



**HAL**  
open science

# Estimation from Moments Measurements for Amyloid Depolymerisation

Aurora Armiento, Marie Doumic, Philippe Moireau, Human Rezaei

► **To cite this version:**

Aurora Armiento, Marie Doumic, Philippe Moireau, Human Rezaei. Estimation from Moments Measurements for Amyloid Depolymerisation. *Journal of Theoretical Biology*, 2016, 397, pp.68 - 88. 10.1016/j.jtbi.2016.02.037 . hal-01248255

**HAL Id: hal-01248255**

**<https://hal.science/hal-01248255>**

Submitted on 24 Dec 2015

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

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

# Estimation from Moments Measurements for Amyloid Degradation

Aurora Armiento <sup>\*</sup>   Marie Doumic <sup>†</sup>   Philippe Moireau <sup>‡</sup>   H. Rezaei <sup>§</sup>

January 15, 2016

## Abstract

Estimating reaction rates and size distributions of protein polymers is an important step for understanding the mechanisms of protein misfolding and aggregation, a key feature for amyloid diseases. This study aims at setting this framework problem when the experimental measurements consist in the time-dynamics of a moment of the population (*i.e.* for instance the total polymerised mass, as in Thioflavine T measurements, or the second moment measured by Static Light Scattering). We propose a general methodology, and we solve the problem theoretically and numerically in the case of a depolymerising system. We then apply our method to experimental data of degrading oligomers, and conclude that smaller aggregates of ovPrP protein should be more stable than larger ones. This has an important biological implication, since it is commonly admitted that small oligomers constitute the most cytotoxic species during prion misfolding process.

**Keywords:** Amyloid, prion, protein stability, oligomer, transport equation, state estimation, inverse problem, data assimilation

## Introduction

Protein aggregation is a key feature of a large range of diseases, called *amyloid* diseases, among which we can quote Alzheimer's, Parkinson's, Huntington's, transmissible spongiform encephalopathies (or prion diseases - *e.g.* Creutzfeldt–Jakob's, Kuru, bovine spongiform encephalopathy/madcow), etc [20, 23].

This category of diseases takes its name from the protein fibrils, called *amyloids*, which are formed during the disease and accumulate into the tissue. Their formation arise from

---

<sup>\*</sup> Univ Paris Diderot, Sorbonne Paris Cité, Lab. J.L. Lions, UMR CNRS 7598, Inria , Paris, France

<sup>†</sup>Sorbonne Universités, Inria, UPMC Univ Paris 06, Lab. J.L. Lions UMR CNRS 7598, Paris, France

<sup>‡</sup>Inria and Université Paris-Saclay, Campus de l'École Polytechnique, 91128 Palaiseau, France

<sup>§</sup>Virologie et Immunologie Moléculaires, Institut National de la Recherche Agronomique, F-78352 Jouy-en-Josas, France

misfolded versions of proteins present naturally in the body, each disease having its specific precursor protein (*e.g.* APP for Alzheimer’s, PrP for Prion,  $\beta_2m$  for haemodialysis-associated amyloidosis). While their accumulation in organs is characteristic for the disease, the reason for their association as well as their role in tissue damages are still unclear. Moreover, their aggregation mechanisms - most probably specific for each protein involved - are at the moment largely unknown.

The main reasons for so many open questions to remain, despite both the longstanding interest raised in the biological, biophysical and biochemical communities, and the major importance of amyloid diseases for public health, are twofold. First, the number of possible chain-reactions involved is huge, possibly infinite - as the size of aggregates is. Hence model design and discrimination is very complex, and conclusions made on a specific protein are hardly translatable to another one. Second, the most common experimental devices can measure averaged quantities on the polymerised proteins, such as the total polymerised mass (Thioflavine T measurements [5]) or the average size of polymers (Static Light Scattering (SLS) [27]). How such measurements may be used to estimate reaction rates (which may also be an infinity) and size distribution of aggregates, and thus to select the major mechanisms, is an emerging field of inverse problems with few theoretical progress [1] and positive results on experimental data [29, 30].

To contribute to this new field, this article focus on one of the major concerns in pathologies due to protein misassembly and aggregation: the determination of oligomer size distribution. It has been reported that – while amyloid fibrils present low biological activity – oligomers and small assemblies are the cytopathogenic elements [26, 13]. Depending on the type of pathology and the protein involved, oligomers could either be involved into the pathway of amyloid fibrils formation or be associated to an independent pathway, which only leads to the formation of oligomers. Oligomer size characterisation can play a key role in distinguishing between these pathways. Therefore, the investigation on size distribution remains the first step to understand how oligomers are formed, their biological activity and their biophysical characterisation to finally design therapeutic strategies.

This question - how to estimate size distributions - leads us to setting a framework problem and studying it, both theoretically and numerically, in one of its simplest possible version. We then apply our method to experimental data, using the time-dependent average size of polymers (measured by SLS) to reconstruct the oligomer initial size distribution. We compare our estimation to the experimental estimation obtained by chromatography and discuss the implications of our results. Eventually, we discuss the new problems and possibilities opened-up by these results, and how this methodology could easily be adapted to other models and experiments.

## Mathematical Setting

Since protein aggregates can reach extremely large average sizes, we adopt here a continuous framework [22] and denote  $x \in (0, \infty)$  the *size* of an aggregate, *i.e.*  $x$  represents the (rescaled) quantity of monomers contained in a given polymer. We thus call  $u(x, t)$  the concentration of polymers of size  $x$  at time  $t$  (see [2] for a discussion and theoretical

justification of the use of a continuous size variable rather than a discrete one).

One of the techniques most widely used is the measurement of Thioflavin T (ThT) fluorescence, [5], which provides measurements of the total polymerised mass, *i.e.* a linear transformation of the first moment of the concentration function

$$z_{\text{tht}}(t) = c_1 \int_0^\infty xu(x, t)dx.$$

The Static Light Scattering (SLS) technique, [27], could give us an affine transformation of the second moment

$$z_{\text{sls}}(t) = c_1 \int_0^\infty x^2u(x, t)dx + c_2,$$

where  $c_1 \geq 0$ ,  $c_2 \in \mathbb{R}$ .

The framework problem we want to contribute stands: Under which assumptions (and limitations) is it possible to estimate the reaction rates and/or the initial size distribution, from a time measurement  $z_{\text{tht}}(t)$  or  $z_{\text{sls}}(t)$ ?

As a first simplifying assumption, we model the primary reactions involved in the evolution of polymers with the Lifshitz-Slyozov system, that is one of the most common polymerising/depolymerising model. In this system, polymers (or clusters, in another application context) can only grow by monomer addition, with a size-dependent reaction rate  $a(x)$ , and depolymerise by monomer loss, with a reaction rate  $b(x)$ . This results in the following system

$$\begin{cases} \frac{\partial}{\partial t}u(x, t) + \frac{\partial}{\partial x} \left( (a(x)v(t) - b(x))u(x, t) \right) = 0, & x \in [0, \ell], t \geq 0, \\ u(\ell, t) = 0, \\ u(x, 0) = u_0(x), \end{cases} \quad (1)$$

where  $\ell \in (0, \infty]$  is the upper bound of polymer sizes. We assume here  $\ell < \infty$ , in contrast with the initial Lifshitz-Slyozov model [14]. The function  $v(t)$  is the concentration of monomers in the cuvette and is directly related to polymer concentration from the following mass conservation law

$$v(t) + \int_0^\ell xu(x, t)dx = v(0) + \int_0^\ell xu_0(x)dx > 0 \quad \forall t \geq 0. \quad (2)$$

When applied to amyloid formation, this model may be seen as a qualitative model taking into account what biologists call *primary pathway* and neglecting, as a first approach, *secondary pathways* such as fragmentation or coalescence [8]. Note that there are many other possible applications of this model, such as phase transition, which was the original application for which it had been designed [14].

The problem now stands: Measuring  $z_{\text{sls}}(t)$  or  $z_{\text{tht}}(t)$ , or more generally the time-dependence of a  $n$ -th moment defined by  $\int_0^\ell x^n u(x, t)dx$ , with  $u(x, t)$  solution of System (1)(2), what may be possibly estimated among the unknown quantities, *i.e.* the initial state  $u_0(x)$  and the parameter functions  $a(x)$  and  $b(x)$ ?

This problem in its full generality is both nonlinear and highly ill-posed. Hence, we proceed to further simplifications and study the state estimation of a model of pure depolarisation. Assuming to start with no monomers, *i.e.*  $v(0) = 0$ , we can neglect the polymerisation term, at least during the beginning of the reaction – see Figure 12 for measurements of such an experiment. The model then becomes

$$\begin{cases} \frac{\partial}{\partial t}u(x, t) - \frac{\partial}{\partial x}(b(x)u(x, t)) = 0, & x \in [0, \ell], t \geq 0, \\ u(\ell, t) = 0, \\ u(x, 0) = u_0(x). \end{cases} \quad (3)$$

The state estimation problem may be formulated as follows:

(IP) How to estimate  $u_0$  – the initial condition of System (3) – from the given *a priori* knowledge of  $b(x)$  and measurement  $\int_0^\ell x^n u(x, t) dx$ ?

In order to settle a general framework, easy to adapt to more complex problems in the future, we introduce below the notations for the standard state-space formalism used for dynamical systems.

### State space formalism

We introduce the state space  $\mathcal{U} = \mathbf{L}^2([0, \ell])$  equipped with its natural norm and introduce the state variable  $u$  standing for the function

$$u : \begin{cases} [0, \tau] & \longrightarrow & \mathcal{U} \\ t & \longmapsto & u(t) : \begin{cases} [0, \ell] & \longrightarrow & \mathbb{R} \\ x & \longmapsto & u(x, t). \end{cases} \end{cases}$$

Then we rewrite System (3) in the state-space form

$$\begin{cases} \frac{du}{dt} = Au, \\ u(0) = u_0, \end{cases} \quad (4)$$

where  $A$  is the linear functional operator – called model operator –

$$A : \begin{cases} \mathcal{D}(A) \subset \mathbf{L}^2([0, \ell]) & \longrightarrow & \mathbf{L}^2([0, \ell]), \\ f & \longmapsto & \partial_x(bf), \end{cases}$$

of domain

$$\mathcal{D}(A) = \{f \in \mathbf{H}^1([0, \ell]) \mid f(\ell) = 0\}.$$

Assuming  $b' \in \mathbf{L}^\infty([0, \ell])$ , we easily prove that there exists  $\lambda$  such that the operator  $A - \lambda Id$  is dissipative, hence  $A$  is the generator of a strongly continuous semigroup  $\mathbb{T}_t$  – see for instance [3] for an introduction to such concepts.

We formalise our measurement procedure by introducing the observation space  $\mathcal{Z} = \mathbb{R}$  and a so-called observation operator associated in our case to the  $n$ -th momentum of a given state variable

$$C : \begin{cases} \mathcal{U} & \longrightarrow & \mathcal{Z}, \\ u & \longmapsto & \int_0^\ell x^n u(x) dx, \end{cases} \quad (5)$$

which is a time-invariant linear bounded operator with  $\|C\| \leq \ell^{n+\frac{1}{2}}$ . In the following sections, we will use the notation  $C^n$  when we need to stress the dependence on the order of the moment.

Note that the observation operator is defined independently of the model  $A$ . However – by taking into account the model dynamics – we easily write the relation between observations and the initial condition of polymer concentration. To do so, we introduce the operator  $\Psi_\tau \in \mathcal{L}(\mathcal{U}, \mathbf{L}^2([0, \tau], \mathcal{Z}))$

$$\Psi_\tau : \begin{cases} \mathcal{U} & \longrightarrow & \mathbf{L}^2([0, \tau]), \\ u_0 & \longmapsto & C\mathbb{T}_\tau u_0, \end{cases}$$

since in our case  $\mathbf{L}^2([0, \tau], \mathcal{Z}) = \mathbf{L}^2([0, \tau])$ .

Let us now denote by  $z(t)$  the observations at our disposal. We can say that the observations are related to a target solution  $\check{u}$  of System (3) up to some measurement errors – *i.e.* observation noise –  $\chi$ . Formally, we have

$$z = C\check{u} + \chi.$$

Using the various operators introduced, we formulate our inverse problem in two equivalent forms. In the more classical inverse-problem formulation, our objective appears as

Inverting  $\Psi_\tau$  to reconstruct  $\check{u}_0$  from the given measurement  $z$  generated through time  $t \in [0, \tau]$ .

In a more data assimilation form we aim at

Estimating  $\check{u}_0$  from given measurements  $z$  generated through time  $t \in [0, \tau]$ , knowing the model dynamics  $A$  and the model of observation operator  $C$ .

In Section 1, we consider the specific case where the depolymerisation rate  $b(x)$  is constant: we show that the problem is equivalent to the estimation of the  $(n + 1)$ -th derivative of the measurement, so that we can use (for instance) a kernel regularisation method for which we recall the standard convergence results. This gives us some light on what we could expect for convergence in more general cases. In Section 2, we turn to the variational formulation, recall its intrinsic links with the previous regularisation method, and extend it to non constant  $b(x)$ . We then illustrate our results by numerical simulations in Section 3. We apply our method, together with a statistical study for the measurement noise, to analyse the experimental data in Section 4. All this exploratory study leads us to sketch perspectives for future work and open problems.

# 1 First Approach: kernel regularisation

In this section we assume to have a constant depolymerisation rate  $b(x) = b > 0$ . We know that in this case, the solution of System (3) is given by  $u(x, t) = u_0(x + bt)$ . We have, by a simple change of variable, that

$$\forall t > 0, \quad C^n u(t) = \int_0^\ell x^n u(x, t) dx = \int_{bt}^\ell (x' - bt)^n u_0(x') dx',$$

and therefore

$$\Psi_\tau^n : \begin{cases} \mathbf{L}^2([0, \ell]) & \longrightarrow & \mathbf{L}^2([0, \tau]), \\ u_0 & \longmapsto & \left( t \rightarrow \int_{bt}^\ell (x - bt)^n u_0(x) dx \right), \end{cases} \quad (6)$$

where  $n$  can be avoided when not necessary. We easily see that

$$\text{Ran} \Psi_\tau^n = \left\{ u \in \mathbf{H}^{n+1}([0, \tau]), u_0(\tau) = \dots = u_0^{(n)}(\tau) = 0 \right\}.$$

Deriving recursively  $\Psi_\tau u_0$ , we obtain

$$\frac{d^{n+1}}{dt^{n+1}} (\Psi_\tau u_0) = (-b)^{n+1} n! u_0(bt) \quad \text{for } n \geq 0,$$

so that we have the following explicit formula for  $u_0$

$$\boxed{u_0(x) = \frac{1}{n!(-b)^{n+1}} \frac{d^{n+1}}{dt^{n+1}} (\Psi_\tau u_0) \left( \frac{x}{b} \right), \quad \text{for } n > 0.} \quad (7)$$

In the previously seen formalism, we model an additive noise as follows: we call  $\varepsilon$  the upper bound for the noise level measured in a Sobolev space  $\mathbf{W}^{-s,p}([0, \tau])$ -norm, and we assume

$$\|\chi\|_{\mathbf{W}^{-s,p}([0, \tau])} \leq \varepsilon. \quad (8)$$

The choice for the parameters  $s$  and  $p$  depends on the kind of noise ( $s = 0$  for a deterministic noise,  $s = \frac{1}{2}$  and  $p = 2$  for a deterministic equivalent of a gaussian white noise [17]). The ill-posedness of the problem comes from the fact that the noisy measurement  $z$  is in general not differentiable, so that we cannot use directly Equality (7) to solve our problem. This is a classical linear ill-posed problem of order  $\delta_{IP} = n + 1$  in the scale  $\mathbf{W}^{k,p}$ , see [12]. Before applying Formula (7), we need to regularise our measurement  $z$ . A classical regularisation method consists in convolving the measurement with a mollifier sequence, method called kernel density estimation for the statistical problem of estimating the density from an i.i.d. sample [28]. Thanks to classical results, we know that the regularity of the convolution depends on the regularity of both the measurement and the kernel. Let us take a kernel function  $\rho \in \mathcal{C}_c^\infty(\mathbb{R})$ , such that

$$\int_{\mathbb{R}} \rho(x) dx = 1, \quad \int_{\mathbb{R}} x^k \rho(x) dx = 0, \quad \text{for } 1 \leq k \leq m. \quad (9)$$

We define the family of mollifiers  $\rho_\alpha$  by

$$\rho_\alpha = \frac{1}{\alpha} \rho\left(\frac{x}{\alpha}\right), \quad (10)$$

depending on the parameter  $\alpha > 0$ . Our estimation of the initial condition is carried out by the function

$$\hat{u}_0^{\varepsilon, \alpha} = \frac{d^{n+1}}{dx^{n+1}} \rho_\alpha * \left( \frac{1}{n!(-b)^{n+1}} z\left(\frac{x}{b}\right) \right)$$

where the convolution operator  $*$  is defined by  $f * g(x) = \int_{\mathbb{R}} g(x') f(x - x') dx'$ . The accuracy of our approximation shall depend on the noise level  $\varepsilon$ , on the regularity of the kernel family, on the parameter  $\alpha$  and on the order of the derivative, that is  $n + 1$ . Classically, we obtain an optimal upper bound for the accuracy of the estimation as stated in the following proposition.

