain twice larger than for the inverse problem. In the numerical tests we take $L = 4$, and we show the numerical solution N only on the interval $[0, L]$ since it is uniformly small on $[L, 2L]$.

We solve the inverse problem by the different methods on a regular grid of $I_1 + 1$ points on $[0, L]$, with $\Delta x_1 = L/I_1$. This grid is taken ten times finer than the grid used for the direct problem, *i.e.* we take $I_1 = 10I$. Since we have chosen L large enough so that $N(x \geqslant L) \approx 0$, we have always obtained that indeed $H(x \geqslant L) \approx 0$.

As before, we denote by H^Q and H^f the solution data H obtained respectively by the quasi-reversibility method of Section 3.4 and by the first filtering approach (from zero) of Section 3.3. We also define a solution H^{fQ} by mixing both methods, *i.e.* by solving the following equation:

$$\begin{cases} \alpha \frac{\partial}{\partial x}(B_{\varepsilon,\alpha} N_\varepsilon)(y) + 4B_{\varepsilon,\alpha}(y) N_\varepsilon(y) - B_{\varepsilon,\alpha}(x) N_\varepsilon(x) = \left(\frac{\partial}{\partial x} N_\varepsilon + \lambda_{\varepsilon,\alpha} N_\varepsilon \right) * \rho_\alpha(x), \quad x \geqslant 0, \\ \\ B_{\varepsilon,\alpha}(x = 0) N_\varepsilon(x = 0) = 0, \end{cases} \tag{31}$$

where $\lambda_{\varepsilon,\alpha}$ is defined by

$$\lambda_{\varepsilon,\alpha} = \frac{\int_0^\infty N_\varepsilon * \rho_\alpha dx}{\int_0^\infty x N_\varepsilon * \rho_\alpha dx + \frac{\alpha}{4} \int_0^\infty N_\varepsilon * \rho_\alpha dx}. \tag{32}$$

The relative error is measured, as seen in Theorem 2.1 and in Theorem 5.1 of [1], by

$$\delta^Q = \frac{||BN_\varepsilon - H^Q||^2_{l^2}}{||N_\varepsilon||_{l^2}}, \qquad \delta^f = \frac{||BN_\varepsilon - H^f||^2_{l^2}}{||N_\varepsilon||_{l^2}}, \qquad \delta^{fQ} = \frac{||BN_\varepsilon - H^{fQ}||^2_{l^2}}{||N_\varepsilon||_{l^2}}.$$

We have divided by $||N_\varepsilon||_{L^2}$ and not by $||N||_{H^2}$ because in practice we only know the entry data with noise.

In order to illustrate the accuracy of our method, we also compare it to a naive way (brute force) of considering the equation. Namely, we approximate $\frac{\partial}{\partial x}N(x)$ by a second-order Euler scheme without regularization. It gives a solution H^b by the same formula (30), where we simply take $\alpha = 0$.

The Direct Problem. We have first tested the direct problem for various division rates B. Three different solutions N for three given division rates B are depicted in Figure 1 with 800 grid points.

In the particular case when B is constant, we can go further and evaluate the computational error. Then, we know that $\lambda = B$ and the exact solution N_{exact} can be explicitly calculated, as shown in [16, 3], by the formula:

$$N_{\text{exact}}(x) = \bar{N} \sum_{n=0}^{\infty} \alpha_n e^{-2^n Bx}, \tag{33}$$

where the coefficients are defined recursively by $\alpha_0 = 1$ and $\alpha_n = (-1)^n \frac{2\alpha_{n-1}}{2^n - 1}$, and $\bar{N}$ is chosen to ensure the mass one normalization. We take $B = 1$ and obtain the continuous curve of Figure 1. We can measure here the relative error by

$$\delta^D = \frac{\|N - N_{\text{exact}}\|_{l^1}}{\|N_{\text{exact}}\|_{l^1}},$$

where N represents the numerical solution of Section 3.1. We choose this norm because for B constant, the solution of the adjoint problem is $\phi = 1$ and the General Relative Entropy Principle ([18, 3]) gives us that this quantity decreases along the time iterations. Still for 800 points, we obtain $\delta^D = 7.7.10^{-3}$.

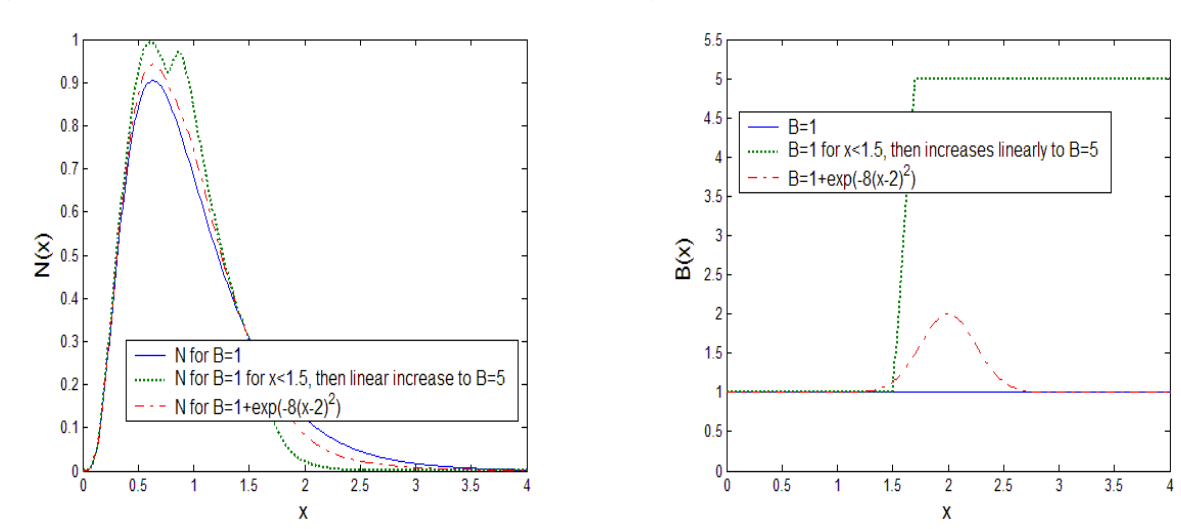

Figure 1: Solutions N (left) obtained by the numerical resolution of Section 3.1 for the direct problem with three different division rates B (right).

The noiseless case ($\varepsilon = 0$). In the simplest case where the data is perfectly known, *i.e.* for $\varepsilon = 0$, we verify that the different schemes allow us to recover B. Since the precision of the data is directly

linked to the number of points used in the scheme, we run the codes with 1.000 points for the direct problem (below, we will take only 100 points).

We test several values of α and we use the three functions B of Figure 1 for each method for the inverse problem. The error estimate is found to depend on the method used but not significantly on the division rate B. Therefore we have drawn in Figure 2 the average error estimates for the three division rates B. In Figure 3 we have depicted the products $B.N$ in the case $B = 1$ and $\alpha = 0.01$ (other cases are similar): it shows that the precision obtained is satisfactory. In Figures 4, 5 and 6 we have drawn the approximations of B in each of the three cases, calculated only for $N > 0.01$ (indeed, for N too small the division leads to insignificant results on B).

Not surprisingly, the brute force method reveals to be satisfactory, with an error estimate of $\delta^b = 1.3.10^{-2}$, since we are in the case where N is very regular. The filtering method can reach this level of error for $\alpha = 10^{-2}$ but cannot go further. However, both the quasi reversibility method and the mixed method given by Equation (31) improve it with minimum values $\delta^Q = 6.9.10^{-3}$ and $\delta^{fQ} = 6.5.10^{-3}$ reached for $\alpha = 10^{-2}$.

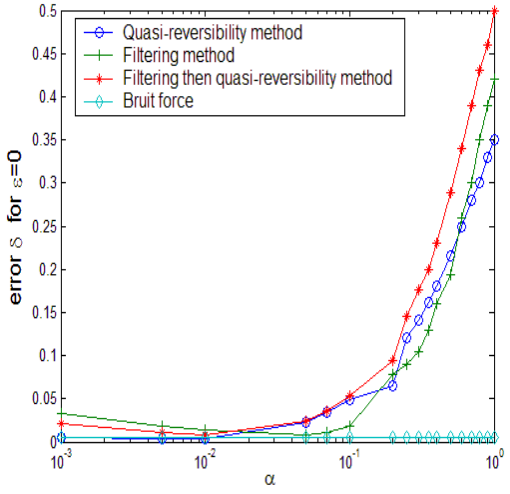

Figure 2: For $\varepsilon = 0$, numerical errors obtained with the different methods for the inverse problem.

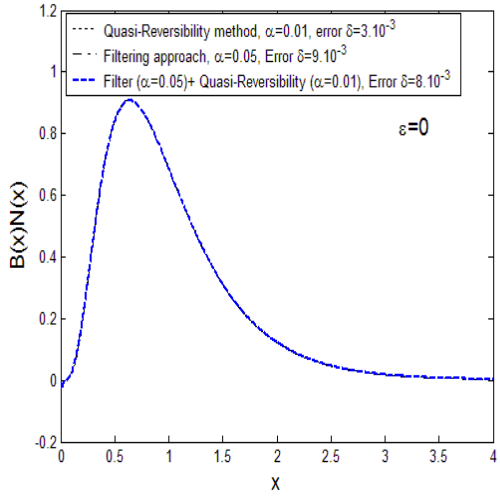

Figure 3: Numerical reconstruction of $B.N$ obtained by each method for the inverse problem when $B = 1$, $\varepsilon = 0$ and $\alpha = 0.01$.

Link between the noise level ε and the regularization parameter α. For noise levels $\varepsilon = 0.01$, $\varepsilon = 0.05$ and $\varepsilon = 0.1$ respectively, the Figures 7, 8 and 9 give the curves ε as a function of α for the three inverse methods. We compare the reconstructed division rates B in Figures 10 and 11.

Each of the error curves presents a minimum for an optimal value of α, as expressed by estimate (13) for instance. In Figures 12, 13, 14 and 15, we have compared three curves, drawn in a log-log scale: $\sqrt{\varepsilon}$ to serve as a reference curve, $f(\varepsilon) = \min_{\alpha} \delta(\alpha, \varepsilon)$, and $g(\varepsilon) = \arg\min_{\alpha} \delta(\alpha, \varepsilon)$. One can see that for each method, these three curves have comparable slopes ($\frac{1}{2}$ on a log-log scale): they show that even though the combination of filtering and quasi-reversibility method improves the optimal errors in absolute value, it does not change the order of convergence of the approximation, which remains of order $O(\sqrt{\varepsilon})$. Figure 15 gives also the convergence of the filtering method for much smaller values

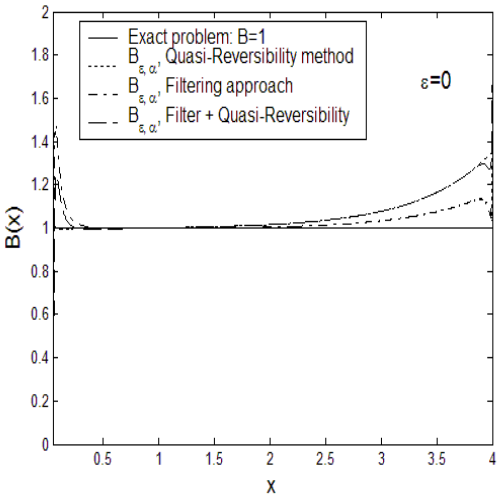

Figure 4: Reconstructed division rate B using the three inverse methods, for $\varepsilon = 0$, $\alpha = 0.01$ with N computed from $B = 1$.

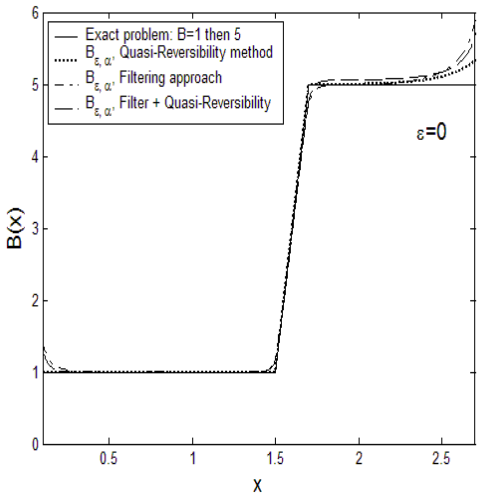

Figure 5: Reconstructed division rate B, for $\varepsilon = 0$, $\alpha = 0.01$ and a jump $B = 1$ to 5 as in Figure 1

of ε (for which an increased number of 500 points has been taken, in order to avoid numerical bias): the comparison with $\sqrt{\epsilon}$ is there particularly evident, and we have obtained similar curves for the two other methods. The speed of convergence is though in complete accordance with the theoretical estimate (13).

Influence of the choice of λ_0 instead of $\lambda_{\varepsilon,\alpha}$. To evaluate the influence of the error term due to the distance $|\lambda_{\varepsilon,\alpha} - \lambda_0|$, we compare the curves obtained respectively by taking on the inverse code the exact λ_0 or the value $\lambda_{\varepsilon,\alpha}$ expressed by the balance laws. They are drawn in Figure 13 for the quasi-reversibility method. They show that even though the *a priori* knowledge of λ_0 improves the error in absolute value, it does not change the order of convergence of the scheme. Thus it is in complete accordance with Estimate (13).

5 Conclusions

We have considered size-structured equations connected to several areas of biology from cell division to prion proliferation by aggregation and fragmentation. We have addressed the numerical efficiency of some inverse problem solution methods to tackle the problem of recovering the division rate from the size distribution of cells. The latter involves a dilation equation with a singular right-hand side that needs regularization for actual implementation. For that purpose, we have introduced a filtering method and proved its convergence for noisy data. This method brings in an operator that has a non-trivial kernel and we have selected a numerical approximation that is able to recover the natural solution we want to reach.

The implementation of the inverse algorithm, based on the filtering method, confirms the convergence analysis. In particular, there is an optimal regularization parameter as can be seen in the graphs of Figures 7, 8 or 9 for instance. Comparison with a quasi-reversibility method introduced

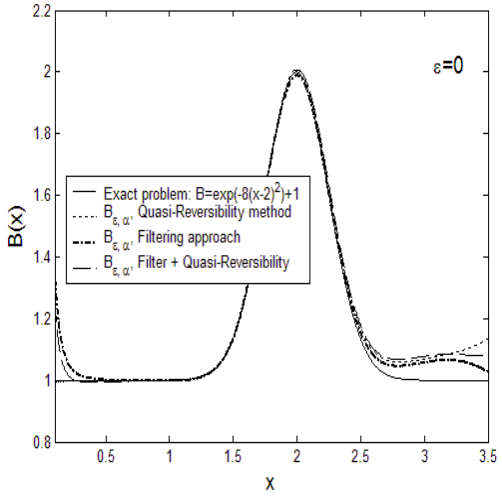

Figure 6: Reconstructed division rate B, for $\varepsilon = 0$, $\alpha = 0.01$ and $B = 1 + exp(-8(x-2)^2)$.

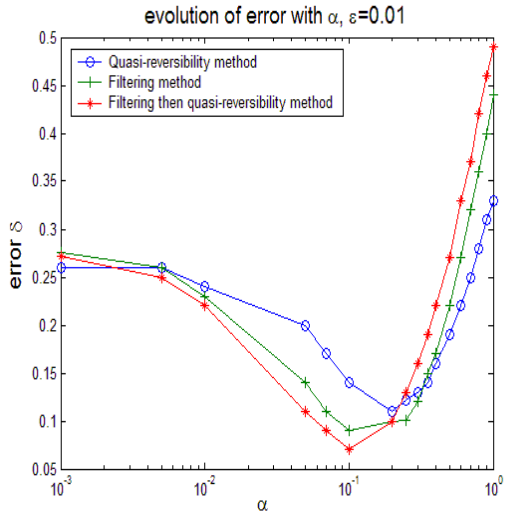

Figure 7: Numerical error when $\varepsilon = 0.01$ for the different methods.

earlier leads to the conclusion that a combination of filtering and quasi-reversibility methods seems to be more efficient because the oscillations are reduced, but without improving the rate of convergence.

We also analyzed the impact of using the exact value of λ_0 or the λ_ε on the different solutions of the inverse problem. In our simulations, the difference between using λ_0 or λ_ε seemed to be immaterial as far as the accuracy of method is concerned. This is in perfect accordance with the theoretical estimate (13).

The above remarks open several directions for continuation and extension of the present work. On the practical side, the present work sets the stage for the use of experimental data either from the existing literature or from more recent biological experiments. On the theoretical side, the possibility of improving the convergence by combining the filtering and quasi-reversibility methods should be investigated further.

Finally, we point out that although the Tikhonov method is more standard, we did not study it so far because it seems more time consuming. Indeed, iterations are needed to solve both the direct problem and the inverse one. To overcome such difficulty a completely new theory has to be developed so as to suit the particular structure of our model. This provides yet another direction for future work.

Acknowledgments

The authors were supported by the CNPq-INRIA agreement INVEBIO. JPZ was supported by CNPq under grants 302161/2003-1 and 474085/2003-1. JPZ and BP are thankful to the RICAM special semester and to the International Cooperation Agreement Brazil-France. Part of this work was conducted during the Special Semester on Quantitative Biology Analyzed by Mathematical Methods, October 1st, 2007 -January 27th, 2008, organized by RICAM, Austrian Academy of Sciences.

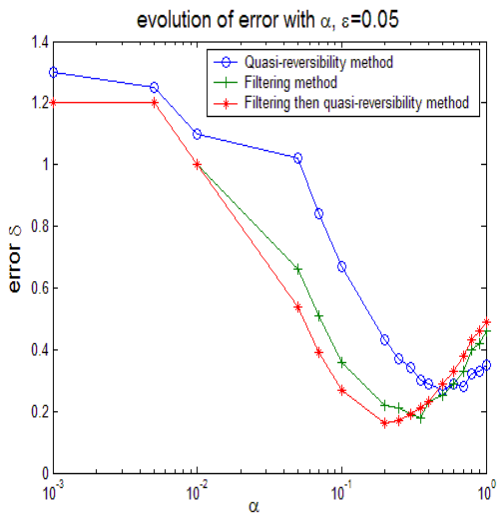

Figure 8: Numerical error when $\varepsilon = 0.05$.

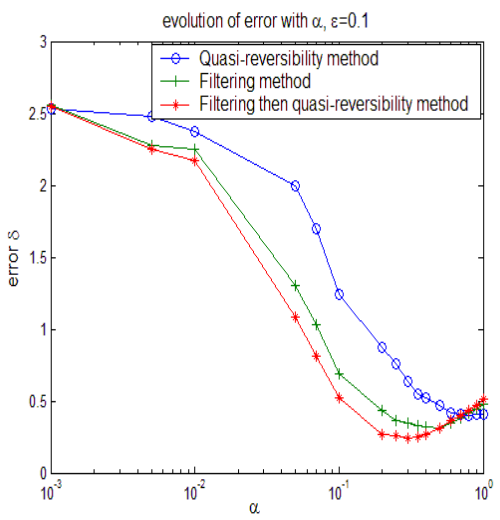

Figure 9: Numerical error when $\varepsilon = 0.1$.

A Well-Posedness of Functional Equation Associated to the Inverse Problem.

We have seen that the regularization method for the inverse problem relies mostly on solving the equation

$$4H(2x) - H(x) = L(x), \qquad x \geq 0. \tag{34}$$

Eventhough this equation is formally very simple, its analysis reveals some complexity. It may admit several solutions in general. Among them, we can mention two with simple representation formulas (we leave to the reader to check they are indeed formally solutions)

$$H^{(1)}(x) = \sum_{n=1}^{+\infty} 2^{-2n} L(2^{-n}x). \tag{35}$$

$$H^{(2)}(x) = -\sum_{n=0}^{+\infty} 2^{2n} L(2^{n}x), \tag{36}$$

To clarify this issue and motivate our choice of a solution, we first state general results concerning solutions to (34) and then come back to our original problem (5).

We first mention the following

Proposition A.1 *Let $L \in L^2(\mathbb{R}_+, x^p dx)$, with $p \neq 3$, then there exists a unique solution $H \in L^2(\mathbb{R}_+, x^p dx)$ to (34) and*

- *for $p < 3$, this solution is given explicitly by the formula (35). Furthermore, for $1 \leq q \leq \infty$, if $L \in L^q(\mathbb{R}_+)$ then $H^{(1)} \in L^q(\mathbb{R}_+)$.*
- *for $p > 3$, this solution is given explicitly by the formula (36).*

Because we look for an integrable function H (the number of cells is supposedly finite), the function $H^{(1)}$ is preferable (take $q = 1$). It also behaves better near $x \approx 0$ because the weight $p < 3$ imposes that $H^{(1)}$ vanishes at 0 as we expect.

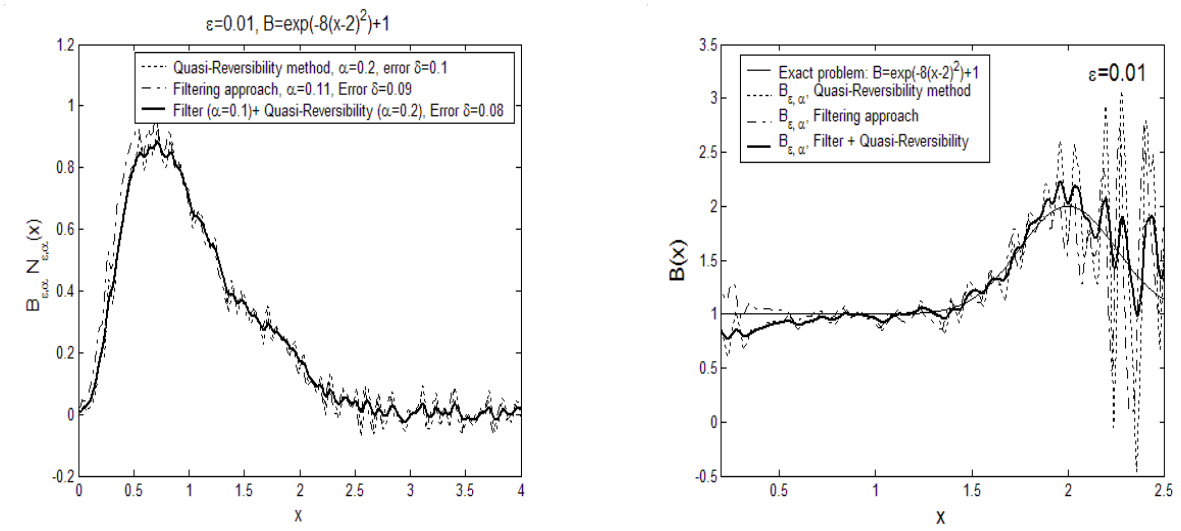

Figure 10: In the case $\varepsilon = 0.01$, $\alpha = 0.05$, $B = 1 + e^{-8(x-2)^2}$, numerical solution $B.N$ (left) and B (right) by the different methods.

From the point of view of exact solutions of the direct problem, we find that $H = BN$ and L belong to all spaces $L^2(\mathbb{R}_+, x^p dx)$ for all $p \in \mathbb{R}$. Therefore, the two solutions coincide and in principle we could choose any of them. In practice, errors on the data L are better handled by $H^{(1)}$ than by $H^{(2)}$ for the afore mentioned reason. Notice indeed that these two solutions are different in general. One can check for instance that for $L = 0$, there is a singular distributional solution $\delta'_{x=0}$. Furthermore,

Lemma A.2 *The solutions to (34) with $L = 0$ in $\mathcal{D}'(0,\infty)$ have the form $\frac{f(\log(x))}{x^2}$ with $f \in \mathcal{D}'(\mathbb{R})$ a* $\log(2)-$ *periodic distribution.*

Proof of Proposition A.1: We consider the Hilbert space $X = L^2(\mathbb{R}_+, x^p dx)$ and we simply apply the Lax-Milgram theorem to a properly chosen bilinear form.

Case 1, $p < 3$. We solve the equation in the variable $y = 2x$ that is (20). and consider the bilinear form $a(u, v)$ on $X \times X$ defined by

$$a(u,v) = 4\int_0^\infty u(y)v(y)y^p dy - \int_0^\infty u(\frac{y}{2})v(y)y^p dy.$$

This form is obviously continuous and it remains to prove that it is coercive. We have

$$a(u,u) = 4\int_0^\infty u(y)^2 y^p dy - \int_0^\infty u(\frac{y}{2})u(y)y^p dy \geqslant (4 - 2^{\frac{p+1}{2}}) \int_0^\infty u(y)^2 y^p dy,$$

and it is indeed coercive as long as $\alpha = 4 - 2^{\frac{p+1}{2}}$ is positive which holds true for $p < 3$. The Lax-Milgram Theorem asserts that there is a unique $H \in X$ such that $a(H, \cdot) = (L, \cdot)$, where $(\cdot, \cdot\cdot)$ denotes the inner product in X, that is a solution of (20).

Case 2, $p > 3$. We work in the variable x and consider the continuous bilinear form $b(u, v)$ on $X \times X$ defined by

$$b(u,v) = -4\int_0^\infty u(2x)v(x)x^p dx + \int_0^\infty u(x)v(x)x^p dx.$$

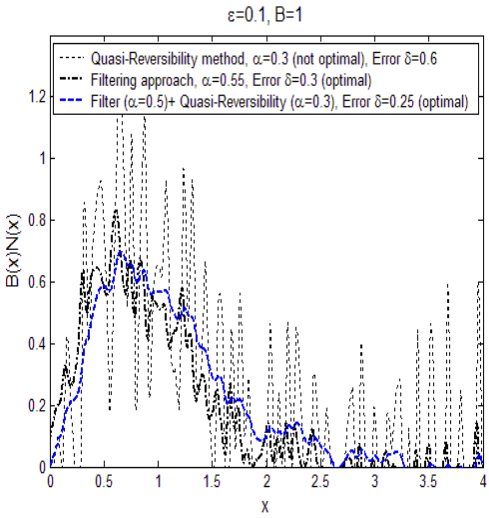

Figure 11: In the case $\varepsilon = 0.1$ and $B = 1$, the numerical solution $B.N$ by the different methods.

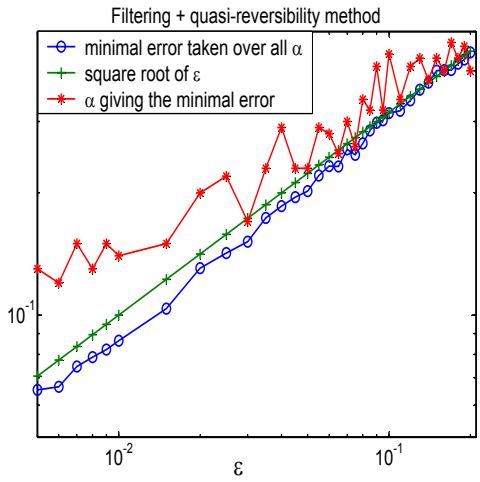

Figure 12: Filtering and quasi-reversibility method: minimal error as a function of the noise level ε for the optimal value α, with a comparison to the theoretical curve $\sqrt{\varepsilon}$.

