



# Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application

Benjamin Favetto, Adeline Samson

► **To cite this version:**

Benjamin Favetto, Adeline Samson. Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application. *Scandinavian Journal of Statistics*, Wiley, 2010, 37 (2), pp.200:220. .

**HAL Id: hal-00324133**

**<https://hal.archives-ouvertes.fr/hal-00324133v2>**

Submitted on 5 Aug 2009

**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.

Parameter estimation for a bidimensional  
partially observed Ornstein-Uhlenbeck process  
with biological application

Running headline : Estimation for partially observed  
Ornstein-Uhlenbeck process

Benjamin Favetto\* and Adeline Samson\*

\* MAP5, University Paris Descartes, France

## Abstract

We consider a bidimensional Ornstein-Uhlenbeck process to describe the tissue microvascularisation in anti-cancer therapy. Data are discrete, partial and noisy observations of this stochastic differential equation (SDE). Our aim is the estimation of the SDE parameters. We use the main advantage of a one-dimensional observation to obtain an easy way to compute the exact likelihood using the Kalman filter recursion. We also propose a recursive computation of the gradient and hessian of the log-likelihood based on Kalman filtering, which allows to implement an easy numerical maximisation of the likelihood and the exact maximum likelihood estimator (MLE). Furthermore, we establish the link between the observations and an ARMA process and we deduce the asymptotic properties of the MLE. We show that this ARMA property can be generalised to a higher dimensional underlying Ornstein-Uhlenbeck diffusion. We compare this estimator with the one obtained by the well-known EM algorithm on simulated data.

**Key Words:** ARMA process, EM algorithm, Hidden Markov Model, Kalman filter, Maximum likelihood estimation, Ornstein-Uhlenbeck process, Partial observations

## 1 Introduction

Stochastic continuous-time models are a useful tool to describe biological or physiological systems based on continuous evolution (see *e.g.* Ditlevsen and De Gaetano (2005), Ditlevsen *et al.* (2005), Picchini *et al.* (2006)). The biological context of this work is the modeling of tissue microvascularisation in anti-cancer therapy. This microcirculation is usually modeled by a bidimensional deterministic differential system which describes the circulation of a contrast agent between two compartments (see Brochot *et al.* (2006) and appendix A). However, this deterministic model is unable to capture the random fluctuations observed along time. In this paper, we consider a stochastic version of this system to take into account random variations around the deterministic solution by adding a Brownian motion on each compartment. This leads to a bidimensional stochastic differential equation (SDE) defined as:

$$\begin{cases} dP(t) &= (\alpha a(t) - (\lambda + \beta)P(t) + (k - \lambda)I(t))dt + \sigma_1 dW_1(t) \\ dI(t) &= (\lambda P(t) - (k - \lambda)I(t))dt + \sigma_2 dW_2(t) \end{cases} \quad (1)$$

where  $P(t)$  and  $I(t)$  represent contrast agent concentrations in each compartment,  $a(t)$  is an input function assumed to be known,  $\alpha$ ,  $\beta$ ,  $\lambda$  and  $k$  are unknown positive parameters,  $W_1$  and  $W_2$  are two independent Brownian motions on  $\mathbb{R}$ , and  $\sigma_1$ ,  $\sigma_2$  are the constant diffusion terms. We assume that  $(P(0), I(0))$  is a random variable independent of  $(W_1, W_2)$ . In our biological context, only the sum  $S(t) = P(t) + I(t)$  can be measured. So (1) is changed into:

$$\begin{cases} dS(t) &= (\alpha a(t) - \beta S(t) + \beta I(t)) dt + \sigma_1 dW_1(t) + \sigma_2 dW_2(t) \\ dI(t) &= (\lambda S(t) - kI(t)) dt + \sigma_2 dW_2(t) \end{cases} \quad (2)$$

Noisy and discrete measurements  $(y_i, i = 0, \dots, n)$  of  $S(t)$  are performed at times  $t_0 = 0 < t_1 < \dots < t_n = T$ . The observation model is thus:

$$y_i = S(t_i) + \sigma \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1)$$

where  $(\varepsilon_i)_{i=0, \dots, n}$  are assumed to be independent and  $\sigma$  is the unknown standard deviation of the Gaussian noise. To evaluate the effect of the treatment on a patient, it is of importance to have a proper estimation of all unknown parameters from this data set. The aim of this paper is to investigate this problem both theoretically and numerically on simulated data.

Parametric inference for discretely observed general SDEs has been widely investigated. Genon-Catalot and Jacod (1993) and Kessler (1997) propose estimators based on minimization of suitable contrasts and study the asymptotic distribution of these estimators when the sampling interval tends to zero as the number of observations tends to infinity. For fixed sampling interval, Bibby and Sørensen (1995) propose martingale estimating functions. In a biological context, Ditlevsen *et al.* (2005) propose an estimation method based on simulation. Picchini *et al.* (2008) propose estimators based on the Hermite expansion of the transition densities. When combining the case of discrete, partial and noisy observations, parameter estimation is a more delicate statistical problem. In this context, it is classical to estimate the unobserved signal (filtering) (see *e.g.* Cappé *et al.* (2005)). However, our aim is the estimation of SDE parameters. In this paper, we use the main advantage of a one-dimensional observation  $y$  and the Gaussian framework of all distributions to obtain an easy way to compute the exact likelihood. For this, we solve and discretize the SDE (2). Then we use the Kalman filter recursion to compute the exact likelihood as proposed by Pedersen (1994) and implemented in the Danish Technical University

project CTSM. We also obtain a recursive computation of the exact gradient and hessian of the log-likelihood based on Kalman filtering, which allows us to implement an easy numerical maximisation of the likelihood using a gradient method and to compute the exact maximum likelihood estimator. The exact observed Fisher information matrix is also directly obtained. As our model is a hidden Markov model, we develop a second approach based on the EM algorithm, which is widely used in this context since the so-called complete likelihood (observed, unobserved) is generally explicit whereas the exact likelihood (observed) is generally not explicit. This method has been first proposed by Shumway and Stoffer (1982) and Segal and Weinstein (1989). Segal and Weinstein (1989) claim that the EM algorithm is computationally more efficient than the Kalman filters. Thus we compare the EM algorithm and the Kalman filter approach in our context.

In Section 2, we study the SDE. We detail in Section 3 the computation of the exact likelihood, the score and hessian functions. We present the EM method in Section 4. In Section 5, we establish the link between the observations and an ARMA process. This allows to deduce the asymptotic properties of the maximum likelihood estimator. Section 6 contains numerical results based on simulated data. This allows to compare the two estimation methods. Appendix A describes briefly the biological background. Appendix B, C, D, E and F contain some proofs and auxiliary results. In particular, the ARMA property can be generalised to a higher dimensional underlying Ornstein-Uhlenbeck diffusion.

## 2 Study of the stochastic differential equation

Introducing  $U(t) = (S(t), I(t))'$  where  $'$  denotes the transposed matrix, (2) can be written in a matrix form:

$$\begin{cases} dU(t) &= (F(t) + G U(t))dt + \Sigma dW(t), U(0) = U_0 \\ y_i &= J U(t_i) + \sigma \varepsilon_i \end{cases}$$

where  $J = (1 \ 0)$  and

$$F(t) = \begin{pmatrix} \alpha a(t) \\ 0 \end{pmatrix}, G = \begin{pmatrix} -\beta & \beta \\ \lambda & -k \end{pmatrix}, dW(t) = \begin{pmatrix} dW_1(t) \\ dW_2(t) \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1 & \sigma_2 \\ 0 & \sigma_2 \end{pmatrix}$$

The process  $(U(t))$  is a bidimensional Ornstein-Uhlenbeck diffusion, which can be explicitly solved. From the biological context (see Appendix A), the parame-

ters satisfy  $\beta, k, \lambda > 0$  and  $\lambda < k$ . This implies that  $G$  is diagonalizable with two distinct negative eigenvalues. Setting  $d = (\beta - k)^2 + 4\beta\lambda > 0$ , the eigenvalues of  $G$  are distinct and equal to:

$$\mu_1 = \frac{-(\beta + k) - \sqrt{d}}{2} \quad \text{and} \quad \mu_2 = \frac{-(\beta + k) + \sqrt{d}}{2}$$

The diagonal matrix  $D$  of eigenvalues and the matrix  $P$  of eigenvectors are:

$$D = \begin{pmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 1 \\ \frac{\beta - k - \sqrt{d}}{2\beta} & \frac{\beta - k + \sqrt{d}}{2\beta} \end{pmatrix} \quad \text{with} \quad D = P^{-1}GP.$$

**Proposition 1** *Let  $X(t) = P^{-1}U(t)$  and  $\Gamma = (\Gamma^{kj})_{1 \leq k, j \leq 2} = P^{-1}\Sigma$ . Then, for  $t, h \geq 0$ , we have:*

$$X(t+h) = e^{Dh}X(t) + B(t, t+h) + Z(t, t+h) \quad (3)$$

where

$$B(t, t+h) = e^{D(t+h)} \int_t^{t+h} e^{-Ds} P^{-1}F(s) ds \quad (4)$$

$$Z(t, t+h) = e^{D(t+h)} \int_t^{t+h} e^{-Ds} \Gamma dW_s. \quad (5)$$

Therefore, the conditional distribution of  $X(t+h)$  given  $X(s), s \leq t$  is

$$\mathcal{N}_2(e^{Dh}X(t) + B(t, t+h), R(t, t+h))$$

where

$$R(t, t+h) = \left( \frac{e^{(\mu_k + \mu_l)h} - 1}{\mu_k + \mu_l} (\Gamma\Gamma')^{kl} \right)_{1 \leq k, l \leq 2} \quad (6)$$

If  $a(t) \equiv c \geq 0$  with  $c$  a constant,  $(X(t))$  has a Gaussian stationary distribution with mean equal to

$$M = -D^{-1}P^{-1}F$$

and covariance matrix equal to

$$V = \left( \frac{1}{-(\mu_k + \mu_l)} (\Gamma\Gamma')^{kl} \right)_{1 \leq k, l \leq 2}$$

*Proof.* See Appendix B.

