



Bayesian inversion of a diffusion model with application to biology

Jean-Charles Croix¹ · Nicolas Durrande² · Mauricio A. Alvarez³

Received: 24 January 2020 / Revised: 6 March 2021 / Accepted: 26 May 2021 /

Published online: 6 July 2021

© The Author(s) 2021

Abstract

A common task in experimental sciences is to fit mathematical models to real-world measurements to improve understanding of natural phenomenon (reverse-engineering or inverse modelling). When complex dynamical systems are considered, such as partial differential equations, this task may become challenging or ill-posed. In this work, a linear parabolic equation is considered as a model for protein transcription from mRNA. The objective is to estimate jointly the differential operator coefficients, namely the rates of diffusion and self-regulation, as well as a functional source. The recent Bayesian methodology for infinite dimensional inverse problems is applied, providing a unique posterior distribution on the parameter space continuous in the data. This posterior is then summarized using a Maximum a Posteriori estimator. Finally, the theoretical solution is illustrated using a state-of-the-art MCMC algorithm adapted to this non-Gaussian setting.

Keywords Bayesian inverse problems · Diffusion equation · Functional MCMC · Gaussian processes

Mathematics Subject Classification 62G05 · 62P10 · 35R30 · 62F15

✉ Jean-Charles Croix
j.croix@sussex.ac.uk

Nicolas Durrande
nicolas@proowler.io

Mauricio A. Alvarez
mauricio.alvarez@sheffield.ac.uk

¹ School of Mathematical and Physical Sciences, University of Sussex, Falmer, Brighton BN1 9QH, UK

² Prowler.io, Cambridge CB2 1LA, UK

³ Department of Computer Science, University of Sheffield, Sheffield S10 2TN, UK

1 Introduction

The problem of diffusion in a porous media, which is ubiquitous in Physics, Engineering and Biology, is usually represented by the following partial differential equation (with constant diffusion and damping without transport):

$$\begin{aligned} \frac{\partial z}{\partial t}(x, t) + \lambda z(x, t) - D \Delta z(x, t) &= f(x, t), \quad \forall (x, t) \in \Omega \times]0, T], \\ z(x, t) &= 0, \quad \forall (x, t) \in \Omega \times \{t = 0\}, \\ z(x, t) &= 0, \quad \forall (x, t) \in \partial\Omega \times]0, T], \end{aligned} \quad (1)$$

where the spatial domain is an open set $\Omega \subset \mathbb{R}^n$ ($n \leq 3$) and the final time is $T \in \mathbb{R}^+$ (other initial and boundary conditions are possible). In real world applications, the quantity of interest z (hereafter called the solution of Eq. 1) is typically the concentration of some chemical and evolves from a null initial state under three distinct mechanisms: (a) direct variation in concentration, given by the source f , (b) diffusion at rate D , (c) production or depletion at a rate λ . Different hypotheses on the parameters lead to a well-defined solution [well-posedness in the sense of Hadamard Hadamard (1902)] but only a particular case will be dealt with here. Besides the traditional computation of the solution from the parameters, one can use this model for the determination of an optimal control (e.g. source leading to the minimization of a particular cost functional) or the identification of parameters from partial knowledge of the solution in an inverse setting and the latter is the objective of this paper. The motivation comes from a challenging identification problem in Biology where the objective is to infer jointly self-regulation and diffusion rates with the source f given a limited number of noisy observations of the solution z . Note that given their physical interpretation, the parameters $u = (\lambda, D, f)$ must all be non-negative (in an obvious sense).

This problem has already been solved using various approaches. In Becker et al. (2013), the authors use a system of ordinary differential equations instead of Eq. 1, and minimize a discrete version of a least-square penalty functional, while confidence intervals on parameters are given by bootstrapping. In a Bayesian setting, alternative methods have been based on Latent Force Models (Alvarez et al. 2013; Särkkä et al. 2018). These approaches assume that the source can be modelled with Gaussian Processes (Rasmussen and Williams 2006). In particular, if f is taken to be such a process and if the decay and the diffusion are constants, z is Gaussian as well (since z depends linearly in f). The two constant parameters λ and D can then be optimized as hyper-parameters through standard likelihood maximization (Lopez-Lopera et al. 2019). Major drawbacks of this approach are both the difficulty to enforce positiveness of the reconstructed source and the absence of uncertainty quantification on λ, D .

In this work, a more general methodology is applied, based on the recent advent of Bayesian Inverse Problems (Stuart 2010) for infinite dimensional spaces. This has the advantage of dealing with the ill-posedness while fully integrating the quantification of uncertainties in a unified approach. Moreover, the possibility to include previous physical constraints in the prior will be particularly meaningful. The rest of this paper

is organized as follows: Sect. 2 presents all the mathematical analysis underlying the Bayesian inversion, namely sufficient regularity of the forward operator (mapping the parameter (λ, D, f) to the PDE solution z), existence and uniqueness of the posterior under a simple class of priors, and finally uniqueness of the associated posterior. Section 3 focuses on an adaptation of a geometric Markov Chain Monte-Carlo algorithm, well-defined in function spaces. Finally, Sect. 4 contains all implementation details, as well as a methodology to tune hyper-parameters of the model. All numerical results are based on a real-world dataset related to the developmental biology of the *Drosophila* *Melanogaster*.

2 Bayesian inversion

As previously announced, the goal in this work is to infer a source term f (mRNA concentration) jointly with rates of diffusion D and decay λ (i.e. the parameter $u = (\lambda, D, f)$) from noisy and partial measurements of the solution z (gap protein concentration). This problem is ill-posed for multiple reasons: (a) the parameter u is infinite dimensional and only finite data are available, (b) the solution map is not injective and (c) observations are noisy. The typical approach to alleviate this issue is to regularize the problem, usually adding more constraints with Tikhonov-Philips functionals, to ensure uniqueness and continuity w.r.t. the observations (Isakov 2017; Schuster et al. 2012). Doing so, the regularized solution will be compatible with the dataset. Additionally, a particularly valuable information is a representation of all parameters u that would lead to similar data, giving precise statement on how the dataset is informative (Ghanem et al. 2017; Biegler et al. 2011; Sullivan 2015). One approach consists in treating these 2 objectives sequentially, first regularizing then quantifying the resulting uncertainty. However, the Bayesian methodology for inverse problems (Stuart 2010) and more recently Dashti and Stuart (2015) is precisely tailored to complete both tasks at once in an elegant manner. One particularity of these recent contributions is to tackle inverse problems directly in function spaces, postponing discretization at the very end for implementation purposes, which leads to robust algorithms w.r.t. discretization dimension. Indeed, finite approximations of probability measures may be absolutely continuous while their infinite counterparts are mutually singular. This becomes particularly troublesome in MCMC sampling for instance (Cotter et al. 2013).

In essence, instead of searching for one particular parameter solving the regularized problem, this approach considers conditional probability measures of parameters given observations. Namely, given a prior distribution (see Sect. 2.2) and few technical conditions on the forward operator (see Sect. 2.1), Bayes theorem applies and exhibits a unique posterior distribution (see Sect. 2.3), which is continuous in the data [w.r.t. Hellinger metric (Stuart 2010)]. Finally, one may summarize information from this posterior distribution with point estimators such as expected value or modes (Sect. 2.4). All these steps will now be presented.

2.1 Forward model analysis

The first step is to detail precisely the required regularity of the solution map from Eq. 1. Using common variational techniques from PDE theory [see Evans (1998) or Brezis (2011)], one can show that this equation has a unique weak solution (see Proposition 1 which proof is in the appendix) given $u = (\lambda, D, f)$ in a domain \mathcal{U} that will be explicitated later on. Moreover, this solution evolves smoothly when the parameter varies. Without any loss of generality and keeping in mind the biological application, the underlying physical domain will be $\Omega =]0, L[$ with $L \in \mathbb{R}^+$, representing the anterior-posterior axis of a *Drosophila* embryo.

Proposition 1 Let $0 < \lambda_M, 0 < D_m \leq D_M, \mathcal{U} = [0, \lambda_M] \times [D_m, D_M] \times \mathcal{C}([0, T] \times [0, L]; \mathbb{R})$ with the norm $\|u\|_{\mathcal{U}} = |\lambda| + |D| + \|f\|_{\infty}$, then for all $u \in \mathcal{U}$, Eq. 1 has a unique weak solution z , continuous on the domain $[0, L] \times [0, T]$, defining the map:

$$z : u \in \mathcal{U} \rightarrow z(u) \in C([0, T] \times [0, L]; \mathbb{R}).$$

Moreover, this map has the following properties:

1. (Energy estimate): it satisfies the following estimate $\forall u \in \mathcal{U}$,

$$\|z(u)\|_{\infty} \leq C \|f\|_{\infty},$$

where $C > 0$ is a constant independent of u ,

2. (Local Lipschitz continuity): $\forall u$ in the interior of \mathcal{U} , $\forall r > 0$ such that $\mathcal{B}(u, r) \subset \mathcal{U}$, $\exists L(u, r) > 0$,

$$\forall (u_1, u_2) \in \mathcal{B}(u, r) \times \mathcal{B}(u, r), \|z(u_1) - z(u_2)\|_{\infty} \leq L(u, r) \|u_1 - u_2\|_{\mathcal{U}}$$

3. (Differentiability): it is twice Fréchet differentiable in the interior of \mathcal{U} .