The same calculation leads us to:

$$b(u,u) \geqslant \alpha \int_0^\infty u(x)^2 x^p dx, \qquad \text{with} \quad \alpha = 1 - 2^{\frac{3-p}{2}} > 0,$$

and the same conclusion holds.

To check formulae (36) and (35), it remains to prove that these solutions belong to the corresponding spaces:

$$\|H^{(1)}\|_{L^2(\mathbb{R}_+,x^p dx)} \leqslant \sum_{n=1}^{\infty} 2^{-2n} ||L(2^{-n}x)||_{L^2(\mathbb{R}_+,x^p dx)} = \sum_{n=1}^{\infty} 2^{\frac{n}{2}(p-3)} ||L(x)||_{L^2(\mathbb{R}_+,x^p dx)}.$$

This sum converges *iff* $p > 3$. In the same way, we write:

$$||H^{(2)}||_{L^2(x^p dx)} \leqslant \sum_{n=0}^{\infty} 2^{2n} ||L(2^n x)||_{L^2(\mathbb{R}_+,x^p dx)} = \sum_{n=0}^{\infty} 2^{\frac{n}{2}(3-p)} ||L(x)||_{L^2(\mathbb{R}_+,x^p dx)},$$

which converges *iff* $p > 3$. □

Proof of Lemma A.2: When $L = 0$, we first define $\mathcal{H} \in \mathcal{D}'(0,\infty)$ as the second antiderivative of H, and notice that it should verify

$$\mathcal{H}(2x) = \mathcal{H}(x).$$

We perform the change of variables $y = \log(x)$ and notice that, if $\mathcal{H} \in \mathcal{D}'(0,\infty)$, it is equivalent to look for solutions $f \in \mathcal{D}'(\mathbb{R})$ of

$$f\big(y + \log(2)\big) = f(y). \tag{37}$$

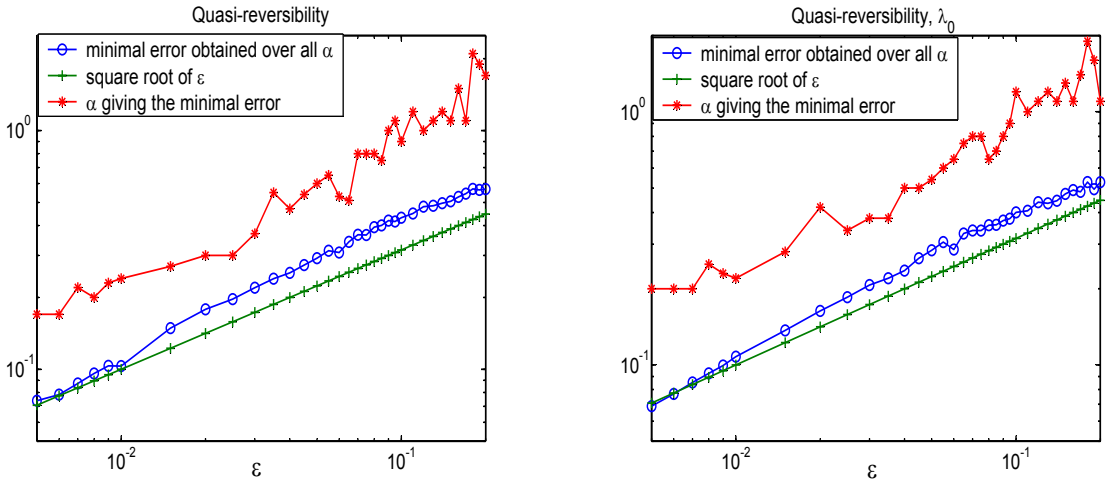

Figure 13: Quasi-reversibility method, with (left) $\lambda_{\varepsilon,\alpha}$ given by relation (9) or (right) *a priori* knowledge of λ_0 : minimal error and optimal regularization parameter α as functions of the noise level ε, with a comparison to the theoretical curve $\sqrt{\varepsilon}$. We see that the *a priori* knowledge of λ_0 does not improve the speed of convergence of the scheme.

Hence, all the solutions in $\mathcal{D}'(0,\infty)$ are given by $\frac{f\big(\log(x)\big)}{x^2}$, where $f \in \mathcal{D}'(\mathbb{R})$. □

To conclude this Appendix, we come back to our original problem (5) and draw the consequences in terms of B, not H.

Theorem A.3 *Let $N \in L^2(\mathbb{R}_+)$, with $N(x) > 0$ for $x > 0$. Let $L \in L^2(\mathbb{R}_+)$. There exists a unique $B \in L^2(\mathbb{R}_+, N^2dx)$ solution of*

$$4B(2x)N(2x) - B(x)N(x) = L(x). \tag{38}$$

Proof: The theorem follows directly from Proposition A.1 for $p = 0$, and since $N > 0$, we can define $B = H/N$ for $B \in L^2(\mathbb{R}_+, N^2dx)$. □

This theorem shows that we can find a solution B of (5) for all N and all λ, this is the basis of our algorithm. However, if we want that the solution B belongs to the space $L^1(\mathbb{R}_+; xN(x)dx)$, integration of (38) multiplied by x shows that L has to satisfy the condition

$$\int_0^\infty xL(x)dx = 0.$$

Applying this to Equation (5), we recover that $\lambda_0 = \int_0^\infty N(x)dx / \int_0^\infty xN(x)dx$. In the case of Equations (6) and (11) respectively, we get formulae (9) and (12), which discrete versions are expressed by (29) and (27).

In view of these considerations, it is better to use a discrete scheme defined by a matrix A that preserves a similar discrete property. Namely, for all $H = (H_i)$, we should have $\sum_i i(AH)_i = 0$, in other words the vector of components i belongs to the kernel of the adjoint of A. Indeed, this property yields the (discrete) regularity $H \in L^1(\mathbb{R}_+; xdx)$.

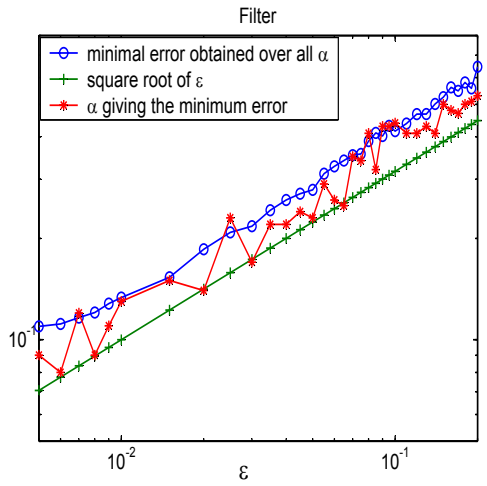

Figure 14: Filtering method for standard levels of noise.

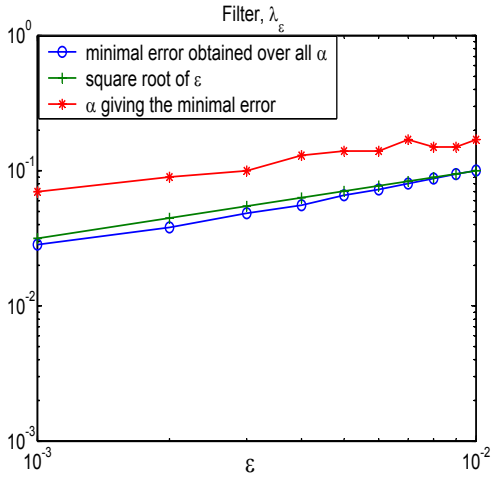

Figure 15: Filtering method for smaller values of ε and increased number of points.

References

[1] Benoît Perthame and Jorge P. Zubelli. On the inverse problem for a size-structured population model. *Inverse Problems*, 23(3):1037–1052, 2007.

[2] J. A. J. Metz and O. Diekmann. Formulating models for structured populations. In *The dynamics of physiologically structured populations (Amsterdam, 1983)*, volume 68 of *Lecture Notes in Biomath.*, pages 78–135. Springer, Berlin, 1986.

[3] Benoît Perthame. Transport equations arising in biology. In *Frontiers in Mathematics*, Frontiers in Mathematics. Birkhauser, 2007.

[4] Heinz W. Engl, William Rundell, and Otmar Scherzer. A regularization scheme for an inverse problem in age-structured populations. *J. Math. Anal. Appl.*, 182(3):658–679, 1994.

[5] Mats Gyllenberg, Andrei Osipov, and Lassi Päivärinta. The inverse problem of linear age-structured population dynamics. *J. Evol. Equ.*, 2(2):223–239, 2002.

[6] William Rundell. Determining the birth function for an age structured population. *Math. Population Stud.*, 1(4):377–395, 397, 1989.

[7] Michael Pilant and William Rundell. Determining a coefficient in a first-order hyperbolic equation. *SIAM J. Appl. Math.*, 51(2):494–506, 1991.

[8] J.A. Carrillo and T. Goudon. A numerical study on large-time asymptotics of the Lifshitz-Slyozov system. *J. Sci. Comput.*, 20(1):69–113, 2004.

[9] Jean-François Collet, Thierry Goudon, Frédéric Poupaud, and Alexis Vasseur. The Beker-Döring system and its Lifshitz-Slyozov limit. *SIAM J. Appl. Math.*, 62(5):1488–1500, 2002.

[10] Jean-François Collet, Thierry Goudon, and Alexis Vasseur. Some remarks on large-time asymptotic of the Lifshitz-Slyozov equations. *J. Statist. Phys.*, 77(1-2):139–152, 1999.

[11] Philippe Laurençot. Convergence to self-similar solutions for a coagulation equation. *Z. Angew. Math. Phys.*, 56(3):398–411, 2005.

[12] Philippe Laurençot and Stéphane Mischler. Liapunov functionals for Smoluchowski's coagulation equation and convergence to self-similarity. *Monatsh. Math.*, 146(2):127–142, 2005.

[13] V. Calvez, N. Lenuzza, D. Oelz, J.-P. Deslys, F. Mouthon, P. Laurent, and B. Perthame. Bimodality, prion aggregates infectivity and prediction of strain phenomenon, arXiv : 0802.2024 (2008).

[14] J. Masel, V.A.A. Jansen, and M.A. Nowak. Quantifying the kinetic parameters of prion replication. *Biophys. Chem*, 77:139–152, 1999.

[15] Meredith L. Greer, Laurent Pujo-Menjouet, and Glenn F. Webb. A mathematical analysis of the dynamics of prion proliferation. *J. Theoret. Biol.*, 242(3):598–606, 2006.

[16] Benoît Perthame and Lenya Ryzhik. Exponential decay for the fragmentation or cell-division equation. *J. Differential Equations*, 210(1):155–177, 2005.

[17] Philippe Michel. Existence of a solution to the cell division eigenproblem. *Model. Math. Meth. Appl. Sci.*, 16(suppl. issue 1):1125–1153, 2006.

[18] Philippe Michel, Stéphane Mischler, and Benoît Perthame. General relative entropy inequality: an illustration on growth models. *J. Math. Pures Appl. (9)*, 84(9):1235–1260, 2005.

[19] Johann Baumeister and Antonio Leitão. *Topics in inverse problems*. Publicações Matemáticas do IMPA. [IMPA Mathematical Publications]. Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 2005. 25º Colóquio Brasileiro de Matemática. [25th Brazilian Mathematics Colloquium].

[20] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[21] Robert Lattès. Non-well-set problems and the method of quasi reversibility. In *Functional Analysis and Optimization*, pages 99–113. Academic Press, New York, 1966.

[22] R. Lattès and J.-L. Lions. *Méthode de quasi-réversibilité et applications*. Travaux et Recherches Mathématiques, No. 15. Dunod, Paris, 1967.

[23] François Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws, and well-balanced schemes for sources*. Frontiers in Mathematics. Birkhäuser, 2004.

[24] Randall J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Frontiers in Mathematics. Cambridge University Press, 2002.

[25] Edwige Godlewski and Pierre-Arnaud Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Applied Mathematical Sciences, vol. 118. Springer, 1996.

[26] Denis Serre. *Matrices: Theory and Applications*. TELOS, 2002.

[27] Gilbert Strang. Wavelets and dilation equations: a brief introduction. *SIAM Review*, 31(4):614–627, 1989.

4 Statistical Inverse [11]

NONPARAMETRIC ESTIMATION OF THE DIVISION RATE OF A SIZE-STRUCTURED POPULATION *

M. DOUMIC†, M. HOFFMANN‡, P. REYNAUD-BOURET§, AND V. RIVOIRARD¶

Abstract. We consider the problem of estimating the division rate of a size-structured population in a nonparametric setting. The size of the system evolves according to a transport-fragmentation equation: each individual grows with a given transport rate, and splits into two offsprings of the same size, following a binary fragmentation process with unknown division rate that depends on its size. In contrast to a deterministic inverse problem approach, as in [24, 5], we take in this paper the perspective of statistical inference: our data consists in a large sample of the size of individuals, when the evolution of the system is close to its time-asymptotic behavior, so that it can be related to the eigenproblem of the considered transport-fragmentation equation (see [23] for instance). By estimating statistically each term of the eigenvalue problem and by suitably inverting a certain linear operator (see [5]), we are able to construct a more realistic estimator of the division rate that achieves the same optimal error bound as in related deterministic inverse problems. Our procedure relies on kernel methods with automatic bandwidth selection. It is inspired by model selection and recent results of Goldenschluger and Lepski [14, 15].

Key words. Lepski method, Oracle inequalities, Adaptation, Aggregation-fragmentation equations, Statistical inverse problems, Nonparametric density estimation, Cell-division equation

AMS subject classifications. 35A05, 35B40, 45C05, 45K05, 82D60, 92D25, 62G05, 62G20

1. Introduction.

1.1. Motivation. Structured models have long served as a representative deterministic model used to describe the evolution of biological systems, see for instance [20] or [21] and references therein. In their simplest form, structured models describe the temporal evolution of a population structured by a biological parameter such as size, age or any significant *trait*, by means of an evolution law, which is a mass balance at the macroscopic scale. A paradigmatic example is given by the transport-fragmentation equation in cell division, that reads

$$\begin{cases} \dfrac{\partial}{\partial t}n(t,x) + \dfrac{\partial}{\partial x}\big(g_0(x)n(t,x)\big) + B(x)n(t,x) = 4B(2x)n(t,2x), \quad t \geq 0, \quad x \geq 0, \\ gn(t,x=0) = 0, \quad t > 0, \\ n(t=0,x) = n^0(x), \quad x \geq 0. \end{cases} \tag{1.1}$$

The mechanism captured by Equation (1.1) can be described as a mass balance equation (see [1, 21]): the quantity of cells $n(t,x)$ of size x at time t is fed by a transport term $g_0(x)$ that accounts for growth by nutrient uptake, and each cell can split into two offsprings of the same size according to a division rate $B(x)$. Supposing $g_0(x) = \kappa g(x)$, where we suppose a given model for the growth

*The research of M. Doumic is partly supported by the Agence Nationale de la Recherche, Grant No. ANR-09-BLAN-0218 TOPPAZ. The research of M. Hoffmann is partly supported by the Agence Nationale de la Recherche, Grant No. ANR-08-BLAN-0220-01 MADCOF. The research of P. Reynaud-Bouret and V. Rivoirard is partly supported by the Agence Nationale de la Recherche, Grant No. ANR-09-BLAN-0128 PARCIMONIE.

†INRIA Rocquencourt, projet BANG, Domaine de Voluceau, BP 105, 781153 Rocquencourt, France. email: marie.doumic-jauffret@inria.fr

‡ENSAE-CREST and CNRS-UMR 8050, 3, avenue Pierre Larousse, 92245 Malakoff Cedex, France. email: marc.hoffmann@ensae.fr

§CNRS-UMR 6621 and Université de Nice Sophia-Antipolis, Laboratoire J-A Dieudonné, Parc Valrose, 06108 Nice cedex 02, France. email: patricia.reynaud-bouret@unice.fr

¶CEREMADE, CNRS-UMR 7534, Université Paris Dauphine, Place Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France. INRIA Paris-Rocquencourt, projet Classic. email: Vincent.Rivoirard@dauphine.fr

rate $g(x)$ known up to a multiplicative constant $\kappa > 0$, and experimental data for $n(t,x)$, the problem we consider here is to recover the division rate $B(x)$ and the constant κ.

In [24], Perthame and Zubelli proposed a deterministic method based on the asymptotic behavior of the cell amount $n(t,x)$: indeed, it is known (see *e.g.* [22, 23]) that under suitable assumptions on g and B, by the use of the *general relative entropy principle* (see [20]), one has

$$\int_0^\infty \left| n(t,x)e^{-\lambda t} - \langle n^0, \phi\rangle N(x)\right| \phi(x)\, dx \underset{t\to\infty}{\longrightarrow} 0 \tag{1.2}$$

where $\langle n^0, \phi\rangle = \int n^0(y)\phi(y)dy$ and ϕ is the adjoint eigenvector (see [22]). The density N is the first eigenvector, and (λ, N) the unique solution of the following eigenvalue problem

$$\begin{cases} \kappa \dfrac{\partial}{\partial x}\big(g(x)N(x)\big) + \lambda N(x) = 4BN(2x) - BN(x), \quad x > 0, \\ \\ B(0)N(0) = 0, \qquad \int N(x)dx = 1, \qquad N(x) \geq 0, \qquad \lambda > 0. \end{cases} \tag{1.3}$$

Moreover, under some supplementary conditions, this convergence occurs exponentially fast (see [23]). Hence, in the rest of this article, we work under the following analytical assumptions.

Assumption 1 (Analytical assumptions)**.**

1. *For the considered nonnegative functions g and B and for $\kappa > 0$, there exists a unique eigenpair (λ, N) solution of Problem (1.3).*
2. *This solution satisfies, for all $p \geq 0$, $\int x^p N(x)dx < \infty$ and $0 < \int g(x)N(x)dx < \infty$.*
3. *The functions N and gN belong to $\mathcal{W}^{s+1}$ with $s \geq 1$, and in particular $\|N\|_\infty < \infty$ and $\|(gN)'\|_2 < \infty$. ($\mathcal{W}^{s+1}$ denotes the Sobolev space of regularity $s+1$ measured in $\mathbb{L}^2$-norm.)*
4. *We have $g \in \mathbb{L}^\infty(\mathbb{R}_+)$ with $\mathbb{R}_+ = [0, \infty)$.*

Hereafter $\|\bullet\|_2$ and $\|\bullet\|_\infty$ denote the usual $\mathbb{L}^2$ and $\mathbb{L}^\infty$ norms on $\mathbb{R}_+$. Assertions 1 and 2 are true under the assumptions on g and B stated in Theorem 1.1 of [3], under which we also have $N \in \mathbb{L}^\infty$. Assertion 3 is a (presumably reasonable) regularity assumption, necessary to obtain rates of convergence together with the convergence of the numerical scheme. Assertion 4 is restrictive, but mandatory in order to apply our statistical approach.

Thanks to this asymptotic behavior provided by the entropy principle (1.2), instead of requiring time-dependent data $n(t,x)$, which is experimentally less precise and more difficult to obtain, the inverse problem becomes: How to recover (κ, B) from observations on (λ, N) ? In [24, 5], as generally done in deterministic inverse problems (see [6]), it was supposed that experimental data were preprocessed into an approximation N_ε of N with an *a priori* estimate of the form $\|N - N_\varepsilon\| \leq \varepsilon$ for a suitable norm $\|\bullet\|$. Then, recovering B from N_ε becomes an inverse problem with a certain degree of ill-posedness. From a modelling point of view, this approach suffers from the limitation that knowledge on N is postulated in an abstract and somewhat arbitrary sense, that is not genuinely related to experimental measurements.

1.2. The statistical approach. In this paper, we propose to overcome the limitation of the deterministic inverse problems approach by assuming that we have n data, each data being obtained from the measurement of an individual cell picked at random, after the system has evolved for a long time so that the approximation $n(t,x) \approx N(x)e^{\lambda t}$ is valid. This is actually what happens if one observes cell cultures in laboratory after a few hours, a typical situation for *E. Coli* cultures for instance, provided, of course, that the underlying aggregation-fragmentation equation is valid.

Each data is viewed as the outcome of a random variable X_i, each X_i having probability distribution $N(x)dx$. We thus observe $(X_1, \ldots, X_n)$, with

$$\mathbb{P}(X_1 \in dx_1, \ldots, X_n \in dx_n) = \prod_{i=1}^n N(x_i)dx_i,$$

and where $\mathbb{P}(\bullet)$ hereafter denotes probability[1]. We assume for simplicity that the random variables X_i are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that they are stochastically independent. Our aim is to build an estimator of $B(x)$, that is a function $x \rightsquigarrow \hat{B}_n(x, X_1, \ldots, X_n)$ that approximates the true $B(x)$ with optimal accuracy and nonasymptotic estimates. To that end, consider the operator

$$(\lambda, N) \rightsquigarrow \mathfrak{T}(\lambda, N)(x) := \kappa \frac{\partial}{\partial x}\big(g(x)N(x)\big) + \lambda N(x), \quad x \geq 0. \tag{1.4}$$

From representation (1.3), we wish to find B, solution to $\mathfrak{T}(\lambda, N) = \mathcal{L}(BN)$, where

$$\mathcal{L}\big(\varphi\big)(x) := 4\varphi(2x) - \varphi(x), \tag{1.5}$$

based on statistical knowledge of (λ, N) only. Suppose that we have preliminary estimators $\hat{L}$ and $\hat{N}$ of respectively $\mathfrak{T}(\lambda, N)$ and N, and an approximation $\mathcal{L}_k^{-1}$ of $\mathcal{L}^{-1}$. Then we can reconstruct B in principle by setting formally

$$\widehat{B} := \frac{\mathcal{L}_k^{-1}(\hat{L})}{\widehat{N}}.$$

This leads us to distinguish three steps that we briefly describe here. The whole method is fully detailed in Section 2.

The first and principal step is to find an optimal estimator $\hat{L}$ for $\mathfrak{T}(\lambda, N)$. To do so, the main part consists in applying twice the Goldenschluger and Lepski's method [15] (GL for short). This method is a new version of the classical Lepski method [10, 11, 12, 13]. Both methods are adaptive to the regularity of the unknown signal and the GL method furthermore provides with an oracle inequality. For the unfamiliar reader, we discuss *adaptive properties* later on, and explain in details the *GL method* and *the oracle point of view* in Section 2.

1. First, we estimate the density N by a kernel method, based on a kernel function K. We define $\hat{N} = \hat{N}_{\hat{h}}$ where $\hat{N}_h$ is defined by (2.1) and the bandwidth $\hat{h}$ is selected automatically by (2.3) from a properly-chosen set $\mathcal{H}$. (see Section 2.1 for more details). A so-called oracle inequality is obtained in Proposition 2 measuring the quality of estimation of N by $\hat{N}$. Notice that this result, which is just a simplified version of [15], is valid for estimating any density, since we have only assumed to observe an $n-$sample of N, so that this result can be considered *per se.*
2. Second, we estimate the density derivative (up to g) $D = \frac{\partial}{\partial x}(gN)$, again by a kernel method with the same kernel K as before, and select an optimal bandwidth $\tilde{h}$ given by Formula (2.6) similarly. This defines an estimator $\hat{D} := \hat{D}_{\tilde{h}}$ where $\hat{D}_h$ is specified by (2.4), and yields an oracle inequality for $\hat{D}$ stated in Proposition 3. In the saemway as for N, this result has an interest *per se* and is not a direct consequence of [15].

From there, it only remains to find estimators of λ and κ. To that end, we make the following *a priori* (but presumably reasonable) Assumption 2 on the existence of an estimator $\hat{\lambda}_n$ of λ.

Assumption 2 (Assumption on $\hat{\lambda}_n$)**.** *For some $q > 1$ we have*

$$\varepsilon_{\lambda,n} = \big(\mathbb{E}[|\hat{\lambda}_n - \lambda|^q]\big)^{1/q} < \infty, \qquad R_{\lambda,n} = \mathbb{E}[\hat{\lambda}_n^{2q}] < \infty.$$

Indeed, in practical cell culture experiments, one can track n individual cells that have been picked at random through time. By looking at their evolution, it is possible to infer λ in a classical

[1] In the sequel, we denote by $\mathbb{E}(\bullet)$ the expectation operator with respect to $\mathbb{P}(\bullet)$ likewise.

parametric way, via an estimator $\hat{\lambda}_n$ that we shall assume to possess from now on[2]. Based on the following simple equality

$$\kappa = \lambda \rho_g(N) \text{ where } \rho_g(N) = \frac{\int_{\mathbb{R}_+} xN(x)dx}{\int_{\mathbb{R}_+} g(x)N(x)dx}, \tag{1.6}$$

obtained by multiplying (1.3) by x and integrating by part, we then define an estimator $\hat{\kappa}_n$ by (2.8). Finally, defining $\hat{L} = \hat{\kappa}_n \hat{D} + \hat{\lambda}_n \hat{N}$ ends this first step. The second step consists in the formal inversion of $\mathcal{L}$ and its numerical approximation: For this purpose, we follow the method proposed in [5] and recalled in Section 2.4. To estimate $H := BN$, we state

$$\hat{H} := \mathcal{L}_k^{-1}(\hat{L}) \tag{1.7}$$

where $\mathcal{L}_k^{-1}$ is defined by (2.10) on a given interval $[0, T]$. A new approximation result between $\mathcal{L}^{-1}$ and $\mathcal{L}_k^{-1}$ is given by Proposition 4. The third and final step consists in setting $\hat{B} := \frac{\hat{H}}{\hat{N}}$, clipping this estimator in order to avoid explosion when N becomes too small, finally obtaining

$$\tilde{B}(x) := \max(\min(\hat{B}(x), \sqrt{n}), -\sqrt{n}). \tag{1.8}$$

1.3. Rates of convergence. Because of the approximated inversion of $\mathcal{L}$ on $[0, T]$, we will have access to error bounds only on $[0, T]$. We set $\|f\|_{2,T}^2 = \int_0^T f^2(x)dx$. for the $\mathbb{L}^2$-norm restricted to the interval $[0, T]$. If the fundamental (yet technical) statistical result is the oracle inequality for $\hat{H}$ stated in Theorem 1 (see Section 2.5), the relevant part with respect to existing works in the non-stochastic setting [5, 24] is its consequence in terms of rates of convergence. For presenting them, we need to assume that the kernel K has regularity and vanishing moments properties.

Assumption 3 (Assumptions on K)**.** *The kernel K is differentiable with derivative K'. Furthermore, $\int K(x)dx = 1$ and $\|K\|_2$ and $\|K'\|_2$ are finite. Finally, there exists a positive integer m_0 such that $\int K(x)x^p dx = 0$ for $p = 1, \ldots, m_0 - 1$ and $I(m_0) := \int |x|^{m_0} K(x)dx$ is finite.*

Then our proposed estimators satisfy the following properties.