### 3 Parameter estimation by maximum likelihood

Our aim is to estimate the unknown parameters  $\alpha, \beta, \lambda, k, \sigma_1, \sigma_2$  and  $\sigma$  from observations  $y_{0:n} = (y_0, \dots, y_n)$ . As the law of  $((X(t)), \varepsilon_i, i = 0, \dots, n)$  is Gaussian, the likelihood of  $y_{0:n}$  can be explicitly evaluated. However, the direct maximization of this likelihood requires the inversion of a matrix of dimension  $2(n+1) \times 2(n+1)$  (the covariance matrix of  $(X(t_i))$ ). This inversion can be numerically instable. In this section, we present an alternative method for the computation of the exact likelihood based on Kalman filtering, which does not require any matrix inversion. This is due to the fact that data are one-dimensional. Moreover, it is worth stressing that we need not come back to the initial process  $(U(t))$  for computing the likelihood. Indeed, as  $(U(t))$  is not observed, we can use either  $(U(t))$  or any other transformation of  $(U(t))$  even involving unknown parameters. As  $(X(t))$  is simpler, we consider the following transformed model:

$$\begin{cases} dX(t) &= (DX(t) + P^{-1}F(t))dt + \Gamma dW_t, X(0) = P^{-1}U_0 = X_0 \\ y_i &= JPX(t_i) + \sigma\varepsilon_i \end{cases} \quad (7)$$

Given the particular form of our vector  $J = (1 \ 0)$  and the fact that the eigenvectors can be chosen up to a proportionality constant, we have

$$H = JP = (1 \ 1).$$

It is especially interesting for further computations of the gradient and hessian of the likelihood that  $H$  does not depend of any unknown parameter. From model (7) and (3)-(6), we deduce the following discrete-time evolution system where  $X_i = X(t_i)$ :

$$\begin{cases} X_i &= A_i X_{i-1} + B_i + \eta_i, \quad \eta_i \sim \mathcal{N}(0, R_i) \\ y_i &= HX_i + \sigma\varepsilon_i \end{cases} \quad (8)$$

where  $A_i = \exp(D(t_i - t_{i-1}))$ ,  $B_i = B(t_{i-1}, t_i)$ ,  $R_i = R(t_{i-1}, t_i)$ .

### 3.1 Computation of the exact likelihood

This discrete model is a hidden Markov model (HMM) (Cappé *et al.*, 2005):  $(X_i)$  is a hidden Markov chain on  $\mathbb{R}^2$  and, conditionally on  $(X_i)$ , observations  $(y_i)$  are independent. Genon-Catalot and Laredo (2006) study the maximum likelihood estimator for general HMM. They specialize the exact likelihood in the case where the unobserved Markov chain is a Gaussian one-dimensional AR(1) process. We generalize this computation to the case where the unobserved Markov chain is a bidimensional AR(1) process. Let  $\phi$  denote the vector of unknown parameters and  $y_{0:i} = (y_0, \dots, y_i)$  the vector of observations until time  $t_i$ . By recursive conditioning, it is sufficient to compute the distribution of  $y_i$  given  $y_{0:i-1}$ :

$$L(\phi, y_{0:n}) = p(y_0; \phi) \prod_{i=1}^n p(y_i | y_{0:i-1}; \phi).$$

But the conditional law of  $y_i$  given  $y_{0:i-1}$  can be evaluated by

$$p(y_i | y_{0:i-1}; \phi) = \int p(y_i | X_i; \phi) p(X_i | y_{0:i-1}; \phi) dX_i$$

Then, as the innovation noise  $\eta_i$  of the hidden Markov chain, and the observation noise  $\varepsilon_i$  are Gaussian variables, by elementary computations on Gaussian laws, we are able to get the law of  $y_i$  given  $y_{0:i-1}$  if we know the mean and covariance of the Gaussian conditional law of  $X_i$  given  $y_{0:i-1}$ . This conditional distribution can be exactly computed using Kalman recursions as proposed by Pedersen (1994) and implemented in the Danish Technical University project CTSM. This computation is described below.

#### 3.1.1 Kalman filter

To ease the reading, the parameter  $\phi$  is omitted. The Kalman filter is an iterative procedure which computes recursively the following conditional distributions

$$\begin{aligned} \mathcal{L}(X_i | y_{0:i-1}) &= \mathcal{N}_2(\hat{X}_i^-, P_i^-) \quad (\text{prediction}) \\ \mathcal{L}(X_i | y_{0:i}) &= \mathcal{N}_2(\hat{X}_i, P_i) \quad (\text{filter}) \end{aligned}$$

where

$$\begin{aligned} \hat{X}_i^- &= \mathbb{E}(X_i | y_{0:i-1}) & \text{and} & & P_i^- &= \mathbb{E}((X_i - \hat{X}_i^-)(X_i - \hat{X}_i^-)') \\ \hat{X}_i &= \mathbb{E}(X_i | y_{0:i}) & \text{and} & & P_i &= \mathbb{E}((X_i - \hat{X}_i)(X_i - \hat{X}_i)') \end{aligned}$$



Let us assume that the law of  $X_0$  is Gaussian. Initial values for the algorithm are:

$$X_0 \sim \mathcal{N}(\hat{X}_0^-, P_0^-)$$

with  $\hat{X}_0^- = 0$ ,  $P_0^- = 0$  (from theoretical point of view, one can choose the stationary distribution  $\hat{X}_0^- = M$ ,  $P_0^- = V$  without changes). Next we have the recursive formulae obtained using (8):

$$\begin{aligned} \hat{X}_i^- &= A_i \hat{X}_{i-1}^- + B_i, & P_i^- &= A_i P_{i-1}^- A_i' + R_i, & i \geq 1 \\ \hat{X}_i &= \hat{X}_i^- + K_i(y_i - H \hat{X}_i^-), & P_i &= (I - K_i H) P_i^-, & i \geq 0 \end{aligned} \quad (9)$$

where  $K_i = P_i^- H' (H P_i^- H' + \sigma^2)^{-1}$  (see *e.g.* Cappé *et al.* (2005)).

### 3.1.2 Computation of the exact likelihood of the observations

The conditional distribution of  $y_i$  given  $y_{0:i-1}$  is Gaussian and one-dimensional. Let  $m_i(\phi) = \mathbb{E}_\phi(y_i | y_{0:i-1})$  and  $V_i(\phi) = \text{Var}_\phi(y_i | y_{0:i-1})$  denote its mean and variance which are given using (8) by

$$m_i(\phi) = H \hat{X}_i^-, \quad V_i(\phi) = H P_i^- H' + \sigma^2$$

where  $\hat{X}_i^-$  and  $P_i^-$  depend on  $\phi$ . The exact likelihood of  $y_{0:n}$  is thus equal to

$$L(\phi, y_{0:n}) = \prod_{i=0}^n \frac{1}{\sqrt{2\pi V_i(\phi)}} \exp\left(-\frac{1}{2} \frac{(y_i - m_i(\phi))^2}{V_i(\phi)}\right). \quad (10)$$

Relations (9) imply that there exist two functions  $F_\phi$  and  $G_\phi$  such that

$$m_i(\phi) = F_\phi(m_{i-1}(\phi)), \quad V_i(\phi) = G_\phi(V_{i-1}(\phi)) \quad (11)$$

These iterative relations are used to compute the derivatives of the log-likelihood.

### 3.2 Computation of the maximum likelihood estimator

Pedersen (1994) and the Danish Technical University project CTSM propose to approximate the maximum likelihood estimator (MLE) using a quasi-Newton maximisation method based on the approximation of the gradient and the hessian of the log-likelihood. In our model, we show that it is possible to compute the exact MLE. We use a conjugate gradient method, which relies on the explicit knowledge of the gradient and hessian of the log-likelihood. Both can be exactly

computed using formula (10) and observing that the derivatives of  $m_i(\phi)$  and  $V_i(\phi)$  can be explicitly and recursively computed by derivating formulae (11).

### 3.2.1 New parametrization

In order to simplify the derivatives of (11), from now on, we assume that observation times are equally spaced and set

$$\Delta = t_i - t_{i-1}, \forall i = 1, \dots, n.$$

Hence we have  $A_i = A$ ,  $R_i = R$ . For the sake of simplicity we set  $a(t) \equiv c \geq 0$ , where  $c$  is a known constant, corresponding to an intravenous injection in our biological framework. Hence we have  $B_i = B = -(I-A)D^{-1}P^{-1}F = (I-A)M$ . Set  $Z_i = X_i - M$  and  $m = HM$ . Therefore, model (8) becomes

$$\begin{cases} Z_i &= AZ_{i-1} + \eta_i, \quad \eta_i \sim \mathcal{N}(0, R) \\ y_i &= HZ_i + m + \sigma\varepsilon_i \end{cases} \quad (12)$$

The exact likelihood (10) of  $y_{0:n}$  is thus equal to

$$L(\phi, y_{0:n}) = \prod_{i=0}^n \frac{1}{\sqrt{2\pi V_i(\phi)}} \exp\left(-\frac{1}{2} \frac{(y_i - m - H\hat{Z}_i^-(\phi))^2}{V_i(\phi)}\right). \quad (13)$$

Moreover, instead of  $\phi = (\alpha, \beta, \lambda, k, \sigma_1, \sigma_2, \sigma^2)$ , we propose a new parametrization fitted with the discretization. We consider  $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6)$  where  $\theta_i = e^{\mu_i \Delta}$ ,  $i = 1, 2$  and  $\theta_3, \theta_4$  and  $\theta_5$  are explicit functions of  $\mu_1, \mu_2, \sigma_1, \sigma_2$  and  $\Delta$  such that

$$A = A(\theta) = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix} \quad \text{and} \quad R = R(\theta) = \begin{pmatrix} \theta_3 & \theta_5 \\ \theta_5 & \theta_4 \end{pmatrix}$$

and  $\theta_6 = m$ . We set  $\vartheta = (\theta, \sigma^2)$ . Our aim is to maximize the likelihood  $L(\vartheta, y_{0:n})$  with respect to  $\vartheta$ . Given an estimation  $\hat{\vartheta}$ ,  $\hat{\phi}$  can be obtained by solving numerically the equation  $f(\hat{\phi}) = \hat{\vartheta}$  where  $f$  is the mapping  $\phi \rightarrow f(\phi) = \vartheta$  (see Appendix C for details). Note that, later on, we will see that only six out of the seven parameters can be consistently estimated.