**PROPOSITION 1**

Let  $1 \leq p < \infty$ ,  $n \in \mathbb{N}$ ,  $0 \leq s < 1$  and let  $\check{u}_0 \in \mathbf{W}^{m+1,p}([0, \ell])$  with  $m$  defined as in Equation (9). Let  $\Psi_\tau \check{u}_0 \in \mathbf{W}^{m+n+2,p}([0, \tau])$  defined in Equation (6). Let  $z \in \mathbf{W}^{-s,p}([0, \tau])$  a measurement of the  $n$ -th momentum  $\Psi_\tau \check{u}_0$  such that  $\tau \geq \frac{\ell}{b}$  and

$$\|z - \Psi_\tau \check{u}_0\|_{\mathbf{W}^{-s,p}([0, \tau])} \leq \varepsilon.$$

Let us define

$$\check{u}_0(x) = \frac{1}{n!(-b)^{n+1}} \frac{d^{n+1}}{dt^{n+1}} \Psi_\tau \check{u}_0\left(\frac{x}{b}\right), \quad (11)$$

Let  $\rho$  defined by Equation (9) and  $\rho_\alpha$  by Equation (10), with  $\alpha \in (0, 1)$ . We define

$$\hat{u}_0^{\varepsilon, \alpha} = \frac{d^{n+1}}{dx^{n+1}} \rho_\alpha * \left( \frac{1}{n!(-b)^{n+1}} z\left(\frac{x}{b}\right) \right) \quad (12)$$

as an approximation of  $\check{u}_0$ . Then the following estimation is of optimal order in the sense of [12]

$$\|\hat{u}_0^{\varepsilon, \alpha} - \check{u}_0\|_{\mathbf{L}^p([0, \ell])} \leq \Theta \left( \frac{\varepsilon}{\alpha^{n+s+1}} + \alpha^{m+1} \right) = F_\varepsilon(\alpha), \quad (13)$$

where the constant  $\Theta$  depends on  $\|\Psi_\tau \check{u}_0\|_{\mathbf{W}^{m+n+2,p}([0, \tau])}$ ,  $\|x^{m+1} \rho\|_{\mathbf{L}^1(\mathbb{R})}$ ,  $\|\rho^{(n)}\|_{\mathbf{L}^1(\mathbb{R})}$ ,  $\|\rho^{(n+1)}\|_{\mathbf{L}^1(\mathbb{R})}$ .

For the sake of completeness, the proof of this proposition is recalled in Appendix 6. This gives us an *a priori* method to choose the parameter  $\alpha$ : Aiming at the smallest approximation error – we select the  $\alpha$  that minimises  $F_\varepsilon(\alpha)$ . The *a priori* optimal choice for  $\alpha$  is the minimiser of the convex function  $F_\varepsilon(\alpha)$

$$\alpha_{opt} = O\left(\varepsilon^{\frac{1}{n+m+2+s}}\right). \quad (14)$$



By this choice, we obtain an estimation  $\hat{u}_0^{\varepsilon, \alpha \text{opt}}$  such that

$$\boxed{\|\hat{u}_0^{\varepsilon, \alpha \text{opt}} - \check{u}_0\|_{L^p([0, \ell])} = O\left(\varepsilon^{\frac{m+1}{n+m+2+s}}\right)}. \quad (15)$$

In the case of a variable depolymerisation rate  $b(x)$ , computations are not so easy and in general we do not have such an explicit relation between measurements and initial condition. This is part of the reasons why we now turn to data assimilation approaches.

## 2 Second Approach: a data assimilation variational approach

In this section, we propose to base our inverse problem solving strategy on the so-called 4d-Var approach as named by [15]. The principle consists in minimising – hence the variational designation – with respect to the initial condition a least-square criterion  $\mathcal{J}$  combining the discrepancy between the actual data and the simulation, with additional regularisation terms accounting for the confidence in the model.

The advantage of this method lies in its very general formalism that leads to high flexibility in the choice of the model operator or the observation operator.

Typically, we decompose  $\check{u}_0$  as the sum of a known *a priori*  $u_o$ , and an unknown variation  $\check{\xi}$  representing the uncertain part of our initial concentration

$$\check{u}_0 = u_o + \check{\xi}. \quad (16)$$

As  $\check{\xi}$  is unknown, the trajectory  $\{\check{u}(t), t \in [0, \tau]\}$  cannot be obtained directly. However, we can parametrise the dynamics (4) with respect to any guess  $\xi$  of  $\check{\xi}$ . We denote by  $\{u_{|\xi}(t), t \in [0, \tau]\}$  the resulting state trajectory knowing the guess  $\xi$

$$\begin{cases} \dot{u}_{|\xi} = Au_{|\xi} \\ u_{|\xi}(0) = u_o + \xi. \end{cases} \quad (17)$$

We then write the criterion to minimise

$$\mathcal{J}_\tau(\xi) = \frac{1}{2} \langle \xi, P_0^{-1} \xi \rangle_{\mathcal{U}} + \frac{1}{2} \int_0^\tau \gamma |z - C(u_{|\xi})|^2 dt. \quad (18)$$

The isomorphism on  $\mathcal{U}$ , namely  $P_0$ , and the scalar  $\gamma$  are weights on the natural norm on  $\mathcal{U}$  and  $\mathcal{Z}$ , respectively. These weights are defined in accordance with the level of confidence into our *a priori* on the initial condition and the *measurement* – typically based on an *a priori* evaluation of the noise  $\chi$ . Note that contrarily to the kernel regularisation method, if the space for the noise is less regular than  $L^2$ , this method cannot be used directly: prior regularisation on the measurement is needed. On the contrary, this method provides a unique minimiser for general rates  $b(x)$  as soon as the direct problem is well-posed.

Our objective is to minimise  $\mathcal{J}_\tau$  under the constraint of the model dynamics (17). We thus introduce the so-called *adjoint variable*  $q_{|\xi, \tau}$  as the Lagrange multiplier associated

with the dynamical constraint (17). The adjoint variable is then solution – see for instance [7] – of the dynamics

$$\begin{cases} \dot{q}_{|\xi, \tau} + A^* q_{|\xi, \tau} = -\gamma C^* (z - C u_{|\xi}), & t \in [0, \tau] \\ q_{|\xi, \tau}(\tau) = 0, \end{cases} \quad (19)$$

where  $A^*$  is the adjoint of the model operator defined by

$$\begin{aligned} A^* : \mathcal{D}(A^*) \subset \mathbf{L}^2([0, \ell]) &\longrightarrow \mathbf{L}^2([0, \ell]) \\ f &\longmapsto -b(x) \partial_x f \end{aligned}$$

with domain

$$\mathcal{D}(A^*) = \{f \in \mathbf{H}^1([0, \ell]) \mid f(0) = 0\}$$

and  $C^*$  is the adjoint of the observation operator  $C$  defined by (5), hence

$$\begin{aligned} C^* : \mathbb{R} &\longrightarrow \mathbf{L}^2([0, \ell]) \\ r &\longmapsto f_r : x \mapsto x^n r. \end{aligned}$$

Therefore, the adjoint system reads in strong formulation

$$\begin{cases} \frac{\partial}{\partial t} q_{|\xi, \tau}(x, t) - b(x) \partial_x q_{|\xi, \tau}(x, t) \\ \qquad \qquad \qquad = -\gamma x^n \left( z - \int_0^\ell x'^n \bar{u}(x', t) dx' \right), & x \in [0, \ell], t \in [0, \tau] \\ q_{|\xi, \tau}(0, t) = 0, \\ q_{|\xi, \tau}(x, \tau) = 0. \end{cases} \quad (20)$$

Using the adjoint variable, a standard computation allows to characterise  $\bar{\xi}_{|\tau} = \arg \min_{\xi} \mathcal{J}_{\tau}$  as

$$\bar{\xi}_{|\tau} = P_0 \bar{q}_{|\tau}(0),$$

where  $\bar{q}_{|\tau}$  is the adjoint variable associated to the  $\bar{u}_{|\tau} = u_{|\bar{\xi}_{|\tau}}$ , hence leading to a famous both-end problem formulation [7]

$$\begin{cases} \dot{\bar{u}}_{|\tau} = A \bar{u}_{|\tau}, & t \in [0, \tau] \\ \dot{\bar{q}}_{|\tau} + A^* \bar{q}_{|\tau} = -\gamma C^* (z - C \bar{u}_{|\tau}), & t \in [0, \tau] \\ \bar{u}_{|\tau}(0) = u_{\circ} + P_0 \bar{q}_{|\tau}(0), \\ \bar{q}_{|\tau}(\tau) = 0. \end{cases} \quad (21)$$

## 2.1 Equivalence with the kernel regularisation method

According to a classical interpretation, we can read the second term of the criterion as the ordinary least-square data fitting term, while the first term is often considered as a regularisation term by choosing

$$P_0 = \frac{1}{\beta} \text{Id},$$

with  $\beta$  small enough so that

$$\beta \|\bar{\xi}_\tau\|^2 \ll \gamma \int_0^\tau \|\chi\|^2 dt.$$

Minimising the criterion  $\mathcal{J}_\tau$  is then equivalent to minimise for  $\alpha^2 = \frac{\beta}{\gamma}$

$$\min_{\xi} \left\{ \alpha^2 \|\xi\|_{\mathbf{L}^2([0,\ell])}^2 + \|z(t) - \Psi_\tau(\xi)\|_{\mathbf{L}^2([0,\tau])}^2 \right\},$$

where clearly appears the classical Tikhonov regularisation.

Moreover, we can consider different criteria by changing the state space  $\mathcal{U}$  or considering different  $P_0$ . For instance, when choosing

$$P_0 = \frac{1}{\beta} \text{Id}, \quad \mathcal{U} = \mathbf{H}^s([0, \ell]),$$

the variational method is equivalent to the *generalized* Tikhonov method where we minimise

$$\min_{\xi} \left\{ \alpha \|\xi\|_{\mathbf{H}^s([0,\ell])}^2 + \|z(t) - \Psi_\tau(\xi)\|_{\mathbf{L}^2([0,\tau])}^2 \right\}.$$

Note that  $s > -\frac{n+1}{2}$  is necessary for this minimisation to be a regularising method - see for instance the analysis of Tikhonov's regularisation in Hilbert scales in [6], and below the comments on the observability condition.

To give some insight into the links between the two regularisation methods, let us take the case of classical Tikhonov regularisation,  $b$  constant,  $u_o = 0$  with  $\tau \geq \frac{\ell}{b}$ .

We recall that  $\Psi_\tau(\xi)(t) = \int_{bt}^\ell (y - bt)^n \xi(y) dy$ . The operator  $\Psi_\tau$  is injective, compact, and with dense image when taken from  $L^2([0, \ell])$  to  $L^2([0, \tau])$  with  $\tau = \ell/b$ . Its adjoint operator is

$$\Psi_\tau^*(v)(x) = \int_0^{\frac{y}{b}} (y - bt)^n v(t) dt.$$

This provides us with the following result.

**PROPOSITION 2**

For any  $z \in L^2([0, \tau])$ , there exists a unique minimiser  $\bar{\xi}$  for  $J(\xi)$  defined by

$$J(\xi) = \frac{\alpha^2}{2} \|\xi\|_{\mathbf{L}^2([0,\ell])}^2 + \frac{1}{2} \int_0^\tau |z(t) - \Psi_\tau(\xi)|^2 dt,$$

and  $\bar{\xi} \in \mathbf{H}^{n+1}([0, \ell])$ . If moreover  $\check{\xi} \in \mathbf{H}^{n+1}([0, \ell])$  with  $\check{\xi}(0) = \dots = \check{\xi}^{(n)}(0) = 0$ , we have the following estimate

$$\|\check{\xi} - \bar{\xi}\|_{\mathbf{L}^2([0,\ell])} \leq \frac{1}{\alpha} \|z - \Psi_\tau(\check{\xi})\|_{\mathbf{L}^2([0,\tau])} + \alpha \|\check{\xi}\|_{\mathbf{H}^{n+1}([0,\ell])}.$$

We recognise here the case  $s = 0$  and  $n = m$  of Proposition 1, by denoting  $\alpha = \tilde{\alpha}^{n+1}$ . For the sake of completeness, we sketch out the proof in Appendix 6.

## 2.2 Observability Condition

In data assimilation, the well-posedness or ill-posedness of the inverse problems is characterised by a so-called observability condition translating that there is enough information in the data to reconstruct the initial condition. Typically, this condition is of the form

There exists a time  $\tau_0$  and a constant  $\Theta > 0$  such that, for all  $u \in C([0, \tau], \mathcal{U})$  solution of System (4), we have

$$\forall \tau > \tau_0, \quad \int_0^\tau |Cu(t)|^2 dt > \Theta \|u_0\|_{\mathcal{U}}^2. \quad (22)$$

Following our previous computation, we remark that when  $\mathcal{U} = \mathbf{L}^2([0, \ell])$  equipped with its natural norm Inequality (22) cannot be satisfied. Let us see this in the  $b$ -constant case.

$$\int_0^\tau |Cu|^2 dt = \int_0^\tau \left[ \int_0^\ell x^n u(x, t) dx \right]^2 dt = \int_0^\tau \left[ \int_{bt}^\ell x^n u_0(x + bt) dx \right]^2 dt$$

If  $n = 0$ , we call  $F(bt) = \int_{bt}^\ell u_0(x + bt) dx$ . The observability condition then reads

$$\int_0^\tau F(s)^2 ds \geq \Theta \int_0^\ell F'(s)^2 ds.$$

Counter examples proving that this cannot be uniformly the case for any  $F$  are well-known, take for instance any mollifier sequence  $\rho_\alpha = \frac{1}{\alpha} \rho(\frac{x}{\alpha})$ , where  $\rho \in C_c^\infty((0, \min\{\ell, \tau\}))$ .