Proposition 1. *Under Assumptions 1, 2 and 3, let us assume that $R_{\lambda,n}$ and $\sqrt{n}\epsilon_{\lambda,n}$ are bounded uniformly in n and specify $\mathcal{L}_k^{-1}$ with $k = n$ f. Assume further that the family of bandwidth $\mathcal{H} = \tilde{\mathcal{H}} = \{D^{-1} : D = D_{\min}, ..., D_{\max}\}$ depends on n is such that $1 \leq D_{\min} \leq n^{1/(2m_0+1)}$ and $n^{1/5} \leq D_{\max} \leq n^{1/2}$ for all n. Then $\hat{H}$ satisfies, for all $s \in [1; m_0 - 1]$*

$$\mathbb{E}\left[\|\hat{H} - H\|_{2,T}^q\right] = O\big(n^{-\frac{qs}{2s+3}}\big), \tag{1.9}$$

Furthermore, if the kernel K is Lipschitz-regular, if there exists an interval $[a, b]$ in $(0, T)$ such that

$$[m, M] := [\inf_{x\in[a,b]} N(x), \sup_{x\in[a,b]} N(x)] \subset (0, \infty), \qquad Q := \sup_{x\in[a,b]} |H(x)| < \infty,$$

and if $\ln(n) \leq D_{\min} \leq n^{1/(2m_0+1)}$ and $n^{1/5} \leq D_{\max} \leq (n/\ln(n))^{1/(4+\eta)}$ for some $\eta > 0$, then $\hat{B}$ satisfies, for all $s \in [1, m_0 - 1]$,

$$\mathbb{E}\left[\|(\tilde{B} - B)1_{[a,b]}\|_2^q\right] = O\big(n^{-\frac{qs}{2s+3}}\big). \tag{1.10}$$

[2]Mathematically sepaking, this only amounts to enlarge the probability space to a rich enough structure that captures this estimator. We do not pursue that here.

1.4. Remarks and comparison to other works. 1) Let us establish formal correspondences between the methodology and results when recovering B from (1.3) from the point of view of statistics or PDE analysis. ¡ After renormalization, we obtain the rate $n^{-s/(2s+3)}$ for estimating B, and this corresponds to ill-posed inverse problems of order 1 in nonparametric statistics. We can make a parallel with additive deterministic noise following Nussbaum and Pereverzev [19] (see also [16] and the references therein). Suppose we have an approximate knowledge of N and λ up to deterministic errors $\zeta_1 \in \mathbb{L}^2$ and $\zeta_2 \in \mathbb{R}$ with noise level $\varepsilon > 0$: we observe

$$N_\varepsilon = N + \varepsilon\zeta_1, \quad \|\zeta_1\|_2 \leq 1, \tag{1.11}$$

and $\lambda_\varepsilon = \lambda + \varepsilon\zeta_2, \ \ |\zeta_2| \leq 1$. From the representation

$$B = \frac{\mathcal{L}^{-1}\mathfrak{T}(N,\lambda)}{N},$$

where $\mathfrak{T}(N,\lambda)$ is defined in (1.4), we have that the recovery of $\mathfrak{T}(N,\lambda)$ is ill-posed in the terminology of Wahba [26] for it involves the computation of the derivative of N. Since $\mathcal{L}$ is bounded with an inverse bounded in $\mathbb{L}^2$ and the dependence in λ is continuous, the overall inversion problem is ill-posed of degree $a = 1$. By classical inverse problem theory for linear cases[3], this means that if $N \in \mathcal{W}^s$, the optimal recovery rate in $\mathbb{L}^2$-error norm should be $\varepsilon^{s/(s+a)} = \varepsilon^{s/(s+1)}$ (see also the work of Doumic, Perthame and collaborators [24, 5]).

Suppose now that we replace the deterministic noise ζ_1 by a random Gaussian *white noise*: we observe

$$N_\varepsilon = N + \varepsilon\mathbb{B} \tag{1.12}$$

where $\mathbb{B}$ is a Gaussian white noise, *i.e.* a random distribution in $\mathcal{W}^{-1/2}$ that operates on test functions $\varphi \in \mathbb{L}^2$ and such that $\mathbb{B}(\varphi)$ is a centered Gaussian variable with variance $\|\varphi\|_2^2$. Model (1.12) serves as a representative toy model for most stochastic error models such as density estimation or signal recovery in the presence of noise. Let us formally introduce the α-fold integration operator $\mathcal{I}^\alpha$ and the derivation operator ∂. We can rewrite (1.12) as

$$N_\varepsilon = \mathcal{I}^1(\partial N) + \varepsilon\mathbb{B}$$

and applying $\mathcal{I}^{1/2}$ to both side, we (still formally) equivalently observe

$$Z_\varepsilon := \mathcal{I}^{1/2}N_\varepsilon = \mathcal{I}^{3/2}(\partial N) + \varepsilon\mathcal{I}^{1/2}\mathbb{B}.$$

We are back to a deterministic setting, since in this representation, we have that the noise $\varepsilon\mathcal{I}^{1/2}\mathbb{B}$ is in $\mathbb{L}^2$. In order to recover ∂N from Z_ε, we have to invert the operator $\mathcal{I}^{3/2}$, which has degree of ill-posedness $3/2$. We thus obtain the rate

$$\varepsilon^{s/(s+3/2)} = \varepsilon^{2s/(2s+3)} = n^{-s/(2s+3)}$$

for the calibration $\varepsilon = n^{-1/2}$ dictated by (1.12) when we compare our statistical model with the deterministic perturbation (see for instance [18] for establishing formally the correspondence $\varepsilon = n^{-1/2}$ is a general setting). This is exactly the rate we find in Proposition 1: the deterministic error model and the statistical error model coincide to that extent[4].

[3] although here the problem is nonlinear, but that will not affect the argument.

[4] The statistician reader willl note that the rate $n^{-s/(2s+3)}$ is also the minimax rate of convergence when estimating the derivative of a density, see [7].

2) The estimators $\hat{H}$ and $\hat{B}$ do not need the exact knowledge of s as an input to recover this optimal rate of convergence. We just need to know an upper bound $m_0 - 1$ to choose the regularity of the kernel K. This capacity to obtain the optimal rate without knowing the precise regularity is known in statistics as adaptivity in the minimax sense (see [25] for instance for more details). It is close in spirit to what the discrepancy principle can do in deterministic inverse problems [6]. However, in the deterministic framework, one needs to know the level of noise ε, which is not realistic in practice. In our statistical framework, this level of noise is linked to the size sample n through the correspondence $\varepsilon = n^{-1/2}$.

3) Finally, note that the rate is polynomial and no extra-logarithmic terms appear, as it is often the case when adaptive estimation is considered (see [10, 11, 12, 13]).

The next section explains in more details the GL approach and presents our estimators to a full extent, including the fundamental oracle inequalities. It also elaborates on the methodology related to oracle inequality. The main advantage of oracle inequalities is that they hold nonasymptotically (in n) and that they guarantee an optimal choice of bandwidth with respect to the selected risk. Section 3 is devoted to numerical simulations that illustrate the performance of our method. Proofs are delayed until Section 4.

2. Construction and properties of the estimators.

2.1. Estimation of N by the GL method. We first construct an estimator of N. A natural approach is a kernel method, which is all the more appropriate for comparisons with analytical methods (see [5] for the deterministic analogue). The kernel function K should satisfy the following assumption, in force in the sequel.

Assumption 4 (Assumption on the kernel density estimator)**.** *$K : \mathbb{R} \to \mathbb{R}$ is a continuous function such that $\int K(x)dx = 1$ and $\int K^2(x)dx < \infty$.*

For $h > 0$ and $x \in \mathbb{R}$, define

$$\hat{N}_h(x) := \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), \tag{2.1}$$

where $K_h(x) = h^{-1}K(h^{-1}x)$. Note in particular that $\mathbb{E}(\hat{N}_h) = K_h \star N$, where $\star$ denotes convolution. We measure the performance of $\hat{N}_h$ via its squared integrated error, *i.e.* the average $\mathbb{L}^2$ distance between N and $\hat{N}_h$. It is easy to see that

$$\mathbb{E}[\|N - \hat{N}_h\|_2] \leq \big\|N - K_h \star N\big\|_2 + \mathbb{E}[\|K_h \star N - \hat{N}_h\|_2],$$

with

$$\begin{aligned}
\mathbb{E}[\|K_h \star N - \hat{N}_h\|_2^2] &= \frac{1}{n^2}\,\mathbb{E}\Big[\int \Big[\sum_{i=1}^{n}\Big(K_h(x - X_i) - \mathbb{E}\left(K_h(x - X_i)\right)\Big)\Big]^2 dx\Big] \\
&= \frac{1}{n^2}\int \sum_{i=1}^{n}\mathbb{E}\Big[\Big(K_h(x - X_i) - \mathbb{E}\left(K_h(x - X_i)\right)\Big)^2\Big]dx \\
&\leq \frac{1}{n}\,\mathbb{E}\Big[\int K_h^2(x - X_1)dx\Big] = \frac{\big\|K_h\big\|_2^2}{n} = \frac{\big\|K\big\|_2^2}{nh}.
\end{aligned}$$

Applying the Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}[\|N - \hat{N}_h\|_2] \leq \|N - K_h \star N\|_2 + \frac{1}{\sqrt{nh}}\big\|K\big\|_2.$$

The first term corresponds to a bias term, it decreases when $h \to 0$. The second term corresponds to a variance term, which increases when $h \to 0$. If one has to choose h in a family $\mathcal{H}$ of possible bandwidths, the best choice is $\bar{h}$ where

$$\bar{h} := \mathrm{argmin}_{h\in\mathcal{H}} \big\{ \big\| N - K_h \star N \big\|_2 + \frac{1}{\sqrt{nh}} \big\| K \big\|_2 \big\}. \tag{2.2}$$

This ideal compromise $\bar{h}$ is called the "oracle": it depends on N and then cannot be used in practice. Hence one wants to find an automatic (data-driven) method for selecting this bandwidth. The Lepski method [10, 11, 12, 13] is one of the various theoretical adaptive methods available for selecting a density estimator. In particular it is the only known method able to select a bandwidth for kernel estimators. However the method do not usually provide a non asymptotic oracle inequality. Recently, Goldenschluger and Lepski [14] developed powerful probabilistic tools that enable to overcome this weakness and that can provide with a fully data-driven bandwidth selection method. We give here a practical illustration of their work: how should one select the bandwidth for a given kernel in dimension 1?

The main idea is to estimate the bias term by looking at several estimators. The method consists in setting first, for any x and any $h, h' > 0$,

$$\hat{N}_{h,h'}(x) := \frac{1}{n} \sum_{i=1}^{n} (K_h \star K_{h'})(x - X_i) = (K_h \star \hat{N}_{h'})(x).$$

Next, for any $h \in \mathcal{H}$, define

$$\begin{aligned} A(h) &:= \sup_{h'\in\mathcal{H}} \big\{ \|\hat{N}_{h,h'} - \hat{N}_{h'}\|_2 - \frac{\chi}{\sqrt{nh'}} \|K\|_2 \big\}_+ \\ &= \sup_{h'\in\mathcal{H}} \big\{ \max\big\{0, \|\hat{N}_{h,h'} - \hat{N}_{h'}\|_2 - \frac{\chi}{\sqrt{nh'}} \|K\|_2 \big\} \big\}, \end{aligned}$$

where, given $\varepsilon > 0$, we set $\chi := (1+\varepsilon)(1 + \|K\|_1)$. The quantity $A(h)$ is actually a good estimator of $\big\|N - K_h \star N\big\|_2$ up to the term $\big\|K\big\|_1$ (see (4.2) and (4.3) in Section 4). The next step consists then in setting

$$\hat{h} := \arg\min_{h\in\mathcal{H}} \big\{ A(h) + \frac{\chi}{\sqrt{nh}} \|K\|_2 \big\}, \tag{2.3}$$

and our final estimator of N is obtained by putting $\hat{N} := \hat{N}_{\hat{h}}$. Let us specify what we are able to prove at this stage.

Proposition 2. *Assume $N \in \mathbb{L}^\infty$ and work under Assumption 4. If $\mathcal{H} \subset \{D^{-1}, D = 1, \dots, D_{\max}\}$ with $D_{\max} = \delta n$ for $\delta > 0$, then, for any $q \geq 1$,*

$$\mathbb{E}\left[\|\hat{N} - N\|_2^{2q}\right] \leq C(q)\chi^{2q} \inf_{h\in\mathcal{H}} \big\{ \|K_h \star N - N\|_2^{2q} + \frac{\big\|K\big\|_2^{2q}}{(hn)^q} \big\} + C_1 n^{-q},$$

where $C(q)$ is a constant depending on q and C_1 is a constant depending on q, ε, δ, $\|K\|_2$, $\|K\|_1$ and $\|N\|_\infty$.

The previous inequality is called an oracle inequality, for we have $\mathbb{E}[\|\hat{N} - N\|_2] \leq (\mathbb{E}[\|\hat{N} - N\|_2^{2q}])^{1/(2q)}$ and $\hat{h}$ is performing as well as the oracle $\bar{h}$ up to some multiplicative constant. In that sense, we are able to select the *best bandwidth* within our family $\mathcal{H}$.

Remark 1. *As compared to the results of Goldenschluger and Lepski in [14], we do not consider the case where $\mathcal{H}$ is an interval and we do not specify K except for Assumption 4. This simpler method is more reasonable from a numerical point of view, since estimating N is only a preliminary*

step. The probabilistic tool we use here is classical in model selection theory (see Section 4 and [17]) and actually, we do not use directly [14]. In particular the main difference is that, in our specific case, we are able to get $\max(\mathcal{H})$ *fixed whereas Goldenschluger and Lepski [14] require* $\max(\mathcal{H})$ *to tend to 0 with* n*. The price to pay is that we obtain a uniform bound (see Lemma 1 in Section 4.3) which is less tight, but that will be sufficient for our purpose.*

2.2. Estimation of $\frac{\partial}{\partial x}\big(g(x)N(x)\big)$ by the GL method. The previous method can of course be adapted to estimate

$$D(x) := \frac{\partial}{\partial x}\big(g(x)N(x)\big).$$

We adjust here the work of [15] to the setting of estimating a derivative. We again use kernel estimators with more stringent assumptions[5] on K.

Assumption 5 (Assumption on the kernel of the derivative estimator)**.** *The function* K *is differentiable,* $\int K(x)dx = 1$ *and* $\int (K'(x))^2 dx < \infty$*.*

For any bandwidth $h > 0$, we define the kernel estimator of D as

$$\begin{array}{rcl} \hat{D}_h(x) & := & \frac{1}{n}\sum_{i=1}^n g(X_i)K'_h(x - X_i) \\ & = & \frac{1}{nh^2}\sum_{i=1}^n g(X_i)K'\big(\frac{x-X_i}{h}\big). \end{array} \tag{2.4}$$

Indeed

$$\begin{aligned} \mathbb{E}(\hat{D}_h(x)) &= \int K'_h(x-u)g(u)N(u)du \\ &= \big(K'_h \star (gN)\big)(x) = \big(K_h \star (gN)'\big)(x). \end{aligned}$$

Again we can look at the integrated squared error of $\hat{D}$. We obtain the following upper bound:

$$\mathbb{E}[\|\hat{D}_h - D\|_2] \le \|D - K_h \star D\|_2 + \mathbb{E}[\|K_h \star D - \hat{D}_h\|_2],$$

with

$$\begin{aligned} \mathbb{E}[\|K_h \star D - \hat{D}_h\|_2^2] &= \frac{1}{n^2}\,\mathbb{E}\Big[\int \Big[\sum_{i=1}^n \Big(g(X_i)K'_h(x-X_i) - \mathbb{E}\left(g(X_i)K'_h(x-X_i)\right)\Big)\Big]^2 dx\Big] \\ &= \frac{1}{n^2}\int \sum_{i=1}^n \mathbb{E}\Big[\Big(g(X_i)K'_h(x-X_i) - \mathbb{E}\left(g(X_i)K'_h(x-X_i)\right)\Big)^2\Big]dx \\ &\le \frac{1}{n}\,\mathbb{E}\big[\int g^2(X_1){K'_h}^2(x-X_1)dx\big] \\ &\le \frac{\|g\|_\infty^2\|K'_h\|_2^2}{n} = \frac{\|g\|_\infty^2\|K'\|_2^2}{nh^3}. \end{aligned}$$

Hence, by Cauchy-Schwarz inequality

$$\mathbb{E}[\|D - \hat{D}_h\|_2] \le \big\|D - K_h \star D\big\|_2 + \frac{1}{\sqrt{nh^3}}\big\|g\big\|_\infty\big\|K'\big\|_2.$$

Once again, there is a bias-variance decomposition, but now the variance term is of order $\frac{1}{\sqrt{nh^3}}\big\|K'\big\|_2\big\|g\big\|_\infty$. We therefore define the oracle by

$$\bar{\bar{h}} := \mathrm{argmin}_{h\in\tilde{\mathcal{H}}}\left\{\big\|D - K_h \star D\big\|_2 + \frac{1}{\sqrt{nh^3}}\big\|g\big\|_\infty\big\|K'\big\|_2\right\}. \tag{2.5}$$

[5] For sake of simplicity we use the same kernel to estimate N and D but this choice is not mandatory.

Now let us apply the GL method in this case. Let $\tilde{\mathcal{H}}$ be a family of bandwidths. We set for any $h, h' > 0$,

$$\hat{D}_{h,h'}(x) := \frac{1}{n}\sum_{i=1}^{n} g(X_i)(K_h \star K_{h'})'(x - X_i)$$

and

$$\tilde{A}(h) := \sup_{h' \in \tilde{\mathcal{H}}} \big\{ \|\hat{D}_{h,h'} - \hat{D}_{h'}\|_2 - \frac{\tilde{\chi}}{\sqrt{nh'^3}} \|g\|_\infty \|K'\|_2 \big\}_+,$$

where, given $\tilde{\varepsilon} > 0$, we put $\tilde{\chi} := (1 + \tilde{\varepsilon})(1 + \|K\|_1)$. Finally, we estimate D by using $\hat{D} := \hat{D}_{\tilde{h}}$ with

$$\tilde{h} := \mathrm{argmin}_{h \in \tilde{\mathcal{H}}} \big\{ \tilde{A}(h) + \frac{\tilde{\chi}}{\sqrt{nh^3}} \|g\|_\infty \|K'\|_2 \big\}. \tag{2.6}$$

As before, we are able to prove an oracle inequality for $\hat{D}$.

Proposition 3. *Assume $N \in \mathbb{L}^\infty$. Work under Assumption 5. If $\tilde{\mathcal{H}} = \{D^{-1}, D = 1 \ldots, \tilde{D}_{\max}\}$, with $\tilde{D}_{\max} = \sqrt{\tilde{\delta} n}$ for $\tilde{\delta} > 0$, then for any $q \geq 1$,*

$$\mathbb{E}\left[\|\hat{D} - D\|_2^{2q}\right] \leq \tilde{C}(q)\tilde{\chi}^{2q} \inf_{h \in \tilde{\mathcal{H}}} \big\{ \|K_h \star D - D\|_2^{2q} + \big(\frac{\|g\|_\infty \|K'\|_2}{\sqrt{nh^3}}\big)^{2q} \big\} + \tilde{C}_1 n^{-q},$$

where $\tilde{C}(q)$ is a constant depending on q and $\tilde{C}_1$ is a constant depending on q, $\tilde{\varepsilon}$, $\tilde{\delta}$, $\|K'\|_2$, $\|K'\|_1$, $\|g\|_\infty$ and $\|N\|_\infty$.

2.3. Estimation of κ (and λ). As mentioned in the introduction, we will not consider the problem of estimating λ and we work under Assumption 2: an estimator $\hat{\lambda}_n$ of λ is furnished by the practitioner prior to the data processing for estimating B. It becomes subsequently straightforward to obtain an estimator of κ by estimating $\rho_g(N)$, see the form of (1.6). We estimate $\rho_g(N)$ by

$$\hat{\rho}_n := \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n g(X_i) + c}, \tag{2.7}$$

where $c > 0$ is a (small) tuning constant[6]. Next we simply put

$$\hat{\kappa}_n = \hat{\lambda}_n \hat{\rho}_n. \tag{2.8}$$

2.4. Approximated inversion of $\mathcal{L}$. From (2.7), the right-hand side of (1.3) is consequently estimated by

$$\hat{\kappa}_n \hat{D} + \hat{\lambda}_n \hat{N}.$$

It remains to formally apply the inverse operator $\mathcal{L}^{-1}$. However Given φ, the dilation equation

$$\mathcal{L}\big(\psi\big)(x) = 4\psi(2x) - \psi(x) = \varphi(x), \quad x \in \mathbb{R}_+ \tag{2.9}$$

admits in general infinitely many solutions, see Doumic *et al.* [5], Appendix A. Nevertheless, if $\varphi \in \mathbb{L}^2$, there is a unique solution $\psi \in \mathbb{L}^2$ to (2.9), see Proposition A.1. in [5], and moreover it defines a continuous operator $\mathcal{L}^{-1}$ from $\mathbb{L}^2$ to $\mathbb{L}^2$. Since K_h and gK_h' belong to $\mathbb{L}^2$, one can define

[6] In practice, one can take $c = 0$.

a unique solution to (2.9) when $\varphi = \hat{\kappa}_n \hat{D} + \hat{\lambda}_n \hat{N}$. This inverse is not analytically known but we can only approximate it via the fast algorithm described below.

Given $T > 0$ and an integer $k \geq 1$, we construct a linear operator $\mathcal{L}_k^{-1}$ that maps a function $\varphi \in \mathbb{L}^2$ into a function with compact support in $[0, T]$ as follows. Consider the regular grid on $[0, T]$ with mesh $k^{-1}T$ defined by

$$0 = x_{0,k} < x_{1,k} < \cdots < x_{i,k} := \tfrac{i}{k} T < \ldots < x_{k,k} = T.$$

We set

$$\varphi_{i,k} := \frac{k}{T} \int_{x_i,k}^{x_{i+1,k}} \varphi(x) dx \ \text{ for } \ i = 0, \ldots, k-1,$$

and define by induction the sequence [7]

$$H_{i,k}(\varphi) := \frac{1}{4}(H_{i/2,k}(\varphi) + \varphi_{i/2,k}), \quad i = 0, \ldots, k-1,$$

what gives, for $i = 0$ and $i = 1$, $H_0(\varphi) := \frac{1}{3}\varphi_{0,k}$, and $H_1(\varphi) := \frac{4}{21}\varphi_{0,k} + \frac{1}{7}\varphi_{1,k}$. Finally, we define

$$\mathcal{L}_k^{-1}(\varphi)(x) := \sum_{i=0}^{k-1} H_{i,k}(\varphi) 1_{[x_{i,k}, x_{i+1,k})}(x). \tag{2.10}$$

As stated in the introduction, we eventually estimate $H = BN$ by

$$\hat{H} = \mathcal{L}_k^{-1}(\hat{\kappa}_n \hat{D} + \hat{\lambda}_n \hat{N}).$$

The stability of the inversion is given by the fact that $\mathcal{L}_k^{-1} : \mathbb{L}^2 \to \mathbb{L}^2$ is continuous, see Lemma 4 in Section 4.3, and by the following approximation result between $\mathcal{L}^{-1}$ and $\mathcal{L}_k^{-1}$.

Proposition 4. *Let $T > 0$ and $\varphi \in \mathcal{W}^1$. Let $\mathcal{L}^{-1}(\varphi)$ denote the unique solution of (2.9) belonging to $\mathbb{L}^2$. We have for $k \geq 1$:*

$$\|\mathcal{L}_k^{-1}(\varphi) - \mathcal{L}^{-1}(\varphi)\|_{2,T} \leq C \frac{T}{\sqrt{k}} \|\varphi\|_{\mathcal{W}^1}, \quad \text{with } C < \frac{1}{\sqrt{6}}.$$

Hence, $\mathcal{L}_k^{-1}$ behaves nicely over sufficiently smooth functions. Moreover the estimation of N and the estimators $\hat{\kappa}_n$ and $\hat{\lambda}_n$ are essentially regular. Finally we estimate B as stated in (1.8). The overall behaviour of the estimator is finally governed by the quality of estimation of the derivative D, which determines the accuracy of the whole inverse problem in all these successive steps.

2.5. Oracle inequalities for $\hat{H}$. We are ready to state our main result, namely the oracle inequality fulfilled by $\hat{H}$.

Theorem 1. *Work under Assumptions 1 and 2 and let K a kernel satisfying Assumptions 4 and 5. Define $\mathcal{H} \subset \{D^{-1}, D = 1 \ldots, D_{\max}\}$ with $D_{\max} = \delta n$ for $\delta > 0$ and $\tilde{\mathcal{H}} \subset \{D^{-1} : \ D = 1, ..., \tilde{D}_{\max}\}$ with $\tilde{D}_{\max} = \sqrt{\tilde{\delta} n}$ for $\tilde{\delta} > 0$. For $k \geq 1$ and $T > 0$, let us define $\mathcal{L}_k^{-1}$ by (2.10) on the interval $[0, T]$. Finally, define the estimator*

$$\hat{H} = \mathcal{L}_k^{-1}(\hat{\lambda}_n \hat{N}_{\hat{h}} + \hat{\kappa}_n \hat{D}_{\tilde{h}}),$$

where $\hat{N}_{\hat{h}}$ and $\hat{D}_{\tilde{h}}$ are kernel estimators defined respectively by (2.1) and (2.4), and where we have selected $\hat{h}$ and $\tilde{h}$ by (2.3) and (2.6). Moreover take $\hat{\kappa}_n$ as defined by (2.7) and (2.8) for some $c > 0$.

[7] for any sequence $u_i, i = 1, 2, \ldots$, let $u_{i/2} := u_{i/2}$ if i is even and $\frac{1}{2}(u_{(i-1)/2} + u_{(i+1)/2})$ otherwise.

The following upper bound holds for any n :

$$\mathbb{E}\left[\left\|\hat{H}-H\right\|_{2,T}^{q}\right] \leq C_1\Big\{\sqrt{R_{\lambda,n}}\inf_{h\in\tilde{\mathcal{H}}}\left[\|K_h\star D-D\|_2^q+\Big(\frac{\left\|g\right\|_\infty\|K'\|_2}{\sqrt{nh^3}}\Big)^q\right]$$
$$+\inf_{h\in\mathcal{H}}\left[\|K_h\star N-N\|_2^q+\Big(\frac{\left\|K\right\|_2}{\sqrt{nh}}\Big)^q\right]+\varepsilon_{\lambda,n}^q+\Big((\|N\|_{\mathcal{W}^1}+\|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\Big)^q\Big\}+C_2n^{-\frac{q}{2}},$$

where C_1 is a constant depending on q, g, N, ε, $\tilde{\varepsilon}$, $\left\|K\right\|_1$ and c; and C_2 is a constant depending on q, g, N, $\tilde{\varepsilon}$, ε, δ, $\tilde{\delta}$, $\left\|K\right\|_2$, $\left\|K\right\|_1$, $\left\|K'\right\|_2$, $\left\|K'\right\|_1$.

1) Note that the upper bound quantifies the additive cost of each step used in the estimation method. The first part is an oracle bound on the estimation of D times the size of the estimator of λ. The second part is the oracle bound for N. Of course, the results are sharp only if $\hat{\lambda}$ is good, which can be seen through $\varepsilon_{\lambda,n}$. Finally, the bound is also governed by the approximated inversion through the only term where k appears. The last term is just a residual term, that will be in most of the cases negligible with respect to the other terms. In particular since all the previous errors are somehow unavoidable, this means that, as far as our method is concerned, our upper bound is the best possible that can be achieved in order to select the different bandwidths h, up to multiplicative constants. Moreover, one can see how the limitation in k influences the method and how large k needs to be chosen to guarantee that the main error comes from the fact that one estimates a derivative.