### 3.2.2 Computation of the exact gradient and hessian of the log-likelihood

Let  $W_i(\vartheta) = y_i - \vartheta_6 - H\hat{Z}_i^-(\vartheta)$  and  $l_{0:i}(\vartheta) = \log L(\vartheta, y_{0:i})$ . Using (10), it comes:

$$l_{0:i}(\vartheta) = l_{0:i-1}(\vartheta) - \frac{1}{2} \log(2\pi V_i(\vartheta)) - \frac{1}{2} \frac{W_i(\vartheta)^2}{V_i(\vartheta)}. \quad (14)$$

Thus for  $i = 1, \dots, n$ ,  $q = 1, \dots, 7$ :

$$\frac{\partial l_{0:i}}{\partial \vartheta_q}(\vartheta) = \frac{\partial l_{0:i-1}}{\partial \vartheta_q}(\vartheta) - \frac{1}{2} \frac{1}{V_i(\vartheta)} \frac{\partial V_i}{\partial \vartheta_q}(\vartheta) - \frac{W_i(\vartheta)}{V_i(\vartheta)} \frac{\partial W_i(\vartheta)}{\partial \vartheta_q} + \frac{1}{2} \frac{W_i(\vartheta)^2}{V_i(\vartheta)^2} \frac{\partial V_i(\vartheta)}{\partial \vartheta_q} \quad (15)$$

where

$$\begin{aligned} \frac{\partial V_i(\vartheta)}{\partial \vartheta_q} &= H \frac{\partial P_i^-(\vartheta)}{\partial \vartheta_q} H', 1 \leq q \leq 6, \quad \frac{\partial V_i(\vartheta)}{\partial \sigma^2} = H \frac{\partial P_i^-(\vartheta)}{\partial \sigma^2} H' + 1 \\ \frac{\partial W_i(\vartheta)}{\partial \vartheta_q} &= -\frac{\partial m}{\partial \vartheta_q} - H \frac{\partial \hat{X}_i^-(\vartheta)}{\partial \vartheta_q}, 1 \leq q \leq 7 \end{aligned}$$

Furthermore, the derivatives of  $\hat{X}_i^-(\vartheta)$  and  $P_i^-(\vartheta)$  can be obtained using Kalman recursions (see appendix D). With a more cumbersome computation, second order derivatives of  $l_{0:n}(\vartheta)$  can be analogously deduced from (15). For  $i = 1, \dots, n$ ,  $q, r = 1, \dots, 7$ ,

$$\begin{aligned} \frac{\partial^2 l_{0:i}}{\partial \vartheta_r \partial \vartheta_q}(\vartheta) &= \frac{\partial^2 l_{0:i-1}}{\partial \vartheta_r \partial \vartheta_q}(\vartheta) - \frac{1}{2} \frac{1}{V_i(\vartheta)} \frac{\partial^2 V_i}{\partial \vartheta_r \partial \vartheta_q}(\vartheta) + \frac{1}{2} \frac{1}{V_i^2(\vartheta)} \frac{\partial V_i}{\partial \vartheta_r}(\vartheta) \frac{\partial V_i}{\partial \vartheta_q}(\vartheta) \\ &\quad - \frac{1}{2} \left( 2 \frac{W_i(\vartheta)}{V_i(\vartheta)} \frac{\partial^2 W_i(\vartheta)}{\partial \vartheta_r \partial \vartheta_q} - \frac{W_i(\vartheta)^2}{V_i(\vartheta)^2} \frac{\partial^2 V_i(\vartheta)}{\partial \vartheta_r \partial \vartheta_q} \right) \\ &\quad - \left( \frac{\partial W_i(\vartheta)}{\partial \vartheta_r} \frac{\partial W_j(\vartheta)}{\partial \vartheta_q} \frac{1}{V_i(\vartheta)} - \frac{W_i(\vartheta)}{V_i(\vartheta)^2} \frac{\partial W_i(\vartheta)}{\partial \vartheta_r} \frac{\partial V_i(\vartheta)}{\partial \vartheta_q} \right) \\ &\quad + \left( \frac{\partial W_i(\vartheta)}{\partial \vartheta_q} \frac{\partial V_i(\vartheta)}{\partial \vartheta_r} \frac{W_i(\vartheta)}{V_i(\vartheta)^2} - \frac{W_i(\vartheta)^2}{V_i(\vartheta)^4} \frac{\partial V_i(\vartheta)}{\partial \vartheta_r} V_i(\vartheta) \frac{\partial V_i(\vartheta)}{\partial \vartheta_q} \right) \end{aligned} \quad (16)$$

where (see appendix D for details)

$$\frac{\partial^2 V_i(\vartheta)}{\partial \vartheta_r \partial \vartheta_q} = H \frac{\partial^2 P_i^-(\vartheta)}{\partial \vartheta_r \partial \vartheta_q} H' \quad \text{and} \quad \frac{\partial^2 W_i(\vartheta)}{\partial \vartheta_r \partial \vartheta_q} = -H \frac{\partial^2 \hat{X}_i^-(\vartheta)}{\partial \vartheta_r \partial \vartheta_q}.$$

Hence, we obtain an explicit expression of  $\left( -\frac{\partial^2 l_{0:n}}{\partial \vartheta_r \partial \vartheta_q}(\vartheta) \right)_{1 \leq q, r \leq 7}$ .

### 3.2.3 Maximisation of the exact likelihood

To compute the maximum likelihood estimator, the conjugate gradient algorithm is applied to minimize  $l_{0:n}^-(\vartheta) = -l_{0:n}(\vartheta)$  (see Stoer and Bulirsch (1993)). Let  $\nabla l^-$  denote the gradient of  $l_{0:n}^-$  and  $Hess l^-$  its hessian evaluated by (15)-(16). Starting with an arbitrary initial vector  $\vartheta_0$ , we set as descent direction  $u_0 = \vartheta_0$ . At iteration  $k$ , given  $\vartheta_k$  and  $u_k$ , the parameter and descent direction are updated by

$$\vartheta_{k+1} = \vartheta_k - \frac{\langle u_k, \nabla l^-(\vartheta_k) \rangle}{\langle u_k, Hess l^-(\vartheta_k) u_k \rangle} u_k, \quad u_{k+1} = -\nabla l^-(\vartheta_{k+1}) + \frac{\|\nabla l^-(\vartheta_{k+1})\|}{\|\nabla l^-(\vartheta_k)\|} u_k.$$

Classical stopping conditions are used. The sequence  $(\vartheta_k)_k$  converges towards the maximum of the likelihood  $l_{0:n}(\vartheta)$ .

## 4 Parameter estimation by Expectation Maximization algorithm

An alternative method to estimate  $\vartheta = (\theta, \sigma^2)$  is the Expectation Maximization (EM) algorithm, proposed by Dempster *et al.* (1977), see also Shumway and Stoffer (1982) and Segal and Weinstein (1989). The EM algorithm is a classical approach to estimate parameters of models with non-observed or incomplete data, especially it is widely used for HMMs. In our case, the non-observed data are the  $(Z_i)$ 's, the complete data are the  $(y_i, Z_i)$ 's. The principle is to maximize

$$\vartheta \rightarrow Q(\vartheta | \vartheta_*) = \mathbb{E}(\log p(y_{0:n}, Z_{0:n}; \vartheta) | y_{0:n}; \vartheta_*)$$

with  $Z_{0:n} = (Z_0, \dots, Z_n)$ . This is often easier than the maximization of the observed data log-likelihood since the log-likelihood of the complete data is generally simpler. Moreover, according to Wu (1983), as our model is an exponential family, the EM estimate sequence  $(\vartheta_k)_k$  converges towards a (local) maximum of the data likelihood. The EM algorithm uses two steps: the Expectation step (E-step) and the Maximization step (M-step). Starting with an initial value  $(\vartheta_0)$ , the  $k$ -th iteration is

- E-step: evaluation of  $Q_k(\vartheta) = Q(\vartheta | \vartheta_k)$
- M-step: update of  $\vartheta_k$  by  $\vartheta_{k+1} = \arg \max Q_k(\vartheta)$ .