However, it would have been possible to have the observability condition if we would have chosen different metrics. For example, let us consider the case of very regular observations in  $\mathbf{H}^{n+1}([0, \tau], \mathcal{Z})$  with the seminorm

$$|f|_{\mathbf{H}^{n+1}([0, \tau])} = \int_0^\tau \left| \frac{d^{n+1} f}{dt^{n+1}} \right|^2 dt$$

and state space  $\mathcal{U} = \mathbf{L}^2([0, \ell])$ . Thanks to Equation (7), we can easily find a constant  $\Theta$  that satisfies,  $\forall \tau > \tau_0 = \frac{\ell}{b}$ , the observability condition

$$\int_0^\tau \left| \frac{d^{n+1} \Psi_\tau u_0(t)}{dt^{n+1}} \right|^2 dt \geq \Theta \|u_0\|_{\mathbf{L}^2([0, \ell])}^2,$$

associated to the criterion

$$J(\xi) = \frac{\gamma_\xi}{2} \|\xi\|_{\mathbf{L}^2([0, \ell])}^2 + \frac{\gamma_z}{2} \|z(t) - \Psi_\tau u_0(t)\|_{\mathbf{H}^{n+1}([0, \tau])}^2.$$

Alternatively, we can satisfy the observability condition in the case of less regular initial condition in  $\mathcal{U} = \mathbf{H}^{-(n+1)}([0, \ell])$  and observations in  $\mathbf{L}^2([0, \tau])$ . The criterion thus reads

$$J(\xi) = \frac{\gamma\xi}{2} \|\xi\|_{\mathbf{H}^{-(n+1)}([0, \ell])}^2 + \frac{\gamma z}{2} \|z(t) - \Psi_\tau u_0(t)\|_{\mathbf{L}^2([0, \tau])}^2.$$

Therefore, the inequality of the observability condition becomes

$$\int_0^\tau |\Psi_\tau u_0(t)|^2 \geq \Theta \|u_0\|_{\mathbf{H}^{-(n+1)}([0, \ell])}^2.$$

According to Equation (7), we can rewrite this inequality as

$$\int_0^\tau |\Psi_\tau u_0(t)|^2 \geq \frac{\Theta}{n!(-b)^{n+1}} \left\| \frac{d^{n+1}}{dt^{n+1}} (\Psi_\tau u_0) \right\|_{\mathbf{H}^{-(n+1)}([0, \tau])}^2.$$

It is easy to prove that

$$\left\| \frac{d^{n+1}}{dt^{n+1}} \Psi_\tau u_0 \right\|_{\mathbf{H}^{-(n+1)}([0, \tau])} \leq \|\Psi_\tau u_0\|_{\mathbf{L}^2([0, \tau])},$$

and we can satisfy the observability condition with  $\Theta = n!(-b)^{n+1}$  and  $\tau_0 = \frac{\ell}{b}$ .

In these two cases (the space  $[0, \tau] \rightarrow \mathcal{Z}$  equipped with a very regular norm, or on the contrary the very weak assumption on the regularity of the initial state  $\mathcal{U}$ ), the observability condition shows that the problem is well-posed for  $\tau \geq \frac{\ell}{b}$  and so there is no need for either a regularisation or an *a priori* information. However they cannot be used for real applications since we do not observe  $z(t)$  in such a regular space, and we want to reconstruct regular initial states.

## 3 Numerical Analysis

### 3.1 Model discretisation

In this section we describe the numerical implementation and the comparison between the two approaches.

We set the space domain to  $[0, \ell] = [0, 200]mer$ . We define the notation *mer* for monomer which is the fundamental unit aggregating into oligomers. We set the time domain to  $[0, \tau] = [0, 100]min$  and the transport velocity to  $b = 2min^{-1}$ .

We present two cases associated to two initial concentration conditions: the gaussian function  $u_{0g} = e^{\frac{1}{2} - \frac{(x-100)^2}{20^2}} \mu M$  and the characteristic function  $u_{0ch} = I_{[70, 130]} \mu M$ . For the sake of simplicity, in the following of this section we omit the units.

We consider a uniform space grid  $0 = x_0 < \dots < x_{N_x} = \ell$ , with a constant space step  $\delta x$ . By evaluating the continuous initial conditions on this grid, we obtain the vector

$$\check{u}_0 = (\check{u}_{0,j})_{0 \leq j \leq N_x} = \check{u}_0(x_j).$$

We fix a time discretisation  $t_0 < \dots < t_{N_t}$  of the time domain  $[0, \tau]$  with a constant time step  $\delta t$ . We call  $u_j^k$  the approximation of  $u(x_j, t_k)$ . The cluster concentrations, at time  $t_k$ , are approximated by the vector  $u^k = (u_j^k)_j$ . To compute these quantities, we refer to the discrete model

$$\begin{cases} u^{k+1} = A_{k+1|k} u^k, & \text{for } k \in \mathbb{N} \\ u^0 = u_0. \end{cases} \quad (23)$$

The expression of the discrete model operator  $A_{k+1|k}$  depends on the numerical scheme which is adopted to discretise the transport equation of System (3). For the upwind scheme it is

$$A_{k+1|k} = \mathbb{1}_{N_x} + \delta t b D, \quad (24)$$

where the discrete differential operator  $D$  is such that

$$(Du^k)_j = \frac{u_{j+1}^k - u_j^k}{\delta x} \quad \text{if } b > 0 \quad (Du^k)_j = \frac{u_j^k - u_{j-1}^k}{\delta x} \quad \text{if } b < 0.$$

We can also use a numerical scheme with higher approximation order such as a Lax-Wendroff scheme. The discrete model operator associated to this scheme is

$$A_{k+1|k} = \mathbb{1}_{N_x} + \frac{b\delta t}{\delta x} D_x^c + \frac{b^2\delta t^2}{2\delta x^2} D_{xx},$$

where

$$(D_x^c u^k)_j = \frac{u_{j+1}^k - u_{j-1}^k}{2\delta x} \quad \text{and} \quad (D_{xx} u^k)_j = \frac{u_{j+1}^k - 2u_j^k + u_{j-1}^k}{2\delta x^2}.$$

We choose space and time steps satisfying the *Courant–Friedrichs–Levy (CFL) condition*  $|\frac{b\delta t}{\delta x}| \leq 1$  that ensures the stability of the schemes [16].

### 3.2 Synthetic data generation

To test our inversion strategies, we generate synthetic observations. In this respect we fix uniform grids on  $[0, \tau]$  and  $[0, \ell]$  with discretisation steps much smaller than the ones considered solving the inverse problem. Specifically, we take the time step  $\delta t = 10^{-3}$  and space step  $\delta x = 2 \cdot 10^{-3}$ . We use the discrete model (23) with  $\check{u}_0$  as initial condition to compute the sequence  $(\check{u}^k)_{1 \leq k \leq N_{\text{obs}}}$ . Consequently, we compute the observations thanks to the discrete observation operator

$$C_k^{(n)} = \delta x \begin{pmatrix} \frac{x_0^n}{2} & x_1^n & \dots & x_{N_x-1}^n & \frac{x_{N_x}^n}{2} \end{pmatrix}$$

obtained by using the trapezoidal rule to approximate the space integral appearing in the continuous definition. We remark that – since the continuous observation operator  $C$  is time independent –  $C_k$  does not depend on  $k$ .

We consider synthetic observations of the form

$$z_k = C_k \check{u}_k + \chi_k, \quad k = 0, \dots, N_{\text{obs}}, \quad (25)$$

where  $\chi_k = \varepsilon\omega_k$  and the values  $\omega_k$  are randomly generated according to the standard gaussian distribution. As we can see in [18], this construction produces a white gaussian noise on the observations such that heuristically

$$\|z - C\check{u}\|_{\mathbf{H}^{-\frac{1}{2}}([0,\tau])} \leq \varepsilon.$$

Consequently, we can take the  $\varepsilon$  as the noise level in  $\mathbf{H}^{-\frac{1}{2}}([0,\tau])$ .

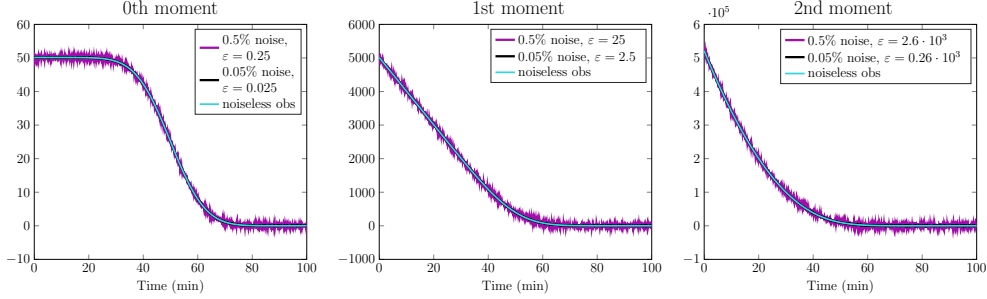


Figure 1: Three moments of the state function  $u$  having dynamics (23) where  $u_{0g} = e^{\frac{1}{2} - \frac{(x-100)^2}{20^2}}$  and  $b = 2$ .

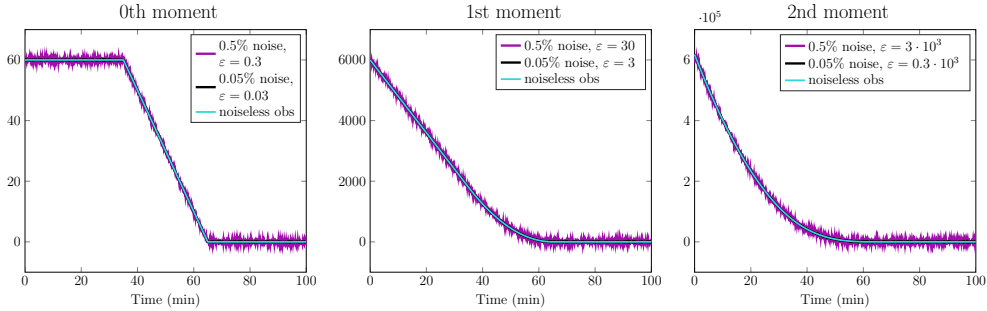


Figure 2: Three moments of the state function  $u$  having dynamics (23) where  $u_{0ch} = I_{[70,130]}$  and  $b = 2$ .

In Figures 1 - 2, we present the synthetic observations associated to the gaussian function  $u_{0g}$  and the characteristic function  $u_{0ch}$ , respectively. In both cases we have computed the first three moments of the state function  $u$ . Moreover, we consider noised observations. The noise corresponds to errors of 0.05% and 0.5%. The corresponding values of the noise level  $\varepsilon$  have been reported in figure legends.

### 3.3 Numerical Simulations: kernel regularisation method

In this section we present some examples of numerical initial condition estimation by the kernel regularisation method. We recall that the estimation is given by the function  $\hat{u}_0^{\varepsilon,\alpha}$

defined in Equation (12). The parametrised kernel family  $\rho_\alpha$  is defined by  $\rho_\alpha(x) = \frac{1}{\alpha}\rho(\frac{x}{\alpha})$ . The kernel  $\rho$  is chosen as the gaussian kernel,  $\rho = \frac{1}{0.3\sqrt{2\pi}}e^{-\frac{x^2}{2(0.3)^2}}$ . According to these choices, the coefficient  $m$  – defined in Equation (9) – is equal to 1. We compute the  $(n + 1)$ -th derivative of the convolution between  $z$  and the regularisation kernel  $\rho_\alpha$  as the convolution between  $z$  and the  $(n + 1)$ -th derivative of  $\rho_\alpha$ . The derivative can be either analytically computed or approximated by finite differences. In the examples of this section we have considered the analytic expression of kernel derivatives.

To compute the discrete convolution we need two vectors. One vector is the set of measurements  $z$ . The other vector is obtained by evaluating the derivative of the kernel function,  $\frac{d^{n+1}\rho_\alpha}{dx^{n+1}}$ , on a discrete grid. First of all we approximate the kernel  $\rho$  by  $\tilde{\rho} = \rho I_{[-2,2]}$ , where  $I_{[-2,2]}$  is the characteristic function for the domain  $[-2, 2]$ . We remark that when we numerically compute the integral of  $\tilde{\rho}$  we obtain 1, which is the same value as the integral of  $\rho$  over  $\mathbb{R}$ . Hence, approximating  $\rho$  by  $\tilde{\rho}$  we make an error smaller than machine-precision. Consequently, we consider the support of  $\rho_\alpha$  and its derivatives to be included in  $[-2\alpha, 2\alpha]$ . We evaluate  $\frac{d^{n+1}\rho_\alpha}{dx^{n+1}}$  over the grid  $-2\alpha = x_1 < \dots < x_r = 2\alpha$  with the same discretisation step,  $\delta t$ , as the measurement time grid.

The discrete convolution of two vectors  $x, y$  of lengths  $m$  and  $h$ , respectively, is the vector  $w$  such that  $w_k = \sum_j x_j y_{k-j+1}$ , for  $k = 1, \dots, m + h + 1$ . For every component  $k$  the index sum  $j$  varies between  $\max(1, k + 1 - h)$  and  $\min(k, m)$ . This algorithm is equivalent to extend with zeros the vectors for indices  $j$  outside the range  $[\max(1, k + 1 - h), \min(k, m)]$ .

We remark that zero-padding the kernel derivative vector is equivalent to evaluating the function  $\frac{d^{n+1}\rho_\alpha}{dx^{n+1}}$  outside the domain  $[-2\alpha, 2\alpha]$ . Moreover, extending with zeros the measurements for times bigger than the observation time is coherent with the biological interpretation: a depolymerising system in which all polymers have been reduced into monomers cannot change its state.

On the other side, adding zeros for negative times would lead to a bad reconstruction of the initial condition on the left border. Our idea is to extend the observation data for negative times in  $[-\tau_\alpha, 0]$ . The positive value  $\tau_\alpha$  is such that the component  $w_k$  – corresponding to the left border value of the estimation  $\hat{u}_0^{\varepsilon, \alpha}$  – is computed as a complete sum. To this purpose, we fix  $\tau_\alpha$  bigger than the length of the kernel domain, specifically  $4\alpha$ . To consider negative times, we extend the definition of the initial condition for negative sizes by  $u_0(x) = 0$  if  $x < 0$ . The  $n$ -th moment for negative times reads

$$C^n u(t) = \int_{bt}^{\ell} (x - bt)^n u_0(x) dx = \int_0^{\ell} (x - bt)^n u_0(x) dx.$$

We notice that the  $n$ -th moment is a polynomial of degree  $n$  in  $t$ . Assuming there is a size  $x_{min} > 0$  such that the support of  $u_0$  is included in  $[x_{min}, \ell]$ , we obtain that, for every  $t \leq \frac{x_{min}}{b}$ ,  $C^n u(t) = \int_{x_{min}}^{\ell} (x - bt)^n u_0(x) dx$ . To conclude, we assume  $x_{min} = 10$  in our numerical examples. We fit the observations relative to times in the range  $[0, 5]$  with an  $n$  degree polynomial. We present in Figure 3 an example of extension of observation data



for the first three moments in the case of gaussian initial condition  $u_{0g}$ . We discretise the arbitrarily chosen negative domain  $[-10, 0]$  with a time step  $\delta t$ . We evaluate the polynomial fit on this grid and we use these data to extend our observations.

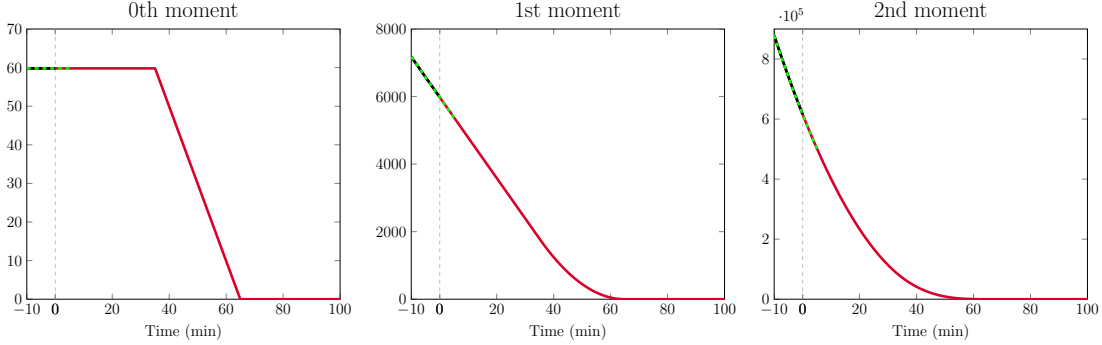


Figure 3: From the left to the right the 0th, 1st and 2nd moment relative to the initial condition  $u_{0g}$  and depolymerisation rate  $b = 2$ . In red the moments for positive times, in black the moments for negative times in  $[-10, 0]$ , in green the polynomial function fitting the red curve in  $[0, 5]$  and used to extend the data on  $[-10, 0]$ .