2) The result holds for any n. In particular, we expect the method to perform well for small sample size, see the numerical illustration below. This also shows that we are able to select a good bandwidth as far as the kernel K is fixed, even if there is no assumption on the moments of K and consequently on the approximation properties of K. In the next simulation section, we focus on a Gaussian kernel which has only one vanishing moment (hence one cannot really consider minimax adaptation for regular function with it) but for which the bandwidth choice is still important in practice. The previous result guarantees an optimal bandwidth choice even for this kernel, up to some multiplicative constant.

3) From this oracle result, we can easily deduce the rates of convergence of Proposition 1 at the price of further assumptions on the kernel K, *i.e.* Assumption 3 defined in the Introduction. Section 1.2.2 of [25] recalls how to build compactly supported kernels satisfying Assumption 3. If m_0 satisfies Assumption 3, then for any $s \leq m_0$, for any $f \in \mathcal{W}^s$,

$$\|K_h \star f - f\|_2 \leq C\|f\|_{\mathcal{W}^s}h^s,$$

where C is a constant that can be expressed by using K and m_0 (see Theorem 8.1 of [9]). Now, it is sufficient to choose $h = D^{-1}$ of order $n^{-1/(2s+3)}$ to obtain (1.9). The complete proof of Proposition 1 is delayed until the last section.

3. Numerical illustration. Let us illustrate our method through some simulations.

3.1. The numerical protocol. First, we need to build up simulated data: to do so, we depart from given g, κ and B on a regular grid $[0, dx, \ldots, X_M]$ and solve the direct problem by the use of the so-called *power algorithm* to find the corresponding density N and the principal eigenvalue λ (see for instance [5] for more details). We check that N almost vanishes outside the interval $[0, X_M]$; else, X_M has to be increased. The density N being given on this grid, we approximate it by a spline function, and build a n-sample by the rejection sampling algorithm. For the sake of simplicity, we do not simulate an approximation on λ and keep the exact value, thus leading, in the estimate of Theorem 1, to $R_{\lambda,n} = \lambda^{2q}$ and $\hat{\lambda}_n = \lambda$.

We then follow step by step the method proposed here and detailed in Section 2.

1. The GL method for the choice of $\hat{N} = \hat{N}_{\bar{h}}$. We take the classical Gaussian kernel $K(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$, set $D_{\max} = n$ and limit ourselves to a logarithmic sampling $\mathcal{H} = \{1, 1/2, \ldots, 1/9, 1/10, 1/20, \ldots, 1/100, 1/200, \ldots, 1/n\}$ in order to reduce the cost of computations (The GL method is indeed the most time-consuming step in the numerical protocol).
2. The GL method for the choice of $\hat{D}_h$. The procedure is similar except that we choose here $D_{max} = \sqrt{n}$. The selected bandwidths $\bar{h}$ and $\tilde{h}$ can be different. We check that the GL method does not select an extremal point of $\mathcal{H}$.
3. The choice of $\hat{\kappa}_n$, as defined by (2.8).
4. The numerical scheme described in Section 2.4 and in [5] for the inversion of $\mathcal{L}$.
5. The division by $\hat{N}$ and definition of $\hat{B}$ as described in (1.8).

At each step, we compare, in $\mathbb{L}^2$-norm, the reconstructed function and the original one: $\hat{N}$ *vs* N, $\frac{\partial}{\partial x}(g\hat{N})$ *vs* $\frac{\partial}{\partial x}(gN)$, $\hat{H}$ *vs* BN and finally $\hat{B}$ *vs* B.

3.2. Results on simulated data. We first test the three cases simulated in [5] in which the numerical analysis approach was dealt with. Namely on the interval $[0, 4]$, we consider the cases where $g \equiv 1$ and first $B = B_1 \equiv 1$, second $B(x) = B_2(x) = 1$ for $x \leq 1.5$, then linear to $B_2(x) = 5$ for $x \geq 1.7$. This particular form is interesting because due to this fast increase on B, the solution N is not that regular and exhibits a 2-peaks distribution (see Figure 3.1). Finally, we test $B(x) = B_3(x) = \exp(-8(x-2)^2) + 1$.

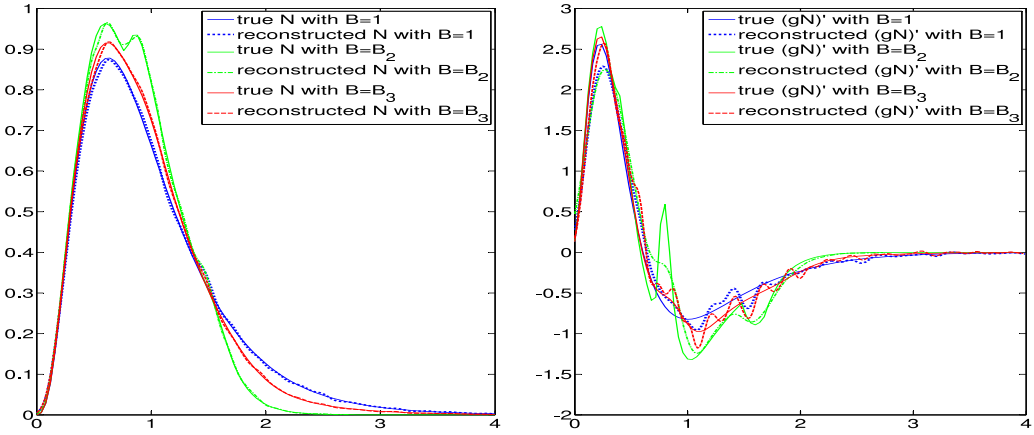

FIG. 3.1. *Reconstruction of N (left) and of $\frac{\partial}{\partial x}(gN)$ (right) obtained with a sample of $n = 5.10^4$ data, for three different cases of division rates B.*

In Figures 3.1 and 3.2, we show the simulation results with $n = 5.10^4$ (a realistic value for *in vitro* experiments on *E. Coli* for instance) for the reconstruction of N, $\frac{\partial}{\partial x}(gN)$, BN and B.

One notes that the solution can well capture the global behavior of the division rate B, but, as expected, has more difficulties in recovering fine details (for instance, the difference between B_1 and B_3) and also gives much more error when B is less regular (case of B_2). One also notes that even if the reconstruction of N is very satisfactory, the critical point is the reconstruction of its derivative. Moreover, for large values of x, even if N and its derivative are correctly reconstructed, the method fails in finding a proper division rate B. This is due to two facts: first, N vanishes, so the division by N leads to error amplification. Second, the values taken by $B(x)$ for large x have little influence on the solutions N of the direct problem: whatever the values of B, the solutions N will not vary much, as shown by Figure 3.1 (left). A similar phenomenon occurred indeed when solving

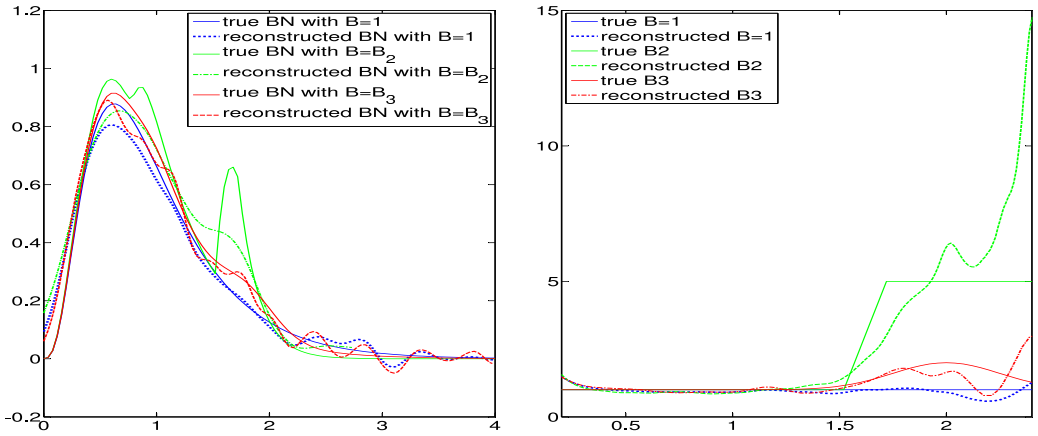

FIG. 3.2. *Reconstruction of BN (left) and of B (right) obtained with a sample of $n = 5.10^4$ data, for three different cases of division rates B.*

the deterministic problem in [5] (for instance, we refer to Fig. 10 of this article for a comparison of the results).

We also test a case closer to biological true data, namely the case $B(x) = x^2$ and $g(x) = x$. The results are shown on Figures 3.3 and 3.4 for n-samples of size 10^3, 5.10^3, 10^4 and 5.10^4.

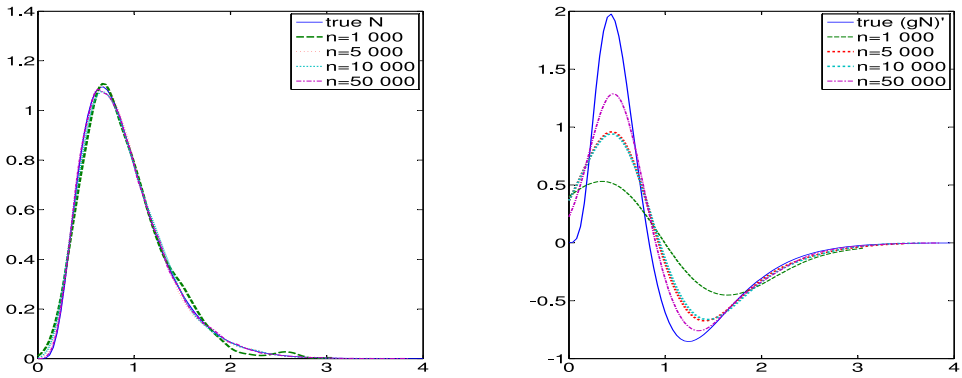

FIG. 3.3. *Reconstruction of N (left) and of $\frac{\partial}{\partial x}(gN)$ (right) obtained for $g(x) = x$ and $B(x) = x^2$, for various sample sizes.*

One notes that reconstruction is already very good for N when $n = 10^3$, unlike the reconstruction of $\frac{\partial}{\partial x}(gN)$ that requires much more data.

Finally, in Table 3.5 we give average error results on 50 simulations, for $n = 1000$, $g \equiv B \equiv 1$. We display the relative errors in $\mathbb{L}^2$ norms, (defined by $||\phi - \hat{\phi}||_{\mathbb{L}^2}/||\phi||_{\mathbb{L}^2}$), and their empirical variances. In Table 3.6, for the case $g(x) = x$ and $B(x) = x^2$, we give some results on standard errors for various values of n, and compare them to $n^{-1/5}$, which is the order of magnitude of the expected final error on BN, since with a Gaussian kernel we have $s = 1$ in Proposition 1. We see that our numerical results are in line with the theoretical estimates: indeed, the error on H is roughly twice as large as $n^{-1/5}$.

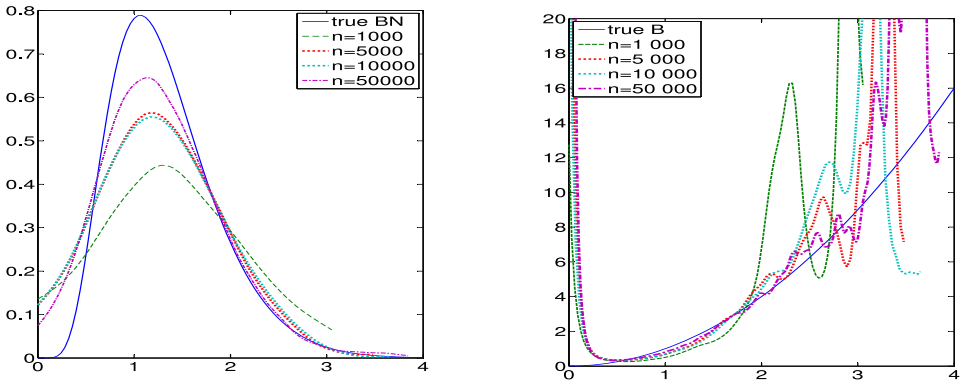

FIG. 3.4. *Reconstruction of BN (left) and of B (right) obtained for $g(x) = x$ and $B(x) = x^2$, for various sample sizes.*

Error on N: average	Variance	Error on $\frac{\partial}{\partial x}(gN)$: average	Variance
0.088	0.15	0.51	0.28
Error on BN: average	Variance	Average $\hat{h}$	Average $\tilde{h}$
0.39	0.29	0.12	0.40

FIG. 3.5. *Average error results on* 50 *simulations, for* $n = 1000$, $g \equiv B \equiv 1$.

4. Proofs. In Section 4.1, we first give the proofs of the main results of Section 2. This allows us, in Section 4.2, to prove the results of Section 2.5, which require the collection of all the results of Section 2, *i.e.* the oracle-type inequalities on the one hand and a numerical analysis result on the other hand. This illustrates the subject of our paper that lies at the frontier between these fields. Finally, we state the technical lemmas used in Section 4.1, whose proofs can be found in the companion paper [4]. These technical tools are concerned with probabilistic results, namely concentration and Rosenthal-type inequalities that are often the main bricks to establish oracle inequalities, and also the boundedness of $\mathcal{L}_k^{-1}$. In the sequel, the notation $\square_{\theta_1,\theta_2,\dots}$ denotes a generic positive constant depending on $\theta_1, \theta_2, \dots$ (the notation $\square$ simply denotes a generic positive absolute constant). It means that the values of $\square_{\theta_1,\theta_2,\dots}$ may change from line to line.

4.1. Proofs of the main results of Section 2.

Proof of Proposition 2. For any $h^* \in \mathcal{H}$, we have:

$$\begin{aligned}\|\hat{N} - N\|_2 &\le \|\hat{N}_{\hat{h}} - N_{\hat{h},h^*}\|_2 + \|N_{\hat{h},h^*} - \hat{N}_{h^*}\|_2 + \|\hat{N}_{h^*} - N\|_2 \\ &\le A_1 + A_2 + A_3,\end{aligned}$$

with

$$A_1 := \|\hat{N}_{\hat{h}} - N_{\hat{h},h^*}\|_2 \le A(h^*) + \frac{\chi}{\sqrt{n\hat{h}}}\|K\|_2,$$

$$A_2 := \|N_{\hat{h},h^*} - \hat{N}_{h^*}\|_2 \le A(\hat{h}) + \frac{\chi}{\sqrt{nh^*}}\|K\|_2$$

and

$$A_3 := \|\hat{N}_{h^*} - N\|_2.$$

n	$n^{-\frac{1}{5}}$	$\hat{h}$	$\tilde{h}$	error on N	error on D	error on H
10^3	0.25	0.1	0.5	0.06	0.68	0.42
5.10^3	0.18	0.07	0.3	0.03	0.45	0.28
10^4	0.16	0.08	0.3	0.035	0.46	0.29
5.10^4	0.11	0.04	0.2	0.014	0.31	0.19

FIG. 3.6. *Some results on standard errors for various values of n, case $g(x) = x$ and $B(x) = x^2$.*

We obtain

$$\begin{aligned}\|\hat{N} - N\|_2 &\leq A(h^*) + \frac{\chi}{\sqrt{n\hat{h}}}\|K\|_2 + A(\hat{h}) + \frac{\chi}{\sqrt{nh^*}}\|K\|_2 + \|\hat{N}_{h^*} - N\|_2 \\ &\leq 2A(h^*) + \frac{2\chi}{\sqrt{nh^*}}\|K\|_2 + \|\hat{N}_{h^*} - N\|_2. \end{aligned} \tag{4.1}$$

Since we have

$$\begin{aligned}A(h^*) &= \sup_{h' \in \mathcal{H}} \left\{\|\hat{N}_{h^*,h'} - \hat{N}_{h'}\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2\right\}_+ \\ &\leq \sup_{h' \in \mathcal{H}} \Big\{\Big\{\|\hat{N}_{h^*,h'} - \mathbb{E}[\hat{N}_{h^*,h'}] - \big(\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}]\big)\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2\Big\}_+ \\ &\qquad + \|\,\mathbb{E}[\hat{N}_{h^*,h'}] - \mathbb{E}[\hat{N}_{h'}]\|_2\Big\} \end{aligned} \tag{4.2}$$

and for any x and any h'

$$\begin{aligned}\mathbb{E}\left[\hat{N}_{h^*,h'}(x)) - \mathbb{E}(\hat{N}_{h'}(x)\right] &= \int (K_{h^*} \star K_{h'})(x-u)N(u)du - \int K_{h'}(x-v)N(v)dv \\ &= \int\int K_{h^*}(x-u-t)K_{h'}(t)N(u)dtdu - \int K_{h'}(x-v)N(v)dv \\ &= \int\int K_{h^*}(v-u)K_{h'}(x-v)N(u)dudv - \int K_{h'}(x-v)N(v)dv \\ &= \int K_{h'}(x-v)\Big(\int K_{h^*}(v-u)N(u)du - N(v)\Big)dv, \end{aligned}$$

we derive

$$\|\,\mathbb{E}(\hat{N}_{h^*,h'}) - \mathbb{E}(\hat{N}_{h'})\|_2 \leq \|K\|_1\|E_{h^*}\|_2, \tag{4.3}$$

where

$$E_{h^*}(x) := (K_{h^*} \star N)(x) - N(x), \quad x \in \mathbb{R}_+$$

represents the approximation term. Combining (4.1), (4.2) and (4.3) entails

$$\|\hat{N} - N\|_2 \leq \|\hat{N}_{h^*} - N\|_2 + 2\|K\|_1\|E_{h^*}\|_2 + \frac{2\chi}{\sqrt{nh^*}}\|K\|_2 + 2\zeta_n,$$

with

$$\begin{aligned}\zeta_n &:= \sup_{h' \in \mathcal{H}} \left\{\|\hat{N}_{h^*,h'} - \mathbb{E}[\hat{N}_{h^*,h'}] - \big(\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}]\big)\|_2 - \frac{\chi}{\sqrt{nh'}}\|K\|_2\right\}_+ \\ &= \sup_{h' \in \mathcal{H}} \left\{\|K_{h^*} \star \big(\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}]\big) - (\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}])\|_2 - \frac{(1+\varepsilon)(1+\|K\|_1)}{\sqrt{nh'}}\|K\|_2\right\}_+ \\ &\leq (1+\|K\|_1) \sup_{h' \in \mathcal{H}} \left\{\|\hat{N}_{h'} - \mathbb{E}[\hat{N}_{h'}]\|_2 - \frac{(1+\varepsilon)}{\sqrt{nh'}}\|K\|_2\right\}_+. \end{aligned}$$

Hence

$$\mathbb{E}\left[\|\hat{N}-N\|_2^{2q}\right] \le \square_q\big(\mathbb{E}\left[\|\hat{N}_{h^*}-N\|_2^{2q}\right] + \big\|K\big\|_1^{2q}\|E_{h^*}\|_2^{2q} + \chi^{2q}\frac{\|K\|_2^{2q}}{(nh^*)^q} + (1+\big\|K\big\|_1)^{2q}\,\mathbb{E}[\xi_n^{2q}]\big),$$

where

$$\xi_n = \sup_{h'\in\mathcal{H}}\big\{\|\hat{N}_{h'}-\mathbb{E}(\hat{N}_{h'})\|_2 - \frac{(1+\varepsilon)}{\sqrt{nh'}}\|K\|_2\big\}_+.$$

Now, we have:

$$\begin{aligned}\mathbb{E}\left[\|\hat{N}_{h^*}-N\|_2^{2q}\right] &\le 2^{2q-1}\big(\mathbb{E}\left[\|\hat{N}_{h^*}-\mathbb{E}[\hat{N}_{h^*}]\|_2^{2q}\right] + \|\,\mathbb{E}[\hat{N}_{h^*}) - N\|_2^{2q}\right]\\ &\le 2^{2q-1}\big(\mathbb{E}\left[\|\hat{N}_{h^*}-\mathbb{E}[\hat{N}_{h^*}]\|_2^{2q}\right] + \|E_{h^*}\|_2^{2q}\big).\end{aligned}$$

Then, by setting

$$Kc_{h^*}(X_i,x) := K_{h^*}(x-X_i) - \mathbb{E}(K_{h^*}(x-X_1)),$$

we obtain

$$\begin{aligned}\mathbb{E}\left[\|\hat{N}_{h^*}-\mathbb{E}[\hat{N}_{h^*}]\|_2^{2q}\right] &= \mathbb{E}\left[\Big(\int\Big(\frac{1}{n}\sum_{i=1}^n Kc_{h^*}(X_i,x)\Big)^2dx\Big)^q\right]\\ &\le \frac{2^{q-1}}{n^{2q}}\Big(\mathbb{E}\left[\Big(\sum_{i=1}^n\int Kc_{h^*}^2(X_i,x)dx\Big)^q\right]\\ &\qquad + \mathbb{E}\left[\Big|\sum_{1\le i,j\le n\ i\ne j}\int Kc_{h^*}(X_i,x)Kc_{h^*}(X_j,x)dx\Big|^q\right]\Big).\end{aligned}$$

Since

$$\begin{aligned}\int Kc_{h^*}^2(X_i,x)dx &= \int\Big(K_{h^*}(x-X_i) - \mathbb{E}\left[K_{h^*}(x-X_1)\right]\Big)^2dx\\ &\le 2\big(\int K_{h^*}^2(x-X_i)dx + \int\Big(\mathbb{E}\left[K_{h^*}(x-X_1)\right]\Big)^2dx\big)\\ &\le 2\big(\|K_{h^*}\|_2^2 + \int\mathbb{E}\left[K_{h^*}^2(x-X_1)\right]dx\big)\\ &\le 4\|K_{h^*}\|_2^2 = \frac{4}{h^*}\|K\|_2^2,\end{aligned}$$

the first term can be bounded as follows

$$\mathbb{E}\left[\big(\int\sum_{i=1}^n Kc_{h^*}^2(X_i,x)dx\big)^q\right] \le \big(\frac{4n}{h^*}\|K\|_2^2\big)^q.$$

For the second term, we apply Theorem 8.1.6 of de la Peña and Giné (1999) (with $2q \geq 2$) combined with the Cauchy-Schwarz inequality:

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\sum_{1\leq i,j\leq n\ i\neq j}\int Kc_{h^*}(X_i,x)Kc_{h^*}(X_j,x)dx\Big|^q\Big]\\
&\leq \Big(\mathbb{E}\Big[\Big|\sum_{1\leq i,j\leq n\ i\neq j}\int Kc_{h^*}(X_i,x)Kc_{h^*}(X_j,x)dx\Big|^{2q}\Big]\Big)^{\frac{1}{2}}\\
&\leq \square_q n^q\Big(\mathbb{E}\Big[\Big|\int Kc_{h^*}(X_1,x)Kc_{h^*}(X_2,x)dx\Big|^{2q}\Big]\Big)^{\frac{1}{2}}\\
&\leq \square_q n^q\Big(\mathbb{E}\Big[\Big|\int Kc^2_{h^*}(X_1,x)dx\Big|^{2q}\Big]\Big)^{\frac{1}{2}} \leq \square_q\big(\frac{4n}{h^*}\|K\|_2^2\big)^q.
\end{aligned}$$

It remains to deal with the term $\mathbb{E}(\xi_n^{2q})$. By Lemma 1 below, we obtain

$$\mathbb{E}[\xi_n^{2q}] \leq \square_{q,\eta,\delta}\big\|K\big\|_2,\big\|K\big\|_1,\big\|N\big\|_\infty\, n^{-q}$$

and the conclusion follows.

Proof of Proposition 3. The proof is similar to the previous one and we avoid most of the computations for simplicity. For any $h_0 \in \tilde{\mathcal{H}}$,

$$\begin{aligned}
\|\hat{D}_{\tilde{h}} - D\|_2 &\leq \|\hat{D}_{\tilde{h}} - \hat{D}_{\tilde{h},h_0}\|_2 + \|\hat{D}_{\tilde{h},h_0} - \hat{D}_{h_0}\|_2 + \|\hat{D}_{h_0} - D\|_2\\
&\leq \tilde{A}_1 + \tilde{A}_2 + \tilde{A}_3,
\end{aligned}$$

with

$$\tilde{A}_1 := \|\hat{D}_{\tilde{h}} - \hat{D}_{\tilde{h},h_0}\|_2 \leq \tilde{A}(h_0) + \frac{\tilde{\chi}}{\sqrt{n\tilde{h}^3}}\|g\|_\infty\|K'\|_2,$$

$$\tilde{A}_2 := \|\hat{D}_{\tilde{h},h_0} - \hat{D}_{h_0}\|_2 \leq \tilde{A}(\tilde{h}) + \frac{\tilde{\chi}}{\sqrt{nh_0^3}}\|g\|_\infty\|K'\|_2$$

and

$$\tilde{A}_3 := \|\hat{D}_{h_0} - D\|_2.$$

Then,

$$\|\hat{D}_{\tilde{h}} - D\|_2 \leq 2\tilde{A}(h_0) + \frac{2\tilde{\chi}}{\sqrt{nh_0^3}}\|g\|_\infty\|K'\|_2 + \|\hat{D}_{h_0} - D\|_2.$$

To study $\tilde{A}(h_0)$, we first evaluate

$$\begin{aligned}
\mathbb{E}[\hat{D}_{h_1,h_2}(x)] - \mathbb{E}[\hat{D}_{h_2}(x)]. &= (K_{h_1}\star K_{h_2}\star(gN)')(x) - (K_{h_2}\star(gN)')(x)\\
&= \int D(t)(K_{h_1}\star K_{h_2})(x-t)dt - \int D(t)K_{h_2}(x-t)dt\\
&= \int D(t)\int K_{h_1}(x-t-u)K_{h_2}(u)dudt - \int D(t)K_{h_2}(x-t)dt\\
&= \int D(t)\int K_{h_1}(v-t)K_{h_2}(x-v)dvdt - \int D(v)K_{h_2}(x-v)dv\\
&= \int K_{h_2}(x-v)\Big(\int D(t)K_{h_1}(v-t)dt - D(v)\Big)dv\\
&= (K_{h_2}\star\tilde{E}_{h_1})(x),
\end{aligned}$$

where we set, for any real number x

$$\begin{aligned}\tilde{E}_{h_1}(x) &:= \int D(t)K_{h_1}(x-t)dt - D(x)\\ &= (K_{h_1} \star D)(x) - D(x). \end{aligned} \tag{4.4}$$

It follows that

$$\begin{aligned}\tilde{A}(h_0) &= \sup_{h\in\tilde{\mathcal{H}}} \big\{\|\hat{D}_{h_0,h} - \hat{D}_h\|_2 - \frac{\tilde{\chi}}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\big\}_+\\ &\le \sup_{h\in\tilde{\mathcal{H}}} \Big\{\big\{\|\hat{D}_{h_0,h} - \mathbb{E}[\hat{D}_{h_0,h}] - \big(\hat{D}_h - \mathbb{E}[\hat{D}_h]\big)\|_2 - \frac{\tilde{\chi}}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\big\}_+\\ &\qquad +\|\,\mathbb{E}[\hat{D}_{h_0,h}] - \mathbb{E}[\hat{D}_h]\|_2\Big\}\\ &\le \sup_{h\in\tilde{\mathcal{H}}} \big\{\|\hat{D}_{h_0,h} - \mathbb{E}[\hat{D}_{h_0,h}] - (\hat{D}_h - \mathbb{E}[\hat{D}_h])\|_2 - \frac{\tilde{\chi}}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\big\}_+ + \|K\|_1\|\tilde{E}_{h_0}\|_2\\ &\le (1+\|K\|_1)\sup_{h\in\tilde{\mathcal{H}}} \big\{\|\hat{D}_h - \mathbb{E}[\hat{D}_h]\|_2 - \frac{(1+\tilde{\varepsilon})}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\big\}_+ + \|K\|_1\|\tilde{E}_{h_0}\|_2, \end{aligned} \tag{4.5}$$

In order to obtain the last line, we use the following chain of arguments:

$$\begin{aligned}\hat{D}_{h_0,h}(x) &= \frac{1}{n}\sum_{i=1}^n g(X_i)\int K'_h(x-X_i-t)K_{h_0}(t)dt\\ &= \int K_{h_0}(t)\Big(\frac{1}{n}\sum_{i=1}^n g(X_i)K'_h(x-X_i-t)\Big)dt\end{aligned}$$

and

$$\mathbb{E}\left[\hat{D}_{h_0,h}(x)\right] = \int K_{h_0}(t)\Big(\int g(u)K'_h(x-u-t)N(u)du\Big)dt,$$

therefore

$$\hat{D}_{h_0,h}(x) - \mathbb{E}\left[\hat{D}_{h_0,h}(x)\right] = \int K_{h_0}(t)G(x-t)dt = K_{h_0}\star G(x),$$

with

$$\begin{aligned}G(x) &= \frac{1}{n}\sum_{i=1}^n g(X_i)K'_h(x-X_i) - \int g(u)K'_h(x-u)N(u)du\\ &= \hat{D}_h(x) - \mathbb{E}\left[\hat{D}_h(x)\right].\end{aligned}$$

Therefore

$$\begin{aligned}\|\hat{D}_{h_0,h} - \mathbb{E}[\hat{D}_{h_0,h}]\|_2 &\le \|K_{h_0}\|_1\|G\|_2\\ &\le \|K\|_1\|\hat{D}_h - \mathbb{E}[\hat{D}_h]\|_2,\end{aligned}$$

which justifies (4.5). In the same way as in the proof of Proposition 2, we can establish the following:

$$\mathbb{E}\left[\|\tilde{E}_{h_0}\|_2^{2q}\right] = \mathbb{E}\left[\|\hat{D}_{h_0} - D\|_2^{2q}\right] \le \square_q\Big(\|\tilde{E}_{h_0}\|_2^{2q} + \Big(\frac{\|g\|_\infty\|K'\|_2}{\sqrt{nh_0^3}}\Big)^{2q}\Big).$$

Finally, we successively apply (4.4), (4.5) and Lemma 1 in order to conclude the proof.