In our model, function  $Q$  has an explicit expression. Recall that we have assumed that observations times are equally spaced and that  $a(t) \equiv c$ . For simplicity, we also set for the initial variable  $Z_0 = 0$  ( $\hat{Z}_0^- = 0$  and  $P_0^- = 0$ ). The complete data log-likelihood is thus equal to:

$$\begin{aligned} \log p(y_{0:n}, Z_{0:n}; \vartheta) &= -\frac{n+1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=0}^n (y_i - \vartheta_6 - HZ_i)^2 \\ &- \frac{1}{2} \sum_{i=1}^n \log(2\pi|R(\vartheta)|) - \frac{1}{2} \sum_{i=1}^n (Z_i - A(\vartheta)Z_{i-1})' R(\vartheta)^{-1} (Z_i - A(\vartheta)Z_{i-1}). \end{aligned}$$

Function  $Q$  consists in taking the conditional expectation given  $y_{0:n}$  under  $\mathbb{P}_{\vartheta_*}$ . This conditional distribution is the so-called smoothing distribution at  $\vartheta_*$ . In our model, it is Gaussian and characterized by  $M_{i|0:n}(\vartheta_*) = \mathbb{E}_{\vartheta_*}(Z_i|y_{0:n})$  and

$$\Sigma_{i|0:n}(\vartheta_*) = \text{Var}_{\vartheta_*}(Z_i|y_{0:n}), \quad \Sigma_{i-1,i|0:n}(\vartheta_*) = \text{Cov}_{\vartheta_*}(Z_{i-1}, Z_i|y_{0:n})$$

These can be obtained through a forward-backward algorithm (see Appendix E). Thus function  $Q$  is equal to:

$$\begin{aligned} Q(\vartheta|\vartheta_*) &= -\frac{n+1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=0}^n [(y_i - \vartheta_6 - HM_{i|0:n}(\vartheta_*))^2 + H\Sigma_{i|0:n}(\vartheta_*)H'] \\ &- \frac{n}{2} \log(2\pi|R(\vartheta)|) - \frac{1}{2} \text{Tr} \{ R(\vartheta)^{-1} [C(\vartheta_*) - T(\vartheta_*)A'(\vartheta) - A(\vartheta)T'(\vartheta_*) + A(\vartheta)S(\vartheta_*)A'(\vartheta)] \} \end{aligned}$$

where

$$\begin{aligned} T(\vartheta_*) &= \sum_{i=1}^n \left( \Sigma_{i-1,i|0:n}(\vartheta_*) + M_{i|0:n}(\vartheta_*)M'_{i-1|0:n}(\vartheta_*) \right) \\ S(\vartheta_*) &= \sum_{i=1}^n \left( \Sigma_{i-1|0:n}(\vartheta_*) + M_{i-1|0:n}(\vartheta_*)M'_{i-1|0:n}(\vartheta_*) \right) \\ C(\vartheta_*) &= \sum_{i=1}^n \left( \Sigma_{i|0:n}(\vartheta_*) + M_{i|0:n}(\vartheta_*)M'_{i|0:n}(\vartheta_*) \right). \end{aligned}$$

The matrices  $A$ ,  $R$ ,  $\theta_6$  and  $\sigma^2$  are updated as

$$\begin{aligned}
A(\vartheta_k) &= \text{diag}(T(\vartheta_{k-1})S^{-1}(\vartheta_{k-1})) \\
R(\vartheta_k) &= \frac{1}{n}(C(\vartheta_{k-1}) - T(\vartheta_{k-1})S^{-1}(\vartheta_{k-1})T'(\vartheta_{k-1})) \\
\vartheta_{6k} &= \frac{1}{n+1} \sum_{i=0}^n (y_i - HM_{i|0:n}(\vartheta_{k-1})) \\
\sigma_k^2 &= \frac{1}{n+1} \sum_{i=0}^n [(y_i - \vartheta_{6k-1} - HM_{i|0:n}(\vartheta_{k-1}))^2 + H\Sigma_{i|0:n}(\vartheta_{k-1})H']
\end{aligned}$$

## 5 Properties of the exact maximum likelihood estimator in the stationary case

Recall that we have assumed that  $a(t) \equiv c$ . Moreover in this paragraph we assume that the initial variable  $X_0$  has the stationary distribution  $\mathcal{N}_2(M, V)$  given in Proposition 1. This implies that the joint process  $(X_i, y_i)$  is strictly stationary. Let  $(y_i)_{i \in \mathbb{Z}}$  be its extension to a process indexed by  $\mathbb{Z}$ .

### 5.1 Link with an ARMA model

We generalize the result of Genon-Catalot *et al.* (2003) to the bidimensional case and also to the multidimensional case (see Appendix F).

**Proposition 2** *Let  $(\tilde{y}_i) = (y_i - \theta_6)$  define the centered process. The process  $(\tilde{y}_i)_{i \in \mathbb{Z}}$  is centered Gaussian and ARMA(2,2).*

*Proof.* Evidently  $(\tilde{y}_i)$  is centered Gaussian. We easily check that

$$\tilde{y}_i - (\theta_1 + \theta_2)\tilde{y}_{i-1} + \theta_1\theta_2\tilde{y}_{i-2} = \xi_i \quad (17)$$

where  $\xi_i$  is defined by

$$\xi_i = HA(\theta)\eta_{i-1} + H\eta_i + \sigma\varepsilon_i - (\theta_1 + \theta_2)H\eta_{i-1} - (\theta_1 + \theta_2)\sigma\varepsilon_{i-1} + \theta_1\theta_2\sigma\varepsilon_{i-2}$$

As the  $(\eta_i)_i$  and  $(\varepsilon_i)_i$  are mutually independent, we get that:

$$\text{Cov}(\xi_i, \xi_{i+k}) = 0, \quad \forall k \geq 3.$$

This implies that  $(\xi_i)$  is MA(2). Hence the result.  $\square$

**Proposition 3** *The spectral density  $f(u, \vartheta)$  of  $(\tilde{y}_i)$  has the explicit form:*

$$f(u, \vartheta) = \sigma^2 + \frac{H(ARA' + (1 + (\theta_1 + \theta_2)^2)R)H' + 2 \cos(u)(HA - (\theta_1 + \theta_2)H)RH'}{1 + (\theta_1 + \theta_2)^2 + \theta_1^2\theta_2^2 - 2(\theta_1 + \theta_2)(1 + \theta_1\theta_2) \cos(u) + 2 \cos(2u)\theta_1\theta_2} \quad (18)$$

with  $A = A(\theta)$ ,  $R = R(\theta)$ .

*Proof.* Let  $\gamma(k) = Cov(\xi_i, \xi_{i+k})$ . Elementary computations show that

$$\begin{aligned} \gamma(0) &= HARA'H' + HRH'(1 + (\theta_1 + \theta_2)^2) + \sigma^2(1 + (\theta_1 + \theta_2)^2 + \theta_1^2\theta_2^2) \\ \gamma(1) &= (HA - (\theta_1 + \theta_2)H)RH' - \sigma^2(\theta_1 + \theta_2)(1 + \theta_1\theta_2) \\ \gamma(2) &= \sigma^2\theta_1\theta_2 \\ \gamma(k) &= 0 \quad \forall k \geq 3 \end{aligned}$$

The spectral density (with respect to  $\frac{du}{2\pi}$ )  $h(u, \vartheta)$  of  $(\xi_i)$  is

$$h(u, \vartheta) = \sum_{n \in \mathbb{Z}} \gamma(n) \exp(-inu) = \gamma(0) + \gamma(1)2 \cos(u) + \gamma(2)2 \cos(2u).$$

For the AR(2) part, we set:  $p(x) = 1 - (\theta_1 + \theta_2)x + \theta_1\theta_2x^2$  (recall that  $\theta_1, \theta_2 < 1$ ).

Then

$$\begin{aligned} f(u, \vartheta) &= \frac{h(u, \vartheta)}{|p(\exp(-iu))|^2} \\ &= \frac{\gamma(0) + \gamma(1)2 \cos(u) + \gamma(2)2 \cos(2u)}{1 + (\theta_1 + \theta_2)^2 + \theta_1^2\theta_2^2 - 2(\theta_1 + \theta_2)(1 + \theta_1\theta_2) \cos(u) + 2 \cos(2u)\theta_1\theta_2} \end{aligned}$$

See Brockwell and Davis (1991) for technical details.  $\square$

The number of parameters which are identifiable on the spectral density is precised by the following proposition

**Proposition 4** *The identifiable quantities are  $\sigma^2$ ,  $\theta_1 + \theta_2$  and  $\theta_1\theta_2$ , and at most two out of three parameters among  $\theta_3, \theta_4$  and  $\theta_5$ .*

*When  $\Delta$  is small, exactly two out of the three paramaters  $\theta_3, \theta_4$  and  $\theta_5$  are identifiable.*

*Proof.* See Appendix G.

## 5.2 Asymptotic properties of maximum likelihood estimator

We now deduce the asymptotic properties of the maximum likelihood estimators when  $(y_i)$  is stationary.

We denote by  $\vartheta^0$  the true value of parameters vector  $(\theta_1, \dots, \theta_6, \sigma^2)$ . We assume that the parameter set  $\Theta$  is an open subset of  $\mathbb{R}^7$ . We denote by  $\vartheta_- = (\theta_1, \dots, \theta_5, \sigma^2) \in \Theta_-$  the projection of  $\vartheta$  on  $\mathbb{R}^6$ .

The following result is classical to estimate the parameter  $\vartheta_6^0$  (see *e.g.* in Brockwell and Davis (1991)).

**Proposition 5** (Mean estimator) *Let  $\bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i$  be the empirical mean. Under the assumption of stationarity of  $(y_i)$ ,  $\bar{y}_n \rightarrow \theta_6^0$  a.s. as  $n \rightarrow \infty$ . Moreover,  $\sqrt{n}(\bar{y}_n - \theta_6^0)$  converges in distribution:*

$$\sqrt{n}(\bar{y}_n - \theta_6^0) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, J(\vartheta_-^0))$$

$$\text{where } J(\vartheta_-^0) = f(0, \vartheta_-^0) = \frac{\gamma(0)+2\gamma(1)+2\gamma(2)}{((1-\theta_1)(1-\theta_2))^2}$$

We now consider the centered process  $(\tilde{y}_i)$ . Consider the two assumptions (which can be checked up to some technicities)

A1  $(u, \vartheta_-) \mapsto f(u, \vartheta_-)$  is a  $\mathcal{C}^3$ -function on a neighborhood of  $[-\pi, \pi] \times \Theta_-$

A2  $\vartheta_- \mapsto f(\cdot, \vartheta_-)$  is one to one

As  $(\tilde{y}_i)$  is a ARMA(2,2) process, its spectral density is positive for every  $(u, \vartheta_-) \in [-\pi, \pi] \times \Theta_-$ .