The proof is standard using methods from PDE and optimal control theory (Evans 1998). The properties stated in proposition 1 will be important for subsequent analysis (Sect. 2.3):

1. The energy estimate is fundamental to establish most of the results concerning the posterior distribution. Indeed, it gives a precise upper bound, useful to show integrability of the yet-to-come likelihood,
2. Local Lipschitz continuity implies measurability of the solution map w.r.t. the Borel σ -algebra and is used in the characterization of posterior modes (Maximum a Posteriori estimators),
3. Second order Fréchet differentiability will be necessary for geometric methods in optimization (research of posterior modes) and sampling (Markov-Chain Monte-Carlo) since they rely on Hessian-type information.

2.2 Choice of a prior distribution

The second step is to choose a prior probability distribution on \mathcal{U} , encoding all knowledge on the problem at hand, while being simple enough to keep the analysis tractable. The prior will be constructed as a product measure, specifying marginal measures $\mu_0^\lambda, \mu_0^D, \mu_0^f$ on all parameters:

$$\mu_0(du) := \mu_0^\lambda(d\lambda) \otimes \mu_0^D(dD) \otimes \mu_0^f(df). \tag{2}$$

For simplicity, both measures for λ and D will be taken uniform on respective intervals $[D_m, D_M]$ and $[0, \lambda_M]$. Now, since f must be non-negative, $\mathbf{1}$ is re-parametrized with the following source term:

$$f^* = \exp(f), \tag{3}$$

where $f \in C([0, T] \times [0, L]; \mathbb{R})$. Selecting a Borel probability measure μ_0^f on $C([0, T] \times [0, L]; \mathbb{R})$ will imply both continuity and positivity of the source f^* almost-surely. The energy estimate is adjusted:

$$\forall u \in \mathcal{U}, \|z(u)\|_\infty \leq C^* \exp(\|f\|_\infty).$$

In this work, μ_0^f will be taken as the Gaussian measure associated with a continuous Gaussian process [see Bogachev (1998) for a presentation of infinite dimensional Gaussian measures] such that f is almost-surely continuous. Remark that the exponential map in Eq. 3 could be replaced with any sufficiently differentiable function from \mathbb{R} to \mathbb{R}^+ (thus keeping the second order Fréchet differentiability of the solution map). Besides, alternative distributions are also possible for f like Besov priors [from Dashti et al. (2013)] or more general convex measures [from Hosseini and Nigam (2017)]. However, this choice is also motivated by practical reasons, since one can build a Gaussian measure μ_{ref} dominating μ_0 :

$$\mu_{ref} = \mathcal{N}(\lambda_{ref}, \sigma_\lambda^2) \otimes \mathcal{N}(D_{ref}, \sigma_D^2) \otimes \mu_0^f.$$

Indeed, choose $(\lambda_{ref}, D_{ref}) \in \mathbb{R}^2$ and $\sigma_\lambda^2, \sigma_D^2 > 0$ then $\mu_0 \ll \mu_{ref}$ with

$$\begin{aligned} \frac{d\mu_0}{d\mu_{ref}}(u) &= \frac{d\mu_0^\lambda}{d\mu_{ref}^\lambda}(\lambda) \frac{d\mu_0^D}{d\mu_{ref}^D}(D) \\ &= \frac{2\pi\sigma_\lambda\sigma_D}{\lambda_M(D_M - D_m)} \exp\left(\frac{(\lambda - \lambda_{ref})^2}{2\sigma_\lambda^2} + \frac{(D - D_{ref})^2}{2\sigma_D^2}\right) \mathbb{1}_{[0, \lambda_{ref}]}(\lambda) \mathbb{1}_{[D_m, D_M]}(D). \end{aligned}$$

The parameters of μ_{ref} are tuned by choosing $\lambda_{ref} = \frac{\lambda_M}{2}, \sigma_\lambda^2 = \frac{\lambda_M^2}{12}, D_{ref} = \frac{D_M - D_m}{2}$ and $\sigma_D^2 = \frac{(D_M - D_m)^2}{12}$ (minimizing Kullback-Leibler divergence of μ_{ref} relative to μ_0). This dominant measure will be critical for posterior modes analysis (Sect. 2.4) and MCMC sampling (Sect. 3.2).

2.3 Posterior distribution

It is now time to show that the particular setting provided so far (forward model and prior distribution) leads to a well defined posterior measure. This is the purpose of Proposition 2 which is a direct application of the theory initially developed in Stuart (2010) [see Dashti and Stuart (2015) for an updated presentation]. In this purpose, consider a dataset $y = (y_i)_{i \in [1, n]} \in \mathbb{R}^n$ consisting of observations from the solution z at different times and locations $(t_i, x_i)_{i \in [1, n]} \in ([0, T] \times [0, L])^n$, produced under the following additive noise model (in vector notations):

$$y = \mathcal{G}(u) + \eta, \quad (4)$$

where $\eta \sim \mathcal{N}(0, \sigma_\eta^2 I_n)$ (I_n being the identity matrix of dimension n) and $\mathcal{G} : \mathcal{U} \rightarrow \mathbb{R}^n$ is the observation operator, mapping directly the PDE parameter u to the value of the associated solution $(z[u](x_i, t_i))_{i \in [1, n]}$ (composition of solution map z with Dirac type measure). Note that it is well-defined since the function z is continuous on the domain for every $u \in \mathcal{U}$. The following proposition, which is again proved in the appendix, establishes the existence, uniqueness and continuity in y of the posterior probability measure μ^y (the solution of the inverse problem), expressing how observations y updated prior beliefs μ_0 on the parameter u .

Proposition 2 Let \mathcal{G} be the observation operator defined in Eq. 4, $y \in \mathbb{R}^n$ a dataset and μ_0 the probability measure defined in Eq. 2, then there exists a unique posterior measure μ^y for $u|y$. It is characterized by the following Radon-Nikodym density w.r.t. μ_0 :

$$\forall u \in \mathcal{U}, \quad \frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp(-\Phi(u; y)),$$

with the negative log-likelihood

$$\Phi(u; y) = \frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|_{\mathbb{R}^n}^2,$$

and the marginal

$$Z(y) = \int_{\mathcal{U}} \exp(-\Phi(u; y)) \mu_0(du).$$

Furthermore, μ^y is continuous in y w.r.t. Hellinger distance.

In fact, Proposition 2 gives two distinct results: a) the existence and uniqueness of a posterior (as long as μ_0 is Radon and $\mu_0(\mathcal{U}) = 1$, which is the case here), b) well-posedness of the Bayesian inverse problem. In particular, the use of a Gaussian prior on f gives sufficient integrability, even under the re-parametrization from Eq. 3 (using Fernique's theorem). If one chooses a different map between f^* and f in Eq. 3, this condition may be considerably relaxed (using something slower than the exponential) and prior measures with lower integrability conditions can be considered.

2.4 Maximum a posteriori estimator

In the previous section, the well-posedness of the Bayesian inverse problem has been proved. However, the posterior distribution is only known up to a multiplicative constant, through its density w.r.t. μ_0 . In the application, μ^y will need to be summarized, which is usually done by the selection of a particular estimator in \mathcal{U} , such the posterior mean or a mode. One consequence of Proposition 2 is that the posterior mean is automatically continuous in the data y [since well-posedness is w.r.t. Hellinger distance, see Dashti and Stuart (2015)]. However, optimality properties (in a decision theoretic context) are not yet well-understood in infinite dimension to the best of our knowledge. This is why posterior modes (or Maximum a Posteriori) are considered instead. Furthermore, they provide a clear link with the classical Tikhonov–Philips regularization [see Dashti et al. (2013); Helin and Burger (2015); Agapiou et al. (2018)] and a useful variational characterization (in case of Gaussian or Besov priors) which will be the cornerstone of the numerical application, see Proposition 3 (which proof is given in the appendix).

Proposition 3 Let μ_0 be the prior probability measure defined in Eq. 2 and μ_{ref} the Gaussian reference measure from Eq. 2.2, then the modes of μ^y are exactly the minimizers of the following (generalized) Onsager-Machlup functional:

$$I(u) := \Phi(u; z) + \frac{1}{2} \|f\|_{\mu_0^f}^2,$$

where $\|\cdot\|_{\mu_0^f}$ is the Cameron–Martin norm associated to μ_0^f .