Proof of Proposition 4. We use the notation and definitions of Section 2.4. We have

$$\|\mathcal{L}_k^{-1}(\varphi) - \mathcal{L}^{-1}(\varphi)\|_{2,T}^2 = \sum_{i=0}^{k-1} \int_{x_{i,k}}^{x_{i+1,k}} \big(H_{i,k} - \mathcal{L}^{-1}(\varphi)(x)\big)^2 dx := \sum_{i=0}^{k-1} L_{i,k}.$$

We prove by induction that for all i, one has $L_{i,k} \leq C^2 \frac{T^2}{k^2} \|\varphi\|_{\mathcal{W}^1}^2$. The result follows by summation over i. We first prove the two following estimates:

$$\int_{x_{i,k}}^{x_{i+1,k}} \big(\varphi_{i,k} - \varphi(x)\big)^2 dx \leq \frac{T^2}{4\pi^2 k^2} \|\varphi\|_{\mathcal{W}^1}^2, \tag{4.6}$$

$$|\varphi_{i+1,k} - \varphi_{i,k}|^2 \leq \frac{T}{k} \|\varphi\|_{\mathcal{W}^1}^2. \tag{4.7}$$

By definition, $\varphi_{i,k}$ is the average of the function φ on the interval $[x_{i,k},\ x_{i+1,k}]$ of size $\frac{T}{k}$. Thus (4.7) is simply Wirtinger inequality applied to $\varphi \in \mathcal{W}^1$ on the interval $[x_{i,k},\ x_{i+1,k}]$.*For*(4.7), we use the Cauchy-Schwarz inequality:

$$\begin{aligned}
|\varphi_{i+1,k} - \varphi_{i,k}|^2 &= \frac{k^2}{T^2} \Big(\int_{x_{i,k}}^{x_{i+1,k}} \big(\varphi(x + \frac{T}{k}) - \varphi(x)\big) dx \Big)^2 = \frac{k^2}{T^2} \Big(\int_{x_{i,k}}^{x_{i+1,k}} \int_{x}^{x+\frac{T}{k}} \varphi'(z) dz\, dx \Big)^2 \\
&\leq \frac{k^2}{T^2} \Big(\int_{x_{i,k}}^{x_{i+1,k}} \sqrt{\frac{T}{k}} \|\varphi\|_{\mathcal{W}^1} dx \Big)^2 = \frac{T}{k} \|\varphi\|_{\mathcal{W}^1}^2.
\end{aligned}$$

We are ready to prove by induction the two following inequalities:

$$L_{i,k} \leq C_1^2 \frac{T^2}{k^2} \|\varphi\|_{\mathcal{W}^1}^2, \tag{4.8}$$

$$|H_{i+1,k} - H_{i,k}|^2 \leq C_2^2 \frac{T}{k} \|\varphi\|_{\mathcal{W}^1}^2. \tag{4.9}$$

for two constants C_1 and C_2 specified later on. First, for $i = 0$, we have

$$L_{0,k} = \int_0^{\frac{T}{k}} |H_{0,k}(\varphi) - \mathcal{L}^{-1}(\varphi)(x)|^2 dx = \int_0^{\frac{T}{k}} \big|\frac{1}{3}\varphi_{0,k} - \mathcal{L}^{-1}(\varphi)(x)\big|^2 dx.$$

We recall (see Proposition A.1. of [5]) that $\mathcal{L}^{-1}(\varphi)(x) = \sum_{n=1}^{\infty} 2^{-2n} \varphi(2^{-n}x)$, and we use the fact that $\frac{1}{3} = \sum_{n=1}^{\infty} 2^{-2n}$ and for $a, b > 0$, $ab \leq \frac{1}{2}(a^2 + b^2)$ in order to write

$$\begin{aligned}
L_{0,k} &= \int_0^{\frac{T}{k}} \big| \sum_{n=1}^{\infty} 2^{-2n} \big(\varphi_{0,k} - \varphi(2^{-n}x)\big) \big|^2 dx \leq \sum_{n,n'=1}^{\infty} 2^{-2n-2n'} \int_0^{\frac{T}{k}} |\varphi_{0,k} - \varphi(2^{-n}x)|^2 dx \\
&\leq \frac{1}{3} \sum_{n=1}^{\infty} 2^{-n} \int_0^{2^{-n}\frac{T}{k}} |\varphi_{0,k} - \varphi(y)|^2 dy \leq \frac{1}{3} \frac{T^2}{4\pi^2 k^2} \|\varphi\|_{\mathcal{W}^1}^2.
\end{aligned}$$

This proves the first induction assumption for $i = 0$, and

$$|H_{1,k} - H_{0,k}|^2 = \big|\frac{1}{7}(\varphi_{1,k} - \varphi_{0,k})\big|^2 \leq \frac{1}{7^2}\frac{T}{k}\|\varphi\|^2_{\mathcal{W}^1},$$

proves the second one. Let us now suppose that the two induction assumptions are true for all $j \leq i-1$, and take $i \geq 1$. Let us first evaluate

$$L_{i,k} = \int_{x_{i,k}}^{x_{i+1,k}} \big(H_{i,k} - \mathcal{L}^{-1}(\varphi)(x)\big)^2 dx = \frac{1}{16}\int_{x_{i,k}}^{x_{i+1,k}} \big(H_{\frac{i}{2},k} + \varphi_{\frac{i}{2},k} - \mathcal{L}^{-1}(\varphi)(\frac{x}{2}) - \varphi(\frac{x}{2})\big)^2 dx.$$

We distinguish the case when i is even and when i is odd. Let i be even: then, by definition

$$L_{i,k} \leq \frac{1}{8}\int_{x_{i,k}}^{x_{i+1,k}} \big(H_{\frac{i}{2},k} - \mathcal{L}^{-1}(\varphi)(\frac{x}{2})\big)^2 dx + \frac{1}{8}\int_{x_{i,k}}^{x_{i+1,k}} \big(\varphi_{\frac{i}{2},k} - \varphi(\frac{x}{2})\big)^2 dx \leq \frac{1}{4}(C_1^2 + \frac{1}{4\pi^2})\frac{T^2}{k^2}\|\varphi\|^2_{\mathcal{W}^1}$$

by the induction assumption and Assertion 4.7 on φ for $j = \frac{i}{2}$. If i is odd, we write by definition

$$L_{i,k} = \frac{1}{16}\int_{x_{i,k}}^{x_{i+1,k}} \left(H_{\frac{i-1}{2},k} + \varphi_{\frac{i-1}{2},k} - \mathcal{L}^{-1}(\varphi)(\frac{x}{2}) - \varphi(\frac{x}{2}) + \frac{1}{2}(H_{\frac{i+1}{2},k} - H_{\frac{i-1}{2},k}) + \frac{1}{2}(\varphi_{\frac{i+1}{2},k} - \varphi_{\frac{i-1}{2},k})\right)^2 dx.$$

Hence, re-organizing terms, we can write

$$\begin{aligned} L_{i,k} \leq{}& \frac{1}{2}\int_{x_{\frac{i-1}{2},k}}^{x_{\frac{i-1}{2}+1,k}} \left(H_{\frac{i-1}{2},k} - \mathcal{L}^{-1}(\varphi)(y)\right)^2 dy + \frac{1}{2}\int_{x_{\frac{i-1}{2},k}}^{x_{\frac{i-1}{2}+1,k}} \left(\varphi_{\frac{i-1}{2},k} - \varphi(y)\right)^2 dy \\ &+ \frac{1}{16}\frac{T}{k}(H_{\frac{i+1}{2},k} - H_{\frac{i-1}{2},k})^2 + \frac{1}{16}\frac{T}{k}(\varphi_{\frac{i+1}{2},k} - \varphi_{\frac{i-1}{2},k})^2 \end{aligned}$$

Putting together the four inequalities above (the estimates for φ and the induction assumptions), we obtain

$$L_{i,k} \leq \frac{T^2}{k^2}\|\varphi\|^2_{\mathcal{W}^1}\left(\frac{C_1^2}{2} + \frac{1}{8\pi^2} + \frac{C_2^2}{16} + \frac{1}{16}\right)$$

and (4.8) is proved. It remains to establishe (4.9). Let us write it for i even (the case of an odd i is similar):

$$|H_{i+1,k} - H_{i,k}|^2 = \frac{1}{16}|H_{\frac{i+1}{2}} - H_{\frac{i}{2}} + \varphi_{\frac{i+1}{2}} - \varphi_{\frac{i}{2}}|^2 = \frac{1}{32}|H_{\frac{i}{2}+1} - H_{\frac{i}{2}} + \varphi_{\frac{i}{2}+1} - \varphi_{\frac{i}{2}}|^2.$$

Hence, as previously, we obtain

$$|H_{i+1,k} - H_{i,k}|^2 \leq \frac{1}{16}\frac{T}{k}\|\varphi\|^2_{\mathcal{W}^1}(C_2^2 + 1).$$

To complete the proof, we remark that $C_2^2 = \frac{1}{15}$ and $C_1^2 = \frac{1}{4\pi^2} + \frac{1}{8}(1 + \frac{1}{15}) < \frac{1}{6}$ are suitable. It is consequently sufficient to take $C = C_1$.

4.2. Proof of Theorem 1 and Proposition 1.

Proof of Theorem 1. It is easy to see that

$$\begin{aligned}\|\widehat{H}-H\|_{2,T} &= \|\mathcal{L}_k^{-1}(\hat{\kappa}\hat{D}+\hat{\lambda}_n\hat{N})-\mathcal{L}^{-1}(\mathcal{L}(BN))\|_{2,T}\\ &\le \|\mathcal{L}_k^{-1}(\hat{\kappa}\hat{D}+\hat{\lambda}_n\hat{N})-\mathcal{L}_k^{-1}(\mathcal{L}(BN))\|_{2,T}\\ &\qquad +\|\mathcal{L}_k^{-1}(\mathcal{L}(BN))-\mathcal{L}^{-1}(\mathcal{L}(BN))\|_{2,T}\\ &\le \|\mathcal{L}_k^{-1}(\hat{\kappa}\hat{D}+\hat{\lambda}_n\hat{N}-(\kappa D+\lambda N))\|_{2,T}\\ &\qquad +\frac{1}{3}\frac{T}{\sqrt{k}}\|\mathcal{L}(BN)\|_{\mathcal{W}^1},\end{aligned}$$

thanks to Proposition 4. Note that $\mathcal{L}(BN)=\kappa(gN)'+\lambda N$ so that we can write

$$\|\mathcal{L}(BN)\|_{\mathcal{W}^1}\le C(\|N\|_{\mathcal{W}^1}+\|gN\|_{\mathcal{W}^2}).$$

We obtain, thanks to Lemma 4 that gives the boundedness of the operator $\mathcal{L}_k^{-1}$:

$$\begin{aligned}\|\widehat{H}-H\|_{2,T} &\le \square\big(\big\|\hat{\kappa}_n\hat{D}-\kappa D\big\|_{2,T}+\big\|\hat{\lambda}_n\hat{N}-\lambda N\big\|_{2,T}+(\|N\|_{\mathcal{W}^1}+\|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\big)\\ &\le \square\big(|\hat{\kappa}_n|\|\hat{D}-D\|_2+|\hat{\lambda}_n|\|\hat{N}-N\|_2+|\hat{\kappa}_n-\kappa|\|D\|_2+|\hat{\lambda}_n-\lambda|\|N\|_2\\ &\qquad +(\|N\|_{\mathcal{W}^1}+\|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\big)\\ &\le \square\big(|\hat{\lambda}_n||\hat{\rho}_n|\|\hat{D}-D\|_2+|\hat{\lambda}_n|\big(\|\hat{N}-N\|_2+|\hat{\rho}_n-\rho_g(N)|\|D\|_2\big)\\ &\qquad +\big(\|N\|_2+|\rho_g(N)|\|D\|_2\big)|\hat{\lambda}_n-\lambda|+(\|N\|_{\mathcal{W}^1}+\|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\big).\end{aligned}$$

Taking expectation and using Cauchy-Schwarz inequality, we obtain for any $q\ge 1$,

$$\begin{aligned}\mathbb{E}[\|\widehat{H}-H\|_{2,T}^q] \le \square_q\Big[\big(\mathbb{E}[\hat{\lambda}_n^{2q}]\big)^{1/2}\Big\{\big(\mathbb{E}[\hat{\rho}_n^{4q}]\big)^{1/4}\big(\mathbb{E}[\|\hat{D}-D\|_2^{4q}]\big)^{1/4}+\big(\mathbb{E}[\|\hat{N}-N\|_2^{2q}]\big)^{1/2}\\ +\|D\|_2^q\big(\mathbb{E}[|\hat{\rho}_n-\rho_g(N)|^{2q}]\big)^{1/2}\Big\}\\ +\big(\|N\|_2+\rho_g(N)\big\|D\big\|_2\big)^q\,\mathbb{E}[|\hat{\lambda}_n-\lambda|^q]+\big((\|N\|_{\mathcal{W}^1}+\|gN\|_{\mathcal{W}^2})\frac{T}{\sqrt{k}}\big)^q\Big].\end{aligned}$$

Now, Lemma 3 gives the behaviour of $\mathbb{E}[|\hat{\rho}_n-\rho_g(N)|^{2q}]$. In particular, we obtain

$$\mathbb{E}[\hat{\rho}_n^{4q}]\le\square_{q,g,N,c}.$$

We finally apply successively Propositions 2 and 3 to obtain the proof of Theorem 1.

Proof of Proposition 1. We have already proved (1.9). It remains to prove (1.10). We introduce the event

$$\Omega_n=\{2\hat{N}(x)\ge m \text{ for any } x\in[a,b]\}.$$

Then, for n larger that Q^2,

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\int_a^b \big(\tilde{B}(x)-B(x)\big)^2 dx\Big)^{\frac{q}{2}}\Big] \\
&= \mathbb{E}\Big[\Big(\int_a^b \big(\tilde{B}(x)-B(x)\big)^2 dx \times \mathbf{1}_{\Omega_n}\Big)^{\frac{q}{2}}\Big] + \mathbb{E}\Big[\Big(\int_a^b \big(\tilde{B}(x)-B(x)\big)^2 dx \times \mathbf{1}_{\Omega_n^c}\Big)^{\frac{q}{2}}\Big] \\
&\le \mathbb{E}\Big[\Big(\int_a^b \big(\hat{B}(x)-B(x)\big)^2 dx \times \mathbf{1}_{\Omega_n}\Big)^{\frac{q}{2}}\Big] + \big(2(b-a)(n+Q^2)\big)^{\frac{q}{2}}\,\mathbb{P}(\Omega_n^c) \\
&\le \mathbb{E}\Big[\Big(\int_a^b \Big(\frac{\hat{H}}{\hat{N}}-\frac{H}{N}\Big)^2 \times \mathbf{1}_{\Omega_n}\Big)^{\frac{q}{2}}\Big] + \big(4n(b-a)\big)^{\frac{q}{2}}\,\mathbb{P}(\Omega_n^c) \\
&\le \mathbb{E}\Big[\Big(\int_a^b \Big(\frac{\hat{H}N-\hat{N}H}{\hat{N}N}\Big)^2 \times \mathbf{1}_{\Omega_n}\Big)^{\frac{q}{2}}\Big] + \big(4n(b-a)\big)^{\frac{q}{2}}\,\mathbb{P}(\Omega_n^c) \\
&\le \square_{q,m,M,Q}\big(\mathbb{E}\big[\big\|\hat{H}-H\big\|_2^q\big] + \mathbb{E}\big[\big\|\hat{N}-N\big\|_2^q\big]\big) + \big(4n(b-a)\big)^{\frac{q}{2}}\,\mathbb{P}(\Omega_n^c).
\end{aligned}$$

The first term of the right hand side is handled by (1.9) and Proposition 2. The second term is handled by Lemma 2 that establishes that $\mathbb{P}(\Omega_n^c) = O(n^{-q})$.

4.3. Technical lemmas. We state a series of technical results, proof of which we omit. The interested reader may find detailed proofs in the preprint [4]. The first ingredient is a concentration inequality. Note that a more general version of this result can be found in [14].

Lemma 1 (A concentration result)**.** *We have the following estimates*

- *Assume that* $\|K\|_2, \|K\|_1, \|g\|_\infty$ *and* $\|N\|_\infty$ *are finite. For every* $q>0$*, introduce the grid* $\mathcal{H} \subset \{1, 1/2, ..., 1/D_{\max}\}$ *and* $D_{\max} = \delta n$ *for some* $\delta > 0$*. Then, for every* $\eta > 0$

$$\mathbb{E}\Big[\sup_{h\in\mathcal{H}}\Big\{\|\hat{N}_h - \mathbb{E}[\hat{N}_h]\|_2 - \frac{(1+\eta)}{\sqrt{nh}}\|K\|_2\Big\}_+^{2q}\Big] \le \square_{q,\eta,\delta\|K\|_2,\|K\|_1,\|N\|_\infty} n^{-q}.$$

- *Assume that* $\|K'\|_2, \|K'\|_1, \|g\|_\infty$ *and* $\|N\|_\infty$ *are finite. For every* $q>0$*, introduce the grid* $\tilde{\mathcal{H}} \subset \{1, 1/2, ..., 1/\tilde{D}_{\max}\}$ *and* $\tilde{D}_{\max} = \sqrt{\tilde{\delta} n}$ *for some* $\tilde{\delta} > 0$*. Then for every* $\eta > 0$

$$\mathbb{E}\Big[\sup_{h\in\tilde{\mathcal{H}}}\Big\{\|\hat{D}_h - \mathbb{E}[\hat{D}_h]\|_2 - \frac{(1+\eta)}{\sqrt{nh^3}}\|g\|_\infty\|K'\|_2\Big\}_+^{2q}\Big] \le \square_{q,\eta,\tilde{\delta},\|K'\|_2,\|K'\|_1,\|g\|_\infty,\|N\|_\infty} n^{-q}.$$

The second result is based on probabilistic arguments as well.

Lemma 2. *Under Assumptions and notations of Proposition 1, if there exists an interval* $[a,b]$ *in* $(0,T)$ *such that*

$$[m,M] := [\inf_{x\in[a,b]} N(x), \sup_{x\in[a,b]} N(x)] \subset (0,\infty), \qquad Q := \sup_{x\in[a,b]} |H(x)| < \infty,$$

and if $\ln(n) \le D_{\min} \le n^{1/(2m_0+1)}$ *and* $n^{1/5} \le D_{\max} \le (n/\ln(n))^{1/(4+\eta)}$ *for some* $\eta > 0$ *fixed, then there exists* C_η*, a constant depending on* η*, such that for* n *large enough,*

$$\mathbb{P}(\Omega_n^c) \le C_\eta n^{-q}.$$

The third technical ingredient studies the behavior of the moments of $\hat{\rho}_n$. It is based on the Rosenthal inequality.

Lemma 3 (Rosenthal-type inequality)**.** *For any* $p \ge 2$*,*

$$\mathbb{E}\big[|\hat{\rho}_n - \rho_g(N)|^p\big] \le \square_{p,g,N,c} n^{-p/2}.$$

Our final technical result is the boundedness of the approximation operator $\mathcal{L}_k^{-1}$.

Lemma 4. *For any function φ, we have:*

$$\left\|\mathcal{L}_k^{-1}(\varphi)\right\|_{2,T} \leq \sqrt{\frac{1}{3}}\|\varphi\|_{2,T}.$$

Acknowledgment: We warmly thank Oleg Lepski for fruitful discussions that eventually led to the choice of his method for the purpose of this article.

REFERENCES

[1] H. T. Banks, Karyn L. Sutton, W. Clayton Thompson, Gennady Bocharov, Dirk Roose, Tim Schenkel and Andreas Meyerhans, *Estimation of Cell Proliferation Dynamics Using CFSE Data*, Bull. of Math. Biol., doi: 10.1007/s11538-010-9524-5.

[2] Baraud, Y. *A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression.* to appear in Bernoulli.

[3] Doumic, M. and Gabriel, P. (2010) *Eigenelements of a General Aggregation-Fragmentation Model.* Math. Models Methods Appl. Sci. 20 (2010), no. 5, 757–783.

[4] Doumic, M., Hoffmann, M., Reynaud-Bouret, P. and Rivoirard, V. (2011) *Nonparametric estimation of the division rate of a size-structured population.* ArXiv preprint.

[5] Doumic, M., Perthame, B. and Zubelli, J. (2009) *Numerical Solution of an Inverse Problem in Size-Structured Population Dynamics.* Inverse Problems, **25**, 25pp.

[6] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Springer Verlag, 1996.

[7] Gasser, T. and Müller, H.G. (1979) *Optimal convergence properties of kernel estimates of derivatives of a density function.* In Lecture Notes in Mathematics **757**. Springer, Berlin, 144–154.

[8] Hall, P., Heyde, C.C., (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.

[9] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov A. (1998) *Wavelets, Approximation and Statistical Applications.* Springer-Verlag, Berlin.

[10] Lepski, O. V. (1990). *One problem of adaptive estimation in Gaussian white noise.* Theory Probab. Appl. **35** 459–470.

[11] Lepski, O. V. (1991). *Asymptotic minimax adaptive estimation. 1. Upper bounds.* Theory Probab. Appl. **36** 645–659.

[12] Lepski, O. V. (1992). *Asymptotic minimax adaptive estimation. 2. Statistical models without optimal adaptation. Adaptive estimators.* Theory Probab. Appl. **37** 468–481.

[13] Lepski, O. V. (1992). *On problems of adaptive estimation in white Gaussian noise.* In Advances in Soviet Mathematics (R. Z. Khasminskii, ed.) **12** 87–106. Amer. Math. Soc., Providence.

[14] Goldenshluger, A., Lepski, O. (2009) *Uniform bounds for norms of sums of independent random functions* arXiv:0904.1950.

[15] Goldenshluger, A., Lepski, O. (2010) *Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality* arXiv:1009.1016.

[16] Mair, B. A. and Ruymgaart, F.H. (1996) *Statistical inverse estimation in Hilbert scales.* SIAM J. Appl. Math. **56**, 1424–1444.

[17] Massart, P. (2007) *Concentration inequalities and model selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. Springer, Berlin.

[18] Nussbaum, M. (1996). *Asymptotic equivalence of density estimation and white noise.* Ann. Statist. **24** 2399-2430.

[19] Nussbaum, M. and Pereverzev, S. (1999) *The degrees of ill-posedness in stochastic and deterministic noise models.* Preprint WIAS 509.

[20] Perthame, B. Transport equations arising in biology (2007). In *Frontiers in Mathematics*, Frontiers in Mathematics. Birckhauser.

[21] Metz, J.A.J. and Dieckmann O. (1986). *Formulating models for structured populations.* In *The dynamics of physiologically structured populations (Amsterdam, 1983)*, Lecture Notes in Biomath., vol. 68, 78–135.

[22] P. Michel, S. Mischler, B. Perthame, *General relative entropy inequality: an illustration on growth models,* J. Math. Pures Appl. (9) 84, 1235–1260 (2005).

[23] B. Perthame and L. Ryzhik, *Exponential decay for the fragmentation or cell-division equation,* J. of Diff. Eqns, **210**, 155–177 (2005).

[24] B. Perthame and J. P. Zubelli. On the inverse problem for a size-structured population model. *Inverse Problems*, 23(3):1037–1052, 2007.

[25] Tsybakov, A. (2004). *Introduction à l'estimation non-paramétrique.* Springer, Berlin.

[26] Wahba, G. (1977) *Practical approximate solutions to linear operator equations when the data are noisy.* SIAM J. Numer. Anal. **14**, 651–667.