**Proposition 6** (Information matrix) *Let  $\tilde{l}_{0:n}(\vartheta_-) = \log L(\vartheta_-, \tilde{y}_{0:n})$  be the log-likelihood of the centered process  $(\tilde{y}_{0:n})$ . Under the assumption A1, we have  $\mathbb{P}_{\vartheta_-^0}$ -a.s.*

$$\lim_{n \rightarrow \infty} \left( -\frac{1}{n} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \tilde{l}_{0:n}(\vartheta_-^0) \right)_{1 \leq i, j \leq 6} = I(\vartheta_-^0) \quad (19)$$

**Proposition 7** (Consistency and asymptotic normality of the MLE)

*Let  $\hat{\theta}_n$  be a maximum likelihood estimator of  $\vartheta_-^0$  based on  $\tilde{y}_{0:n}$ . Under the assumptions A1 and A2,  $\hat{\theta}_n \rightarrow \vartheta_-^0$  a.s. as  $n \rightarrow \infty$ . Moreover, if  $I(\vartheta_-^0)$  is invertible,  $\sqrt{n}(\hat{\theta}_n - \vartheta_-^0)$  converges in distribution:*

$$\sqrt{n}(\hat{\theta}_n - \vartheta_-^0) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, I^{-1}(\vartheta_-^0))$$



*Proof.* This result may be found *e.g.* in Brockwell and Davis (1991).  $\square$

**Remark 1** *As it is done usually, for further numerical considerations, the empirical mean estimator  $\bar{y}_n$  is plugged in the likelihood for parameter  $\theta_6$  in the Kalman-recursion approach. The EM algorithm estimates this parameter during the algorithm iterations.*

## 6 Simulation study

We compare the performances of the two estimation methods on simulated data sets. The exact maximum likelihood estimators and the EM estimators are computed as described in Section 3 and Section 4, respectively. Data are simulated using equally spaced observation times ( $\Delta = 0.2$ ) and  $n = 200$  or  $n = 1000$  observations. Values of parameter  $\theta$  are deduced from values of biological parameters  $(\alpha, \beta, \lambda, k)$  estimated on real data in Thomassin (2008), and  $\sigma_1^2 = 0.5$ ,  $\sigma_2^2 = 0.125$ ,  $c = 50$ :

$$\theta_1 = 0.6, \theta_2 = 0.9, \theta_3 = 0.7, \theta_4 = 0.2, \theta_5 = 0.1, \theta_6 = 20$$

Two levels of observation noise are used:  $\sigma^2 = 1$  or  $\sigma^2 = 3$ . Thousand replications are performed for each design ( $n = 200$  or  $n = 1000$  observations,  $\sigma^2 = 1$  or  $\sigma^2 = 3$ ). The influence of the time scale  $\Delta$  is evaluated on simulated data with  $\Delta = 0.04$  and  $n = 1000$  observations with parameter values equal to

$$\theta_1 = 0.91, \theta_2 = 0.99, \theta_3 = 0.2, \theta_4 = 0.03, \theta_5 = 0.01, \theta_6 = 20$$

Thousand replications are performed for each observation noise ( $\sigma^2 = 1$  or  $\sigma^2 = 3$ ) in this case. Mean estimates and their empirical standard errors are computed on the 1000 replications of each design. The exact standard errors obtained from the asymptotic information matrix (computed by Kalman-based recursions) are also provided.

Identifiability results of Section 5 show that parameters  $\theta_6$  and  $\sigma^2$  are identifiable, and that at most 4 among the five parameters  $(\theta_1, \dots, \theta_5)$  are identifiable. In our biological context, it is reasonable to assume that  $\sigma^2$  is known. Therefore  $\sigma^2$  is fixed to its true value. One parameter among  $(\theta_1, \dots, \theta_5)$  has to be fixed: we choose to fix  $\theta_5$  to its true value and to estimate  $\theta_1, \theta_2, \theta_3, \theta_4$ . For the MLE estimation, parameter  $\theta_6$  is previously estimated by the empirical mean. Then the other parameters are estimated based on the empirically centered observations.

Results are given in Table 1 for  $\Delta = 0.2$  and Table 2 for  $\Delta = 0.2$  and  $\Delta = 0.04$ . Results obtained on one simulated data set ( $n = 200$ ,  $\Delta = 0.2$ ,  $\sigma^2 = 3$ ) are plotted in Figure 1 and show the ability of the method to estimate the trajectories  $(S(t), P(t), I(t))$ . The EM algorithm is around three times quicker than the MLE algorithm. The mean computation time to estimate parameters of a data set with  $n = 200$  observations is around 4 seconds (CPU time) for the EM algorithm and around 13 seconds (CPU time) for the MLE for the same precision of convergence, on an Apple MacPro  $2 \times 3$  Ghz with 5 Go of RAM. The Matlab code are given at <http://www.mi.parisdescartes.fr/~favetto>.

[Figure 1 about here.]

The EM estimates are often less biased for parameters  $\theta_1$  and  $\theta_2$  than the MLE estimates. Variance parameters  $(\theta_3, \theta_4)$  are estimated with less bias by the MLE method,  $\theta_4$  is always estimated with a large bias by the EM algorithm. The standard errors of the EM estimates are lower than those of the MLE estimates. The standard errors reduce with the increase of the number of observations  $n$ . The bias and standard errors of all parameters decrease with the decrease of the observation noise  $\sigma$ . When  $\Delta$  decreases, bias and standard errors decrease for a small observation noise  $\sigma = 1$ . MLE estimates are very satisfactory in this case. The exact Fisher information matrix provides standard errors of the estimates which are close to the empirical ones, especially those obtained with the MLE approach. Note that the theoretical study of the exact MLE allows to deduce the identifiable parameters. On the contrary, the EM algorithm misses completely the problem.

[Table 1 about here.]

[Table 2 about here.]

## 7 Conclusion

The Kalman filter is classical in the field of noisy, discretely and partially observed stochastic differential equations. In this paper, we have shown that it can be used for the estimation of the parameters by maximum likelihood. In the particular case of an Ornstein-Uhlenbeck process, this method computes the exact likelihood, its gradient and hessian. We have also shown that the EM algorithm, which is classical in the field of hidden Markov models, combined with a smoother algorithm can be used for the parameter estimation.

We study some theoretical properties of the model. We show that only six out of the seven parameters are identifiable and we deduce the asymptotic properties of the maximum likelihood estimate. We illustrate the two methods on simulated data. The identifiability problem is confirmed on the simulation study: the observed Fisher information matrix computed by the Kalman method is not invertible when we estimate the seven parameters.

The next step of this work is its application to real data in anti-cancer therapy. This work could also be extended to the case of non-Gaussian observation errors. For a unidimensional Markov chain  $(X_i)$  observed with non Gaussian errors, Ruiz (1994) proposes a quasi-maximum likelihood estimator based on the Kalman filter and shows the normality asymptotic distribution of this estimator. This approach can be extended to our bidimensional model.

## Acknowledgments

The authors thank V. Genon-Catalot for her constructive advices and help. The authors thank C.A. Cuenod, D. Balvay and Y. Rozenholc for their helpful discussions on the biological problem. This project was supported by a grant BQR from University Paris Descartes managed by Y. Rozenholc.

## References

- Bibby, B. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1**, 17–39.
- Brochot, C., Bessoud, B., Balvay, D., Cuenod, C., Siauve, N. and Bois, F. (2006). Evaluation of antiangiogenic treatment effects on tumors' microcirculation by bayesian physiological pharmacokinetic modeling and magnetic resonance imaging. *Magn Reson Imaging*. **24**, 1059–1067.
- Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods*. Springer.
- Cappé, O., Moulines, E. and Ryden, T. (2005). *Inference in Hidden Markov Models (Springer Series in Statistics)*. Springer-Verlag New York, USA.
- Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Jr. R. Stat. Soc. B* **39**, 1–38.

- Ditlevsen, S. and De Gaetano, A. (2005). Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bull. Math. Biol.* **67**, 547–561.
- Ditlevsen, S., Yip, K. and Holstein-Rathlou, N. (2005). Parameter estimation in a stochastic model of the tubuloglomerular feedback mechanism in a rat nephron. *Math Biosc* **194**, 49–69.
- Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.* **29**, 119–151.
- Genon-Catalot, V., Jeantheau, T. and Laredo, C. (2003). Conditional likelihood estimators for hidden Markov models and stochastic volatility models. *Scand. J. Statist.* **30**, 297–316.
- Genon-Catalot, V. and Laredo, C. (2006). Leroux’s method for general hidden Markov models. *Stochastic Process. Appl.* **116**, 222–243.
- Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211–229.
- Pedersen, A. (1994). Statistical analysis of gaussian diffusion processes based on incomplete discrete observations. *Research Report, Department of Theoretical Statistics, University of Aarhus* **297**.
- Picchini, U., Ditlevsen, S. and De Gaetano, A. (2006). Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations. *J. Math. Biol.* **53**, 771–796.
- Picchini, U., Ditlevsen, S. and De Gaetano, A. (2008). Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Mathematical Medicine and Biology* **25**, 141–155.
- Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of econometrics* **63**, 289–306.
- Segal, M. and Weinstein, E. (1989). A new method for evaluating the log-likelihood gradient, the hessian, and the fisher information matrix for linear dynamic systems. *IEEE Trans. Inf. Theory* **35**, 682–687.
- Shumway, R. and Stoffer, D. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Series Anal* **3**, 253–264.