A minimizer of the previous generalized Onsager-Machlup functional will be noted $u_{MAP} = (\lambda_{MAP}, D_{MAP}, f_{MAP})$ and need not be unique. The precise application of this proposition to the biological setting is done in Sect. 4.1, once the prior distribution is fully specified.

2.5 Approximation

The final step in the theoretical analysis of the Bayesian inverse problem is to study its approximation properties, since it will be solved numerically. There are two important things to check: (a) properties of the approximated posteriors (since it is what is available), (b) the consistency of these approximated posteriors. This will be done by projection of the parameter u onto a finite dimensional subspace of \mathcal{U} , constructed with a stochastic basis under μ_0^f [see Okazaki (1986) for a detailed introduction on stochastic bases including Banach spaces]. The second source of approximation is the use of a numerical solver for the PDE solution, but this will be neglected here (but can be conducted in a subsequent work). The next proposition will establish that the posterior distribution is well approximated, giving an estimate of the error when the Hilbert basis is chosen specifically (spectral basis of the covariance operator, namely Karhunen-Loève decomposition) and its proof is in the appendix.

Proposition 4 Let μ_0 be the prior probability measure from 2, \mathcal{G} the observation operator from Eq. 4, $y \in \mathbb{R}^n$ a dataset, $(f_n)_{n \in \mathbb{N}} \subset \mathcal{C}([0, L] \times [0, T]; \mathbb{R})$ a stochastic basis for μ_f^0 . Note $\forall m \in \mathbb{N}$, $P_m f$ the projection of f onto the span of f_1, \dots, f_m and $(\Phi_m)_{m \in \mathbb{N}}$ the following sequence of approximate negative log-likelihoods:

$$\Phi_m(u; y) := \Phi((\lambda, D, P_m f); y),$$

then the sequence $(\Phi_m)_{m \in \mathbb{N}}$ defines a family of posterior measures $(\mu_m^y)_{m \in \mathbb{N}}$, all continuous w.r.t. y and such that

$$\lim_{m \rightarrow \infty} d_{Hell}(\mu^y, \mu_m^y) = 0.$$

3 Metropolis–Hastings algorithm

As it was previously announced, the main motivation for the Bayesian methodology here is quantification of uncertainty, which will be done by sampling the posterior measure. Among the vast catalogue of methods for probability distributions simulation (Sequential Monte-Carlo, Approximate Bayesian Computations, Transport Maps, etc...), Markov chain Monte-Carlo is very popular [MCMC, see Brooks et al. (2011)] and well defined on function spaces Tierney (1998) even though ergodicity analysis of such algorithms is still in its infancy (Hairer et al. 2005, 2007, 2014; Rudolf and Sprungk 2018). After a short presentation of the Metropolis–Hastings algorithm (Sect. 3.1), this section will focus on a state-of-the-art Markov kernel designed to sample from Gaussian measures (Sect. 3.2) and adapt it to the current non-Gaussian prior using the Gaussian dominating measure μ_{ref} .

3.1 Metropolis–Hastings on function spaces

The Metropolis–Hastings algorithm (MH) is a very general (Tierney 1998) method to design Markov chains to sample from a given probability measure. It is based on a two-step process on each iteration:

1. Given a current state $u \in \mathcal{U}$, propose a new candidate v according to a proposal Markov kernel $Q(u, dv)$ (it is a probability distribution on \mathcal{U} for almost any $u \in \mathcal{U}$),
2. Accept the new state v with probability $\alpha(u, v)$ or remain at u .

This algorithm provides a sample distributed under a predefined probability measure μ , if one selects α and Q in a specific way [see Tierney (1998) for a discussion in general state spaces]. For instance, let $v(du, dv) = \mu(du)Q(u, dv)$ and $v^T(du, dv) = v(dv, du)$, the Metropolis–Hastings algorithm typically considers the following acceptance probability:

$$\alpha_{MH}(u, v) = \min \left(1, \frac{dv^T}{dv}(u, v) \right), \tag{5}$$

which, in particular, requires the absolute continuity of v^T w.r.t. v (detailed balance condition of the Markov chain). Contrary to finite dimensional situations, this condition may be difficult to satisfy and a common way to overcome this situation in Bayesian Inverse problems [see Dashti and Stuart (2015); Girolami and Calderhead (2011); Beskos et al. (2017); Cotter et al. (2013); Hairer et al. (2014)] is to select Q reversible w.r.t. μ_0 . Indeed, in this case (with $v_0(du, dv) = \mu_0(du)Q(u, dv)$):

$$\frac{dv^T}{dv}(u, v) = \frac{\frac{dv^T}{dv_0}(u, v)}{\frac{dv}{dv_0}(u, v)} = \frac{\frac{d\mu^y}{d\mu_0}(v)}{\frac{d\mu^y}{d\mu_0}(u)} = \exp(\Phi(u; y) - \Phi(v; y)). \tag{6}$$

In theory, the MH algorithm can be implemented with a large family of proposal kernels Q . In practice however, they need to be as efficient as possible and thus adapted to the problem at hand. Two common desirable properties for Q are:

- to adjust to locally mimic the target distribution μ^y ,
- include a step size to tune acceptance probability to reasonable values.

These two properties may be used to trade-off self-correlation, acceptance rates and convergence speed to high interest areas of the parameter space. Next section presents a proposal Q with both properties to sample from a Gaussian prior distribution.

3.2 Geometric MCMC under Gaussian reference

The specific Markov proposal kernel Q , tailored to sample distributions having a density w.r.t. a Gaussian measure μ_{ref} will now be presented. Most of the recent work on infinite dimensional MCMC methods is based on the following Langevin stochastic differential equation:

$$\frac{du}{dt} = -\frac{1}{2}K(u) \left(\mathcal{C}_{ref}^{-1}(u - u_{ref}) + \nabla_u \Phi(u; y) \right) + \sqrt{K(u)} \frac{dW}{dt}, \tag{7}$$

where $K(u)$ is a (possibly position-dependent) preconditioner, W a cylindrical Brownian motion and $\nabla_u \Phi(u; y)$ the gradient in u of the negative log-likelihood. According to Beskos et al. (2017), a semi-implicit discretization of Eq. 7 leads to a Markov chain with the following kernel:

$$Q(u, dv) = \mathcal{N} \left(\rho(u - u_{ref}) + u_{ref} + \sqrt{1 - \rho^2} \frac{\sqrt{h}}{2} g(u), \sqrt{1 - \rho^2} K(u) \right), \tag{8}$$

where $h > 0$ is a step-size parameter, $\rho = \frac{1-h}{1+h}$ and:

$$g(u) = -K(u) \left[(\mathcal{C}_{ref}^{-1} - K(u)^{-1})(u - u_{ref}) + \nabla_u \Phi(u; y) \right].$$

This dynamic explores the parameter space with a balance between Newton-type descent to zones of high density and Gaussian exploration. The philosophy behind this

kernel is to use alternative Gaussian reference measures locally adapted to the posterior distribution, since it has been recently showed that higher efficiency is obtained from operator weighted proposals [Law (2014) and later generalized in Beskos et al. (2017) and Cui et al. (2016)]. Indeed, highly informative datasets may result in a posterior measure significantly different from the prior in likelihood-informed directions and non-geometric kernels (such as Independent sampler or preconditioned Crank-Nicholson for instance) become ineffective in this case. However, the infinite dimensional manifold Modified Adjusted Langevin Algorithm [∞ -mMALA from Beskos et al. (2017)] considers a specific preconditioner:

$$K(u) = \left(\mathcal{C}_{ref}^{-1} + H_{\Phi}(u) \right)^{-1},$$

where $H_{\Phi}(u)$ is the Gauss–Newton Hessian operator of Φ , which locally adapts to the posterior. This kernel does not preserve the distribution μ^y but is shown to be absolutely continuous w.r.t. the reference measure μ_{ref} , almost-surely in u (under technical assumptions regarding $K(u)$ linked with Feldman-Hajek theorem) and the Radon-Nikodym density is:

$$\frac{dQ(u, dv)}{d\mu_{ref}}(v) = \frac{dN\left(\frac{\sqrt{h}}{2}g(u), K(u)\right)}{dN(0, \mathcal{C})} \left(\frac{v - \rho(u - u_{ref}) - u_{ref}}{\sqrt{1 - \rho^2}} \right),$$

and noting $w = \frac{v - \rho(u - u_{ref}) - u_{ref}}{\sqrt{1 - \rho^2}}$ as it is done in Beskos et al. (2017), it finally comes:

$$\begin{aligned} \ln \left(\frac{dQ(u, dv)}{d\mu_{ref}}(u, v) \right) &= -\frac{h}{8} |K(u)^{-\frac{1}{2}}g(u)|^2 + \frac{\sqrt{h}}{2} \langle K(u)^{-\frac{1}{2}}g(u), K(u)^{-\frac{1}{2}}w \rangle \\ &\quad - \frac{1}{2} \langle w, H_{\Phi}(u)w \rangle + \ln \left(\left| \mathcal{C}_{ref}^{\frac{1}{2}} K(u)^{-\frac{1}{2}} \right| \right). \end{aligned}$$