5 Modelling Protein Polymerization[18]

OPEN ACCESS Freely available online

PLOS ONE

An Efficient Kinetic Model for Assemblies of Amyloid Fibrils and Its Application to Polyglutamine Aggregation

Stéphanie Prigent[1], Annabelle Ballesta[2], Frédérique Charles[3], Natacha Lenuzza[4], Pierre Gabriel[5], Léon Matar Tine[6], Human Rezaei[1], Marie Doumic[2]*

1 Institut National de Recherche Agronomique, Jouy-en-Josas, France, **2** Institut National de Recherche en Informatique et Automatique, Rocquencourt, France, **3** Université Pierre et Marie Curie, Paris, France, **4** Commissariat à l'Energie Atomique, Gif-sur-Yvette, France, **5** Institut National de Recherche en Informatique et Automatique Rhônes-Alpes, Lyon, France, **6** Université Gaston Berger, Saint-Louis, Sénégal

Abstract

Protein polymerization consists in the aggregation of single monomers into polymers that may fragment. Fibrils assembly is a key process in amyloid diseases. Up to now, protein aggregation was commonly mathematically simulated by a polymer size-structured ordinary differential equations (ODE) system, which is infinite by definition and therefore leads to high computational costs. Moreover, this Ordinary Differential Equation-based modeling approach implies biological assumptions that may be difficult to justify in the general case. For example, whereas several ordinary differential equation models use the assumption that polymerization would occur at a constant rate independently of polymer size, it cannot be applied to certain protein aggregation mechanisms. Here, we propose a novel and efficient analytical method, capable of modelling and simulating amyloid aggregation processes. This alternative approach consists of an integro-Partial Differential Equation (PDE) model of coalescence-fragmentation type that was mathematically derived from the infinite differential system by asymptotic analysis. To illustrate the efficiency of our approach, we applied it to aggregation experiments on polyglutamine polymers that are involved in Huntington's disease. Our model demonstrates the existence of a monomeric structural intermediate $\tilde{c}_1$ acting as a nucleus and deriving from a non polymerizing monomer (c_1). Furthermore, we compared our model to previously published works carried out in different contexts and proved its accuracy to describe other amyloid aggregation processes.

Citation: Prigent S, Ballesta A, Charles F, Lenuzza N, Gabriel P, et al. (2012) An Efficient Kinetic Model for Assemblies of Amyloid Fibrils and Its Application to Polyglutamine Aggregation. PLoS ONE 7(11): e43273. doi:10.1371/journal.pone.0043273

Editor: Yoshitaka Nagai, National Center of Neurology and Psychiatry, Japan

Received May 8, 2012; **Accepted** July 18, 2012; **Published** November 13, 2012

Funding: The work was partially financed by the French Agence Nationale de la Recherche, ANR grant 09-BLAN-0218-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: marie.doumic@inria.fr

Introduction

Protein aggregation and misfolding are involved in several fatal human disorders, such as Alzheimer's, Prion, Huntington's diseases [1,2]. Certain types of aggregates display specific structural traits (*e.g.* a $\beta-$sheet enriched secondary structure) that characterize amyloid assemblies. Recent progress in solid state Nuclear Magnetic Resonance (NMR) has led to a better understanding of amyloid assemblies at the molecular level [3]. However, this structural knowledge constitutes only a snapshot of the dynamic processes. Protein aggregation involves a large amount of chain reactions, *e.g.* conformational exchange, nucleation (which is the formation of a first stable intermediate), polymerization by monomer, dimer or $i-$ mer addition, coalescence, depolymerization (by loss of mono, di or oligomers), fragmentation (breakage into two or more polymers), protein degradation.

To explore the dynamics of amyloid assemblies, nucleation/polymerization reaction schemes have been applied, and to model them, ordinary differential equations (ODEs) have been used for many years [4]. An ODE means an equation containing only one independent variable (e.g. the chemical concentration of molecules) and its derivatives. Therefore in the case of polymerization, the number of equations should be at least equal to the number of sub-units constituting the longuest polymer. This value is extremely large in the case of amyloid fibrils (amyloid fibril sizes can go up to 1 μm length [5]), therefore simplifying assumptions are commonly admitted, e.g. constant reaction rates, meaning that polymers of any size behave roughly in the same way [6–9]. Although such assumptions allow the system to be reduced from an infinite set of ODEs to a couple of equations [4,7], assumptions of this nature are difficult to justify biochemically.

We propose here a new and global framework that can be adapted to most protein polymerization reactions. This method relies on partial differential equations (PDEs). In contrast to an ODE, a PDE permits formulation of problems involving functions of several variables. Instead of handling an infinite set of ODEs, we show that under reasonable assumptions, we can derive an equivalent model composed of a small number of ODEs coupled with a single size-structured PDE. The size variable of fibers replaces the infinite number of ODEs. To derive our model, we tune asymptotic methods from previously published works [10,11]. A fully general model, which is much easier to handle both theoretically and numerically, is obtained. It allows much faster computations than for the full ODE set of equations. Moreover, recent analytical tools developed for PDE analysis can be applied.

The obtention of size-distributions of polymers is a fundamental step [12], as it makes it possible to estimate quantitative reaction rates and build a predictive model by the means of recently developed inverse problem techniques [13].

To illustrate our method, we first formally derive the PDE model in a general case, and then apply our method to expanded polyglutamine (PolyQ) diseases. Finally, we compare our results to existing work [7,8].

Results

The Infinite ODE System

Let us first recall how one can write the differential system describing all the reactions that occur during nucleated protein polymerization. We denote c_1 the protein monomeric concentration and c_1^* the one of a misfolded monomeric species which displays the ability to polymerize. c_1 monomers transform into this monomeric species c_1^* at the rate k_I^+, and c_1^* transform back to c_1 at the rate k_I^-.

c_i represents the concentration of polymers made up of i monomers. We assume that polymers and monomers are degraded with a size-dependent degradation rate denoted k_m^i. The misfolded monomers c_1^* are able to polymerize to give rise to a nucleus c_{i_0}, composed of i_0 monomeric units, with the rate k_{on}^N. As proposed by Oosawa and co-authors [4], a nucleus is generated by the addition of an object to highly unstable entities that are too transitory to be observed. The object stabilizing the highly unstable entities can be a monomer (c_1). If we consider a nucleus c_{i_0} with a size i_0, its formation does not consist in a sequential addition of c_1 till c_{i_0} (where it would be represented by $c_1 \rightarrow c_2 \rightarrow c_3 \rightarrow \cdots \rightarrow c_{i_0}$), but follows an i_0 order kinetic (where $i_0 c_1^* \rightarrow c_{i_0}$).

This nucleus can dissociate at the rate k_{off}^N. We make the reasonable assumption that there is an equilibrium between monomers and oligomers [4].

$$c_1 \underset{k_I^-}{\overset{k_I^+}{\rightleftharpoons}} c_1^*, \qquad \underbrace{c_1^* + \cdots + c_1^*}_{i_0} \underset{k_{off}^N}{\overset{k_{on}^N}{\rightleftharpoons}} c_{i_0} \tag{1}$$

Polymers of size i larger than i_0 can polymerize or depolymerize, which is the gain or the loss of a single monomeric unit: the elongating species is assumed here to be c_1^* (our model is easy to adapt to other cases, *e.g* if the elongating species is a dimer or an oligomer [8]). Those reactions occur at the rate k_{on}^i and k_{dep}^i respectively.

$$c_i + c_1 \underset{k_{dep}^{i+1}}{\overset{k_{on}^i}{\rightleftharpoons}} c_{i+1} \tag{2}$$

Polymers can also coalesce with other polymers or break into two smaller polymers. For the sake of simplicity, we assume that a polymer can only break into two pieces at the exact same time (a breakeage into 3 or more pieces is generally much more hazardeous, so that it can be neglected). Coagulation of two polymers of respective size i and j occurs at the rate $k_{col}^{i,j}$. Fragmentation of a polymer of size i gives rise to smaller polymers of size j and $i-j$ (where $2 \leq j \leq i_0$), at the rate $k_{off}^{j,i}$.

$$c_i + c_j \underset{k_{off}^{i,i+j}}{\overset{k_{col}^{i,j}}{\rightleftharpoons}} c_{i+j} \tag{3}$$

We could have kept the same notation for fragmentation and depolymerization, by denoting $k_{off}^{1,i} = k_{off}^{i-1,i} = \frac{1}{2} k_{dep}^i$. We prefered however to distinguish them, because they involve reactions of different kinds, so that the orders of magnitude may appear different.

Let us define $K_{off}^j = \sum_{i=2}^{j-2} k_{off}^{i,j}$. This represents the total rate with which a polymer of size j can break to give smaller polymers. By symmetry we have that $k_{off}^{i,j} = k_{off}^{j-i,j}$ and $k_{col}^{i,j} = k_{col}^{j,i}$.

The following model is the exact deterministic transcription of the previously considered reactions. It could be completed by other reactions (polymerization pathways, other types of conformational exchange, for instance) to adapt to any possible case. The variation $\frac{dc}{dt}$ of the species c_i (or c_1, c_1^*) depends on two phenomena: 1) their rates of consumption, including depolymerization into a smaller polymer (or transformation into c_1 in the case of c_1^*), polymerization into a higher polymer (or transformation into c_1^* in the case of c_1) and degradation k_m, and 2) their rates of production, *i.e.* polymerization from smaller polymer (or transformation from c_1 in the case of c_1^*) and depolymerization from higher polymer (or transformation from c_1^* in the case of c_1). This induces the following equations.

$$\frac{dc_1}{dt} = -k_I^+ c_1 + k_I^- c_1^* - k_m^1 c_1, \tag{4}$$

$$\begin{aligned} \frac{dc_1^*}{dt} = {} & k_I^+ c_1 - k_I^- c_1^* - i_0 k_{on}^N (c_1^*)^{i_0} + i_0 k_{off}^N c_{i_0} - k_m^{1*} c_1^* \\ & - c_1^* \sum_{i \geq i_0}^{\infty} k_{on}^i c_i + \sum_{j=i_0}^{\infty} k_{dep}^j c_j + 2 \sum_{i=2}^{i_0-1} \sum_{j=i_0}^{\infty} i k_{off}^{i,j} c_j, \end{aligned} \tag{5}$$

$$\begin{aligned} \frac{dc_{i_0}}{dt} = {} & k_{on}^N (c_1^*)^{i_0} - k_{off}^N c_{i_0} - k_{on}^{i_0} c_{i_0} c_1^* + k_{dep}^{i_0+1} c_{i_0+1} - k_m^{i_0} c_{i_0} \\ & + 2 \sum_{j=i_0+2}^{\infty} k_{off}^{i_0,j} c_j - K_{off}^{i_0} c_{i_0} - \sum_{j \geq i_0} k_{col}^{i_0,j} c_{i_0} c_j, \end{aligned} \tag{6}$$

$$\begin{aligned} \frac{dc_i}{dt} = {} & c_1^* (k_{on}^{i-1} c_{i-1} - k_{on}^i c_i) - (k_{dep}^i c_i - k_{dep}^{i+1} c_{i+1}) - k_m^i c_i \\ & + 2 \sum_{j=i+2}^{\infty} k_{off}^{i,j} c_j - K_{off}^i c_i + \frac{1}{2} \sum_{i_0 \leq j \leq i-2} k_{col}^{j,i-j} c_j c_{i-j} \\ & - \sum_{j \geq i_0} k_{col}^{i,j} c_i c_j. \end{aligned} \tag{7}$$

This model and variants of it have been extensively studied, either in the general case in the mathematical literature (see

[10,14] and references therein), or when applying simplifying assumptions in the biological literature [4,6–8]. It is an efficient tool to study protein aggregation when the average size of protein i_M is of a reasonable order. However, for long polymer reactions, this modeling technique may be time-consuming and therefore inefficient to understand the underlying complexity. One can notice the resemblance between this infinite ODE model and a coupled PDE [15].

From ODEs to PDE: a New Size-structured Model

We propose here a new size-structured model composed of two ODEs and one PDE in the case of a large average size i_M of polymers - *i.e.*, $i_M \gg 1$. The main idea is to replace the discrete size i of a polymer by a continuous variable $x \in (\varepsilon i, \varepsilon(i+1))$, in which we have defined the small parameter $\varepsilon := \frac{1}{i_M} \ll 1$. In the same way, the densities $(c_i(t))$ are replaced by a continuous-in-size function $c(t,x)$ (see supplementary data S1 for more details). This model can be derived from the infinite set of ODEs if the two following assumptions hold.

First, for most polymer sizes i, there is only a slight difference between what happens for $i-$ mers and for $i+1$-mers. In other terms, even if quantities and reaction rates vary, it occurs in a "continuous" manner, implying only slight changes from one size i to its neighbor sizes $i+1$ and $i-1$ except for a small number of values. For instance, for degradation coefficients k_m^i, it is formalized as: There exists a constant, denoted below $Cst > 0$, such that

$$\text{For all } i \geq i_0, \quad |k_m^{i+1} - k_m^i| \leq \frac{Cst}{i}.$$

This assertion allows a continuous viewpoint on the equations for c_i. It also means that disruptions in the concentrations or in the coefficients can only appear at some specific points, that will have to be identified, and that are meaningful biologically. Though, this assertion appears to be natural since the conformational changes in polymers only occur at specific sizes [16]. Moreover, having a look at experimental size distributions (Figure 1) confirms how natural it is to view the size of polymers as a continuous quantity.

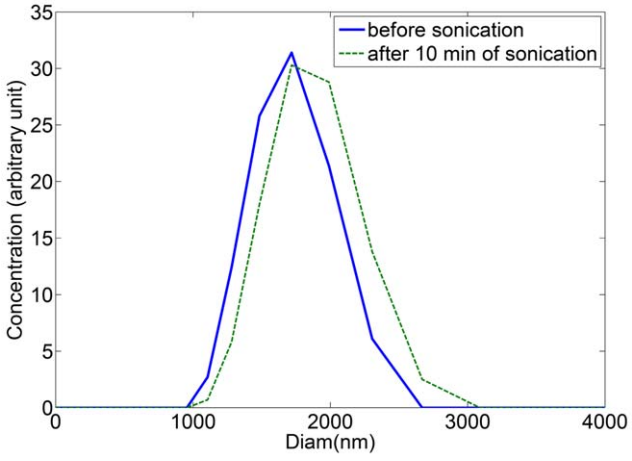

Figure 1. PolyQ41 Fibrils size distribution before (blue plain line) and after (dashed green) 10 min of sonication. The absence of any change in the distribution shows that neither fragmentation nor coalescence occured.
doi:10.1371/journal.pone.0043273.g001

The second and quite standard assumption is that at the beginning of the reaction, when polymer concentrations remain small compared to monomers, polymerization is the main process, whereas fragmentation and coalescence are secondary processes [4,6]. This assumption can be replaced if necessary by a similar one, such as the existence of a dominant polymerization by $j-$mer addition, with $j \ll i_M$ a relatively small oligomer. In such a case, the polymerization terms $k_{on}^i c_1 c_i$ would be replaced by $k_{on}^i c_j c_i$ in the equations, and a similar treatment can apply.

We refer to supplementary data S1 for a rigorous mathematical formulation of these two assertions. They are obtained when the system of equations is rescaled, and this allows us to estimate the relative contribution of each process to the overall dynamics.

Let us turn to the nucleus c_{i_0}. In this equation, the two assertions make it possible to ignore the influence of fragmentation and coalescence. Then as we are in the case where $i_M \gg 1$, the time-dependency of the equation for c_{i_0} is much faster than the one for c_1: it can be written $\frac{d}{dt} c_{i_0} = \frac{1}{\varepsilon} \cdots$ (see supplementary data S1). Hence, it is valid to suppose that it reaches its equilibrium instantaneously, and we can replace Equation (6) by

$$0 = k_{on}^N (c_1^*)^{i_0} - k_{off}^N c_{i_0} - k_{on}^{i_0} c_{i_0} c_1^*.$$

We thus obtain the following equality, which generalizes well-established formulas [6]

$$c_{i_0}(t) = \frac{k_{on}^N (c_1^*)^{i_0}}{k_{off}^N + k_{on}^{i_0} c_1^*}. \tag{8}$$

We can now write the following coupled ODE and PDE system, where i is replaced by a continuous variable x. Differences are replaced by derivatives and sums by integrals.

$$\frac{dc_1}{dt} = -k_I^+ c_1 + k_I^- c_1^* - k_m^1 c_1, \tag{9}$$

$$\begin{aligned} \frac{dc_1^*}{dt} = {} & k_I^+ c_1 - k_I^- c_1^* - \frac{i_0 k_{on}^N (c_1^*)^{i_0+1} k_{on}^{i_0}}{k_{off}^N + k_{on}^{i_0} c_1^*} - k_m^{1*} c_1^* \\ & - c_1^* \int_{x_0}^{\infty} k_{on}(x) c(t,x) dx + \int_{x_0}^{\infty} k_{dep}(x) c(t,x) dx, \end{aligned} \tag{10}$$

$$\begin{aligned} \frac{\partial c}{\partial t} = {} & -c_1^* \frac{\partial}{\partial x}(k_{on}(x) c(t,x)) + \frac{\partial}{\partial x}(k_{dep}(x) c(t,x)) \\ & + 2 \int_x^{\infty} k_{off}(x,y) c(y) dy - K_{off}(x) c(t,x) \\ & + \frac{1}{2} \int_{x_0}^{x} k_{col}(y, x-y) c(t,y) c(t,x-y) dy \\ & - \int_{x_0}^{\infty} k_{col}(x,y) c(t,x) c(t,y) dy - k_m(x) c(t,x), \end{aligned} \tag{11}$$

$$k_{on}(x_0)c(t,x_0)=k_{on}(x_0)\frac{k_{on}^N c_1^{i_0}}{k_{off}^N+k_{on}(x_0)c_1}. \quad (12)$$

Complete rigorous mathematical derivation can be found in supplementary data S1, and also shows that generally the third term in the right-hand side of Equation (10) (the ratio $\frac{i_0 k_{on}^N (c_1^*)^{i_0+1} k_{on}^{i_0}}{k_{off}^N+k_{on}^{i_0}c_1^*}$) is negligible. Even mathematical approximation theorems can be written to validate the model, as is done for instance in [10,11,17].

The advantages are twofold. First, it allows us to investigate numerically, using standard and well-known numerical schemes (see [18]), how a change in the coefficients can influence the overall reaction, and, more specifically, the size distribution. Also, inverse problem techniques could allow size-dependent parameters to be estimated (see for instance [19,20]). Secondly, it is easier to handle mathematically. Theoretical analysis can help us understand the intrinsic mechanisms and formulate new paradigms [21,22].

Application to PolyQ Polymerization

Aggregation of polyglutamine (PolyQ)-containing proteins is responsible for several neurodegenerative disorders including Huntington's disease. We have carried out biophysical analyses to investigate the aggregation kinetics of PolyQ41, which are peptides containing a repetition of 41 glutamine residues per monomer. Such a length of PolyQ repetition per molecule is sufficient to induce aggregation *in vitro* and in transfected cells [23].

Due to its simplicity, PolyQ provides an excellent model system to test our mathematical model. According to the experimental observations (Figure 1), fragmentation can be ignored. Indeed, in Figure 1, the size distribution of PolyQ41 fibrils did not change after 10 min of ultrasound treatments, showing that polymer-to-polymer reactions do not occur.

In order to determine whether coalescence occurs, we monitored simultaneously two types of measurements, polymer size and total polymerized mass. Polymer size was estimated by a static light scattering (SLS) signal. SLS is governed by the weighted average mass of oligomers and therefore highly depends on oligomer size. It can be viewed as a measurement of $I_2(t)=\sum_{i\geq i_0} i^2 c_i=\int x^2 c(t,x)dx$. Total polymerized mass was followed by thioflavine T (ThT) fluorescence. Such fluorescence arises from interactions between ThT and the peculiar structure of amyloids, relatively independently of amyloid size (above a certain size threshold). ThT can be mathematically expressed by $M(t)=\sum_{i\geq i_0} i c_i=\int x c(t,x)dx$. If there were coalescence, the weighted average polymer size would continue to grow even when the total polymerized mass $M(t)$ reached a plateau, so the second moment $I_2(t)$ would continue to grow after the plateau has been reached by $M(t)$. Here, however, both curves reach the plateau roughly simultaneously (see supplementary data S2). Therefore we conclude that coalescence is negligible. As described in Materials and Methods, the spontaneous polymerization of PolyQ41 is prevented by a glutathione s-transferase (GST) tag attached to PolyQ41 peptide. Such experimental system has the advantage of providing a system where only monomeric species are present at time 0, *i.e.* no seeding was required for polymerization: $c_1(t=0)=c_{tot}$, $c_1^*(t=0)=c_i(t=0)=0$. As the GST-polyQ41 does not constitute the pro-aggregative conformer, the PolyQ41 aggregation needs to be ignited by an irreversible enzymatic cleavage (here by thrombin hydrolysis), releasing the GST region apart from PolyQ41. This enzymatic cleavage can be assimilated to an activation process along which the poly Q41 monomer turns into a structurally activated form prone to aggregation. This led us to establish a minimal activation scheme in which the GST-polyQ41, denoted by c_1, is converted into an active form denoted c_1^* with a constant rate k_I^+. The nucleus size i_0, of unknown value, can be equal to 1, 2, 3 or even more. With these assumptions, Model (4)–(7) becomes

$$\frac{dc_1}{dt}=-k_I^+ c_1+k_I^- c_1^*, \quad (13)$$

$$\frac{dc_1^*}{dt}=k_I^+ c_1-k_I^- c_1^*-i_0 k_{on}^N (c_1^*)^{i_0}+i_0 k_{off}^N c_{i_0}-c_1^*\sum_{i\geq i_0} k_{on}^i c_i, \quad (14)$$

$$\frac{dc_{i_0}}{dt}=k_{on}^N (c_1^*)^{i_0}-k_{off}^N c_{i_0}-k_{on}^{i_0} c_{i_0} c_1^* \quad (15)$$

$$\frac{dc_i}{dt}=c_1^*(k_{on}^{i-1} c_{i-1}-k_{on}^i c_i), \quad (16)$$

and we use the continuous version of this model, given by (9)–(12), which becomes

$$\frac{dc_1}{dt}=-k_I^+ c_1+k_I^- c_1^*, \quad (17)$$

$$\begin{aligned}\frac{dc_1^*}{dt}=&k_I^+ c_1-k_I^- c_1^*-\frac{i_0 k_{on}^N k_{on}^{i_0}(c_1^*)^{i_0+1}}{k_{off}^N+k_{on}^{i_0}c_1^*}\\&-c_1^*\int_{x_0}^{\infty} k_{on}(x)c(t,x)dx,\end{aligned} \quad (18)$$

$$\frac{\partial c}{\partial t}=-c_1^*\frac{\partial}{\partial x}(k_{on}(x)c(t,x)), \quad (19)$$

$$k_{on}(x_0)c(t,x_0)=k_{on}(x_0)\frac{k_{on}^N (c_1^*)^{i_0}}{k_{off}^N+k_{on}(x_0)c_1^*}. \quad (20)$$

As an initial approach, we tested piecewise linear polymerization rates. They are linear from k_{on}^{min} to k_{on}^{max} on (x_{i_0},x_1), constantly equal to k_{on}^{max} on (x_1,x_2) and linearly decreasing to zero on (x_2,x_M) with k_{on}^{min} and k_{on}^{max} parameters to be calibrated. We arbitrarily set k_{on}^{min} and x_1, which led to 7 free parameters. We have also tested two different kinds of kinetics when $i_0=1$: first, the special case where there is no nucleus, *i.e.* the polymerization process starts directly from c_1^*, which means $k_{on}^N=k_{on}$ and k_{off}^N is negligible. This reaction scheme was unable to fit properly even a single experimental curve so we abandoned it. Second, the case when the previous model is unchanged but where $i_0=1$: this

means that the nucleus $c_{i_0}=\tilde{c}_1$ is a monomeric species differing only from c_1^* in its conformation. The elongating species remains the intermediate c_1^*. In the following, $i_0=1$ refers to this second case.

The parameters of this model were then estimated by fitting experimental data on PolyQ41 protein polymerization. We performed this in two successive ways. The first consists in fitting separately each experimental curve, corresponding to a given experiment, at a given concentration. The result is that whatever i_0 is, the fit is excellent for any curve, with a measurement error from 0.5 to 2% in L^2 adimensioned norms (see supplementary data S2). It gives almost undistinguishable curves. However, the variability among the optimal coefficients was large, which led us to the second step. This consisted in fitting *simultaneously* all the curves of experiments carried out in identical experimental conditions, but for different concentrations. The global adimensioned error (in L^2-norm) diminished with i_0, and reached its lowest level for $i_0=1$, as shown in Figure 2. For larger values of the nucleus, the error is moreover too large for the model to be acceptable (results shown in supplementary data S2). It gives solid ground to the assumption, already suggested in the literature [24], that the nucleus is of size 1, but with a specific and unconventional nucleation-elongation reaction scheme, where the elongating species c_1^* and the nucleus $c_{i_0}=\tilde{c}_1$ are distinct conformers.

Another result of our simulations is that k_I^- is negligible, thus we can suppose that $c_1=c_0e^{-k_I^+t}$. In the same way, we can compare c_1^* to the solution of the following differential equation

$$\frac{dc_{test}}{dt}=k_I^+c_0e^{-k_I^+t}-i_0k_{on}^Nc_{test}^{i_0},\quad c_{test}(0)=0,$$

i.e., neglect the contribution of polymers in the equation for c_1^*: it fits perfectly for the total duration of the lag phase.

Application to the Knowles et al. Model [7]

As seen for the application to PolyQ, the fully general model (9)–(12) is not yet directly applicable, precisely because of its general character. It can be thought of as the departure point for numerical, biological and mathematical analysis; and it is indeed a powerful way to tackle polymerization issues. To illustrate our approach, we have applied our model to experimental data of amyloid protein aggregation from other authors and we have compared or transposed our model to the recently published models that were accompanying the data [7,8].

In [7], Knowles and coauthors set up a model for polymerization of breakable filament assembly. This model is an analytical approximation that they have applied to (potential) experimental data and compared to exact equations representing the experimental data. For their approximation model, Knowles and coauthors made the following assumptions.