We compute the discrete convolution between the extended observation and the kernel derivative vector. We multiply the resulting vector by  $\delta t$  – to approximate the continuous integral of the convolution – and by  $\frac{1}{(-b)^{n+1}n!}$  in accordance with Equation (12).

We present in Figure 4 and Figure 6 the estimation of the initial conditions  $u_{0g}$  and  $u_{0ch}$ , respectively. We remark that the quality of the estimation decreases when the order of the moment and the noise level increase.

### 3.4 Numerical Simulations: the data assimilation method

We now turn to the variational approach detailed in Section 2. In order to discretise and simulate the two-end problem (21), we rely on a discretised version of the optimal criterion (18) to be minimised under the constraint of the discretised model (23). The resulting time-discretised optimal system can then be proved to converge to time-continuous solution of (21) [7]. Therefore, we decompose the initial condition by defining  $\check{\xi} \in \mathbb{R}^{N_x}$  such that  $\check{u}_0 = u_o + \check{\xi}$  and seek an estimate of  $\check{\xi}$  given by

$$\bar{\xi} = \arg \min_{\xi} J_{N_t}(\xi) = \arg \min_{\xi} \left( \frac{1}{2} \|\xi\|_{P_0^{-1}}^2 + \frac{1}{2} \sum_{k=0}^{N_t} \|z_k - C_k u^k\|_{M_k}^2 \right). \quad (26)$$

The matrix  $M_k$  is the discrete approximation of the operator  $\gamma Id_{\mathcal{Z}}$  and it depends on the quadrature rule chosen to approximate the integral in time. We fix  $M_k = \delta t \gamma I_{N_t}$ .

Furthermore – if we assume that the initial *a priori* approximates the unknown initial condition with the same error on every cluster size – we can take  $P_0 = \frac{1}{\delta x \beta} I_{N_x}$ .

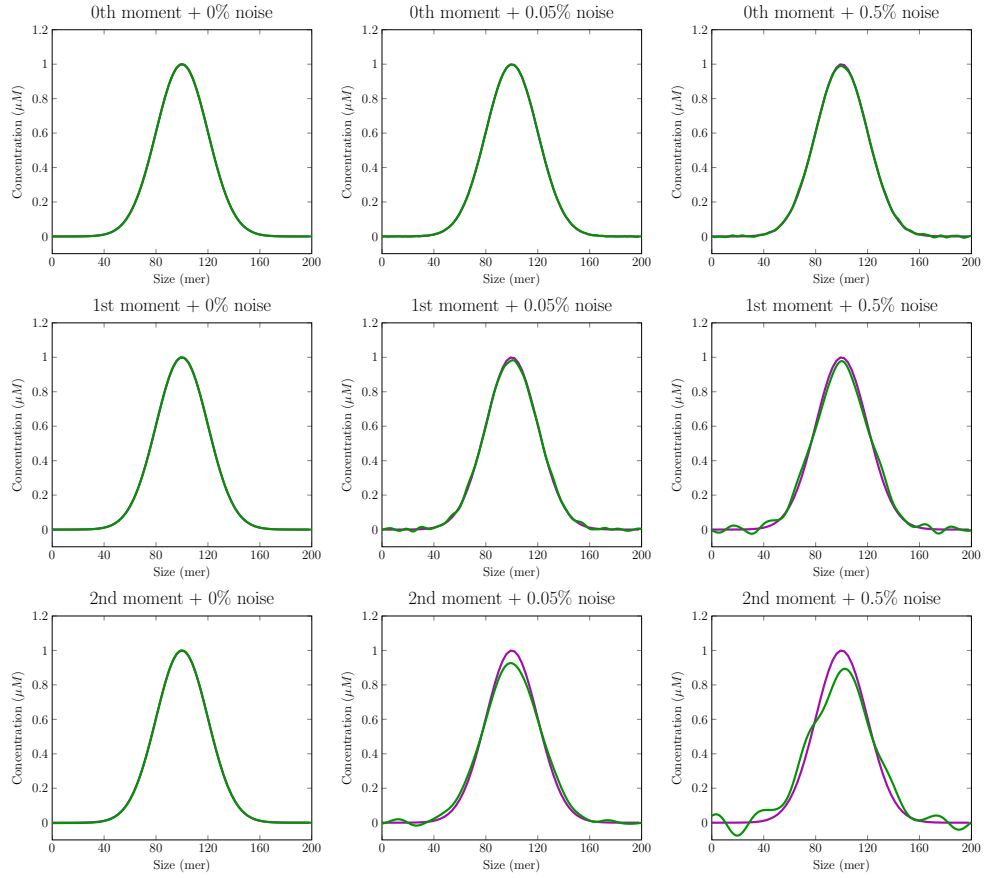


Figure 4: Comparison between the exact gaussian initial condition  $u_{0g}$  (purple line) and the approximations  $\hat{u}_0^{\varepsilon;\alpha}$  (green line) provided by the kernel regularisation method. Each estimation is associated with the measurements in Figure 5.

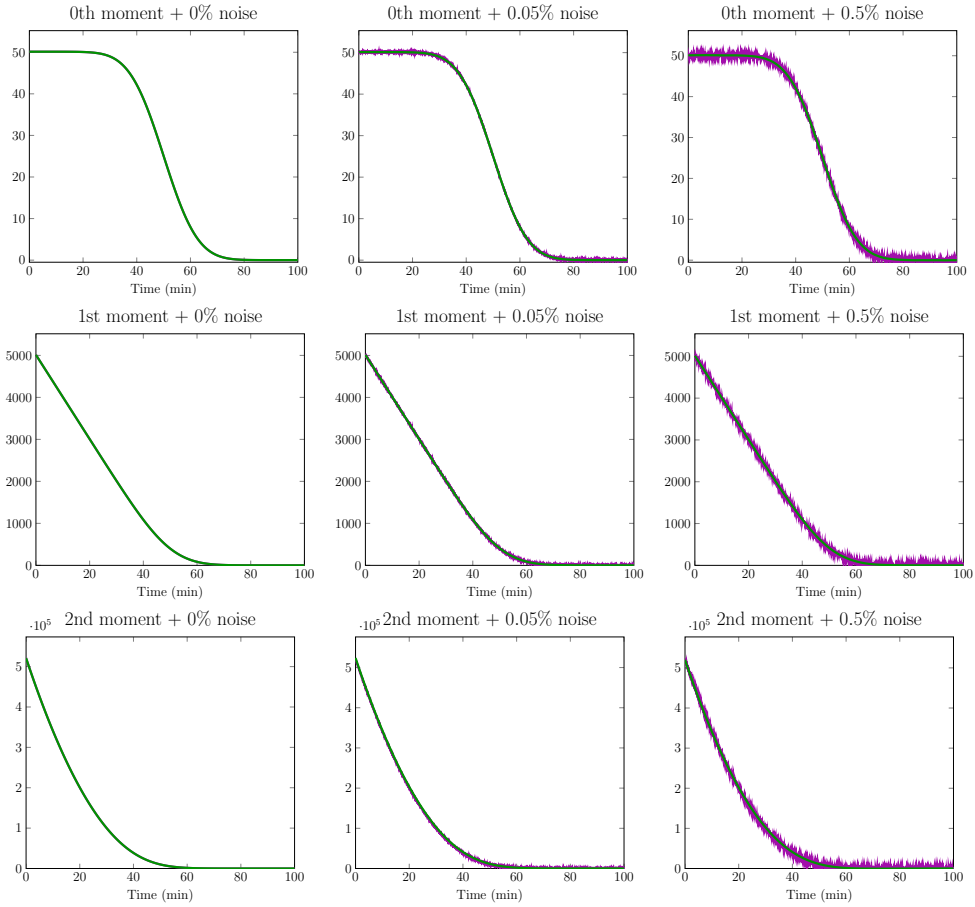


Figure 5: Synthetic observation in the case of gaussian initial condition  $u_{0g}$  (purple line) and relative fit given by the observations generated from the data assimilation estimator. From the top to the bottom by rows, we see the 0th-moment, 1st-moment, 2nd-moment. From the left to the right by columns the noise corresponds to a 0%, 0.05%, 0.5% error.

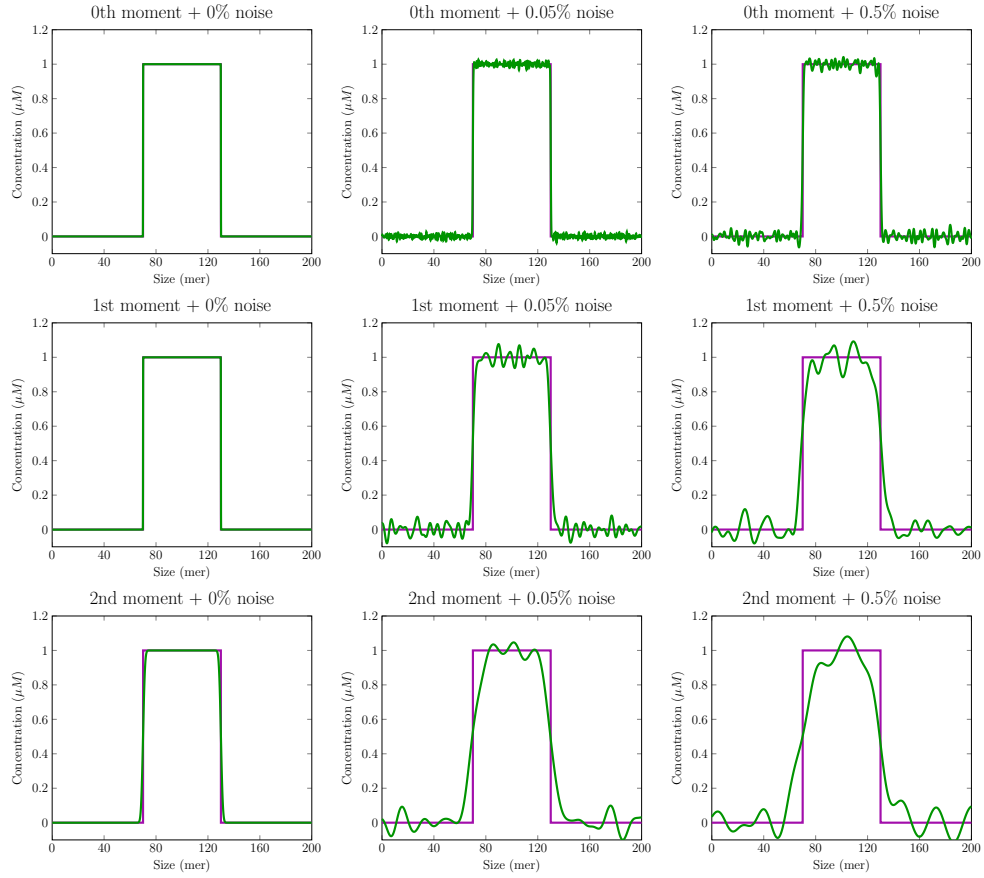


Figure 6: Comparison between the exact initial condition  $u_{0\text{ch}}$  (purple line) and the approximations  $\hat{u}_0^{\varepsilon, \alpha}$  (green line) provided by the data assimilation method. Each estimation is associated with the measurements in Figure 7.

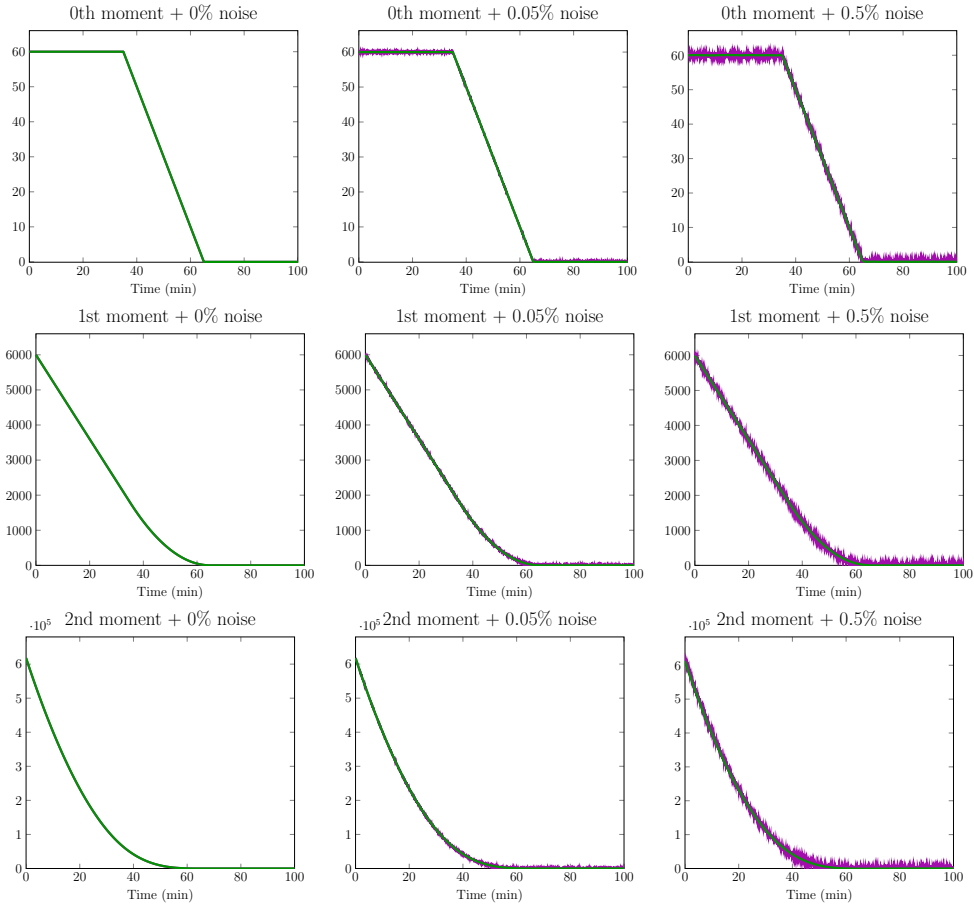


Figure 7: Synthetic data in the case of  $u_{0\text{ch}}$  as initial condition (purple line) and relative fit given by the observations generated from the data assimilation estimator. From the top to the bottom by rows, we see the 0th-moment, 1st-moment, 2nd-moment. From the left to the right by columns the noise corresponds to a 0%, 0.05%, 0.5% error.

These two matrices are classically defined as  $P_0 = \text{Cov}(\check{u}_0 - u_o)$  and  $M_k = \text{Cov}(\chi_k)^{-1}$ . As well explained in [25] – if  $W(t)$  is the time-independent covariance of the continuous white gaussian process  $\chi(t)$  – the covariance of measurement noise in discrete time is  $W_k = \frac{W(t_k)}{\delta t}$ . Since  $M_k = W_k^{-1}$ , we find the relation  $M_k = \delta t M(t) = \delta t \gamma I d_{\mathcal{Z}}$ . Consequently, we define  $\gamma$  as  $\gamma = (\sigma^2)^{-1}$ , where  $\sigma^2$  is the variance of the white gaussian process  $\chi(t)$ . The parameter  $\beta$  is analogously defined as  $\beta^{-1} = \|\check{\xi}\|_{\mathbf{L}^2([0,\ell])}^2$ . In practice, we can numerically estimate the parameter  $\gamma$  – by analysing the noise on the data – while the value of the parameter  $\beta$  reflects the confidence that we have on the *a priori* information on the initial condition.