Stoer, J. and Bulirsch, R. (1993). *Introduction to numerical analysis*, vol. 12 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edn.

Thomassin, I. (2008). Etude des tumeurs annexielles du pelvis feminin en IRM fonctionnelle, mise au point des techniques et applications cliniques. *Physical PhD, Ecole doctorale Sciences et Technologies de l'Information des Télécommunication et des Systèmes*.

Wu, C. (1983). On the convergence property of the EM algorithm. *Ann. Statist.* **11**, 95–103.

The corresponding author for negotiations concerning the manuscript is Adeline Samson (adeline.samson@parisdescartes.fr)

Laboratoire MAP5

Universite Paris Descartes

45 rue des St Peres

75006 Paris

France

## A Physiological model

We focus on the evaluation of anti-angiogenesis treatments in anti-cancer therapy. These treatments take effect on the vascularization of tissue. The *in vivo* evaluation of their efficacy is based on the estimation of the tissue microvascularization parameters. The experiment consists in the injection of a contrast agent to the patient, followed by the recording of a medical images sequence which measures the evolution of the concentration of contrast agent along time. The contrast agent pharmacokinetic is modeled by a bidimensional differential system. The contrast agent pulsates in the plasma and interstitium cells. Let  $a(t)$ ,  $P(t)$  and  $I(t)$  denote respectively the quantity of contrast agent at time  $t$  in the artery, the plasma and the interstitium and  $1 - h$ ,  $V_P$  and  $V_I$  the volume of artery, plasma and interstitium ( $h$  is the hematocrit rate). The initial condition at time  $t_0 = 0$  is  $P(0) = 0, I(0) = 0$ . The contrast agent is injected in vein at time  $t_0$ , transits in the artery and arrives in plasma, with a tissue perfusion flow equal to  $F_{tp}$ . The contrast agent is eliminated from plasma with the perfusion flow  $F_{tp}$ , proportionally to the concentration of contrast agent in plasma. The quantity of contrast agent exchanging from plasma through interstitium is equal to  $K_{trans}$  times the concentration of contrast agent in plasma, where  $K_{trans}$  is

the volume transfer constant. Inversely, the quantity of contrast agent exchanging from interstitium through plasma is equal to  $K_{trans}$  times the concentration of contrast agent in interstitium. Lastly, the two-compartment model is:

$$\begin{cases} \frac{dP(t)}{dt} &= \frac{F_{tp}}{1-h}a(t) - \frac{K_{trans}}{V_P}P(t) + \frac{K_{trans}}{V_I}I(t) - \frac{F_{tp}}{V_P}P(t) \\ \frac{dI(t)}{dt} &= \frac{K_{trans}}{V_P}P(t) - \frac{K_{trans}}{V_I}I(t) \end{cases} \quad (20)$$

For statistical accommodations, we use a new parameterization and set:

$$\alpha = \frac{F_{tp}}{1-h}, \beta = \frac{F_{tp}}{V_P}, \lambda = \frac{K_{trans}}{V_P}, k = \frac{K_{trans}}{V_P} + \frac{K_{trans}}{V_I}$$

Model (20) can thus be transformed as follows:

$$\begin{cases} \frac{dP(t)}{dt} &= \alpha a(t) - \lambda P(t) + (k - \lambda)I(t) - \beta P(t) \\ \frac{dI(t)}{dt} &= \lambda P(t) - (k - \lambda)I(t) \end{cases} \quad (21)$$

## B Proof of Proposition 1

The process  $X(t) = P^{-1}U(t)$  is solution of:

$$dX(t) = (DX(t) + P^{-1}F(t))dt + P^{-1}\Sigma dW_t, \quad X(0) = X_0 = P^{-1}U_0.$$

Applying Ito's formula, we obtain

$$X(t) = e^{Dt}X_0 + e^{Dt} \int_0^t e^{-Ds} P^{-1}F(s)ds + e^{Dt} \int_0^t e^{-Ds} \Gamma dW_s.$$

From this equation, we deduce:

$$X(t+h) = e^{Dh}X(t) + B(t, t+h) + Z(t, t+h)$$

where  $B(t, t+h)$  and  $Z(t, t+h)$  are given in Proposition 1. Using that  $W_1, W_2$  are independent and that  $X_0$  is independent of  $(W_1, W_2)$ , we obtain the conditional law of  $X(t+h)|(X(s), s \leq t)$ .

The stationary distribution can be deduced from equation (3) with  $a(t) = c$ . As the two elements of  $D$  are negative, we have

$$\begin{aligned} \lim_{t \rightarrow +\infty} \mathbb{E}(X(t)) &= \lim_{t \rightarrow +\infty} e^{Dt} \mathbb{E}(X_0) + \lim_{t \rightarrow +\infty} B(0, t) = -D^{-1}P^{-1}F = M \\ \lim_{t \rightarrow +\infty} Var(X(t)) &= \lim_{t \rightarrow +\infty} R(0, t) = \left( \frac{1}{-(\mu_k + \mu_{k'})} (\Gamma \Gamma')^{kk'} \right)_{1 \leq k, k' \leq 2} = V. \end{aligned}$$

If  $X_0 \sim \mathcal{N}_2(M, V)$ , an elementary computation shows that  $(X(t))$  is strictly stationary.  $\square$

## C Link between the two parametrisations

New parameters  $(\theta_1, \dots, \theta_5)$  are given as functions of initial parameters  $(\beta, \lambda, k, \sigma_1, \sigma_2)$  in Section 3.2.1. Now we deduce  $(\beta, \lambda, k, \sigma_1, \sigma_2)$  from  $(\theta_1, \dots, \theta_5)$ . From the definition of  $\mu_1$  and  $\mu_2$  (Section 2), we have

$$\mu_2 - \mu_1 = \sqrt{d}, \quad \mu_1 + \mu_2 = -(\beta + k), \quad \mu_1\mu_2 = \beta(k - \lambda)$$

Thus we rewrite the matrix  $P$  and its inverse

$$P = \begin{pmatrix} 1 & 1 \\ \frac{\mu_1}{\beta} + 1 & \frac{\mu_2}{\beta} + 1 \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} \frac{\mu_2 + \beta}{\mu_2 - \mu_1} & \frac{-\beta}{\mu_2 - \mu_1} \\ -\frac{\mu_1 + \beta}{\mu_2 - \mu_1} & \frac{\beta}{\mu_2 - \mu_1} \end{pmatrix}.$$

Then the covariance matrix  $\Gamma$  is

$$\Gamma = \begin{pmatrix} \sigma_1 \frac{\mu_2 + \beta}{\mu_2 - \mu_1} & \sigma_2 \frac{\mu_2}{\mu_2 - \mu_1} \\ -\sigma_1 \frac{\mu_1 + \beta}{\mu_2 - \mu_1} & -\sigma_2 \frac{\mu_1}{\mu_2 - \mu_1} \end{pmatrix}$$

and finally comes the matrix  $R$

$$R = \begin{pmatrix} \frac{e^{2\mu_1\Delta} - 1}{2\mu_1} (\sigma_1^2 (\frac{\mu_2 + \beta}{\mu_2 - \mu_1})^2 + \sigma_2^2 (\frac{\mu_2}{\mu_2 - \mu_1})^2) & \frac{e^{(\mu_1 + \mu_2)\Delta} - 1}{\mu_1 + \mu_2} (-\sigma_1^2 \frac{\mu_1 + \beta}{\mu_2 - \mu_1} \frac{\mu_2 + \beta}{\mu_2 - \mu_1} - \sigma_2^2 \frac{\mu_1\mu_2}{\mu_2 - \mu_1}) \\ \frac{e^{(\mu_1 + \mu_2)\Delta} - 1}{\mu_1 + \mu_2} (-\sigma_1^2 \frac{\mu_1 + \beta}{\mu_2 - \mu_1} \frac{\mu_2 + \beta}{\mu_2 - \mu_1} - \sigma_2^2 \frac{\mu_1\mu_2}{\mu_2 - \mu_1}) & \frac{e^{2\mu_2\Delta} - 1}{2\mu_2} (\sigma_1^2 (\frac{\mu_1 + \beta}{\mu_2 - \mu_1})^2 + \sigma_2^2 (\frac{\mu_1}{\mu_2 - \mu_1})^2) \end{pmatrix}$$

Define

$$\tilde{\theta}_3 = \sigma_1^2 (\mu_2 + \beta)^2 + \sigma_2^2 \mu_2^2, \quad \tilde{\theta}_4 = \sigma_1^2 (\mu_1 + \beta)^2 + \sigma_2^2 \mu_1^2, \quad \tilde{\theta}_5 = -\sigma_1^2 (\mu_1 + \beta)(\mu_2 + \beta) - \sigma_2^2 \mu_1 \mu_2$$

we have

$$\tilde{\theta}_3 = \theta_3 \frac{2\mu_1(\mu_2 - \mu_1)^2}{\exp(2\mu_1\Delta) - 1}, \quad \tilde{\theta}_4 = \theta_4 \frac{2\mu_2(\mu_2 - \mu_1)^2}{\exp(2\mu_2\Delta) - 1}, \quad \tilde{\theta}_5 = \theta_5 \frac{(\mu_1 + \mu_2)(\mu_2 - \mu_1)^2}{\exp((\mu_1 + \mu_2)\Delta) - 1}.$$