Finally, the acceptance probability associated to the Markov kernel from Eq. 8 is

$$\alpha(u, v) = \min \left(1, \frac{\frac{dQ}{d\mu_{ref}}(v, u) \frac{d\mu^y}{d\mu_0}(v)}{\frac{dQ}{d\mu_{ref}}(u, v) \frac{d\mu^y}{d\mu_0}(u)} \right).$$

This algorithm is well-defined on function spaces (reversibility is ensured w.r.t. μ_0), thus it is robust to discretization as required. The ∞ -mMALA proposal may be computationally expensive, as it requires to compute both gradient $\nabla\Phi$, Gauss–Newton Hessian H_{Φ} and the Cholesky decomposition of $K(u)^{-1}$ at each step. However, different dimension reduction techniques can be used [split in Beskos et al. (2017) or likelihood-informed in Cui et al. (2016)] to reduce the computational burden. A second alternative is to choose a constant preconditioner, located at a (precomputed) posterior mode for instance [similar to HMALA in Cui et al. (2016) and gpCN in Rudolf and Sprungk (2018)], which is done in this work.

4 Numerical application

This section will now introduce the practical implementation of the previous methodology on the problem of reverse-engineering for post-transcriptional gap-gene in *Drosophila Melanogaster*. First, the prior distribution is detailed, as well as a random series representation for f , leading to an approximation of f^* . Moreover, the associated generalized Onsager–Machlup functional is also provided. Then, quantitative results are given on the dataset taken from Becker et al. (2013), consisting in protein concentration measurements irregularly spread in space and time.

4.1 Choice of a continuous Gaussian process

In the previous analysis, μ_0^f has been defined as the probability measure over $\mathcal{C}([0, L] \times [0, T]; \mathbb{R})$ associated with a continuous Gaussian process. In practice, it will be chosen centred for simplicity, thus completely specified by a covariance kernel over the space-time domain $[0, L] \times [0, T]$. The literature on such processes is vast [see Rasmussen and Williams (2006) for instance], and here a tensor product of two Brownian bridges (in time and space) will be used. The associated covariance kernel is given by:

$$K((x, t), (x', t')) = \sigma^2 \frac{4}{LT} \left(\min(x, x') - \frac{xx'}{L} \right) \left(\min(t, t') - \frac{tt'}{T} \right), \quad (9)$$

where the scaling constant is such that the process has maximum variance equal to σ^2 . In particular, this process has a well-known Karhunen–Loeve decomposition (see Bay and Croix) using Schauder-type hat functions. In turn, the process will thus be approximated as follows:

$$\begin{aligned} \tilde{f} &= \sum_{1 \leq i_1, i_2 \leq N} \sigma \sqrt{\lambda_{i_1, i_2}} \xi_i \varphi_{i_1, i_2}, \\ \varphi_{i_1, i_2}(x, t) &= \varphi_{i_1} \left(\frac{x}{L} \right) \varphi_{i_2} \left(\frac{t}{T} \right), \quad \forall (x, t) \in [0, L] \times [0, T], \\ \lambda_{i_1, i_2} &= \lambda_{i_1} \lambda_{i_2}, \end{aligned} \quad (10)$$

where $(\xi_{i_1, i_2})_{1 \leq i_1, i_2 \leq N}$ are i.i.d. $\mathcal{N}(0, 1)$ random variables and φ_i are hat functions on dyadic intervals and the precise weights are given in Bay and Croix. We consider $N = 20$ (400 basis functions) thus the approximated parameter $\tilde{u} = (\lambda, D, \tilde{f})$ is of dimension 402.

4.2 Solution map discretization

The analysis conducted in all previous sections happens to be valid for infinite dimensional quantities. In practice however, one needs to discretize for numerical experiments. In this work, the solution map is approximated using finite elements in

space [FEniCS library in Python, see Alnaes et al. (2015) and Langtangen and Logg (2017)] and finite differences in time. We use 100 finite elements and 30 time steps on a 2018 standard laptop computer.¹ We set $L = 100$ and $T = 100$ is the final time (data-points have been rescaled). All quantities related to negative log-likelihood derivatives (Gradient and Gauss–Newton Hessian matrix) are numerically computed using discrete adjoint methods [see Hinze et al. (2009) or Heinkenschloss (2008)] to keep scalability in N . The initial point in the chain is chosen at the MAP location, obtained by minimization of the functional in Eq. 3 (Prior based initialization results in long burnin phase). Practical optimization is done using L-BFGS-B algorithm from the Scipy library (Byrd et al. 1995).

4.3 Estimation of noise variance

So far, the prior measure as well as the observation model include hyper-parameters playing fundamental roles and needing to be tuned. The complete list is λ_M , D_m , D_M , σ_η^2 and to the best of our knowledge, there are no general methods to estimate them efficiently in this context. Indeed, no closed formulae exists for the likelihood (probability density of $\mathcal{G}(u)$) and cross-validation seems computationally out of reach. However, we provide here an empirical approach to tune the noise variance level σ_η^2 . The other parameters are fixed to arbitrary values. Our approach consists in:

1. Fix $\hat{\sigma}_\eta^2 = 1$.
2. Find u_{MAP} minimizing I using the current noise level σ_η^2 .
3. Update the current variance estimation $\hat{\sigma}_\eta^2 = \frac{1}{n-1} \|y - \mathcal{G}(u_{MAP})\|^2$.
4. Go back to 2.

The algorithm is stopped once the parameter σ_η^2 reaches a stable value, which in this application is $\hat{\sigma}_\eta^2 = 20, 50$. The estimator u_{MAP} is taken as the last computed minimizer of the (discretized) Onsager-Machlup functional I .

4.4 Results

We now turn to our main objective, the inversion and uncertainty quantification of gap-gene protein concentration from Becker et al. (2013). The dataset consists of 508 different measures which are non-uniformly spread in time and space (precise repartition can be seen in Fig. 1). With this estimated value, we compute our initial MAP estimate (numerical minimization of the Onsager-Machlup functional) and use it as initial point in the MCMC sampling. The Markov chain is ran for 21,000 total iterations and the resulting traceplot is given in Fig. 4 for negative log-likelihood, decay, diffusion and first three components of f . The first thousand iterations are used as burnin and according to the autocorrelation function (Fig. 3), we choose to keep one iteration out of two hundreds as posterior sample (thinning). From this, we compute both posterior mean and MAP estimates, the precise values of decay, diffusion, negative log-likelihood and Onsager-Machlup functional being given in Table 1.

¹ All codes are available online at https://github.com/JeanCharlesCroix/2019_Bayesian_Melanogaster.

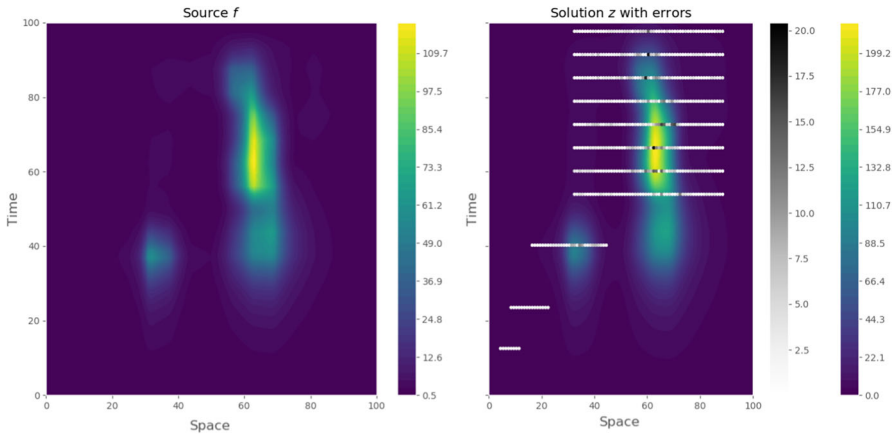


Fig. 1 MAP estimator u_{MAP} obtained from direct optimisation of the Onsager–Machlup functional. Left: estimated source term f_{MAP} (using 400 basis functions). Right: estimated solution $z(u_{MAP})$ with absolute error at data locations. Grey bar represents the level of error between y and $\mathcal{G}(u_{MAP})$

Table 1 Values of decay, diffusion, negative log-likelihood and Onsager–Machlup functional for MAP and posterior mean estimators

Parameter	λ	D	$\Phi(u; z)$	$I(u)$
MAP (optimisation)	0.47	0.62	254.54	416.47
Posterior mean (sampling)	0.47	0.69	253.28	419.29

Additionally to the estimated values, one can also look at the marginal distribution on Fig. 2. These values suggest that there is lack of identifiability between Λ and D (related to the non-injectivity of the forward operator). However, the strength of the Bayesian approach is to quantify this phenomenon, giving here potential values compatible with the data. Concerning the MAP estimator (Fig. 1), we recover both events described in Becker et al. (2013), that is 2 pikes of protein concentrations. The first happens on the anterior part of the embryo in the early experiment ($x = 35, t = 35$). The second is much more intense and happens in the posterior part during the second half of the experiment. The estimated source explains these with an intense and localized increase in concentration. Finally, the uncertainty on both source and solution around data seems to be really low, which provides a good confidence on the level of mRNA at this time and part of the embryo (see Fig. 5). However, the point-wise variance on the solution y remains important before the first observations, which translates into an expected level of uncertainty. This indicates that given the data (and more precisely its time/space repartition), multiple scenarios are valid and our model would require more data to progress in these areas.