To minimise the criterion we use a gradient-descent based optimisation method that – starting from the initial guess  $\xi = 0$  – iteratively attempts to estimate the minimum from the criterion gradient evaluated on the current guess

$$\nabla J_{N_t}(\xi) = P_0^{-1}\xi - (q_{|\xi}^0),$$

where  $q_{|\xi}^0$  is the time-discrete adjoint variable at time 0 solution of the time-discrete system

$$\begin{cases} q_{|\xi}^k = A_{k+1|k} q_{|\xi}^{k+1} - C_k^T M_k (z_k - C_k u_{|\xi}^k), & 0 \leq k \leq N_t \\ q_{|\xi}^{N_t+1} = 0. \end{cases} \quad (27)$$

Note that the time-discrete adjoint variable is the Lagrange multiplier associated with the dynamical constraint (23) in the minimisation of (26). Besides, it is also a time-discretisation of the time-continuous adjoint variable (19).

As numerical synthetic test cases, we present in Figure 8 the estimation of a gaussian initial condition by the variational data assimilation method. Respectively in Figure 10, we estimate a characteristic function. As previously done with the kernel method, the nine estimation curves correspond to the nine observation curves presented in Figure 1 or Figure 2.

## 4 Application on experimental data

Having presented, theoretically investigated and numerically tested our mathematical approach, we are now ready to apply our method to experimental data.

### 4.1 Presentation of the experimental protocol and noise analysis

The data to analyse consist in observations on ovine prion protein oligomers (PrP oligomers), in depolymerising conditions. PrP oligomers are a kind of amyloid deposit generated by the concatenation of monomers forming chains of a few tens of proteins. These structures are relatively small compared to protein polymers, that could be composed by up to thousands of proteins [24].

We refer to Appendix 7 for details on the protocol used to form and make measurements on oligomer systems. We present in Figure 12 an example of Static Light Scattering

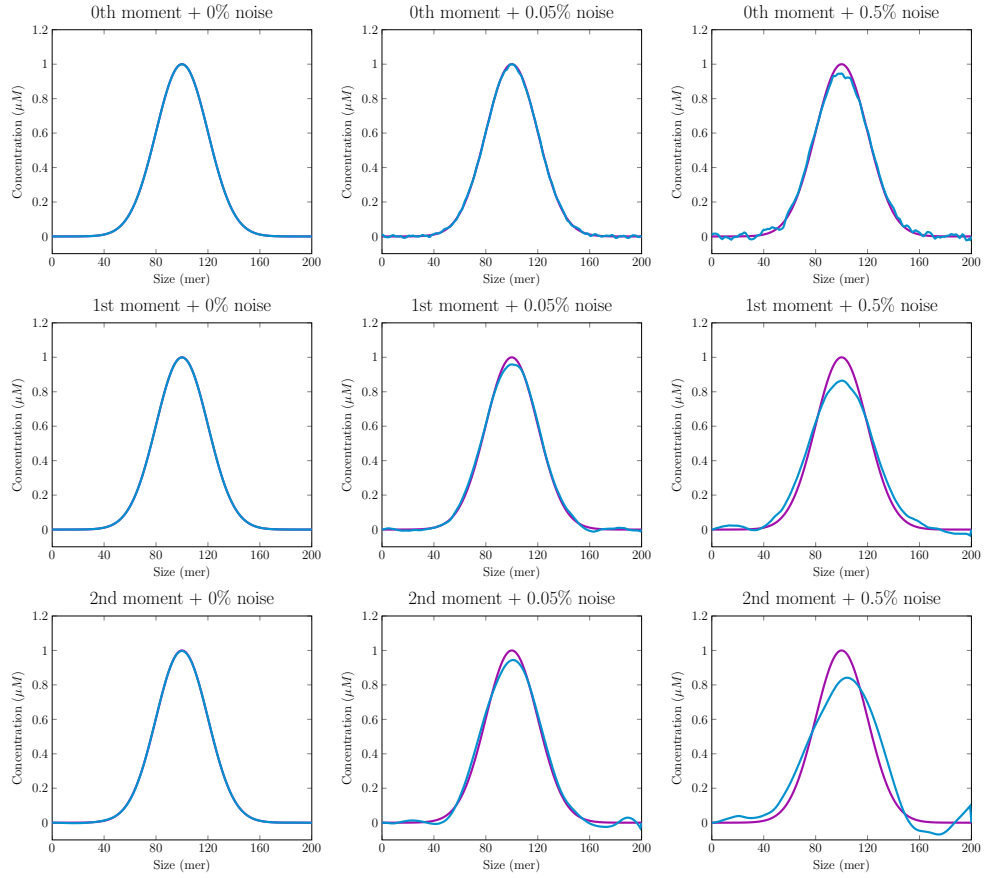


Figure 8: Comparison between the exact gaussian initial condition  $u_{0g}$  (purple line) and the approximations  $\hat{u}_0^\alpha$  (blue line) provided by the data assimilation method. Each estimation is associated with the measurements in Figure 9.

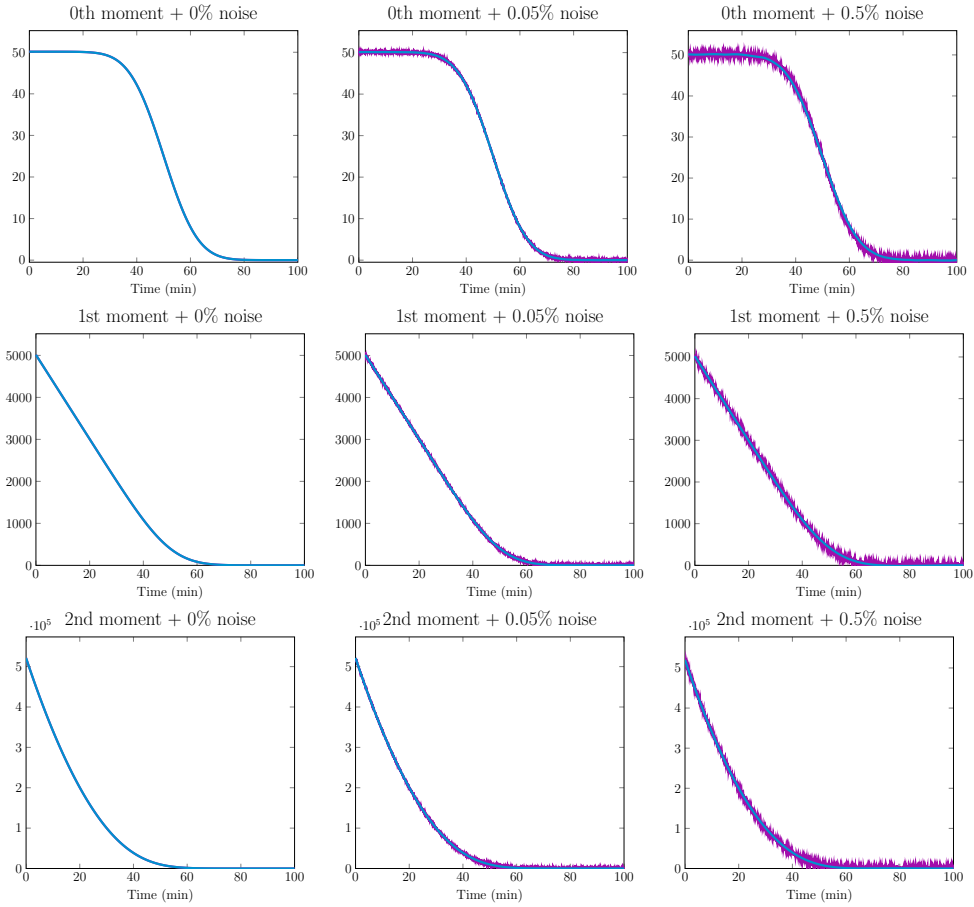


Figure 9: Synthetic observation in the case of gaussian initial condition  $u_{0g}$  (purple line) and relative fit given by the observations generated from the data assimilation estimator. From the top to the bottom by rows, we see the 0th-moment, 1st-moment, 2nd-moment. From the left to the right by columns the noise corresponds to a 0%, 0.05%, 0.5% error.



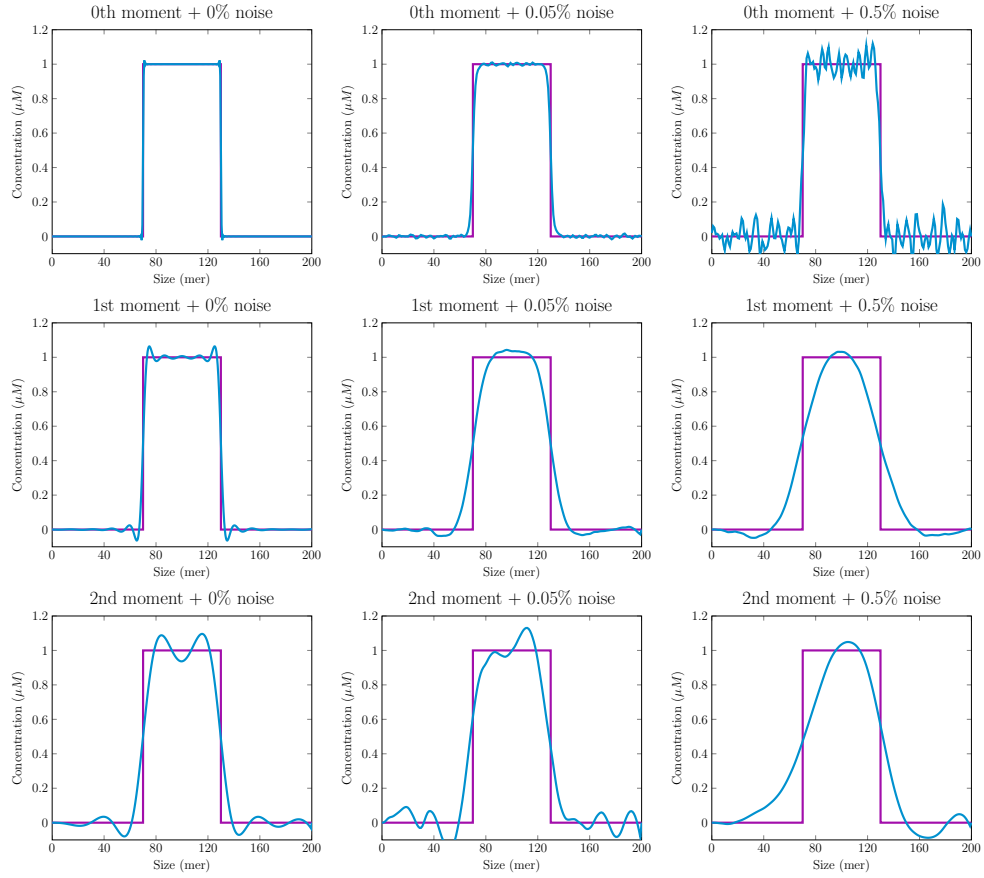


Figure 10: Comparison between the exact initial condition  $u_{0ch}$  (purple line) and the approximations  $\hat{u}_0^\alpha$  (blue line) provided by the data assimilation method. Each estimation is associated with the measurements in Figure 11.

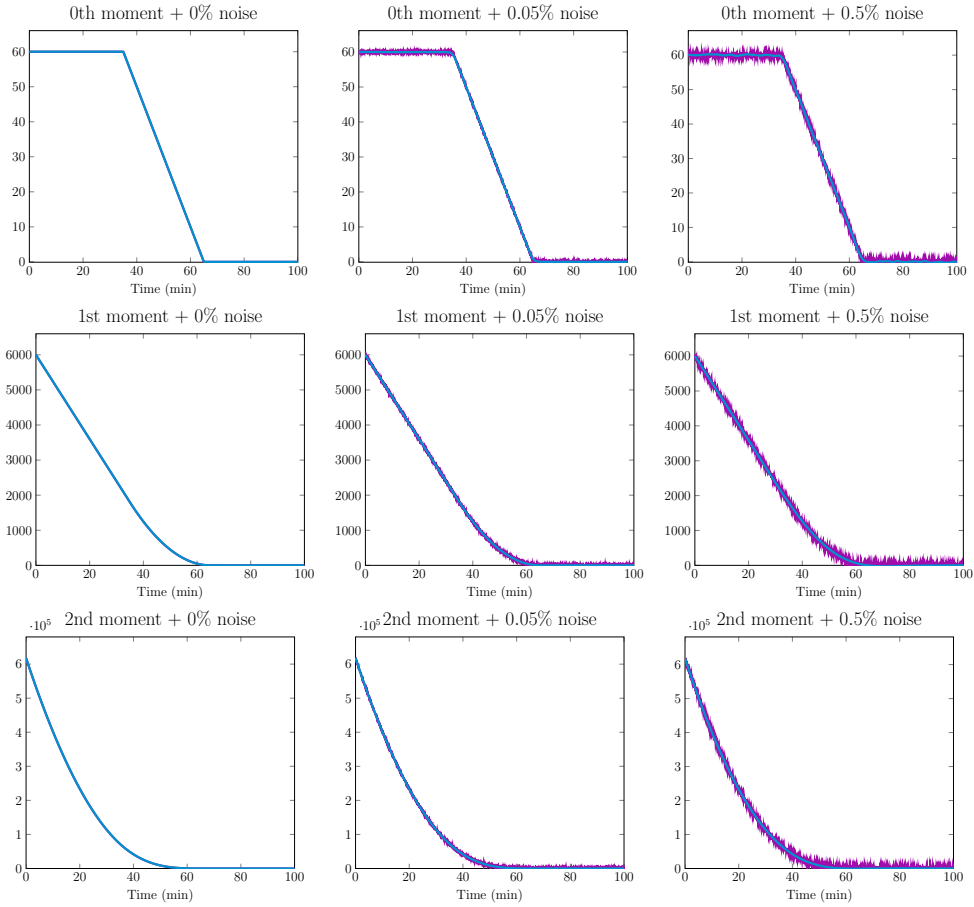


Figure 11: Synthetic data in the case of  $u_{0\text{ch}}$  as initial condition (purple line) and relative fit given by the observations generated from the data assimilation estimator. From the top to the bottom by rows, we see the 0th-moment, 1st-moment, 2nd-moment. From the left to the right by columns the noise corresponds to a 0%, 0.05%, 0.5% error.

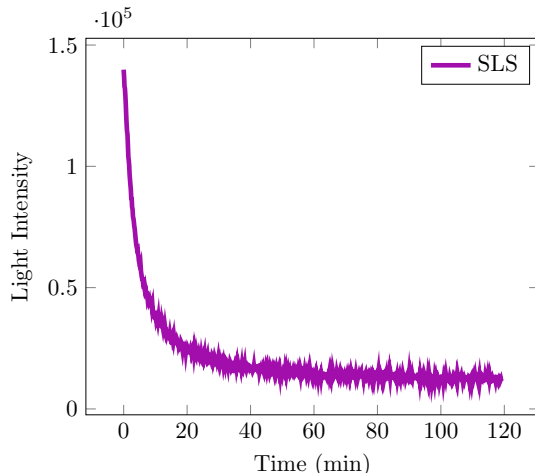


Figure 12: Depolymerisation kinetics of ovPrP oligomers monitored by static light scattering.

measurements on a depolymerising system of PrP oligomers. We recall that the SLS measurement is a linear transformation of the second moment  $z_{SLS}(t) = c_1 C u(t) + c_2$ , with unknown parameters  $c_1, c_2$ . We assume to observe the experiment until all the oligomers are depolymerised into monomers. We fix the parameter  $c_2$  such that the mean of the measurements at end of the observation domain is zero. We thus consider the shifted data  $z_{SLS} - c_2$  as measurements. Solving the inverse problem with this observation data, we estimate the function  $c_1 u_0$ . In the following, we assume that  $c_1 = 1$ .

In order to analyse the measurement noise, we assume it to be a white gaussian additive noise and test this hypothesis. Since we assume that the initial size distribution is a regular function, we expect the corresponding second moment to be a regular function with a smooth graph. For this reason, we start by filtering the data. We use a cubic Savitzky-Golay filter. For more details about this filter see [19].