Notice that  $\mu_1 = \log(\theta_1)/\Delta$  and  $\mu_2 = \log(\theta_2)/\Delta$ . The parameters  $\beta, \lambda, k, \sigma_1^2$

and  $\sigma_2^2$  are solution of the system

$$\left\{ \begin{array}{lcl} k & = & -(\mu_1 + \mu_2 + \beta), \\ \lambda & = & -\left(\frac{\mu_1\mu_2}{\beta} + k\right), \\ (\mu_2 + \beta)^2\sigma_1^2 + \mu_2^2\sigma_2^2 & = & \tilde{\theta}_3, \\ (\mu_1 + \beta)^2\sigma_1^2 + \mu_1^2\sigma_2^2 & = & \tilde{\theta}_4, \\ \sigma_1\beta^2 + \sigma_1(\mu_1 + \mu_2)\beta + \tilde{\theta}_5 + (\sigma_1^2 + \sigma_2^2)\mu_1\mu_2 & = & 0. \end{array} \right.$$

## D Gradient and hessian of the log-likelihood

The gradient and the hessian of the loglikelihood are computed with explicit recursions. We denote  $\vartheta_7 = \sigma^2$ . The first order derivatives of  $A(\vartheta)$  are equal to:

$$\frac{\partial A(\vartheta)}{\partial \vartheta_1} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \frac{\partial A(\vartheta)}{\partial \vartheta_2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \frac{\partial A(\vartheta)}{\partial \vartheta_q} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad q = 3, 4, 5, 6, 7.$$

for  $R(\vartheta)$  we get:

$$\frac{\partial R(\vartheta)}{\partial \vartheta_1} = \frac{\partial R(\vartheta)}{\partial \vartheta_2} = \frac{\partial R(\vartheta)}{\partial \vartheta_6} = \frac{\partial R(\vartheta)}{\partial \vartheta_7} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\frac{\partial R(\vartheta)}{\partial \vartheta_3} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \frac{\partial R(\vartheta)}{\partial \vartheta_4} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \frac{\partial R(\vartheta)}{\partial \vartheta_5} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

and for  $m$  we get:  $\frac{\partial m}{\partial \vartheta_q} = 0, q = 1, \dots, 5, 7$  and  $\frac{\partial m}{\partial \vartheta_6} = 1$ . The second order derivatives of  $A(\vartheta)$  and  $R(\vartheta)$  are null. The first order derivatives of  $\hat{X}_i^-(\vartheta)$  and  $P_i^-(\vartheta)$  with respect to  $\vartheta_q, q = 1, \dots, 7$  can be deduced:

$$\begin{aligned} \frac{\partial \hat{X}_i^-(\vartheta)}{\partial \vartheta_q} &= \frac{\partial A(\vartheta)}{\partial \vartheta_q} \hat{X}_{i-1}(\vartheta) + A(\vartheta) \frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q} \\ \frac{\partial P_i^-(\vartheta)}{\partial \vartheta_q} &= \frac{\partial A(\vartheta)}{\partial \vartheta_q} P_{i-1}(\vartheta) A(\vartheta)' + A(\vartheta) \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} A(\vartheta)' + A(\vartheta) P_{i-1}(\vartheta) \frac{\partial A(\vartheta)'}{\partial \vartheta_q} + \frac{\partial R(\vartheta)}{\partial \vartheta_q}. \end{aligned}$$

Then we get the derivatives of the mean and the covariance of the filter:



$$\begin{aligned}\frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q} &= \frac{\partial \hat{X}_{i-1}^-(\vartheta)}{\partial \vartheta_q} + \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} \frac{H' W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} + \frac{P_{i-1}^-(\vartheta) H'}{V_{i-1}(\vartheta)} \left( \frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_q} - \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} \right) \\ \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} &= \left( I - \frac{P_{i-1}^- H' H}{V_{i-1}(\vartheta)} \right) \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} - \left( \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} \frac{H'}{V_{i-1}(\vartheta)} - \frac{P_{i-1}^-(\vartheta) H'}{V_{i-1}(\vartheta)^2} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \right) H P_{i-1}^-(\vartheta)\end{aligned}$$

The computation of second order derivatives can be deduced. Because second order derivatives of  $A$  and  $R$  are null, we have for  $q, r = 1, \dots, 7$ , :

$$\begin{aligned}\frac{\partial^2 \hat{X}_{i-1}^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} &= A(\vartheta) \frac{\partial^2 \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} + \frac{\partial A(\vartheta)}{\partial \vartheta_r} \frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q} + \frac{\partial A(\vartheta)}{\partial \vartheta_q} \frac{\partial \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_r} \\ \frac{\partial^2 P_{i-1}^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} &= \left( \frac{\partial A(\vartheta)}{\partial \vartheta_r} \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} + \frac{\partial A(\vartheta)}{\partial \vartheta_q} \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_r} \right) A(\vartheta)' + \frac{\partial A(\vartheta)}{\partial \vartheta_r} P_{i-1}(\vartheta) \frac{\partial A(\vartheta)'}{\partial \vartheta_q} \\ &+ A(\vartheta) \left( \frac{\partial^2 P_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} A(\vartheta)' + \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_r} \frac{\partial A(\vartheta)'}{\partial \vartheta_q} + \frac{\partial P_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{\partial A(\vartheta)'}{\partial \vartheta_r} \right) + \frac{\partial A(\vartheta)}{\partial \vartheta_q} P_{i-1}(\vartheta) \frac{\partial A(\vartheta)'}{\partial \vartheta_r}\end{aligned}$$

The second order derivatives of the mean and the covariance of Kalman filter are:

$$\begin{aligned}\frac{\partial^2 \hat{X}_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} &= \frac{\partial^2 \hat{X}_{i-1}^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} + \frac{\partial^2 P_{i-1}^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} \frac{H' W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} + P_{i-1}^-(\vartheta) \frac{H'}{V_{i-1}(\vartheta)} \left( \frac{\partial^2 W_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} \right. \\ &- \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{1}{V_{i-1}(\vartheta)} \frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_r} - \frac{\partial^2 V_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} \frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} + 2 \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_r} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{W_{i-1}(\vartheta)}{V_{i-1}^2(\vartheta)} \\ &- \left. \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_r} \frac{1}{V_{i-1}(\vartheta)} \frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_q} \right) + \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_r} \frac{H'}{V_{i-1}(\vartheta)} \left( \frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_q} - \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} \right) \\ &+ \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} \frac{H'}{V_{i-1}(\vartheta)} \left( \frac{\partial W_{i-1}(\vartheta)}{\partial \vartheta_r} - \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_r} \frac{W_{i-1}(\vartheta)}{V_{i-1}(\vartheta)} \right)\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 P_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} &= \left( I - \frac{P_{i-1}^-(\vartheta) H' H}{V_{i-1}(\vartheta)} \right) \frac{\partial^2 P_{i-1}^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} - \left[ \frac{\partial^2 P_{i-1}^-(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} - \frac{P_{i-1}^-(\vartheta)}{V_{i-1}(\vartheta)} \frac{\partial^2 V_{i-1}(\vartheta)}{\partial \vartheta_q \partial \vartheta_r} \right. \\
&\quad \left. - \left( \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_r} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} + \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_r} \right) \frac{1}{V_{i-1}(\vartheta)} + \frac{2P_{i-1}^-(\vartheta)}{V_{i-1}^2(\vartheta)} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_r} \right] \frac{H' H P_{i-1}^-(\vartheta)}{V_{i-1}(\vartheta)} \\
&\quad - \left( \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_r} H' - \frac{P_{i-1}^-(\vartheta) H'}{V_{i-1}(\vartheta)} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_r} \right) \frac{H}{V_{i-1}(\vartheta)} \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} \\
&\quad - \left( \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_q} H' - P_{i-1}^-(\vartheta) \frac{H'}{V_{i-1}(\vartheta)} \frac{\partial V_{i-1}(\vartheta)}{\partial \vartheta_q} \right) \frac{H}{V_{i-1}(\vartheta)} \frac{\partial P_{i-1}^-(\vartheta)}{\partial \vartheta_r}
\end{aligned}$$

## E Smoother algorithm

The Kalman smoother is calculated recursively with a forward-backward algorithm (see *e.g.* Cappé *et al.* (2005)). The forward algorithm is the classical Kalman filter which computes  $M_{i|0:i-1} = \hat{Z}_i^- = \mathbb{E}(Z_i | y_{0:i-1})$ ,  $\Sigma_{i|0:i-1} = P_i^- = \text{Var}(Z_i | y_{0:i-1})$ ,  $M_{i|0:i} = \hat{Z}_i = \mathbb{E}(Z_i | y_{0:i})$  and  $\Sigma_{i|0:i} = P_i = \text{Var}(Z_i | y_{0:i})$ . Then, in order to calculate  $M_{i|0:n} = \mathbb{E}(Z_i | y_{0:n})$ ,  $\Sigma_{i|0:n} = \text{Var}(Z_i | y_{0:n})$ ,  $\Sigma_{i-1,i|0:n} = \text{Cov}(Z_{i-1}, Z_i | y_{0:n})$ , one performs the set of backward recursions  $i = n, n-1, \dots, 1$ :

$$\begin{aligned}
J_{i-1} &= \Sigma_{i-1|0:i-1} A' (\Sigma_{i|0:i-1})^{-1} \\
M_{i-1|0:n} &= M_{i-1|0:i-1} + J_{i-1} (M_{i|0:n} - M_{i|0:i-1}) \\
\Sigma_{i-1|0:n} &= \Sigma_{i-1|0:i-1} + J_{i-1} (\Sigma_{i|0:n} - \Sigma_{i|0:i-1}) J_{i-1}'
\end{aligned}$$