- Polymerization at a constant rate independent of the size of the polymers,
- no degradation of polymers neither monomers,
- the size of the nucleus is $i_0=2$,
- fragmentation rate depends linearly on the size of the polymer: $k_{off}^{i,j}=k_{off}$ constant,

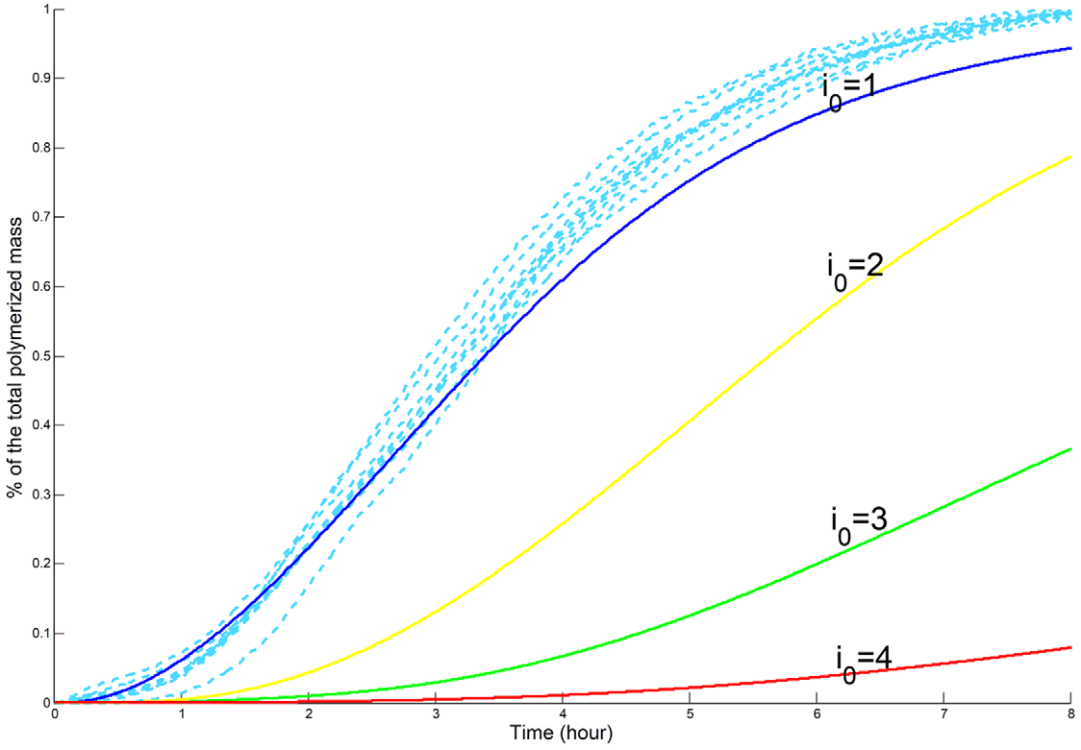

Figure 2. Simulation vs Experiments for Experimental Set 1, for an initial PolyQGST concentration of $100\ \mu M$. The parameters were first estimated for an experimental set of initial concentration $285\ \mu M$, then we compared the experimental measures (dotted lines) for an initial concentration $100\ \mu M$ with the simulations (in solid lines) for $i_0=1,2,3,4$. We see that the smaller i_0 is, the closer the simulation to experimental curves.
doi:10.1371/journal.pone.0043273.g002

- no coalescence,
- nucleation disaggregation occurs with the same rate k_{off} as depolymerization.

With these assumptions, it is well-known that the original ODE system simplifies by summation on a system of 2 non linear coupled ODEs (Equations (3a) and (3b) in [7]), namely:

$$\frac{dP}{dt} = k_{off}(M-(2i_0-1)P)+k_{on}^N(C_0-M)^{i_0}, \quad (21)$$

$$\frac{dM}{dt} = (k_{on}(C_0-M)-i_0(i_0-1)k_{off})P+i_0k_{on}^N(C_0-M)^{i_0} \quad (22)$$

where $M=\sum_{i\geq i_0} ic_i$ represents the total polymerized mass, and $P=\sum_{i\geq i0} c_i$ represents the total number of polymers. They approximate this system by an analytical formula, justified by a fixed point method and shown numerically to give a good approximation. To apply our method, we first look at the average size $i_M(t)$ of polymers, which is given by $i_M=\frac{M(t)}{P(t)}$. It is shown in Figure 3 for the parameter values $k_{on}=10^5M^{-1}s^{-1}$, $k_{off}=2.10^{-8}s^{-1}$, $C_0=5.10^{-6}M$, $k_{on}^N=2.10^{-5}M^{-1}s^{-1}$, $i_0=2$, $M(0)=P(0)=0$. All these values, taken from [7] (fig. 1 of Knowles^{TM} manuscript), directly represent the exact system of (potential) experimental data. We see that our assertion of large polymers is satisfied. Similarly, we check that the range of parameters that they proposed fit to our other assumptions, so that our method can be applied. The assumption on $k_{off}^N=k_{off}$ implies that the nucleus dissociation term in the equations for c_1 and c_{i_0} is negligible: indeed we have $k_{off}c_{i_0}$ to be compared to $c_1k_{on}i_Mc_i$.

We followed their modelling ideas but our method allows us to relax their assumptions in the following sense.

- Polymerization is not necessarily constant, but values $k_{on}(0)>0$ for small polymers of size i close to $i_0 \ll i_M$.
- We neglect degradation of small polymers and of monomers, but we keep a degradation for large polymers,
- $I_0=2$,
- fragmentation rate does not necessarily depend linearly on the size of the polymer, but it is true for small polymers: $k_{off}(x\ll 1,y\ll 1)=k_{off}^0$ constant,
- coalescence is negligible compared to polymerization as long as c_1 remains in the order of (C_0).

With these assumptions, System (9)–(12) can be simplified as follows:

$$\frac{dc_1}{dt} = -i_0k_{on}^Nc_1^{i_0}-c_1\int_0^\infty k_{on}(x)c(t,x)dx, \quad (23)$$

$$\begin{aligned}\frac{\partial c}{\partial t} &= -c_1\frac{\partial}{\partial x}(k_{on}(x)c(t,x))) \\ &+2\int_x^\infty k_{off}(x,y)c(y)dy-K_{off}(x)c(t,x)-k_m(x)c(t,x),\end{aligned} \quad (24)$$

$$k_{on}c_{|x=0} = k_{on}^Nc_1^{i_0-1}. \quad (25)$$

If we take as in [7] k_{off} and k_{on} constant, we recover System (??)(??) by summation, but with the terms $(2i_0-1)P(t)$ and $i_0(i_0-1)k_{off}P(t)$ neglected. Numerical simulations are shown in Figure 3, and we see that this simplification allows a perfect fit with the complete model, fast simulations, and a better understanding of which reaction dominates at any moment (since we have access to size distributions, see Figure 4).

Comments on Size Distributions. For the size parameters taken from [7], fig. 1, we are able to observe the evolution of polymer size distributions over time: see Figure 4. At the beginning of the reaction (in this particular case, for a time between 0 and 5 hours), the average size increases very fast. Then it reaches an equilibrium, and between 6 to 15 hours it reaches an exponential regime during which the whole size distribution, not only the average size, is quite steady. An explanation for this could be taken from [25] for instance. After this period, the average size decreases - and ultimately, the model shows that $M/P\rightarrow i_0+1$ but this would be accomplished only after a very long period of time. A good test for the model proposed by [7] would be to check whether size distribution of polymers resembles such a one-peak distribution. If not, the assumptions would have to be relaxed, *e.g.* by taking variable coefficients [25].

PDE Model Applied to the Xue et al. Model [8]

Xue and colleagues present a new strategy to analyse the self-assembly of misfolded proteins into amyloid fibrils [8]. They analysed fibril length distribution of β2-microglobulin, a protein involved in dialysis-related amyloidosis. Xue and colleagues have developed the following approach. Based on a large data set of experimental growth curves, transitional general parameters of the time-curve, namely the length of the lag phase (T_{lag}) and the slope (k) of the reaction curve at the inflexion point were extracted. Several theoretical models are simulated using the ODE formulation and the theoretical transitional parameters T_{lag} and k were extracted from the numerical growth curve in the same way as for the experimental curve (see Table S2 in Supplementary data S3). Then the best model and its parametrization were determined by comparing the theoretical values with the experimental data through least-squares analysis. This powerful approach is based on the simulation of a full ODE system (with one equation per size of aggregates) for each model investigated and no simplifications were made to reduce the dimension of this system. As a consequence, the method is time-consuming, which limits the number of mechanisms studied and the maximal polymer size (2400 in [8]). In addition, estimation of the best fitting model is based only on general parameters of the curve, which do not seem much sensitive to the distribution of the fragmentation process (see supplementary data S3). To overcome these limitations, we propose transposing their approach using PDE models, allowing for i) faster simulations, ii) no limitation in the size of aggregates, and iii) development of inverse problem techniques ([26,27]) to estimate parameters using the overall time evolution process.

Xue *et al* investigated $\beta 2-$microglobulin growth, using models including different processes: a pre-polymerization step (characterized by either no pre-polymerization, or monomer-dimer equilibrium and dimer addition mechanism, or conformation exchange), an elongation of the aggregates following a one-step function, a linear function or a power function, and a possible

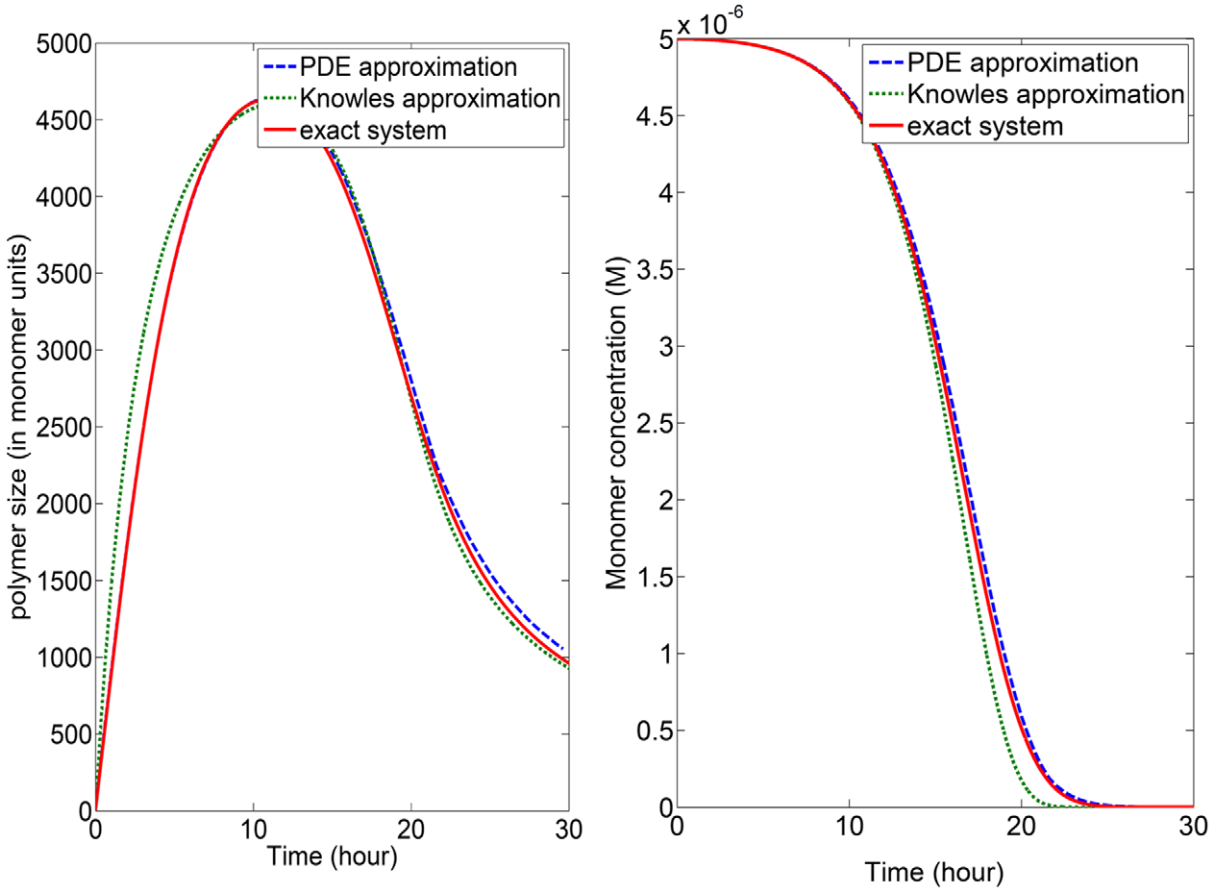

Figure 3. Numerical solution of Equation (21) (22) either using the exact equations, directly representing (potential) experimental data, the PDE approximation, or the analytical approximation proposed in [7]. Left: average size of polymers. Right: monomers concentration. It is clear that the PDE approximation gives excellent results. The parameters used for the exact equations (i.e. values for elongation rate, fragmentation rate and nucleus size) are those from Fig. 1 of [7].
doi:10.1371/journal.pone.0043273.g003

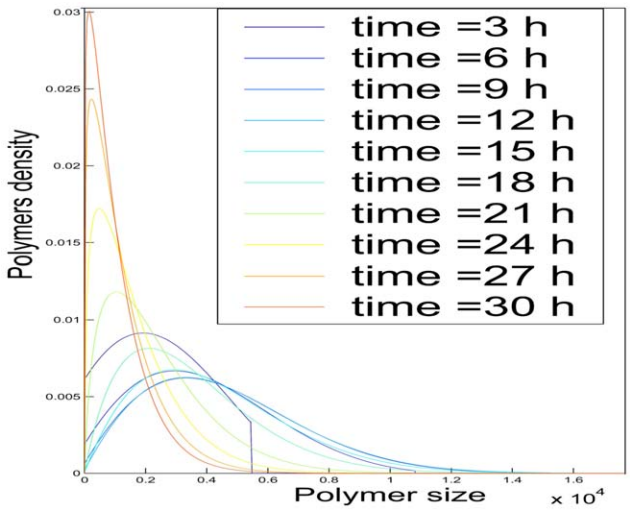

Figure 4. Adimensionned Distributions of the sizes of polymers for various times. To obtain adimensionned distributions of polymer sizes, our model was applied to data taken from Figure 1 of [7]. From 6 to 18 hours one can see that the distribution remains roughly stable.
doi:10.1371/journal.pone.0043273.g004

secondary process such as fragmentation. Their best-fit model is given by:

- No conformational exchange, no coalescence and no degradation of polymers or monomers,
- the size of the nucleus is $i_0 = 2$ and nucleus dissociation occurs only through depolymerization,
- polymerization and depolymerization follow a one-step function with the step at $i = 6$,
- fragmentation into two smaller polymers occurs.

Thus, using the previously introduced notations, the original ODE system can be written

$$\frac{dc_1}{dt} = -i_0 k_{on}^N c_1^{i_0} + i_0 k_{off}^N c_{i_0} - c_1 \sum_{i \geq i_0} k_{on}^i c_i, \tag{26}$$

$$\frac{dc_{i_0}}{dt} = k_{on}^N c_1^{i_0} - k_{off}^N c_{i_0} - k_{on}^{i_0} c_1 c_{i_0} + 2 \sum_{j \geq i_0 + 2} k_{off}^{i_0, j} c_j \tag{27}$$

$$\frac{dc_i}{dt} = c_1(k_{on}^{i-1}c_{i-1} - k_{on}^{i}c_i) - (k_{dep}^{i}c_i - k_{dep}^{i+1}c_{i+1}) + 2\sum_{j\geq i+2} k_{off}^{i,j}c_j - K_{off}^{i}c_i. \quad (28)$$

For the particular choice of fragmentation made in [8], however, fragmentation in polymers of size 1 is close to 0. This ODE system is then formally equivalent to the following PDE system:

$$\frac{dc_1}{dt} = -\frac{i_0 k_{on}^N k_{on}(x_0)c_1^{i_0+1}}{k_{off}^N + k_{on}(x_0)c_1} - c_1\int_{x_0}^{\infty} k_{on}(x)c(t,x)dx, \quad (29)$$

$$\frac{\partial c(t,x)}{\partial t} = -c_1\frac{\partial}{\partial x}(k_{on}(x)c(t,x)) + \frac{\partial}{\partial x}(k_{dep}(x)c(t,x))' + 2\int_{x}^{\infty} k_{off}(x,y)c(t,y)dy - K_{off}(x)c(t,x), \quad (30)$$

$$c(t,x_0) = \frac{k_{on}^N c_1^{i_0}}{k_{off}^N + k_{on}(x_0)c_1}. \quad (31)$$

Due to the shape of the polymerization process, which has a step at $i=6$ (meaning that $k_{on}^i = K_1$ for $i\leq 5$, $k_{on}^i = K_2$ for $i\geq 6$), if the step is high, that is if $K_2 \gg K_1$, it is however preferable to keep all the ODEs occurring for $i\leq 6$ and to set up the PDE (30) only for $i\geq 6$. We then adapt the boundary condition (31) as shown in supplementary data S3. This can also be approximated by the Bishop and Ferrone model [6] by adjusting a nucleus critical size to $i_0 = 6$. Similar work can be done for the different processes studied in [8]. Our study allowed us to enhance their approach by quick investigation of different fragmentation kernels, showing that the shape of the fragmentation does not influence the polymerization dynamics (see supplementary data S3).

Discussion

We proposed a new model (9)–(12) to serve as a global framework to investigate the leading mechanisms of nucleation-elongation processes in amyloid fibrils' assemblies. We applied it to PolyQ41 aggregation, demonstrating experimentally that coalescence and fragmentation were negligible, and predicting by our simulations that the monomer activation was irreversible. Moreover, it highlighted the early step of PolyQ41 nucleus formation and assemblies. With regard to the bibliography, the concept of nucleus in protein aggregation remains obscure. Here the analysis of PolyQ polymerization suggested a kinetic scheme in which c_1^* is at an equilibrium with $\tilde{c}_1$. These two species are monomeric and only differ in their conformation. According to the conventional model of nucleation-elongation process [4], the nucleus is thermodynamically stabilized by the addition of at least one monomer. Here we proposed an unconventional mechanism of nucleation in which the $\tilde{c}_1$ formation constitutes the limiting step in the polymerization process which is stabilized by an interaction with c_1^*. Therefore, the formation of the $[\tilde{c}_1 - c_1^*]$ complex constitutes the first proaggregative species. Furthermore, during the formation of this complex, a structural information exchange should occur between $\tilde{c}_1$ and c_1^*. To reach the formation of a nucleus, two changes of conformation are hence required. The first one arises from the GST-cleavage of c_1 to a conformer c_1^* released as a random coil structure, that is not proaggregative. The second change of conformation is an internal change of the random coil c_1^* into a proaggregative species $\tilde{c}_1$ that is still monomeric.

Our approach also proved highly efficient when applied to previously designed models [7,8], where it can be adapted and used to pursue the research further. We believe it could be applied to many other cases, providing both a unified framework and an efficient way to carry out fast simulations, model discrimination [28], inverse problem methods and analysis.

Materials and Methods

Model Derivation

To derive the continuous model, we first write a rescaled version of the model, that makes use of typical orders of magnitude. Then, quantifying our assumptions, we approximate sums by integrals and differences by derivatives. Finally, from the equation for c_{i_0} we deduce the boundary condition for $c(t,x=x_0)$ (full details in supplementary data S1).

Numerical Implementation

To avoid useless conversions, we implemented the PDE model (9)–(12) with dimensioned numbers, and checked *a posteriori* that the considered orders of magnitude fit the assumptions. We use an explicit upwind scheme - finer methods can be used such as WENO [18].

Parameter Estimation

The parameter estimation was performed by a least-square approach. For $i_0 = 1,2,3,4$, we searched for the optimal set of parameters such that it minimized the quadratic distance between the data points obtained by ThT measures and the simulated curve of the mass, represented by $\int c(t,x)xdx$ in the PDE model or by $\sum ic_i$ in the ODE one. The minimization task was performed by the CMAES algorithm [29]. It was run with 50 different initial parameters sets. Then the optimal solution was used as an initial guess and the minimization algorithm was run again 50 times.

Experimental Results

GST-PolyQ production. The GST-Q41 expression vector was described by Masino et al [30]. GST-polyQ41 fusion protein was produced in E.Coli BL21DE3 and purified by affinity chromatography using Glutathione Sepharose affinity beads (Pharmacia).

Fragmentation experiments. The Fragmentation experiments were performed using an immersion sonotrod oscillating at 40 kHz. The size distributions of polyQ fibrils were monitored before and after sonication by dynamic light scattering (DLS, Wyatt).

Kinetic experiments. All polymerization experiments were performed at 33°C. Aggregation was initiated by thrombin addition (0.5 unit/ÂμM of GST-PolyQ41) leading to the release of PolyQ41 peptide from GST. The aggregation was monitored either by Thioflavine T (Tht) (100 μM) in a 96-well plate fluorescence spectrometer or by a homemade multiwavelength static light scattering/fluorescence system (SLS).

Supporting Information

Figure S1 Parameter estimation considering each curve separately. Time evolution of PolyQ41 polymerized mass for an initial PolyQGST concentration equal to 285 μM. The experimental results are plotted in dotted line and the best-fit curve in solide line. i_0 is set to 3. Best-fit parameters are $k_I^+ = 0.67\ h^{-1}$, $k_I^- = 0$, $k_{on}^N = 7.8.10^2\ M^{-2}h^{-1}$, $k_{off}^N = 5.10^{-2}\ h^{-1}$, $k_{on}^{max} = 1.2.10^9\ M^{-1}h^{-1}$, $i_{max} = 2.10^6$, $x_2 = 0.2\ i_{max}$.
(TIF)

Figure S2 Parameter estimation for Experimental Set 1 when i_0 is set to 3. Time evolution of the adimensioned PolyQ41 polymerized mass for an initial PolyQGST concentration equal to 100 μM (A), 285 μM (B), 420 μM(C). Dotted curves represent experimental results. The solid curve is the best-fit. The global error in L^2 adimensioned norm was equal to 40% and the optimal parameters are very close to those of Figure 1.
(TIF)

Figure S3 Parameter estimation for Experimental Set 1 when i_0 is set to 1. Time evolution of the adimensioned PolyQ41 polymerized mass for an initial PolyQGST concentration equal to 100 μM (A), 285 μM (B), 420 μM (C). Dotted curves represent experimental results. The solid curve is the best-fit. The global error in in L^2 adimensioned norm was equal to 11%. The best-fit parameters are $k_I^+ = 0.65\ h^{-1}$, $k_I^- = 0$, $k_{on}^N = 7.10^{-6}\ M^{-1}h^{-1}$, $k_{off}^N = 5.10^{-2}h^{-1}$, $k_{on}^{max} = 2.3.10^9 M^{-1}h^{-1}$, $x_2 = 0.1\ i_{max}$, $i_{max} = 5.10^6$.
(TIF)

Figure S4 Left: Size distribution of the fragmentation rate for an aggregation of size 20, following a uniform distribution (black) or a mechanical-based distribution (red) of fragmentation. Right: Simulated normalized reaction progress curves of amyloid formation for a uniform distribution (black) and a mechanical-based distribution (red) of fragmentation. See below for the numerical values.
(TIF)

Figure S5 Examples of simulated size distribution of the aggregates for a uniform distribution (black) and a mechanical-based distribution (red) of fragmentation. See above for the numerical values.
(TIF)

Supplementary Data S1 Model derivation from ODE to PDE.
(PDF)

Supplementary Data S2 Application to PolyQ41 polymerization.
(PDF)

Supplementary Data S3 Effect of the fragmentation distribution on the kinetics of the Xue et al. model [1].
(PDF)

Author Contributions

Conceived and designed the experiments: MD HR. Performed the experiments: SP FC HR. Analyzed the data: AB FC NL MD. Contributed reagents/materials/analysis tools: PG LMT NL. Wrote the paper: MD SP HR AB. Contributed equally to the design and supervision of this work: MD HR.

References

1. Shastry BS (2003) Neurodegenerative disorders of protein aggregation. Neurochemistry International 43: 1–7.
2. Ross C, Poirier M (2004) Protein aggregation and neurodegenerative disease. Nature Medicine 10(Suppl): S10–S17.
3. Wasmer C, Lange A, Van Melckebeke H, Siemer AB, Riek R, et al. (2008) Amyloidfibrils of the het-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. Science 319: 1523–1526.
4. Oosawa F, Asakura S (1975) Thermodynamics of the polymerization of protein. Waltham, MA: Academic Press.
5. Dicko C, Kenney J, Vollrath F (2006) Fibrous Proteins: Amyloids, Prions and Beta- proteins, volume 73. Amsterdam: Elsevier. 17–53.
6. Bishop M, Ferrone F (1984) Kinetics of nucleation-controlled polymerization. a perturbation treatment for use with a secondary pathway. Biophysical Journal 46: 631–644.
7. Knowles TPJ, Waudby CA, Devlin GL, Cohen SIA, Aguzzi A, et al. (2009) An Analytical Solution to the Kinetics of Breakable Filament Assembly. Science 326: 1533–1537.
8. Xue WF, Homans SW, Radford SE (2008) Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. PNAS 105: 8926–8931.
9. Masel J, Jansen V, Nowak M (1999) Quantifying the kinetic parameters of prion replication. Biophysical Chemistry 77: 139–152.
10. Doumic M, Goudon T, Lepoutre T (2009) Scaling limit of a discrete prion dynamics model. Communications in Mathematical Sciences 7: 839–865.
11. Collet JF, Goudon T, Poupaud F, Vasseur A (2002) The Becker–Döring system and its Lifshitz–Slyozov limit. SIAM J on Appl Math 62: 1488–1500.
12. Xue WF, Homans SW, Radford SE (2009) Amyloid fibril length distribution quantified by atomic force microscopy single-particle image analysis. Protein engineering design selection PEDS 22: 489–496.
13. Doumic M, Perthame B, Zubelli J (2009) Numerical solution of an inverse problem in sizestructured population dynamics. Inverse Problems 25: 045008.
14. Ball JM, Carr J (1990) The discrete coagulation-fragmentation equations: existence, uniqueness, and density conservation. J Statist Phys 61: 203–234.
15. Wulkow M (1996) The simulation of molecular weight distributions in polyreaction kinetics by discrete galerkin methods. Macromol Theory Simul 5: 396–416.
16. Desai A, Mitchison TJ (1997) Microtubule polymerization dynamics. Annual Review of Cell and Developmental Biology 13: 83–117.
17. Laurençot P, Mischler S (2007) From the discrete to the continuous coagulation-fragmentation equations. Proceedings of the Royal Society of Edinburgh: Section A Mathematics 132: 1219–1248.
18. Gabriel P, Tine LM (2010) High-order WENO scheme for polymerization-type equations. ESAIM Proc 30: 54–70.
19. Doumic M, Maia P, Zubelli J (2010) On the calibration of a size-structured population model from experimental data. Acta Biotheoretica 58(4): 405–413.
20. Banks H, Sutton K, Thompson W, Bocharov G, Roosec D, et al. (2010) Estimation of cell proliferation dynamics using cfse data. Bull of Math Biol 73(1): 116–150.
21. Calvez V, Lenuzza V, Oelz D, Deslys JP, Laurent P, et al. (2009) Size distribution dependence of prion aggregates infectivity. Math Biosci 1: 88–99.
22. Alvarez-Martinez MT, Fontes P, Zomosa-Signoret V, Arnaud JD, Hingant E, et al. (2011) Dynamics of polymerization shed light on the mechanisms that lead to multiple amyloid structures of the prion protein. Biochimica et Biophysica Acta - Proteins and Proteomics 1814: 1305–1317.
23. Scherzinger E, Sittler A, Schweiger K, Heiser V, Lurz R, et al. (1999) Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for huntingtons disease pathology. Proceedings of the National Academy of Sciences of the United States of America 96: 4604–4609.
24. Kar K, Jayaraman M, Sahoo B, Kodali R, Wetzel R (2011) Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. Nature Structural and Molecular Biology 18: 328–336.
25. Calvez V, Lenuzza N, Oelz D, Deslys JP, Laurent P, et al. (2009) Size distribution dependence of prion aggregates infectivity. Math Biosci 1: 88–99.
26. Ackleh AS, Fitzpatrick BG (1997) Modeling aggregation and growth processes in an algal population model: analysis and computations. Journal of Mathematical Biology 35: 480–502.
27. Bortz D, Jackson T, Taylor K, Thompson A, Younger J (2008) Klebsiella pneumoniae occulation dynamics. Bulletin of Mathematical Biology 70: 745–768.
28. Bernacki JP, Murphy RM (2009) Model discrimination and mechanistic interpretation of kinetic data in protein aggregation studies. Biophysical Journal 96: 2871–2887.
29. Hansen N (2006) The CMA evolution strategy: a comparing review. In: Lozano J, Larranaga P, Inza I, Bengoetxea E, editors. Towards a New Evolutionary Computation, volume 192. New York: Springer. 75–102.
30. Masino L, Kelly G, Leonard K, Trottier Y, Pastore A (2002) Solution structure of polyglutamine tracts in gst-polyglutamine fusion proteins. FEBS Letters 513: 267–272.