The difference between the empirical data and the fit gives us an estimation of the noise contribution, see Figure 13a. We run a  $\chi^2$  numerical test to test the null hypothesis of residual points following a gaussian distribution. The test accepts the null hypothesis at the 5% significance level. We then estimate the mean and the standard deviation of the gaussian distribution generating the residual. We estimate the mean at 0 and the standard deviation  $\sigma = 501$ . The purple dotted line in Figure 13b shows the estimated gaussian density function: this leads us to accept our noise model and keep this value of  $\sigma$  as a reasonable estimation of the noise level.

The experimental protocol also includes the separation of oligomers by size, using the Size Exclusion Chromatography (SEC) device. Thanks to this technique we can measure the initial oligomer distribution. We point out that it would not have been possible to make these measurements on fibrils, due to the large size of the aggregates and the limits of the device. We can see in Figure 14 the measurement of the initial oligomer distribution

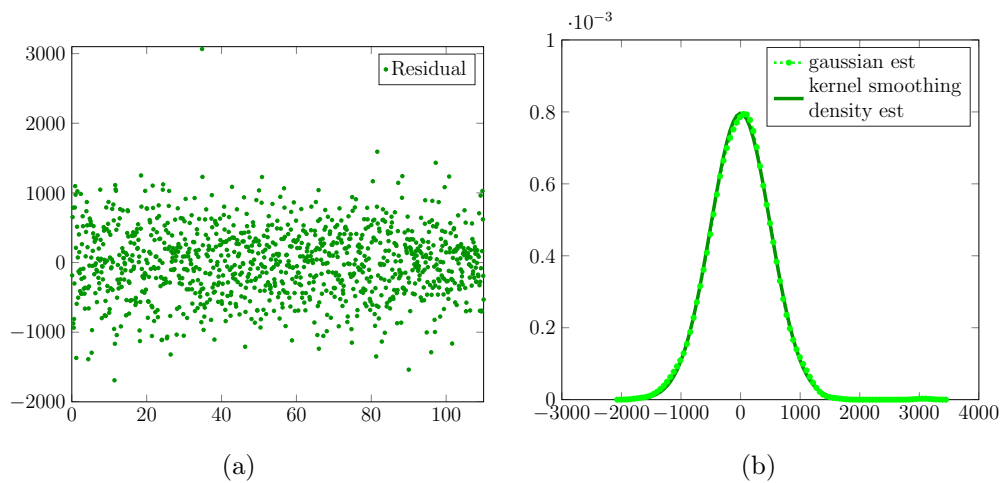


Figure 13: Analysis of the noise on SLS data. We present (left) the residuals obtained as the difference between the SLS data and the cubic Savitzky-Golay filter of the data. In the right figure we present two estimations of the density function associated to the residual data.

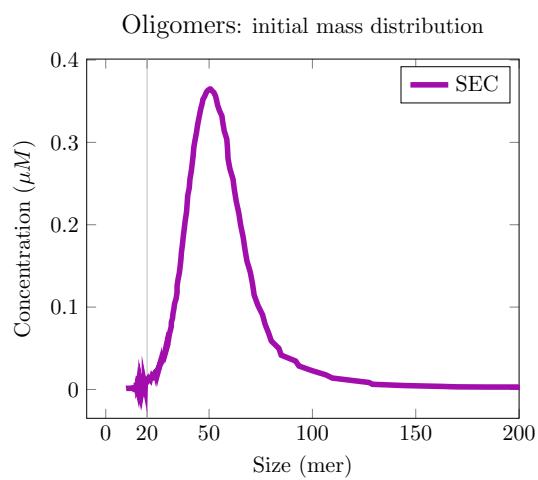


Figure 14: Size-exclusion chromatogram of purified ovPrP oligomers.

associated to the data in Figure 12. We notice that the biggest oligomer size taken into account is  $200mer$ , with an almost zero relative concentration, while the smallest detectable size is  $20mer$ . In this set of data, the most present oligomers have sizes between  $30mer$  and  $100mer$ . We also remark that this distribution has only one peak centred around the size  $50mer$ . Evaluating the noise level on the SEC device is a complex subject, going beyond the scope of this study.

In the following we set up the inverse problem of estimating the initial size distribution by using the SLS measurements only. We then discuss the results obtained when we also take into account the SEC measurement.

## 4.2 Initial state estimation without *a priori*

Oligomer dynamics can be modelled by System (1), [9]. We remark that in this model there are two unknowns: the depolymerisation rate  $b$  and the initial condition  $u_0$ . The approaches that we have presented in this paper are designed to estimate only the initial condition. We thus start our analysis with the simple model of constant backward transport

$$\begin{cases} \frac{\partial}{\partial t}u(x, t) - b\frac{\partial}{\partial x}u(x, t) = 0, \\ u(L, t) = 0, \\ u(x, 0) = u_0(x). \end{cases} \quad (28)$$

Our strategy is to fix an arbitrary value for  $b$  and then perform the initial condition estimation. The resulting estimation depends explicitly on  $b$ , as we have seen in Equation (7). For example, consider two models associated to the rates  $b_1 \neq b_2$ . Equation (7), in the case of a noiseless second moment observation, reads

$$u_{0|b_i}(x) = \frac{1}{2(-b_i)^3} \frac{d^3}{dt^3} z \left( \frac{x}{b_i} \right),$$

for  $i = 1, 2$ . We use the notation  $u_{0|b_i}$  to indicate the solution of the inverse problem when we consider the transport velocity  $b_i$  in the model (28). Eventually, we can notice that

$$u_{0|b_1}(x) = \left( \frac{b_2}{b_1} \right)^3 u_{0|b_2} \left( \frac{b_2}{b_1} x \right). \quad (29)$$

This relation leads us to the conclusion that, when we fix a depolymerisation rate, we obtain a function that differs from the exact one in a linear change of variables and a scaling factor. To illustrate this relation, we show in Figure 15a an example of distributions which produce the same second moment observation, see Figure 15b, evolving with different rates.

We thus fix the depolymerisation rate to the arbitrary value  $b = 2\min^{-1}$ . We consider the experimental time domain  $[0, \tau] = [0, 110]\text{min}$ . Biological considerations lead us to the definition of the size domain  $[0, \ell] = [0, 200]mer$ .

To apply the data assimilation method we need to define the least square criterion. We choose the isomorphism  $P_0$  of the form  $P_0 = \frac{1}{\beta}\text{Id}$ . Consequently, we only need to fix

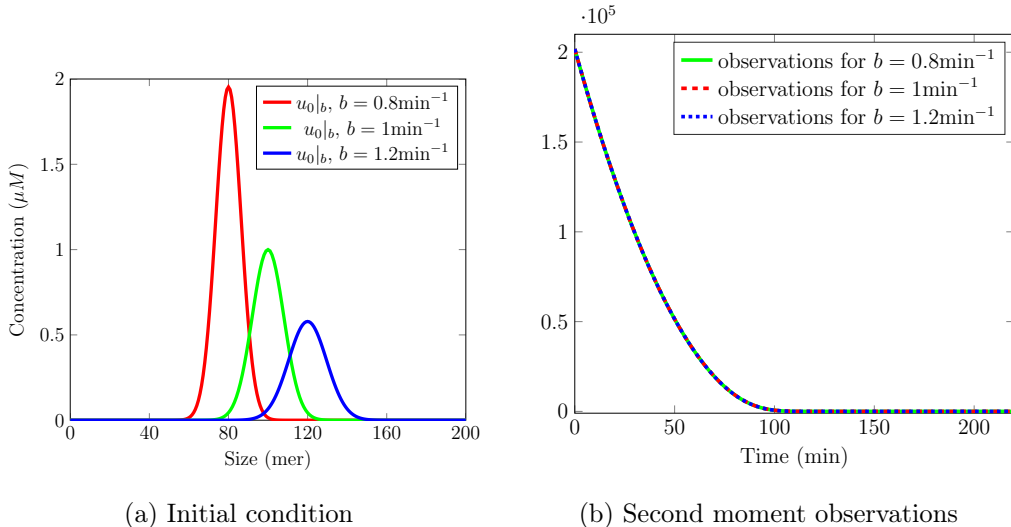


Figure 15: Several choices of initial condition and depolymerisation rate can lead to the same second moment observation.

a value for the regularisation parameters  $\beta$  and  $\gamma$ . As explained before, these parameters are linked by an inversely proportional relation to the confidence on the *a priori* on the initial condition and the noise level on measurements, respectively.

Since for the moment we do not consider additional information on the initial condition, our *a priori* is the zero constant function. In particular, we do not know whether this *a priori* is far or not from the target initial condition. Therefore, we assume to have low confidence on the *a priori* or equivalently we allow the estimations to be far from the *a priori*. This assumption corresponds to the choice of a small value for  $\beta$ . In the following we fix  $\beta = 10^{-2}$ .

We consider  $\gamma \simeq \frac{1}{\sigma^2}$ , where  $\sigma$  is the standard deviation of noise distribution. Consequently, we take  $\gamma = 10^{-6}$ .

We show in Figure 16 the results of data assimilation estimation. We see in Figure 16b that we obtain a good fit of the experimental data. In Figure 16a we have the initial state estimation. We recall that this estimation is associated to the arbitrary choice of  $b$  and the real initial distribution can be a transformation of this function, according to Formula (29). Nevertheless, we can infer interesting features such as the presence of one main peak and the fact that the peak starts from small sizes.

### 4.3 Estimation with *a priori*

In this section we take into account the SEC measurement of Figure 14 to discuss the result of our initial state estimation of Figure 16a.

A first possibility is to admit that the chromatography technique cannot trustfully measure the variation of the distribution but it can nevertheless find the position of the

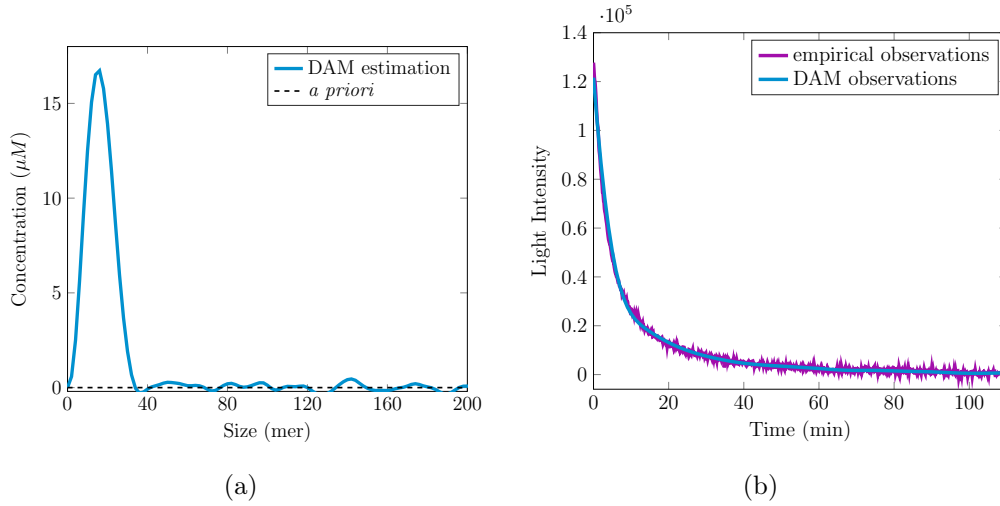


Figure 16: Left: Initial condition estimation by variational approach when we choose  $b = 2 \text{ min}^{-1}$ ,  $\beta = 10^{-2}$ ,  $\gamma = 10^{-6}$ . Right: comparison between the SLS measurements and the observations generated by the observation operator on the state estimation.

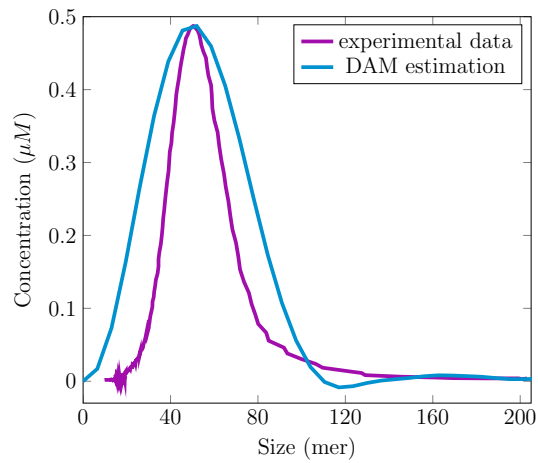


Figure 17: Comparison between the initial condition estimation – associated to the rate  $b = 6.5 \text{ min}^{-1}$  – and the experimental SEC data rescaled to have the same maximum as the estimation.

peak. Hence, we transform the estimation according to Formula (29) to move the peak to the same value as in the chromatography measurements. Since experimental data have been normalised to have integral equal to one, we can define a coefficient to make the two curves have the same maximum. In this way, we could use the SEC to identify the depolymerisation rate – that in this case corresponds to  $b = 6.5\text{min}^{-1}$ . Consequently, we consider the blue curve in Figure 17 as the estimation of the initial condition. However, according to the SEC specification and methodology the difference between the two distributions - the experimental one and the estimated one, see Figure 17 - seems too important to correspond to a noise on the measurement.

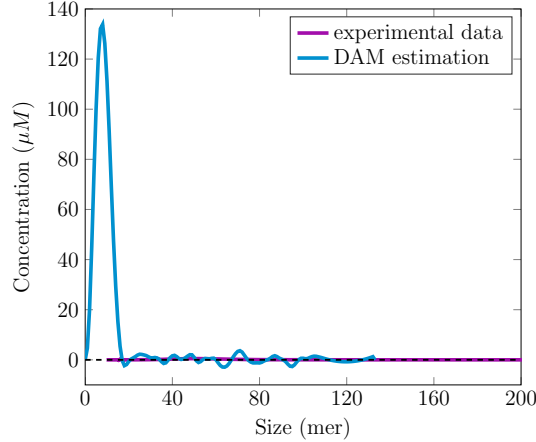


Figure 18: Initial condition estimation associated to the rate  $b = 1\text{min}^{-1}$ . The peak corresponds to sizes less than  $20\text{mer}$ .

A second possible interpretation is to think that the peak of the estimator does not correspond to the peak we can see in SEC measurements and it may concern sizes up to  $20\text{mer}$ , instead. We recall that the SEC device cannot detect aggregates composed by less than 20 monomers. For instance, if we take  $b = 1\text{min}^{-1}$  we would have an initial condition that can illustrate this case.

We present in Figure 18 such an initial condition. We can notice that the maximum of oligomer concentrations for sizes bigger than  $20\text{mer}$  is much smaller than the value of the peak. This would imply that the concentration of oligomers measured by SEC is negligible compared to a high concentration of (hidden) small oligomers: this is barely plausible.

On the contrary, let us assume that we trust completely the SEC data and that those data represent the overall distribution, i.e., we extend the data by zero in the region  $[0, 20]\text{mer}$ . Having fixed the initial condition, our problem can thus be seen as a parameter identification problem. This problem can be presented as

Estimating the depolymerisation rate  $b$ , appearing in the definition of the model dynamics  $A$ , from given measurements  $z$  generated through time  $t \in [0, \tau]$ , knowing the initial condition  $\check{u}_0$  and the model of observation operator  $C$ .