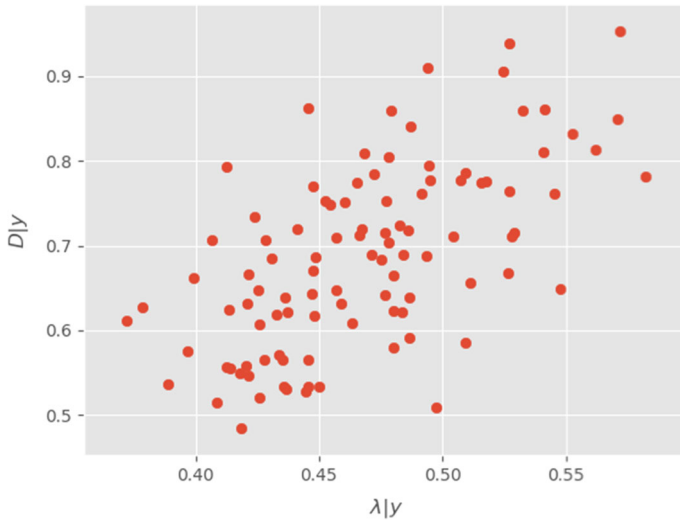


Fig. 2 Samples from the marginal posterior distributions for $\lambda|y$ and $D|y$

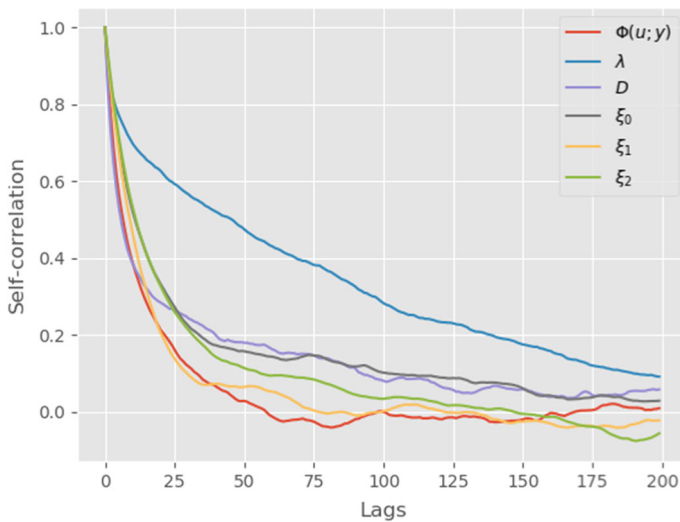


Fig. 3 Self-correlation of $\Phi(u; y)$, λ , D , ξ_0 , ξ_1 and ξ_2 , excluding the first 1000 iterations

5 Conclusions

In this work, we applied the Bayesian inverse problem methodology from Stuart (2010) to a practical biological dynamical system. Doing so, we provide a rigorous and detailed analysis of the forward model, existence and continuity of the posterior measure, characterization of MAP estimates, a consistent approximation and apply a state-of-the-art MCMC methodology. Because the forward MAP is non-linear, the uniqueness of posterior modes is unclear and it appears that local maximas are present.

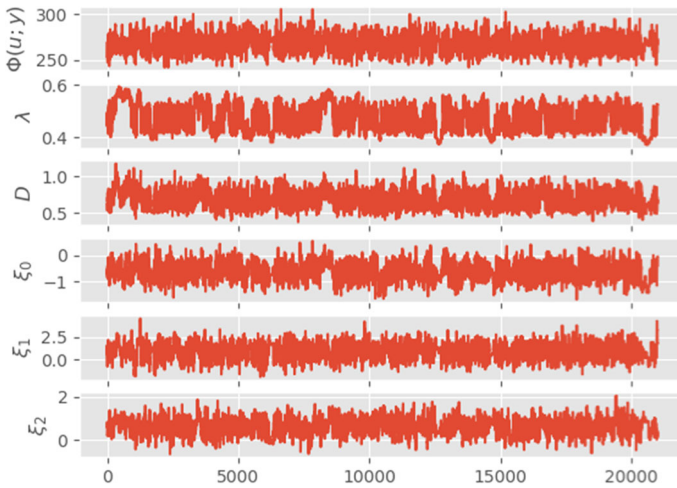


Fig. 4 Trace plots of $\Phi(u; y)$, λ , D , ξ_0 , ξ_1 and ξ_2 (the first 1000 iterations are burned)

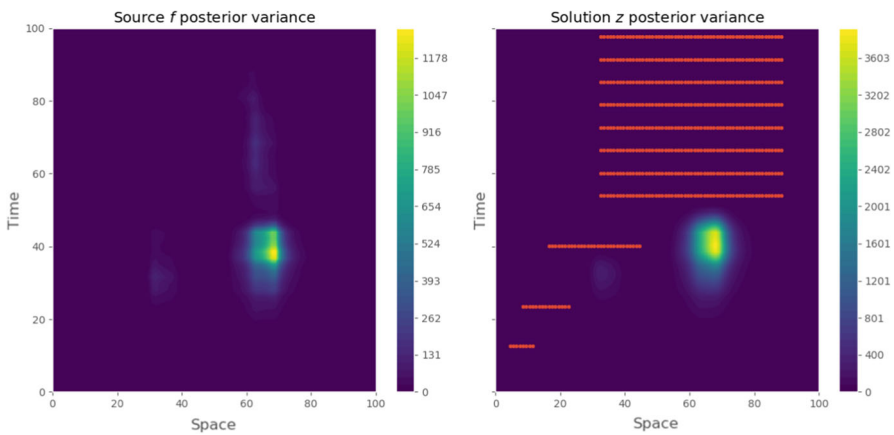


Fig. 5 Posterior point-wise variance of $f^*|y$ (left) and $z(u)|y$ (right). Red dots indicate data locations

Nevertheless, the Bayesian methodology provides both a regularized solution to the problem, while giving a precious quantification of uncertainty. However, the estimation of prior hyper-parameters is still out of reach, giving poor confidence in the estimated variance. This direction requires further research, that we intend to address it in a future work.

Acknowledgements The authors would like to thank Pr. Batton-Hubert, Dr. Bay and Dr. Touboul for their precious comments.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Proofs

Proof of proposition 1 Using standard notations from PDE theory, let $\Omega =]0, L[$, $H = L^2(\Omega)$, $V = H_0^1(\Omega)$, $V^* = H^{-1}(\Omega)$ and

$$W = \left\{ z \in L^2([0, T], V), z' \in L^2([0, T], V^*) \right\},$$

equipped with the norm:

$$\|z\|_W = \left(\|z\|_{L^2([0, T], V)}^2 + \|z'\|_{L^2([0, T], V^*)}^2 \right)^{\frac{1}{2}} \tag{11}$$

1. (Existence, uniqueness and energy estimate) Let $\mathcal{U} = [0, \lambda_M] \times [D_m, D_M] \times \mathcal{C}([0, T] \times [0, L])$ with $\lambda_M > 0$ and $0 < D_m \leq D_M < \infty$, then the differential operator in Eq. 1 is uniformly parabolic. By standard techniques, e.g. Galerkin approximation, one can show the existence and uniqueness of a weak solution $w \in W$ to Eq. 1 [Theorems 3 and 4, section 7.1, Evans (1998)]. Moreover, the regularity of the source term implies an improved regularity for the weak solution, namely $z \in L^2([0, T], H^2(\Omega))$ with $z' \in L^2([0, T], H)$ [Theorem 5, section 7.1, Evans (1998)]. The last result is an energy estimate:

$$\|z\|_{L^2([0, T], H^2(\Omega))} + \|z'\|_{L^2([0, T], H)} \leq C \|f\|_{L^2([0, T] \times [0, L])}. \tag{12}$$