Part IV
References

Bibliography of the author

[1] H.T. Banks, F. Charles, M. Doumic, K.L. Sutton, and W.C. Thompson. Label structured cell proliferation models. *App. Math. Letters*, 23(12):1412–1415, 2010.

[2] V. Calvez, M. Doumic, and P. Gabriel. Self-similarity in a general aggregationfragmentation problem. application to fitness analysis. *Journal de Mathmatiques Pures et Appliques*, 98(1):1 – 27, 2012.

[3] M. Doumic. Etude théorique et simulation numérique de la propagation laser en milieu inhomogène. *PhD. thesis, available on Hal*, 2005.

[4] M. Doumic. Analysis of a population model structured by the cells molecular content. *Math. Model. Nat. Phenom.*, 2(3):121–152, 2007.

[5] M. Doumic. Boundary value problem for an oblique paraxial model of light propagation. *Methods Appl. Anal.*, 16(1):119–137, 2009.

[6] M. Doumic, F. Duboc, F. Golse, and R. Sentis. Simulation of laser beam propagation with a paraxial model in a tilted frame. *Journal of Computational Physics*, 228(3):861–880, 2008.

[7] M. Doumic and P. Gabriel. Eigenelements of a general aggregation-fragmentation model. *Mathematical Models and Methods in Applied Sciences*, 20(05):757, 2009.

[8] M. Doumic, F. Golse, and R. Sentis. A paraxial model for the propagation of light: the boundary value problem for the schrödinger advection equation in a tilted frame. *C.-R. Acad. Sc., Ser. I*, 336:23–28, 2003.

[9] M. Doumic, T. Goudon, and T. Lepoutre. Scaling Limit of a Discrete Prion Dynamics Model. *Comm. in Math. Sc.*, 7(4):839–865, 2009.

[10] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *submitted*, 2012.

[11] M. Doumic, M. Hoffmann, P. Reynaud, and V. Rivoirard. Nonparametric estimation of the division rate of a size-structured population. *SIAM J. on Numer. Anal.*, 50(2):925–950, 2012.

[12] M. Doumic, P.S. Kim, and B. Perthame. Stability analysis of a simplified yet complete model for chronic myelogenous leukemia. *Bulletin of Mathematical Biology*, 72(7):1732–1759, 2010.

[13] M. Doumic, P. Maia, and J.P. Zubelli. On the calibration of a size-structured population model from experimental data. *Acta Biotheoretica*, 58(4):405–413, 2010.

[14] M. Doumic, A. Marcinak-Czochra, B. Perthame, and J.P. Zubelli. A Structured Population Model of Cell Differentiation. *SIAM J. Appl. Math.*, 71(6):1918–1940, 2011.

[15] M. Doumic, B. Perthame, and J.P. Zubelli. Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems*, 25(4):045008, 2009.

[16] M. Doumic and L.M. Tine. Estimating the division rate for the growth-fragmentation equation. *Journal of Mathematical Biology, published online*, 2012.

[17] B. Perthame and J.P. Zubelli. On the inverse problem for a size-structured population model. *Inverse Problems*, 23(3):1037–1052, 2007.

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

[19] M. Doumic T. Bourgeron and M. Escobedo. Estimating the division rate for the self-similar growth-fragmentation equation. in prepare.

References

[20] A S Ackleh, H T Banks, K Deng, and S Hu. Parameter estimation in a coupled system of nonlinear size-structured populations. *Mathematical biosciences and engineering MBE*, 2(2):289–315, 2005.

[21] M. Aizenman and T.A. Bak. Convergence to equilibrium in a system of reacting polymers. *Commun. math. Phys.*, 65(12):203–230, 1979.

[22] O. Arino, E. Sànchez, and G. F. Webb. Necessary and sufficient conditions for asynchronous exponential growth in age structured cell population with quiescence. *J. Math. Anal. and Appl.*, 215:499–513, 1997.

[23] F. Baccelli, D. R. McDonald, and J. Reynier. A mean-field model for multiple tcp connections through a buffer implementing red. *Performance Evaluation*, 49(1-4):77 – 97, 2002.

[24] D. Balagué, J. Cañizo, and P. Gabriel. Fine asymptotics of profiles and relaxation to equilibrium for growth-fragmentation equations with variable drift rates. *Kinetic Related Models*, 2013. in press.

[25] 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.

[26] P. Ballereau, M. Casanova, F. Duboc, D. Dureau, H. Jourdren, P. Loiseau, J. Metral, O. Morice, and R. Sentis. Simulation of the paraxial laser propagation coupled with hydrodynamics in 3d geometry. *Journal of Scientific Computing*, 33:1–24, 2007.

[27] J. Banasiak and W. Lamb. On a coagulation and fragmentation equation with mass loss. *Proc. Roy. Soc. Edinburgh Sect. A*, 136(6):1157–1173, 2006.

[28] H. T. Banks, S. L. Ernstberger, and S. L. Grove. Standard errors and confidence intervals in inverse problems: sensitivity and associated pitfalls. *J. Inverse Ill-Posed Probl.*, 15(1):1–18, 2007.

[29] H T Banks and B.G. Fitzpatrick. Estimation of growth rate distributions in size structured population models. *Quart. Appl. Math.*, 49:215–235, 2005.

[30] H.T. Banks, S.L. Ernstberger, and S. Hu. Sensitivity equations for a size-structured population model. *Quart. Appl. Math.*, 67:627–660, 2009.

[31] H.T. Banks, K.L. Sutton, W.C. Thompson, G. Bocharov, D. Roosec, T. Schenkeld, and A. Meyerhanse. Estimation of cell proliferation dynamics using cfse data. *Bull. of Math. Biol.*, 2010.

[32] V. Bansaye. Proliferating parasites in dividing cells: Kimmels branching model revisited. *Ann. Appl. Probab.*, 18(3):967–996, 2008.

[33] R. Becker and W. Döring. Kinetische behandlung der keimbildung in bersttigten dmpfen. *Annalen der Physik*, 416(8):719–752, 1935.

[34] F. Bekkal Brikci, J. Clairambault, and B. Perthame. Analysis of a molecular structured population model with possible polynomial growth for the cell division cycle. *Math. Comput. Modelling*, 47(7-8):699–713, 2008.

[35] F. Bekkal Brikci, J. Clairambault, B. Ribba, and B. Perthame. An age-and-cyclin-structured cell population model for healthy and tumoral tissues. *J. Math. Biol.*, 57(1):91–110, 2008.

[36] George I. Bell. Cell growth and division: III. Conditions for Balanced Exponential Growth in a Mathematical Model. *Biophysical Journal*, 8(4):431 – 444, 1968.

[37] George I. Bell and Ernest C. Anderson. Cell growth and division: I. a mathematical model with applications to cell volume distributions in mammalian suspension cultures. *Biophysical Journal*, 7(4):329 – 351, 1967.

[38] R E. Bellman. *Adaptive control processes - A guided tour*. Princeton University Press, Princeton, New Jersey, U.S.A., 1961.

[39] M.F. Bishop and F.A. Ferrone. Kinetics of nucleation-controlled polymerization. a perturbation treatment for use with a secondary pathway. *Biophysical Journal*, 46(5):631 – 644, 1984.

[40] D Bonnet and J E Dick. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine*, 3(7):730–737, 1997.

[41] P. Brémaud and L. Massoulié. Stability of nonlinear Hawkes processes. *The Annals of Probability*, 24(3):1563–1588, 1996.

[42] M. J. Cáceres, J. A. Cañizo, and S. Mischler. Rate of convergence to the remarkable state for fragmentation and growth-fragmentation equations. *Journal de Mathematiques Pures et Appliquees*, 96(4):334–362, 2011.

[43] V. Calvez, N. Lenuzza, D. Oelz, J.-P. Deslys, P. Laurent, F. Mouthon, and B. Perthame. Size distribution dependence of prion aggregates infectivity. *Math. Biosci.*, 1:88–99, 2009.

[44] S. Chapman and T.G. Cowling. *The Mathematical Theory of Non-Uniform Gases*. Cambridge University Press, 1970.

[45] G. Chiorino, J. A. J. Metz, D. Tomasoni, and P. Ubezio. Desynchronization rate in cell populations: mathematical modeling and experimental data. *J. Theor. Biol.*, 208:185–199, 2001.

[46] J. Clairambault, S. Gaubert, and T. Lepoutre. Comparison of perron and floquet eigenvalues in age structured cell division cycle models. *Math. Model. Nat. Phenom.*, 4(3):183–2009, 2009.

[47] B. Cloez. Limit theorems for some branching measure-valued processes. *arXiv:1106.0660v2*, 2012.

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

[49] J.F. Collet and S. Hariz. A modified version of the lifshitz-slyozov model. *Applied Mathematics Letters*, 12:81–95, 1999.

[50] Stephen Cooper. Distinguishing between linear and exponential cell growth during the division cycle: Single-cell studies, cell-culture studies, and the object of cell-cycle research. *Theoretical Biology and Medical Modelling*, 3:10, 2006.

[51] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Sciences and Technology*. Springer, 1990.

[52] O. Diekmann, H.J.A.M. Heijmans, and H.R. Thieme. On the stability of the cell size distribution. *Journal of Mathematical Biology*, 19:227–248, 1984.

[53] R.L. Drake. A general mathematical survey of the coagulation equation. *Topics in Current Aerosol Research (Part 2)*, pages 201–376, 1972.

[54] D. Drasdo and S. Höhme. A single-cell-based model of tumor growth in vitro: monolayers and spheroids. *Physical Biology*, 2(3):133, 2005.

[55] N. Echenim, D. Monniaux, M. Sorine, and F. Clément. Multi-scale modeling of the follicle selection process in the ovary. *Mathematical Biosciences*, 198(1):57–79, 2005.

[56] F Eghiaian, T Daubenfeld, Y Quenet, M van Audenhaege, A-P Bouin, G van der Rest, J Grosclaude, and H Rezaei. Diversity in prion protein oligomerization pathways results from domain expansion as revealed by hydrogen/deuterium exchange and disulfide linkage. *Proceedings of the National Academy of Sciences*, 104(18):7414–7419, 2007.

[57] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Springer, 1996.

[58] H.W. Engl, W. Rundell, and O. Scherzer. A Regularization Scheme for an Inverse Problem in Age-Structured Populations. *J. of Math. Anal. and Appl.*, 182:658–679, 1994.

[59] H. Engler, J. Pruss, and G.F. Webb. Analysis of a model for the dynamics of prions ii. *J. of Math. Anal. and App.*, 324(1):98–117, 2006.

[60] M. Escobedo, P Laurençot, S. Mischler, and B. Perthame. Gelation and mass conservation in coagulation-fragmentation models. *J. Differential Equations*, 195(1):143–174, 2003.

[61] M. Escobedo and S. Mischler. Dust and self-similarity for the Smoluchowski coagulation equation. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 23(3):331–362, 2006.

[62] M. Escobedo, S. Mischler, and B. Perthame. Gelation in coagulation and fragmentation models. *Comm. Math. Phys.*, 231(1):157–188, 2002.

[63] M. Escobedo, S. Mischler, and M. Rodriguez Ricard. On self-similarity and stationary problem for fragmentation and coagulation models. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 22(1):99–125, 2005.

[64] J.Z. Farkas. Stability conditions for a nonlinear size-structured model. *Nonlin. Anal. Real World App.*, 6:962–969, 2005.

[65] M.D. Feit and J.A. Fleck. Beam non paraxiality. *J. Opt.Soc.Am. B*, 5:633–640, 1988.

[66] P. Gabriel. The shape of the polymerization rate in the prion equation. *Mathematical and Computer Modelling*, 53(78):1451 – 1456, 2011.

[67] P. Gabriel. Long-time asymptotics for nonlinear growth-fragmentation equations. *Comm. Math. Sc.*, 10(3):787–820, 2012.

[68] T. Gasser and H.G. Müller. Optimal convergence properties of kernel estimates of derivatives of a density function. *Lecture Notes in Mathematics*, 757:144–154, 1979.

[69] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39:1608–1632, 2011.

[70] A. Goldenshluger and O. Lepski. Uniform bounds for norms of sums of independent random functions. *Ann. Probab.*, 39:2318–2384, 2011.

[71] L.M. Greer, P. van den Driessche, L. Wang, and G.F. Webb. Effects of general incidence and polymer joining on nucleated polymerization in a model of prion proliferation. *SIAM Journal on Applied Mathematics*, 68(1):154–170, 2007.

[72] M.L. Greer, L. Pujo-Menjouet, and G.F. Webb. A mathematical analysis of the dynamics of prion proliferation. *J. Theoret. Biol.*, 242(3):598–606, 2006.

[73] A. Groh, J. Krebs, and M. Wagner. Efficient solution of an inverse problem in cell population dynamics. *Inverse Problems*, 27, 2011.

[74] P. Gwiazda and B. Perthame. Invariants and exponential rate of convergence to steady state in the renewal equation. *Markov Processes and Related Fields*, 2:413–424, 2006.

[75] M. Gyllenberg, A. Osipov, and L. Païvärinta. The inverse problem of linear age-structured population dynamics. *Journal of Evolution Equations*, 2(772):223–239, 2002.

[76] M. Gyllenberg, A. Osipov, and L. Pivrinta. On determining individual behaviour from population data. In Dorothee Haroske, Thomas Runst, and Hans-Jrgen Schmeisser, editors, *Function Spaces, Differential Operators and Nonlinear Analysis*, pages 329–339. Birkhuser Basel, 2003.

[77] M. Gyllenberg and G.F. Webb. A Nonlinear Structured Population Model of Tumor Growth With Quiescence. *J. Math. Biol*, 28:671–694, 1990.

[78] A. G. Hawkes and D. Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[79] H.J.A.M. Heijmans. On the stable size distribution of populations reproducing by fission into two unequal parts. *Mathematical Biosciences*, 72(1):19 – 50, 1984.

[80] H.J.A.M. Heijmans. An eigenvalue problem related to cell growth. *Journal of Mathematical Analysis and Applications*, 111:253–280, 1985.

[81] P-E Jabin and B Niethammer. On the rate of convergence to equilibrium in the beckerdring equations. *Journal of Differential Equations*, 191(2):518 – 543, 2003.

[82] K. Kar, M. Jayaraman, B. Sahoo, R. Kodali, and R. Wetzel. Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nature Structural and Molecular Biology*, 18(3):328–336, 2011.

[83] W.O. Kermack and A.G. McKendrick. A contribution to the mathematical theory of epidemics. *Proc. Roy. Society of London, Series A*, 115(772):700–721, 1927.

[84] W.O. Kermack and A.G. McKendrick. Contribution to the mathematical theory of epidemics. ii. the problem of endemicity. *Proc. Roy. Society of London, Series A*, 138(834):55–83, 1932.

[85] P S Kim, P P Lee, and D Levy. Modeling regulation mechanisms of the immune system. *J Theor Biol.*, 246(1):33–69, 2007.

[86] P S Kim, P P Lee, and D Levy. A PDE model for imatinib-treated chronic myelogenous leukemia. *Bull Math Biol.*, 70(7):1994–2016, 2008.

[87] T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson. An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science*, 326(5959):1533–1537, 2009.

[88] M.G. Krein and M.A. Rutman. Linear operators leaving invariant a cone in a banach space. *Amer. Math. Soc. Translation*, 26:128, 1950.

[89] H. E. Kubitschek. Growth during the bacterial cell cycle: Analysis of cell size distribution. *Biophysical Journal*, 9(6):792–809, 1969.

[90] R. Lattès and J.-L. Lions. *Méthode de quasi-réversibilité et applications.* Travaux et Recherches Mathématiques, No.15. Dunod, Paris, 1967.

[91] P. Laurençot and S. Mischler. From the discrete to the continuous coagulation-fragmentation equations. *Proc. Roy. Soc. Edinburgh Sect. A*, 132(5):1219–1248, 2002.

[92] P. Laurençot and B. Perthame. Exponential decay for the growth-fragmentation/cell-division equation. *Comm. Math. Sc.*, 7(2):503–510, 2009.

[93] P. Laurençot and C. Walker. Well-posedness for a model of prion proliferation dynamics. *J. Evol. Equ.*, 7(2):241–264, 2007.

[94] N. Lenuzza. *Modélisation de la réplication des Prions: implication de la dépendance en taille des agrégats de PrP et de l'hétérogénéité des populations cellulaires.* PhD thesis, Paris, 2009.

[95] O.V. Lepski. One problem of adaptive estimation in gaussian white noise. *Theory Probab. Appl.*, 35:459–470, 1990.

[96] O.V. Lepski. Asymptotic minimax adaptive estimation. 1. upper bounds. *Theory Probab. Appl.*, 36:645–659, 1991.

[97] O.V. Lepski. Asymptotic minimax adaptive estimation. 2. statistical models without optimal adaptation. adaptive estimators. *Theory Probab. Appl.*, 37:468–481, 1992.

[98] O.V. Lepski. On problems of adaptive estimation in white gaussian noise. In Khasminskii, editor, *Advances in Soviet Mathematics*, volume 12, pages 87–106. Amer. Math. Soc., Providence, 1992.

[99] I.M. Lifshitz and V.V. Slyozov. The kinetics of precipitation from supersaturated solid solutions. *Journal of Physics and Chemistry of Solids*, 19(12):35 – 50, 1961.

[100] A. Marciniak-Czochra, T. Stiehl, W. Jäger, A. Ho, and W. Wagner. Modeling of asymmetric cell division in hematopoietic stem cells - regulation of self-renewal is essential for efficient repopulation. *Stem Cells Dev.*, 18(3):377–385, 2009.

[101] J. Masel, V.A.A. Jansen, and M.A. Nowak. Quantifying the kinetic parameters of prion replication. *Biophysical Chemistry*, 77(2-3):139 – 152, 1999.

[102] P. Massart. *Concentration inequalities and model selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. Springer, Berlin, 2003.

[103] A.G. McKendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 1926.

[104] J. A. J. Metz and O. Diekmann, editors. *The dynamics of physiologically structured populations*, volume 68 of *Lecture Notes in Biomathematics*. Springer-Verlag, Berlin, 1986. Papers from the colloquium held in Amsterdam, 1983.

[105] P. Michel. Existence of a solution to the cell division eigenproblem. *Math. Models Methods Appl. Sci.*, 16(7, suppl.):1125–1153, 2006.

[106] P. Michel. General relative entropy in a non linear mckendrick model. *Contemporary Mathematics*, 429:205–225, 2007.

[107] P. Michel, S. Mischler, and B. Perthame. General entropy equations for structured population models and scattering. *C. R. Math. Acad. Sci. Paris*, 338(9):697–702, 2004.

[108] P. Michel, S. Mischler, and B. Perthame. General relative entropy inequality: an illustration on growth models. *J. Math. Pures Appl. (9)*, 84(9):1235–1260, 2005.

[109] P. Michel and M.T. Touaoula. Asymptotic behavior for a class of the renewal nonlinear equation with diffusion. *M3AS, published online*, 2012.

[110] F Michor. Mathematical models of cancer stem cells. *J Clin Oncol.*, 26(17):2854–2861, 2008.

[111] F Michor, T P Hughes, Y Iwasa, S Branford, N P Shah, C L Sawyers, and M A Nowak. Dynamics of chronic myeloid leukaemia. *Nature*, 435(7046):1267–1270, 2005.

[112] H. Mohamed and P. Robert. A probabilistic analysis of some tree algorithms. *Annals of Applied Probability*, 15(4):2445–2471, 2005.

[113] P. Moireau and D. Chapelle. Reduced-order Unscented Kalman Filtering with application to parameter identification in large-dimensional systems. *Cont. Optim. and Calc. Variat.*, 17:380–405, 2011.

[114] P. Moireau, D. Chapelle, and P. Le Tallec. Joint state and parameter estimation for distributed mechanical systems. *Computer Methods in Applied Mechanics and Engineering*, 197:659–677, 2008.

[115] P. Moireau, D. Chapelle, and P. Le Tallec. Filtering for distributed mechanical systems using position measurements: Perspectives in medical imaging. *Inverse Problems*, 25(3):035010 (25pp), 2009.

[116] I.M. Navon. Data assimilation for numerical weather prediction: A review. In *Data assimilation for atmospheric, oceanic, hydrologic applications*. Springer, 2009.

[117] 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):467–485 (electronic), 2000.

[118] B. Niethammer and J. J. L. Velázquez. Global well-posedness for an inhomogeneous LSW-model in unbounded domains. *Math. Ann.*, 328(3):481–501, 2004.

[119] B. Niethammer and J. J. L. Velázquez. On the convergence to the smooth self-similar solution in the LSW model. *Indiana Univ. Math. J.*, 55(2):761–794, 2006.

[120] M. Nussbaum. Asymptotic equivalence of density estimation and white noise. *Ann. Statist.*, 24:2399–2430, 1996.

[121] M. Nussbaum and S. Pereverzev. The degrees of ill-posedness in stochastic and deterministic noise models. *Preprint WIAS 509*, 1999.

[122] F Oosawa and S Asakura. *Thermodynamics of the polymerization of protein*. Academic Press, 1975.

[123] K. Pakdaman, B. Perthame, and D. Salort. Dynamics of a structured neuron population. *Nonlinearity*, 23(1):55–75, 2010.

[124] K. Pakdaman, B. Perthame, and D. Salort. Relaxation and self-sustained oscillations in the time elapsed neuron network model. (submitted), 2011.

[125] K. Pakdaman, B. Perthame, and D. Salort. Adaptation and fatigue model for neuron networks and large time asymptotics in a nonlinear fragmentation equation. (submitted), 2012.

[126] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter. Recurrent interactions in spiking networks with arbitrary topology. *Physical Review E*, 85(3):1–7, 2012.

[127] B. Perthame. *Transport equations in biology*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2007.

[128] B. Perthame and L. Ryzhik. Exponential decay for the fragmentation or cell-division equation. *J. Differential Equations*, 210(1):155–177, 2005.

[129] B. Perthame and S.K. Tumuluri. Nonlinear renewal equations. In *Selected topics in cancer modeling*, Model. Simul. Sci. Eng. Technol., pages 65–96. Birkhäuser Boston, Boston, MA, 2008.

[130] B. Perthame and J.P. Zubelli. On the inverse problem for a size-structured population model. *Inverse Problems*, 23(3):1037–1052, 2007.

[131] M. Pilant and W. Rundell. Determining the initial age distribution for an age structured population. *Math Population Stud.*, 3:496–506, 1991.

[132] E T Powers and D L Powers. The kinetics of nucleated polymerizations at high concentrations: Amyloid fibril formation near and above the supercritical concentration. *Biophysical Journal*, 91(1):122 – 132, 2006.

[133] S B Prusiner. Novel proteinaceous infectious particles cause scrapie. *Science*, 216(4542):136–144, 1982.

[134] J. Pruss, L. Pujo-Menjouet, G.F. Webb, and R. Zacher. Analysis of a model for the dynamics of prion. *Dis. Cont. Dyn. Sys. Ser. B*, 6(1):225–235, 2006.

[135] P Reynaud-Bouret and S Schbath. Adaptive estimation for hawkes processes; application to genome analysis. *Annals of Statistics*, 38(5):2781–2822, 2009.

[136] H. Rezaei. Prion protein oligomerization. *Current in Alzheimer research*, 5:572–578, 2008.

[137] I Roeder, M Herberg, and M Horn. An "age"-structured model of hematopoietic stem cell organization with application to chronic myeloid leukemia. *Bull Math Biol.*, 71(3):602–626, 2009.

[138] I Roeder, M Horn, I Glauche, A Hochhaus, M C Mueller, and M Loeffler. Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications. *Nat Med.*, 12(10):1181–1184, 2006.

[139] S I Rubinow. A maturity-time representation for cell populations. *Biophys J.*, 8:1055–1073, 1968.

[140] S I Rubinow. A theory for the age and generation time distribution for a microbial population. *Biophys J.*, 8:1055–1073, 1974.

[141] W. Rundell. Determining the birth function for an age structured population. *Mathematical Population Studies*, 1(4):377–395, 1989.

[142] W. Rundell. Determining the death rate for an age-structured population from census data. *SIAM Journal on Applied Mathematics*, 53(6):1731–1746, 1993.

[143] E Scherzinger, A Sittler, K Schweiger, V Heiser, R Lurz, R Hasenbank, G P Bates, H Lehrach, and E E Wanker. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for huntingtons disease pathology. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8):4604–4609, 1999.

[144] F R Sharpe and A J Lotka. A problem in age-distribution. *Philosophical Magazine*, 21:435–438, 1911.

[145] J.R. Silveira, G.J. Raymond, A.G. Hughson, R.E. Race, V.L. Sim, S.F. Hayes, and B. Caughey. The most infectious prion protein particles. *Nature*, 437(7056):257–261, September 2005.

[146] R. Simha. Kinetics of degradation and size distribution of long chain polymers. *Journal of Applied Physics*, 12:569–578, 1941.

[147] D. Simon. *Optimal State Estimation: Kalman, H^∞, and Nonlinear Approaches*. Wiley-Interscience, 2006.

[148] J.W. Sinko and W. Streifer. A new model for age-size structure of a population. *Ecology*, 48(6):910–918, 1967.

[149] J.W. Sinko and W. Streifer. A model for populations reproducing by fission. *Ecology*, 52(2):330–335, 1971.

[150] M. Smoluchowski. Drei vorträge über diffusion, brownsche molekularbewegung und koagulation von kolloidteilchen. *Physik. Zeitschr.*, 17:557–599, 1916.

[151] M. Smoluchowski. Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen. *Zeitschrift f. physik. Chemie*, 92:129–168, 1917.

[152] E. Stewart, R. Madden, G. Paul, and F. Taddei. Aging and death in an organism that reproduces by morphologically symmetric division. *Curr. Biol.*, 20(12):1099–103, 2010.

[153] T Stiehl and A Marciniak-Czochra. Characterization of stem cells using mathematical models of multistage cell lineages. *Mathematical and Computer Modelling*, 53(78):1505 – 1517, 2011.

[154] G. Strang. Wavelets and dilation equations: A brief introduction. *SIAM Review*, 31(4):614–627, 1989.

[155] E. C. Titchmarsch. *Introduction to the theory of Fourier integrals*. The Clarendon Press, 1937.

[156] G. Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14:651–667, 1977.

[157] P Wang, L Robert, J Pelletier, W L Dang, F Taddei, A Wright, and S Jun. Robust growth of Escherichia coli. *Curr. Biol.*, 20(12):1099–103, 2010.

[158] W-F Xue, S W Homans, and S E Radford. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *PNAS*, 105:8926–8931, 2008.