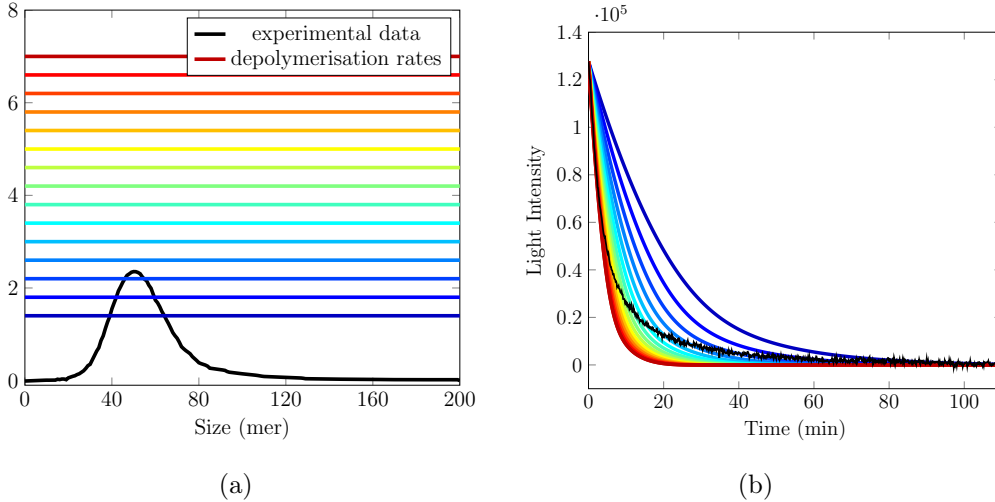


Figure 19: Influence of the depolymerisation rate on the measurement. Left: the initial condition (black line), given by the SEC data, and several constant functions  $b$ . Right: experimental SLS data (black line) and the second moment observations associated to the  $b$  of the same colour on the left figure.

To have a rough approximation of  $b$ , we run the direct model for several choices of  $b$  and then, for each of these choices, we compare the second moment generated with the operator  $C$  to the SLS data  $z$ . We can see in Figure 19 the result of this analysis when we choose  $b$  as a constant function with values varying between  $1\text{min}^{-1}$  and  $7\text{min}^{-1}$ . We came to the conclusion that, if we assume the initial condition in Figure 19a, the solution of the parameter identification problem is not a constant function.

This conclusion follows from the fact that – whenever we consider two rates  $b_1 < b_2$ , – we have  $(y - b_1t)^n - (y - b_2t)^n > 0$  and consequently

$$z_2 = Cu_2 = \int_{b_2t}^{\ell} (y - b_2t)^n u_0(y) dy \leq \int_{b_1t}^{\ell} (y - b_1t)^n u_0(y) dy = Cu_1 = z_1. \quad (30)$$

Specifically, all the curves start from the same value and then they do not cross anymore. Anyway we can see in Figure 19b that experimental data intersect all the synthetic observations. We deduce that it is not possible to define a constant parameter  $b \in [1, 7]\text{min}^{-1}$ , solution of the parameter identification problem. A sensitivity analysis could also be carried out to gain more insights, see [4, 1]. Furthermore, Inequality (30) implies that any value  $b \in (0, 1) \cup (7, \infty)\text{min}^{-1}$  would lead to an observation far from the experimental data.

We have so concluded that, if we take the curve in Figure 19a as the initial oligomer distribution, we need to consider a size-dependent depolymerisation rate. We discuss a simple case in which we distinguish two depolymerisation rates, one for small aggregates and one for big aggregates. In this case  $b$  is a piecewise function that takes only two values. The discontinuity point for  $b$  has been arbitrarily chosen around  $25\text{mer}$ .

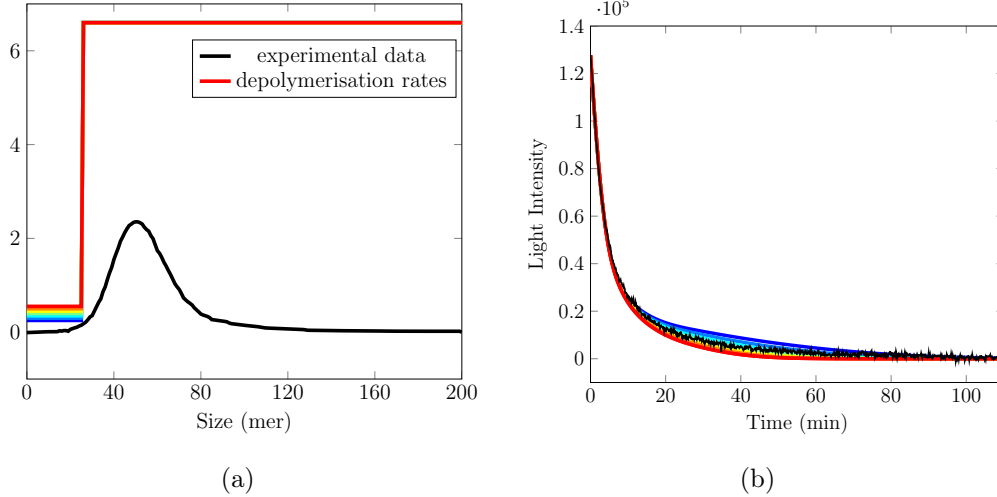


Figure 20: On the left we have the initial condition (black line), given by the SEC data, and several piecewise constant functions  $b(x)$ . On the right we present the experimental SLS data (black line) and the second moment observations associated to the  $b$  of the same colour.

We show in Figure 20 the results of the analysis performed by running the direct model with this hypothesis. After testing several possibilities, we conclude that to fit the data we need a low depolymerisation rate for small aggregates and a high depolymerisation rate for big aggregates. In particular, we can see in Figure 20 that if we consider  $b(x) = b_1 I_{x < 25} + 6.6 I_{x > 25}$  and  $b_1$  varying between  $0.25 \text{min}^{-1}$  and  $0.55 \text{min}^{-1}$  we can well approximate the experimental measurements.

We are aware that this way of proceeding does not solve the parameter identification problem. Nevertheless, it is useful to get some clues on what could be the right solution. To rigorously solve this problem, we could define an initial condition estimation problem associated to the parameter identification problem and then apply the variational approach on this new problem. To define this associated problem, we consider an augmented state  $u^a$  composed by the state function  $u$  and the parameters  $u^a = (u; b)$ . The dynamics of this new variable is given by

$$\dot{u}^a = \begin{pmatrix} \dot{u} \\ \dot{b} \end{pmatrix} = \begin{pmatrix} A(u, b) \\ 0 \end{pmatrix}$$

and the initial condition of the system is  $u^a(0) = (u(0); b(0))$ . We can use the SEC data to fix an *a priori* on the initial condition so that only the parameters remain unknown. Keeping the same notation as before, we call our target  $\xi^a = (0; b)$ . We remark that this new model presents the additional difficulty of the nonlinearity. A common strategy to overcome this difficulty is to approximate this model replacing the model operator by its tangent.

## 4.4 Discussion

Departing from our general methodology, we were able to estimate the *shape* of the initial size distribution of ovPrP oligomers under the assumption of a constant depolymerisation rate. We then compared it to the experimental distribution obtained with the SEC device (Figure 17), and the discrepancy between these two distributions led us to revisit our previous assumptions.

To explain this discrepancy, two hypotheses could be evocated. The first one is a very important underestimation of the amount of small oligomers assemblies by SEC techniques (Figure 18). The second possibility is a higher depolymerisation rate for large PrP assemblies compared to smaller one (Figure 20). According to SEC specification and methodology the first hypothesis can be excluded. Therefore, lower stability of large assemblies appears as the best possible explanation.

The fact that large assemblies present lower stability than small oligomers could have an important biological implication. Indeed it is commonly admitted that small oligomers constitute the most cytotoxic species during prion misfolding process [26]. Therefore the low stability of large assemblies could make an accumulation of lower molecular weight assemblies and contribute to increase the toxicity level.

## 5 Conclusion

In this paper, we have defined the inverse problem of estimating the initial condition of a dynamical system – whose evolution is described by a transport equation – given the measurements of the second moment over time. We have described two possible approaches to solve this problem. The first belongs to the family of kernel regularisation methods, and allowed us to define a good strategy that exploits all the features of the model and deals precisely with the regularity of the functions. We have introduced this strategy in a case of constant transport velocity both to provide the method guidelines and to give an insight into the properties of the model and the relations between the state function, its moments and the transport velocity. However, since it has been designed for this very specific case, this method lacks of flexibility.

The second approach belongs to the family of data assimilation methods. The inverse problem is written in terms of operators and we obtain a general formalism that can be applied directly on a variety of models. We have seen how this approach is equivalent to the first approach in the case of constant transport velocity and we refer to this second approach to address the more general case of variable velocity and a priori information on the data. We have also briefly explained in the previous section how we can use the same strategy to solve a problem of parameter identification, see also [21]. With no more effort we can treat the case of multiple measurements. It would be enough to define an observation operator that – applied on the state function – returns the concatenation of the measurements. This perspective is particularly interesting since we have seen above an example of how by SEC and by SLS it is possible to get several measurements on the same system and the more observations we have the better we can estimate the solution

of the inverse problem.

The two methods have been numerically tested and compared on synthetic data. The results of these numerical estimations are in agreement with theoretical estimates. We remark that the estimations we get can take negative values. This is an expected behaviour because we look for an estimation in  $\mathbf{L}^p$ -norm and we do not enforce positivity constraints. A possible improvement for the variational approach would be to either do a constrained optimisation or parametrise the state function to guarantee its positivity at all times.

In the last section we have presented and discussed our inverse problem methodology applied to experimental data of ovPrP oligomers, and this study exemplified the flexibility of the data assimilation framework. We were led to the conclusion that most probably the smaller polymers are more stable than the larger ones. To support this conclusion, further experiments have to be carried out. The simultaneous measurement of the first moment, *i.e.* of the total polymerised mass (by ThT), and of the second moment (by SLS) should be much more informative, and would lead to interesting extensions of our approach. This is a direction for future work.

## Acknowledgments

The research of M. Doumic and part of the research of A. Armiento are supported by the ERC Starting Grant SKIPPER<sup>AD</sup> (number 306321).

## Appendices

### 6 Mathematical proofs

In order to prove Proposition 1 we first recall a classical lemma on convolution products.

#### LEMMA 3

Let  $n \in \mathbb{N}$ ,  $p \geq 1$ ,  $\alpha \in (0, 1)$ , the function  $\rho \in \mathcal{C}_c^\infty(\mathbb{R})$  and the coefficient  $m$  satisfying Assumptions (9). We define the function  $\rho_\alpha(x) = \frac{1}{\alpha} \rho(\frac{x}{\alpha})$ .

i) If the function  $f$  is in  $\mathbf{W}^{1,p}(\mathbb{R}_+)$ , we have

$$\|f - \rho_\alpha * f\|_{\mathbf{L}^p(\mathbb{R}_+)} \leq c_1 \alpha \|f\|_{\mathbf{W}^{1,p}(\mathbb{R}_+)}, \quad (31)$$

where  $c_1 = \|x\rho\|_{\mathbf{L}^1(\mathbb{R})}$

ii) Let  $n \leq m$ , if the function  $f$  is in  $\mathbf{W}^{n+1,p}(\mathbb{R}_+)$ , we have

$$\|f - \rho_\alpha * f\|_{\mathbf{L}^p(\mathbb{R}_+)} \leq c_2 \alpha^{n+1} \|f\|_{\mathbf{W}^{n+1,p}(\mathbb{R}_+)}, \quad (32)$$

where  $c_2 = \frac{1}{n!} \|x^{n+1}\rho(x)\|_{\mathbf{L}^1(\mathbb{R})}$ .

iii) Furthermore, we have

$$\|\rho_\alpha * f^{(n)}\|_{\mathbf{L}^p(\mathbb{R}_+)} = \|\rho_\alpha^{(n)} * f\|_{\mathbf{L}^p(\mathbb{R}_+)} \leq c_3 \alpha^{-n} \|f\|_{\mathbf{L}^p(\mathbb{R}_+)}, \quad (33)$$

where  $c_3 = \|(\rho^{(n)})_\alpha\|_{\mathbf{L}^1(\mathbb{R})}$ .

iv) Given  $s \in [0, 1)$ , if the function  $f$  is in  $\mathbf{W}^{-s,p}(\mathbb{R}_+)$  and  $\rho, \rho' \in \mathbf{L}^1(\mathbb{R})$ , we have

$$\|\rho_\alpha * f\|_{\mathbf{L}^p(\mathbb{R}_+)} \leq c_4 \alpha^{-s} \|f\|_{\mathbf{W}^{-s,p}(\mathbb{R}_+)}, \quad (34)$$

where  $c_4$  depends on  $\|\rho\|_{\mathbf{L}^1(\mathbb{R})}, \|\rho'\|_{\mathbf{L}^1(\mathbb{R})}$ .

v) Given  $s \in [0, 1)$ , if the function  $f$  is in  $\mathbf{W}^{-s,p}(\mathbb{R}_+)$  and  $\rho^{(n)}, \rho^{(n+1)} \in \mathbf{L}^1(\mathbb{R}_+)$ , we have

$$\|\rho_\alpha * f^{(n)}\|_{\mathbf{L}^p(\mathbb{R}_+)} \leq c_5 \alpha^{-(n+s)} \|f\|_{\mathbf{W}^{-s,p}(\mathbb{R}_+)}. \quad (35)$$

where  $c_5$  depends on  $\|\rho^{(n)}\|_{\mathbf{L}^1(\mathbb{R})}, \|\rho^{(n+1)}\|_{\mathbf{L}^1(\mathbb{R})}$ .

Let us now state and prove Proposition 1.

**PROPOSITION 4 (PROPOSITION 1)**

Let  $1 \leq p < \infty$ ,  $n \in \mathbb{N}$ ,  $0 \leq s < 1$  and the function  $\Psi_\tau u_0$  defined in Equation (6). Let  $\Psi_\tau \check{u}_0 \in \mathbf{W}^{m+n+2,p}([0, \tau])$ , with  $m$  defined as in Equation (9). Let  $z \in \mathbf{W}^{-s,p}([0, \tau])$  a measurement of the  $n$ -th momentum  $\Psi_\tau \check{u}_0$  such that  $\tau \geq \frac{\ell}{b}$  and

$$\|z - \Psi_\tau \check{u}_0\|_{\mathbf{W}^{-s,p}([0, \tau])} \leq \varepsilon.$$

The following relation holds true

$$\check{u}_0(x) = \frac{1}{n!(-b)^{n+1}} \frac{d^{n+1}}{dt^{n+1}} \Psi_\tau \check{u}_0 \left( \frac{x}{b} \right). \quad (36)$$

Let  $\rho$  defined by Equation (9) and  $\rho_\alpha$  by Equation (10), with  $\alpha \in (0, 1)$ . We consider

$$\hat{u}_0^{\varepsilon, \alpha} = \frac{d^{n+1}}{dx^{n+1}} \rho_\alpha * \left( \frac{1}{n!(-b)^{n+1}} z \left( \frac{x}{b} \right) \right) \quad (37)$$

as approximation of  $\check{u}_0$ . Then the following estimation is of optimal order in the sense of [12]

$$\|\hat{u}_0^{\varepsilon, \alpha} - \check{u}_0\|_{\mathbf{L}^p([0, \ell])} \leq \Theta \left( \frac{\varepsilon}{\alpha^{n+s+1}} + \alpha^{m+1} \right) = F_\varepsilon(\alpha), \quad (38)$$

where the constant  $\Theta$  depends on  $\|\Psi_\tau \check{u}_0\|_{\mathbf{W}^{m+n+2,p}([0, \tau])}, \|x^{m+1} \rho\|_{\mathbf{L}^1(\mathbb{R})}, \|\rho^{(n)}\|_{\mathbf{L}^1(\mathbb{R})}, \|\rho^{(n+1)}\|_{\mathbf{L}^1(\mathbb{R})}$ .

*Proof.* We start by defining the function

$$\hat{u}_0^\alpha = \rho_\alpha * u_0.$$

By the triangle inequality for  $\mathbf{L}^p$ -norm we have

$$\|\hat{u}_0^{\varepsilon,\alpha} - u_0\|_{\mathbf{L}^p([0,\ell])} \leq \|\hat{u}_0^\alpha - u_0\|_{\mathbf{L}^p([0,\ell])} + \|\hat{u}_0^{\varepsilon,\alpha} - \hat{u}_0^\alpha\|_{\mathbf{L}^p([0,\ell])}.$$

Let us consider the two terms on the right-hand separately.