Since D is bounded below by $D_m > 0$, the constant C only on the coercivity property of the differential operator, thus D_m and does not depend on D . This in turn implies the announced energy estimate, by the two continuous embeddings: $L^2([0, T] \times [0, L])$ in $\mathcal{C}([0, T] \times [0, L])$ and z into $\mathcal{C}([0, T] \times [0, L])$ [Theorem 4, section 5.9, Evans (1998)]:

$$\|z\|_\infty \leq C' \|f\|_\infty. \tag{13}$$

2. (Local Lipschitz continuity in u , weaker norm) Let u in the interior of \mathcal{U} , $r > 0$ such that $\mathcal{B}(u, r) \subset \mathcal{U}$ and $(u_1, u_2) \in \mathcal{B}(u, r) \times \mathcal{B}(u, r)$. There exists two unique solutions with respect to u_1 and u_2 such that $\forall i \in \{1, 2\}, \forall v \in L^2([0, T], V)$ and for almost-every t in $[0, T]$:

$$\langle z'_i(t), v(t) \rangle_{V^*, V} + \lambda_i \langle z_i(t), v(t) \rangle_H + D_i \langle z_i(t), v(t) \rangle_V = \langle f_i(t), v(t) \rangle_{V^*, V},$$

which, using subtraction and rearranging the terms, leads to:

$$\begin{aligned} &\langle z'_1(t) - z'_2(t), v(t) \rangle_{V^*,V} + \lambda_1 \langle z_1(t) - z_2(t), v(t) \rangle_H + D_1 \langle z_1(t) - z_2(t), v(t) \rangle_V \\ &= \langle f_1(t) - f_2(t), v(t) \rangle_{V^*,V} + (\lambda_2 - \lambda_1) \langle z_2(t), v(t) \rangle_H \\ &\quad + (D_2 - D_1) \langle z_2(t), v(t) \rangle_V. \end{aligned} \tag{14}$$

Now, let $v = z_1 - z_2$ in Eq. 14, drop the term $\lambda_1 \|z_1 - z_2\|_H^2$ in the left-hand side, use Cauchy–Schwarz and Poincaré’s inequalities in the right-hand side, then:

$$\begin{aligned} &\langle z'_1(t) - z'_2(t), z_1(t) - z_2(t) \rangle_{V^*,V} + D_1 \|z_1(t) - z_2(t)\|_V^2 \\ &\leq (\|f_1 - f_2\|_{V^*} + c|\lambda_2 - \lambda_1| \|z_2\|_V + |D_2 - D_1| \|z_2\|_V) \|z_1 - z_2\|_V, \end{aligned}$$

where $c \geq 0$ is Poincaré’s constant. Integrating both sides between 0 and T , using the identity $\frac{d}{dt} (\frac{1}{2} \|z_1(t) - z_2(t)\|_H^2) = \langle z'_1(t) - z'_2(t), z_1(t) - z_2(t) \rangle_{V^*,V}$ and $z_1(0) = z_2(0) = 0$ gives:

$$\begin{aligned} D_1 \|z_1 - z_2\|_{L^2([0,T],V)}^2 &\leq \int_0^T \|f_1 - f_2\|_{V^*} \|z_1 - z_2\|_V dt \\ &\quad + (c|\lambda_2 - \lambda_1| + |D_2 - D_1|) \int_0^T \|z_2\|_V \|z_1 - z_2\|_V dt. \end{aligned} \tag{15}$$

Now, it remains to use the identity $ab \leq \frac{a^2}{2\epsilon^2} + \frac{\epsilon^2 b^2}{2}$, that is

$$\begin{aligned} D_1 \|z_1 - z_2\|_{L^2([0,T],V)}^2 &\leq \frac{1}{D_1} \|f_1 - f_2\|_{L^2([0,T],V^*)}^2 + \frac{D_1}{4} \|z_1 - z_2\|_{L^2([0,T],V)}^2 \\ &\quad + \frac{(c|\lambda_2 - \lambda_1| + |D_2 - D_1|)^2}{D_1} \|z_2\|_{L^2([0,T],V)}^2 + \frac{D_1}{4} \|z_1 - z_2\|_{L^2([0,T],V)}^2 \end{aligned}$$

which gives

$$\|z_1 - z_2\|_{L^2([0,T],V)}^2 \leq C(u, r) \|u_1 - u_2\|_{\mathscr{W}}^2, \tag{16}$$

with $C(u, r) \geq 0$. Let’s rewrite equality 14 this way:

$$\begin{aligned} &\langle z'_1(t) - z'_2(t), v(t) \rangle_{V^*,V} \\ &= -\lambda_1 \langle z_1(t) - z_2(t), v(t) \rangle_H - D_1 \langle z_1(t) - z_2(t), v(t) \rangle_V \\ &\quad + \langle f_1(t) - f_2(t), v(t) \rangle_{V^*,V} + (\lambda_2 - \lambda_1) \langle z_2(t), v(t) \rangle_H \\ &\quad + (D_2 - D_1) \langle z_2(t), v(t) \rangle_V. \end{aligned} \tag{17}$$

Applying Cauchy–Schwarz and Poincaré’s inequalities in the right-hand side of Eq. 17 gives:

$$\begin{aligned} &\langle z'_1(t) - z'_2(t), v(t) \rangle_{V^*,V} \\ &\leq ((\lambda_1 + D_1) \|z_1(t) - z_2(t)\|_V + \|f_1(t) - f_2(t)\|_{V^*} + c|\lambda_2 - \lambda_1| \\ &\quad - \lambda_1 \|z_2(t)\|_V) \|v(t)\|_V + |D_2 - D_1| \|z_2(t)\|_V \|v(t)\|_V. \end{aligned} \tag{18}$$

It remains to integrate, take the supremum for v in the unit ball of $L^2([0, T], V)$ and use the previous estimate from Eq. 16 to conclude that

$$\|z'_1 - z'_2\|_{L^2([0, T], V^*)}^2 \leq C'(u, r) \|u_1 - u_2\|_{\mathcal{U}}^2$$

where $C'(u, r) \geq 0$.

- (Local Lipschitz continuity in u , stronger norm). The previous analysis will be used to provide a similar result under a stronger norm. Indeed, the solutions are both satisfying $z_i \in L^2([0, T], H^2(\Omega))$ with $z'_i \in L^2([0, T], V)$, $\forall i \in \{1, 2\}$. Consider again Eq. 16, then it is equivalent to:

$$\begin{aligned} & \lambda_1 \langle z_1(t) - z_2(t), v(t) \rangle_H + D_1 \langle z_1(t) - z_2(t), v(t) \rangle_V \\ &= \langle f_1(t) - f_2(t) + z'_2(t) - z'_1(t) + (\lambda_2 - \lambda_1)z_2(t) \\ & \quad - (D_2 - D_1)(z_2)_{xx}(t), v(t) \rangle_H. \end{aligned} \tag{19}$$

Using theorem 5 p. 323 in Evans (1998), one has:

$$\begin{aligned} & \|z_1(t) - z_2(t)\|_{H^2(\Omega)}^2 \\ & \leq C \left(\|z_1(t) - z_2(t)\|_{L^2(\Omega)}^2 + \|f_1(t) - f_2(t) + z'_2(t) - z'_1(t) \right. \\ & \quad \left. + (\lambda_2 - \lambda_1)z_2(t) - (D_2 - D_1)(z_2)_{xx}(t)\|_{L^2(\Omega)} \right), \end{aligned} \tag{20}$$

which in turn, using integration, triangle inequality and the estimates obtained before leads to

$$\|z_1 - z_2\|_{L^2([0, T], H^2(\Omega))} \leq C''(u, r) \|u_1 - u_2\|_{\mathcal{U}}, \tag{21}$$

with $C''(u, r) \geq 0$. Consider one last time the weak formulation from Eq. 14, equivalent to (using integration by parts):

$$\begin{aligned} & \langle z'_1(t) - z'_2(t), v(t) \rangle_H + \lambda_1 \langle z_1(t) - z_2(t), v(t) \rangle_H - D_1 \langle z_1(t) - z_2(t), v(t) \rangle_H \\ &= \langle f_1(t) - f_2(t), v(t) \rangle_H + (\lambda_2 - \lambda_1) \langle z_2(t), v(t) \rangle_H \\ & \quad - (D_2 - D_1) \langle z_2(t), v(t) \rangle_H. \end{aligned} \tag{22}$$

Now, choosing $v(t) = z'_1(t) - z'_2(t)$ gives:

$$\begin{aligned} & \|z'_1(t) - z'_2(t)\|_H^2 + \lambda_1 \frac{d}{dt} \left(\frac{1}{2} \|z_1(t) - z_2(t)\|_H^2 \right) \\ &= \langle f_1(t) - f_2(t), z'_1(t) - z'_2(t) \rangle_H + (\lambda_2 - \lambda_1) \langle z_2(t), z'_1(t) - z'_2(t) \rangle_H \\ & \quad - (D_2 - D_1) \langle (z_2)_{xx}(t), z'_1(t) - z'_2(t) \rangle_H + D_1 \langle (z_1)_{xx}(t) \\ & \quad - (z_2)_{xx}(t), z'_1(t) - z'_2(t) \rangle_H. \end{aligned} \tag{23}$$

Using integration and Cauchy–Schwarz inequality, one obtains

$$\|z'_1(t) - z'_2(t)\|_{L^2([0,T],H)} \leq C'''(u, r) \|u_1 - u_2\|_{\mathcal{U}}, \tag{24}$$

which concludes the proof on the local Lipschitz continuity. Finally, the continuous embedding of $\mathcal{C}([0, L] \times [0, T], \mathbb{R})$ into the more regular space, gives the result.