The first term is  $\|\hat{u}_0^\alpha - u_0\|_{\mathbf{L}^p([0,\ell])} = \|\rho_\alpha * u_0 - u_0\|_{\mathbf{L}^p([0,\ell])}$ . By using Inequality (32), we obtain

$$\|\hat{u}_0^\alpha - u_0\|_{\mathbf{L}^p([0,\ell])} \leq \gamma_1 \alpha^{(m+1)} \|u_0\|_{\mathbf{W}^{m+1,p}([0,\ell])},$$

with  $\gamma_1 = \frac{1}{m!} \|x^{(m+1)} \rho\|_{\mathbf{L}^1(\mathbb{R})}$ . While the second term is

$$\begin{aligned} \|\hat{u}_0^{\varepsilon,\alpha} - \hat{u}_0^\alpha\|_{\mathbf{L}^p([0,\ell])} &= \|\rho_\alpha * \hat{u}_0^\varepsilon - \rho_\alpha * u_0\|_{\mathbf{L}^p([0,\ell])} \\ &= \|\rho_\alpha * (\hat{u}_0^\varepsilon - u_0)\|_{\mathbf{L}^p([0,\ell])} \\ &= \frac{1}{n!(-b)^{n+1}} \left\| \rho_\alpha * \frac{d^{n+1}}{dt^{n+1}} (z - \Psi_\tau \check{u}_0) \right\|_{\mathbf{L}^p([0,\ell])}. \end{aligned}$$

By recalling Inequality(35) , with  $f = z - \Psi_\tau \check{u}_0$ , we have

$$\|\hat{u}_0^{\varepsilon,\alpha} - \hat{u}_0^\alpha\|_{\mathbf{L}^p([0,\ell])} \leq \frac{1}{n!(-b)^{n+1}} \gamma_2 \alpha^{-(n+1+s)} \|z - \Psi_\tau \check{u}_0\|_{\mathbf{W}^{-s,p}([0,\ell])}.$$

In conclusion, we obtain

$$\begin{aligned} \|\hat{u}_0^{\varepsilon,\alpha} - u_0\|_{\mathbf{L}^p([0,\ell])} &\leq \gamma_1 \alpha^{(m+1)} \|u_0\|_{\mathbf{W}^{m+1,p}([0,\ell])} + \frac{1}{n!(-b)^{n+1}} \gamma_2 \alpha^{-(n+1+s)} \|z - \Psi_\tau \check{u}_0\|_{\mathbf{W}^{-s,p}([0,\tau])} \\ &\leq \gamma_1 \frac{1}{n!(-b)^{n+1}} \alpha^{(m+1)} \|\Psi_\tau \check{u}_0\|_{\mathbf{W}^{n+m+2,p}([0,\tau])} \\ &\quad + \frac{1}{n!(-b)^{n+1}} \gamma_2 \alpha^{-(n+1+s)} \|z - \Psi_\tau \check{u}_0\|_{\mathbf{W}^{-s,p}([0,\tau])} \\ &\leq \Theta \left( \alpha^{m+1} + \frac{\varepsilon}{\alpha^{n+1+s}} \right), \end{aligned}$$

where  $\Theta = \frac{1}{n!(-b)^{n+1}} \max\{\gamma_1 \|\Psi_\tau \check{u}_0\|_{\mathbf{W}^{n+m+2,p}([0,\tau])}, \gamma_2\}$ . ■

We now turn to Proposition 2.

**PROPOSITION 5 (PROPOSITION 2)**

For any  $z \in L^2([0, \tau])$ , there exists a unique minimiser  $\bar{\xi}$  for  $J(\xi)$  defined by

$$J(\xi) = \frac{\alpha^2}{2} \|\xi\|_{L^2([0,\ell])}^2 + \frac{1}{2} \int_0^\tau |z(t) - \Psi_\tau(\xi)|^2 dt,$$

and  $\bar{\xi} \in \mathbf{H}^{n+1}([0, \ell])$ . If moreover  $\check{\xi} \in \mathbf{H}^{n+1}([0, \ell])$  with  $\check{\xi}(0) = \dots = \check{\xi}^{(n)}(0) = 0$ , we have the following estimate

$$\|\check{\xi} - \bar{\xi}\|_{\mathbf{L}^2([0,\ell])} \leq \frac{1}{\alpha} \|z - \Psi_\tau(\check{\xi})\|_{\mathbf{L}^2([0,\tau])} + \alpha \|\check{\xi}\|_{\mathbf{H}^{n+1}(0,\ell)}.$$

*Proof.* This result is based on general inequalities for Tikhonov method, that we recall below.

**LEMMA 6 (ESTIMATES FOR TIKHONOV REGULARIZATION)**

Let  $K : \mathcal{U} \rightarrow \mathcal{Y}$  a compact injective operator between two Hilbert spaces  $\mathcal{U}$  and  $\mathcal{Y}$ , with norms  $\|\cdot\|_{\mathcal{U}}$  and  $\|\cdot\|_{\mathcal{Y}}$ ,  $K^* : \mathcal{Y} \rightarrow \mathcal{U}$  its adjoint (relatively to these norms) and  $K^\dagger$  its Moore-Penrose pseudo-inverse. Let  $y_\varepsilon \in \mathcal{Y}$ . Let  $u_{\varepsilon,\alpha} \in \mathcal{U}$  the unique solution of

$$(K^*K + \alpha^2)u_{\varepsilon,\alpha} = K^*y_\varepsilon. \quad (39)$$

Then  $u_{\varepsilon,\alpha}$  is also the unique minimiser of the following functional

$$J_K(u) := \frac{1}{2}\|Ku - y_\varepsilon\|_{\mathcal{Y}}^2 + \frac{\alpha^2}{2}\|u\|_{\mathcal{U}}^2.$$

Moreover, if  $u_{\varepsilon,\alpha} \in \text{Ran}(K^*)$  and we have

$$\|Ku_{\varepsilon,\alpha}\|_{\mathcal{Y}} \leq \|y_\varepsilon\|_{\mathcal{Y}}, \quad \|u_{\varepsilon,\alpha}\|_{\mathcal{U}} \leq \frac{1}{\alpha}\|y_\varepsilon\|_{\mathcal{Y}}.$$

If moreover  $y \in \text{Ran}(K)$ , denoting  $Ku = y$  and  $u_\alpha$  the solution to (39) with  $y_\varepsilon = y$ , we have  $u_\alpha \in \text{Ran}(K^*K)$  and

$$\|u_\alpha\|_{\mathcal{U}} \leq \|x\|_{\mathcal{U}},$$

If moreover  $u \in \text{Ran}(K^*)$ , we have

$$\|u_\alpha - u\|_{\mathcal{U}} \leq \alpha\|K^{*\dagger}u\|_{\mathcal{Y}},$$

if moreover  $x \in \text{Ran}(K^*K)$ , we have

$$\|u_\alpha - u\|_{\mathcal{U}} \leq \alpha^2\|(K^*K)^\dagger u\|_{\mathcal{U}}.$$

We now apply this result to  $\mathcal{U} = \mathbf{L}^2([0, \ell])$ ,  $\mathcal{Y} = \mathbf{L}^2([0, \tau])$  and  $K = \Psi_\tau$ . We take  $u = \check{\xi}$  and  $y_\varepsilon$  to be the measurement function  $t \in [0, \tau] \rightarrow z(t)$ , and  $y = \Psi_\tau \check{\xi}$ .

$$\text{Ran}\Psi_\tau^n = \left\{ u \in \mathbf{H}^{n+1}([0, \tau]), u(\tau) = \dots = u^{(n)}(\tau) = 0 \right\},$$

and

$$\text{Ran}\Psi_\tau^{*n} = \left\{ u \in \mathbf{H}^{n+1}([0, \ell]), u(0) = \dots = u^{(n)}(0) = 0 \right\}.$$

Lemma 6 gives us that  $u_{\varepsilon,\alpha} = \bar{\xi}$  is the unique minimizer for  $J$ . We decompose as is well-known

$$\|\check{\xi} - \bar{\xi}\|_{\mathbf{L}^2([0, \ell])} \leq \|\check{\xi} - u_\alpha\|_{\mathbf{L}^2([0, \ell])} + \|u_\alpha - u_{\varepsilon,\alpha}\|_{\mathbf{L}^2([0, \ell])}.$$

In Proposition 2, the assumptions on  $\check{\xi}$  mean that  $\check{\xi} \in \text{Ran}(K^*)$ , hence

$$\|u_\alpha - \check{\xi}\|_{\mathcal{U}} \leq \alpha\|K^{*\dagger}\check{\xi}\|_{\mathcal{Y}} \leq \alpha\|\check{\xi}\|_{\mathbf{H}^{n+1}([0, \ell])}.$$

Concerning the term  $\|\check{\xi} - u_\alpha\|_{\mathbf{L}^2([0, \ell])}$ , we apply the second inequality of Lemma 6 to Equation (39) with  $y_\varepsilon$  replaced by  $y_\varepsilon - y$ , for which  $u_{\varepsilon,\alpha} - u_\alpha$  is a solution, and find

$$\|\check{\xi} - u_\alpha\|_{\mathbf{L}^2([0, \ell])} \leq \frac{1}{\alpha}\|y_\varepsilon - y\|_{\mathcal{Y}} = \frac{1}{\alpha}\|z - \Psi_\tau(\check{\xi})\|_{\mathbf{L}^2([0, \tau])}.$$

This ends the proof. ■

## 7 Oligomer formation protocol

The protocol to form and make measurements on oligomer systems – previously described in [10] – consists in inducing a partial unfolding of full-length ovine PrP protein by thermal treatment. This partial unfolding leads to generation of three distinct oligomers that can be purified by size exclusion chromatography, for further investigation we refer to [11]. The conversion of PrP into the oligomeric form is performed in  $20mM$  sodium citrate buffer (pH 3.40). The PrP – at a final concentration of  $50\mu M$  – is incubated in a Perkin Elmer GenAmp2400 thermocycler at  $65^{\circ}C$  for two hours. Homogeneous fractions of oligomers are then collected after separation by size exclusion chromatography (SEC), as first described in [10]. The SEC is performed at  $20^{\circ}C$  using a TSK 4000SW ( $7mm * 600mm$ ) gel-filtration column (Interchim, Montluçon, France) with  $20mM$  sodium citrate (pH 3.35). Protein elution is monitored by UV absorption at  $280nm$ . The size distribution of oligomer assemblies has been determined by the SEC device coupled with the static light scattering device. Depolymerisation kinetics are performed with an in-lab device using  $407nm$  laser beams in a  $2mm$ -path-length quartz cuvette. Kinetic experiments are performed according to a standardise methodology, as reported in [11]:  $72^{\circ}C$  in  $20mM$  sodium citrate buffer (pH 3.40). The oligomer concentration has been fixed at  $3\mu M$ .

## References

- [1] H. T. Banks, M. Doumic, C. Kruse, S. Prigent, and H. Rezaei. Information content in data sets for a nucleated-polymerisation model. *Journal of Biological Dynamics*, 9(1):172–197, January 2015.
- [2] H. T. Banks, Marie Doumic-Jauffret, and Carola Kruse. Efficient numerical schemes for nucleation-aggregation models: Early steps. Mar 2014.
- [3] H.T. Banks. *A Functional Analysis Framework for Modeling, Estimation and Control in Science and Engineering*. CRC Press, 2012.
- [4] H.T. Banks and D.M. Bortz. A parameter sensitivity methodology in the context of hiv delay equation models. *Journal of Mathematical Biology*, 50(6):607–625, 2005.
- [5] Kiersten M. Batzli and Brian J. Love. Agitation of amyloid proteins to speed aggregation measured by ThT fluorescence: A call for standardization. *Materials Science and Engineering: C*, 48(0):359 – 364, 2015.
- [6] Johann Baumeister and Antonio Leitão. *Topics in inverse problems*. Publicações Matemáticas do IMPA. [IMPA Mathematical Publication]. Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 25<sup>o</sup> colóquio brasileiro de matemática. [25th brazilian mathematics colloquium] edition, 2005.



- [7] Alain Bensoussan. *Filtrage Optimal des Systèmes Lineaires*. Méthodes Mathématiques de l'informatique. Dunod, 1971.
- [8] M. F. Bishop and F. A. Ferrone. Kinetics of nucleation-controlled polymerization. a perturbation treatment for use with a secondary pathway. *Biophys J.*, 46(5):631–644, November 1984.
- [9] M. Doumic, T. Goudon, and T. Lepoutre. Scaling limit of a discrete prion dynamics model. *Communications in Mathematical Sciences*, 7(4):839–865, 2009.
- [10] F. Eghiaian, T. Daubenfeld, Y. Quenet, M. van Audenhaege, AP. 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. *Proc Natl Acad Sci U S A.*, 104(18):7414–7419, may 2007.
- [11] F. Eghiaian, T. Daubenfeld, Y. Quenet, M. van Audenhaege, AP. Bouin, G. van der Rest, J. Grosclaude, and H. Rezaei. The oligomerization properties of prion protein are restricted to the H2H3 domain. *The FASEB Journal*, 24(9):3222–3231, sep 2010.
- [12] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, 1996.
- [13] Barnabas James Gilbert. The role of amyloid  $\beta$  in the pathogenesis of Alzheimer's disease. *J Clin Pathol*, 66:362–366, March 2013.
- [14] Lifshitz I.M. The kinetics of precipitation from supersaturated solid solutions. *Journal of physics and chemistry of solids*, 19:35–50, 1961.
- [15] Francois-Xavier Le Dimet and O Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus A*, 38(2):97–110, 1986.
- [16] Randall J. LeVeque. *Finite-Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.
- [17] M. Nussbaum. Asymptotic equivalence of density estimation and white noise. *Ann. Statist.*, 24:2399–2430, 1996.
- [18] Michael Nussbaum and Sergej V Pereverzev. *The Degree of Ill Posedness in Stochastic and Deterministic Noise Model*. WIAS, 1999.
- [19] Sophocles J. Orfanidis. *Introduction to Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [20] Sian-Yang Ow and Dave E. Dunstan. A brief overview of amyloids and alzheimer's disease. *Protein Science*, 23(10):1315–1331, 2014.

- [21] Antoine Perasso, Béatrice Laroche, Yacine Chitour, and Suzanne Touzeau. Identifiability analysis of an epidemiological model in a structured population. *Journal of Mathematical Analysis and Applications*, 374(1):154 – 165, 2011.
- [22] 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), 11 2012.
- [23] Marina Ramirez-Alvarado, Jane S. Merkel, and Lynne Regan. A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. *Proceedings of the National Academy of Sciences*, 97(16):8979–8984, 2000.
- [24] Human Rezaei, Frédéric Eghiaian, Javier Perez, Bénédicte Doublet, Yvan Choiset, Thomas Haertle, and Jeanne Grosclaude. Sequential Generation of Two Structurally Distinct Ovine Prion Protein Soluble Oligomers Displaying Different Biochemical Reactivities. *Journal of Molecular Biology*, 347:665–679, apr 2005.
- [25] D Simon. *Optimal State Estimation: Kalman, H Infinity, And Nonlinear Approaches*. Springer, 2006.
- [26] S. Simoneau, H. Rezaei, N. Salès, G. Kaiser-Schulz, M. Lefebvre-Roque, C. Vidal, JG. Fournier, J. Comte, F. Wopfner, J. Grosclaude, H. Schätzl, and C. Lasmézas. In vitro and in vivo neurotoxicity of prion protein oligomers. *PLoS Pathog*, 3(8), Aug 2007.
- [27] Daniel Some. Light-scattering-based analysis of biomolecular interactions. *Biophysical Reviews*, 5(2):147–158, 2013.
- [28] M.P. Wand and M.C. Jones. *Kernel Smoothing*. CRC Press, 1994.
- [29] 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.
- [30] W-F Xue and S E Radford. An imaging and systems modeling approach to fibril breakage enables prediction of amyloid behavior. *Biophys. Journal*, 105:2811–2819, 2013.