4. (Fréchet differentiability) Now, we deal with the Fréchet differentiability. Let $u = (\lambda, D, f)$ in the interior of \mathcal{U} , $h^u = (h^\lambda, h^D, h^f) \in \mathcal{U}$ such that $u + h^u \in \mathcal{U}$ and $h^z \in W$, then $\forall v \in L^2([0, T], V)$, we will use the following notation:

$$\begin{aligned} \langle F(z, u), v \rangle &= \int_0^T \langle z'(t), v(t) \rangle_{V^*,V} dt + \lambda \int_0^T \langle z(t), v(t) \rangle_V dt \\ &+ D \int_0^T \langle z(t), v(t) \rangle_H dt - \int_0^T \langle f(t), v(t) \rangle_{V^*,V} dt, \end{aligned} \tag{25}$$

then

$$\langle F(z + h^z, u + h^u) - F(z, u), v \rangle = \langle F'(z, u)[h^z, h^u], v \rangle + c(h^u, h^z),$$

with (using triangle, Poincaré’s and Cauchy-Schwarz inequalities):

$$\begin{aligned} |c(h^u, h^z)| &= \left| h^\lambda \int_0^T \langle h^z(t), v(t) \rangle_H dt + h^D \int_0^T \langle h^z(t), v(t) \rangle_V dt \right| \\ &\leq C \|v\|_{L^2([0,T],V)} \|h^u\|_{\mathcal{U}} \|h^z\|_{L^2([0,T],V)} \\ &\leq C \|v\|_{L^2([0,T],V)} \|h^u\|_{\mathcal{U}} \|h^z\|_W \end{aligned}$$

where $C \geq 0$ is a (new) constant independent from u and:

$$\begin{aligned} \langle F'(z, u)[h^z, h^u], v \rangle &= \int_0^T \langle (h^z)'(t), v(t) \rangle_{V^*,V} dt + h^\lambda \int_0^T \langle z(t), v(t) \rangle_H dt + h^D \int_0^T \langle z(t), v(t) \rangle_V dt \\ &+ \lambda \int_0^T \langle h^z(t), v(t) \rangle_H dt + D \int_0^T \langle h^z(t), v(t) \rangle_V dt - \int_0^T \langle h^f(t), v(t) \rangle_{V^*,V} dt. \end{aligned}$$

Using again Cauchy–Schwarz and Poincaré’s inequalities, we have:

$$|\langle F'(z, u)[h^z, h^u], v \rangle| \leq C \|(h^u, h^z)\| \|v\|_{L^2([0,T],V)}$$

with $C \geq 0$ another constant independent of (h^z, h^u) thus $F'(z, u)$ is linear and bounded, which means that F is Fréchet-differentiable on $W \times \mathcal{U}$. Consider F_z the

partial derivative of F w.r.t. its first variable and $h^z \in W$, then $\forall v \in L^2([0, T], V)$:

$$\begin{aligned} \langle F_z(z, u)h^z, v \rangle &= \int_0^T \langle (h^z)'(t), v(t) \rangle_{V^*, V} dt \\ &\quad + \lambda \int_0^T \langle h^z(t), v(t) \rangle_H dt + D \int_0^T \langle h^z(t), v(t) \rangle_V dt. \end{aligned}$$

This is the weak form associated to the same evolution problem (since F is linear in z), which thus has a unique weak solution $h^z \in W$, thus F_z^{-1} exists. The previous energy estimate gives that F_z^{-1} is bounded. Because F is differentiable and F_z^{-1} exists and is bounded, the implicit function theorem applies and z is differentiable on \mathcal{U} for the norm of W . The second order differentiability is obtained similarly so is the final embedding into the supremum norm. □

Proof of proposition 2 In this proof, the conclusion will follow both theorems 4.3 and 4.5 from Dashti and Stuart (2015), which require to show that the following assumption is satisfied: $\Phi(u; y) \in \mathcal{C}(\mathcal{U} \times \mathbb{R}^n; \mathbb{R})$, there exists two positive functions $M_i : \mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$, monotonic non-decreasing in each variable, with $M_2 > 0$, such that $\forall r > 0, \forall u \in \mathcal{U}, \forall y, y_1, y_2 \in \mathcal{B}(0, r)$:

1. $\Phi(u; y) \geq -M_1(r, \|u\|_{\mathcal{U}})$,
2. $|\Phi(u; y_1) - \Phi(u; y_2)| \leq M_2(r, \|u\|_{\mathcal{U}}) \|y_1 - y_2\|_{\mathbb{R}^n}$.

Here, the the negative log-likelihood is:

$$\Phi(u; y) = \frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|_{\mathbb{R}^n}^2, \tag{26}$$

thus it is clear that $\Phi \in \mathcal{C}(\mathcal{U} \times \mathbb{R}^n; \mathbb{R}^+)$ (since \mathcal{G} is continuous) and furthermore $M_1 = 0$ is a valid choice. To derive a valid function M_2 , first observe that:

$$\forall u \in \mathcal{U}, \|\mathcal{G}(u)\|_{\mathbb{R}^n} \leq n \|z(u)\|_\infty \leq C \exp(\|u\|_{\mathcal{U}}).$$

Consequently, $\forall r > 0$ and $\forall y_1, y_2 \in \mathcal{B}(0, r) \subset \mathbb{R}^n, \forall u \in \mathcal{U}$:

$$\begin{aligned} |\Phi(u; y_1) - \Phi(u; y_2)| &= \frac{1}{2\sigma_\epsilon^2} \left| \|y_1\|_{\mathbb{R}^n}^2 - \|y_2\|_{\mathbb{R}^n}^2 + 2(y_2 - y_1, \mathcal{G}(u))_{\mathbb{R}^n} \right| \\ &= \frac{1}{2\sigma_\epsilon^2} |(\|y_1\|_{\mathbb{R}^n} - \|y_2\|_{\mathbb{R}^n})(\|y_1\|_{\mathbb{R}^n} + \|y_2\|_{\mathbb{R}^n}) \\ &\quad + 2(y_2 - y_1, \mathcal{G}(u))_{\mathbb{R}^n}| \\ &\leq \frac{1}{\sigma_\eta^2} (r + \|\mathcal{G}(u)\|_{\mathbb{R}^n}) \|y_1 - y_2\|_{\mathbb{R}^n}, \\ &\leq \frac{1}{\sigma_\eta^2} (r + C \exp(\|u\|_{\mathcal{U}})) \|y_1 - y_2\|_{\mathbb{R}^n}, \\ &\leq M_2(r, \|u\|_{\mathcal{U}}) \|y_1 - y_2\|_{\mathbb{R}^n}. \end{aligned}$$

It remains to show that:

$$(1 + M_2(r, \|u\|_{\mathcal{U}})^2) \in L^1(\mathcal{U}, \mathcal{B}(\mathcal{U}), \mu_0), \tag{27}$$

which is a direct consequence of Fernique’s and Fubini’s theorems, since μ_0^f is Gaussian. Finally, all conditions from Dashti and Stuart (2015) are verified, leading to existence, uniqueness and continuity of the posterior distribution. \square

Proof of proposition 3 In this work, the prior measure μ_0 is non-Gaussian. However, it might be seen as a small departure from a Gaussian measure of reference, μ_{ref} and this provides a useful variational characterization of modes. Indeed, one has

$$\frac{d\mu^y}{d\mu_{ref}}(u) = \frac{d\mu^y}{d\mu_0}(u) \frac{d\mu_0}{d\mu_{ref}}(u) \propto \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2 - \frac{1}{2} \|f\|_{\mu_0^f}^2\right) \mathbb{1}_{[0, \lambda_{ref}]}(\lambda) \mathbb{1}_{\{D_m, D_M\}}(D). \tag{28}$$

Now, since Φ is locally Lipschitz continuous in u (as a consequence of proposition 1), the analysis from section 4.3 in Dashti and Stuart (2015), initially valid under Gaussian priors, apply as is. \square

Proof of proposition 4 First of all, let us show that the sequence of maps $(\Phi_m)_{m \in \mathbb{N}}$ provides a well-defined sequence of posterior measures (μ_m^y) , all continuous in the data w.r.t. Hellinger metric. The fundamental tool here, is the existence of a Banach space $E \subset \mathcal{C}([0, L] \times [0, T])$, continuously embedded, where $(f_m)_m$ is a Schauder basis and such that $\mu_0^f(E) = 1$ [Theorem 2 in Okazaki (1986)]. This implies that μ_0 -almost surely,

$$\forall m \in \mathbb{N}, \|P_m f\|_\infty \leq C \|P_m f\|_E \leq C' \|f\|_E. \tag{29}$$

In particular, since Fernique’s theorem applies for the norm of E , the analysis from Proposition 2 applies to all measures μ_m . Now, to see that the approximation is consistent, we will apply theorem 4.8 in Dashti and Stuart (2015), which requires to show that

$$|\Phi(u; y) - \Phi_m(u; y)| \leq M(\|u\|_{\mathcal{U}}) \psi(m) \tag{30}$$

where $\lim_{m \rightarrow \infty} \psi(m) = 0$ and $1 + M^2(\|u\|_{\mathcal{U}}) \in L^1(\mu_0)$. Now, noting $u_m = (\lambda, D, P_m f)$ it comes:

$$\|\mathcal{G}(u_m)\|_{\mathbb{R}^n} \leq n \|z(u_m)\|_\infty \leq C \exp(\|P_m f\|_\infty) \leq C_1 \exp(C_2 \|f\|_E). \tag{31}$$

This now gives $\forall y \in \mathcal{B}(0, r)$:

$$\begin{aligned} & |\Phi(u; y) - \Phi_m(u; y)| \\ &= \frac{1}{2\sigma_\eta^2} \left| \|G(u)\|_{\mathbb{R}^n}^2 - \|\mathcal{G}(u_m)\|_{\mathbb{R}^n}^2 + 2\langle y, \mathcal{G}(u_m) - \mathcal{G}(u) \rangle_{\mathbb{R}^n} \right| \\ &= \frac{1}{2\sigma_\eta^2} |(\|\mathcal{G}(u)\|_{\mathbb{R}^n} - \|\mathcal{G}(u_m)\|_{\mathbb{R}^n})(\|\mathcal{G}(u)\|_{\mathbb{R}^n} + \|\mathcal{G}(u_m)\|_{\mathbb{R}^n}) \\ &\quad + 2\langle y, \mathcal{G}(u_m) - \mathcal{G}(u) \rangle_{\mathbb{R}^n}| \\ &\leq \frac{1}{\sigma_\eta^2} (r + C_1 \exp(C_2 \|f\|_E) \|\mathcal{G}(u) - \mathcal{G}(u_m)\|_{\mathbb{R}^n}). \end{aligned}$$

Now, one can use the linearity of z in f^* , thus

$$\|\mathcal{G}(u) - \mathcal{G}(u_m)\|_{\mathbb{R}^n} \leq n \|z(u) - z(u_m)\|_\infty \leq C \|\exp(f) - \exp(P_m f)\|_\infty. \quad (32)$$

It remains to see that

$$\|\exp(f) - \exp(P_m f)\|_\infty \leq C \exp(\|f\|_E) \|f - P_m f\|_E \quad (33)$$

and since $\|f - P_m f\|_E \rightarrow 0$, the result is clear as Fernique's theorem in E gives the required integrability. \square

References

- Agapiou S, Burger M, Dashti M, Helin T (2018) Sparsity-promoting and edge-preserving maximum a posteriori estimators in non-parametric Bayesian inverse problems. *Inverse Prob* 34(4)
- Alnaes MS, Blechta J, Hake J, Johansson A, Kehlet B, Logg A, Richardson C, Ring J, Rognes ME, Wells GN (2015) The FEniCS project version 15. *Arch Numer Softw* 3(100):9–23
- Alvarez MA, Luengo-Garcia D, Lawrence ND (2013) Latent forces models using Gaussian processes. *IEEE Trans Pattern Anal Mach Intell* 35(11):1–20
- Bay X, Croix JC Karhunen-Loève decomposition of Gaussian measures on Banach spaces. In: *Probability and mathematical statistics*
- Becker K, Balsa-Canto E, Cicin-Sain D, Hoermann A, Janssens H, Banga JR, Jaeger J (2013) Reverse-engineering post-transcriptional regulation of gap genes in *Drosophila melanogaster*. *PLoS Comput Biol* 9(10)
- Beskos A, Girolami MA, Lan S, Farrell PE, Stuart AM (2017) Geometric MCMC for infinite-dimensional inverse problems. *J Comput Phys* 335:327–351
- Biegler L, Biros G, Ghattas O, Heinkenschloss M, Keyes D, Mallick B, Marzouk YM, Tenorio L, van Bloemen WB, Willcox KE (2011) *Large-scale inverse problems and quantification of uncertainty*. Wiley, New York
- Bogachev VI (1998) *Gaussian measures*. American Mathematical Society, Providence
- Brezis H (2011) *Functional analysis. Sobolev spaces and partial differential equations*. Springer, Berlin
- Brooks S, Gelman A, Jones G, Meng XL (2011) *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton
- Byrd RH, Lu P, Nocedal J, Zhu C (1995) A limited memory algorithm for bound constrained optimization. *SIAM J Sci Comput* 16(5):1190–1208
- Cotter SL, Roberts GO, Stuart AM, White D (2013) MCMC methods for functions: modifying old algorithms to make them faster. *Stat Sci* 28(3):424–446

- Cui T, Law KJH, Marzouk YM (2016) Dimension-independent likelihood-informed MCMC. *J Comput Phys* 304:109–137
- Dashti M, Stuart AM (2015) The bayesian approach to inverse problems
- Dashti M, Law KJH, Stuart AM, Voss J (2013) MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Prob* 29(9)
- Evans LC (1998) Partial differential equations. American Mathematical Society, Providence
- Ghanem R, Higdon D, Owhadi H (2017) Handbook of uncertainty quantification. Springer, Berlin
- Girolami MA, Calderhead B (2011) Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J R Stat Soc Ser B (Stat Methodol)* 73(2):123–214
- Hadamard J (1902) Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin, Princeton*, pp 49–52
- Hairer M, Stuart AM, Voss J, Wiberg P (2005) Analysis of SPDEs arising in path sampling part I: The Gaussian case. *Commun Math Sci* 3(4):587–603
- Hairer M, Stuart AM, Voss J (2007) Analysis of SPDEs arising in path sampling part II: the nonlinear case. *Ann Appl Probab* 17(5–6):1657–1706
- Hairer M, Stuart AM, Vollmer SJ (2014) Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions. *Ann Appl Probab* 24(6):2455–2490
- Heinkenschloss M (2008) Numerical solution to implicitly constrained optimization problems. Technical report
- Helin T, Burger M (2015) Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Prob* 31(8)
- Hinze M, Pinnau R, Ulbrich M, Ulbrich S (2009) Optimization with PDE constraints. Springer, Berlin
- Hosseini B, Nigam N (2017) Well-posed Bayesian inverse problems: priors with exponential tails. *SIAM/ASA J Uncertain Quant* 5(1):436–465
- Isakov V (2017) Inverse problems for partial differential equations, applied mathematical sciences, vol 127. Springer International Publishing, Cham
- Langtangen HP, Logg A (2017) Solving PDEs in python: the FEniCS tutorial I
- Law KJH (2014) Proposals which speed up function-space MCMC. *J Comput Appl Math* 262:127–138
- Lopez-Lopera AF, Durrande N, Alvarez MA (2019) Physically-inspired Gaussian process models for post-transcriptional regulation in *Drosophila*. *IEEE/ACM Trans Comput Biol Bioinform*
- Okazaki Y (1986) Stochastic basis in Fréchet space 383:379–383
- Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. MIT Press, Cambridge
- Rudolf D, Sprungk B (2018) On a generalization of the preconditioned Crank–Nicolson metropolis algorithm. *Found Comput Math* 18(2):309–343
- Särkkä S, Alvarez MA, Lawrence ND (2018) Gaussian process latent force models for learning and stochastic control of physical systems. *IEEE Trans Autom Control*, pp 1–16
- Schuster T, Kaltenbacher B, Hofmann B, Kazimierski KS (2012) Regularization methods in Banach spaces. De Gruyter, Berlin
- Stuart AM (2010) Inverse problems: a Bayesian perspective. *Acta Numerica* 19:451–459
- Sullivan TJ (2015) Introduction to uncertainty quantification. Springer, Berlin
- Tierney L (1998) A note on Metropolis–Hastings kernels for general state spaces. *Ann Appl Probab* 8(1):1–9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.