To calculate  $\Sigma_{i-1,i|0:n}$ , we have

$$\Sigma_{n-1,n|0:n} = (I - K_n H) A \Sigma_{n-1|0:n-1}$$

and the following backward recursions, for  $i = n-1, n-2, \dots, 1$

$$\Sigma_{i-1,i|0:n} = \Sigma_{i|0:i} J_{i-1}' + J_i (\Sigma_{i,i+1|0:n} - A \Sigma_{i|0:i}) J_{i-1}'$$

## F ARMA property of multidimensional process

In our model,  $(y_i)_{i \in \mathbb{Z}}$  is an ARMA(2,2) process and asymptotic properties of the maximum likelihood estimator are derived. This result can be generalised to the case where  $X_i$  is  $p$ -dimensional under weak assumptions. We consider the model:

$$y_i = HX_i + \sigma\varepsilon_i, X_i = AX_{i-1} + \eta_i, X_0 \sim \nu$$

where  $X_i$  is  $p$ -dimensional,  $A$  is a diagonal matrix with diagonal coefficients  $(\theta_k, k = 1, \dots, p)$  such that  $\theta_k \neq \theta_l$  for  $k \neq l$  and  $\theta_i \in (0, 1)$  for  $i = 1 \dots p$ ,  $(\eta_i)_{i \geq 0}$  is a sequence of independent  $\mathcal{N}_p(0, R)$  random variables,  $H$  is a  $(1, p)$ -matrix and  $(\varepsilon_i)$  is a sequence of i.i.d  $\mathcal{N}(0, 1)$  random variables. Up to a transformation of  $(X_i)$ ,  $H$  is equal to  $H = (1 \dots 1 \ 0 \dots 0)$  with its first  $d$  coordinates equal to 1 and its  $p - d$  next coordinates equal to 0 ( $1 \leq d \leq p$ ). Consequently, we observe with additive noise the partial sum of the first  $d$  coordinates of  $X_i$ . Since  $A$  is diagonal and  $\theta_i \in (0, 1)$  for  $i = 1 \dots p$ , the process  $(X_i)_{i \geq 0}$  admits a stationary distribution  $\nu$ . Then with  $X_0 \sim \nu$ , the process  $(X_i)_{i \geq 0}$  is stationary. Denote  $(y_i)_{i \in \mathbb{Z}}$  the extension of  $y$  to  $\mathbb{Z}$  by stationarity.

**Proposition 8** *The process  $(y_i)_{i \in \mathbb{Z}}$  is ARMA( $d, d$ ).*

*Proof.* Denote  $S_j, j = 1, \dots, d$ , the  $j$ -th symmetric function of  $\theta_1, \dots, \theta_d$ :

$$S_j = \sum_{1 \leq i_1 < \dots < i_j \leq d} \theta_{i_1} \dots \theta_{i_j}$$

and  $S_0 = 1$ . Define the polynomial  $P$  such that  $P(\theta_1) = \dots = P(\theta_d) = 0$ :

$$P(x) = \sum_{k=0}^d (-1)^k S_k x^{d-k}$$

Set  $L$  the one-lag operator :  $Ly_i = y_{i-1}, L\varepsilon_i = \varepsilon_{i-1}$ . Set  $\xi_i = P(L)(y_i) = \sum_{k=0}^d (-1)^k S_k y_{i-k}$ . By recursive computation for  $0 \leq k \leq d$ , we have

$$y_{i-k} = HA^{d-k} X_{i-d} + \sum_{j=0}^{d-k-1} HA^j \eta_{i-k-j} + \sigma\varepsilon_{i-k}$$

We deduce

$$\xi_i = HP(A)X_{i-d} + \sum_{k=0}^d (-1)^k S_k \left( \sum_{j=0}^{d-k-1} HA^j \eta_{i-k-j} \right) + \sigma P(L)(\varepsilon_i)$$

But  $HP(A) = (P(\theta_1) \dots P(\theta_d) 0 \dots 0) = (0 \dots \dots 0)$ . Thus  $\xi_i$  only depends on  $(\eta_j, \varepsilon_j)_{j \leq i}$ . Therefore  $(y_i)$  verifies an AR( $d$ ) equation. Moreover, as the  $(\eta_i)_i$  and  $(\varepsilon_i)_i$  are mutually independent, we get that:

$$Cov(\xi_i, \xi_{i+k}) = 0, \quad \forall k \geq d.$$

This implies that  $(\xi_i)$  is MA( $d$ ). Hence the result.

## G First order identifiability

The spectral density can be rewritten as

$$f(u, \vartheta) = \sigma^2 + \frac{d(\vartheta)e^{iu} + c(\vartheta) + d(\vartheta)e^{-iu}}{(1 - (\theta_1 + \theta_2)e^{iu} + \theta_1\theta_2e^{2iu})(1 - (\theta_1 + \theta_2)e^{-iu} + \theta_1\theta_2e^{-2iu})}$$

where  $c(\vartheta) = H(ARA' + (1 + (\theta_1 + \theta_2)^2)R)H'$  and  $d(\vartheta) = (HA - (\theta_1 + \theta_2)H)RH'$ . The equality  $f(u, \vartheta) = f(u, \vartheta') \quad \forall u \in (0, 2\pi)$  implies

$$\sigma^2 = \sigma'^2, \quad \theta_1 + \theta_2 = \theta'_1 + \theta'_2, \quad \theta_1\theta_2 = \theta'_1\theta'_2, \quad c(\vartheta) = c(\vartheta'), \quad d(\vartheta) = d(\vartheta')$$

We deduce that  $\sigma^2$ ,  $\theta_1 + \theta_2$ ,  $\theta_1\theta_2$  are identifiable and that at most two of three parameters among  $\theta_3, \theta_4$  and  $\theta_5$  are identifiable from  $c(\vartheta)$  and  $d(\vartheta)$ .

We can prove that there are exactly two parameters identifiable when  $\Delta$  is small.

Set

$$g(z, \vartheta) = d(\vartheta) + c(\vartheta)z + d(\vartheta)z^2.$$

so that for  $z = e^{iu}$

$$f(z, \vartheta) = \sigma^2 + z \frac{g(z, \vartheta)}{(1 - \theta_1 z)(\theta_1 - z)(1 - \theta_2 z)(\theta_2 - z)}$$

Note that the product of the roots of  $g$  is equal to 1. Thus, if  $\theta_1$  and  $\theta_2$  are not roots of  $g$ , then no simplification is possible in the expression of  $f$  and exactly two out of the three parameters  $\theta_3, \theta_4, \theta_5$  are identifiable (from  $c(\vartheta)$  and  $d(\vartheta)$ ).

The computation of  $g(\theta_1)$  and  $g(\theta_2)$  gives

$$\begin{aligned}
g(\theta_1) &= (\theta_1 - \theta_2)(\theta_3 + \theta_5) + 2\theta_1^3\theta_3 + \theta_1^3\theta_5 \\
&\quad + \theta_1\theta_2^2\theta_3 + \theta_1^2\theta_2\theta_3 + 2\theta_1\theta_2^2\theta_4 + 2\theta_1^2\theta_2\theta_4 + 2\theta_1\theta_2^2\theta_5 + 5\theta_1^2\theta_2\theta_5 \\
g(\theta_2) &= (\theta_2 - \theta_1)(\theta_4 + \theta_5) + 2\theta_2^3\theta_4 + \theta_2^3\theta_5 \\
&\quad + 2\theta_1\theta_2^2\theta_3 + 2\theta_1^2\theta_2\theta_3 + \theta_1\theta_2^2\theta_4 + \theta_1^2\theta_2\theta_4 + 5\theta_1\theta_2^2\theta_5 + 2\theta_1^2\theta_2\theta_5
\end{aligned}$$

It is not straightforward to prove that  $g$  does not vanish in  $\theta_1$  and  $\theta_2$ . But for small  $\Delta$ , by developing  $\theta_1, \dots, \theta_5$  at first order we have

$$\begin{aligned}
\theta_1 - \theta_2 &= \Delta(\mu_1 - \mu_2) + O(\Delta) \\
\theta_3 + \theta_5 &= \Delta\left(\sigma_1^2\left(\frac{\mu_2 + \beta}{\mu_2 - \mu_1}\right)^2 + \sigma_2^2\left(\frac{\mu_2}{\mu_2 - \mu_1}\right)^2 - \sigma_1^2\frac{\mu_1 + \beta}{\mu_2 - \mu_1}\frac{\mu_2 + \beta}{\mu_2 - \mu_1} - \sigma_2^2\frac{\mu_1\mu_2}{\mu_2 - \mu_1}\right) + O(\Delta) \\
\theta_4 + \theta_5 &= \Delta\left(\sigma_1^2\left(\frac{\mu_1 + \beta}{\mu_2 - \mu_1}\right)^2 + \sigma_2^2\left(\frac{\mu_1}{\mu_2 - \mu_1}\right)^2 - \sigma_1^2\frac{\mu_1 + \beta}{\mu_2 - \mu_1}\frac{\mu_2 + \beta}{\mu_2 - \mu_1} - \sigma_2^2\frac{\mu_1\mu_2}{\mu_2 - \mu_1}\right) + O(\Delta)
\end{aligned}$$

Hence it comes

$$g(\theta_1) = O(\Delta^2) + 4C\Delta + o(\Delta), \quad g(\theta_2) = O(\Delta^2) + 4C\Delta + o(\Delta).$$

with

$$C = \theta_3 + \theta_4 + 2\theta_5$$

But  $C$  is a positive constant because

$$C = (1 \quad 1)R(1 \quad 1)'$$

where  $R$  is a covariance matrix thus positive. Hence  $g(\theta_1)$  and  $g(\theta_2)$  are positive when  $\Delta$  is small. We have thus a non degenerate ARMA(2,2) process and exactly five parameters are identifiable.

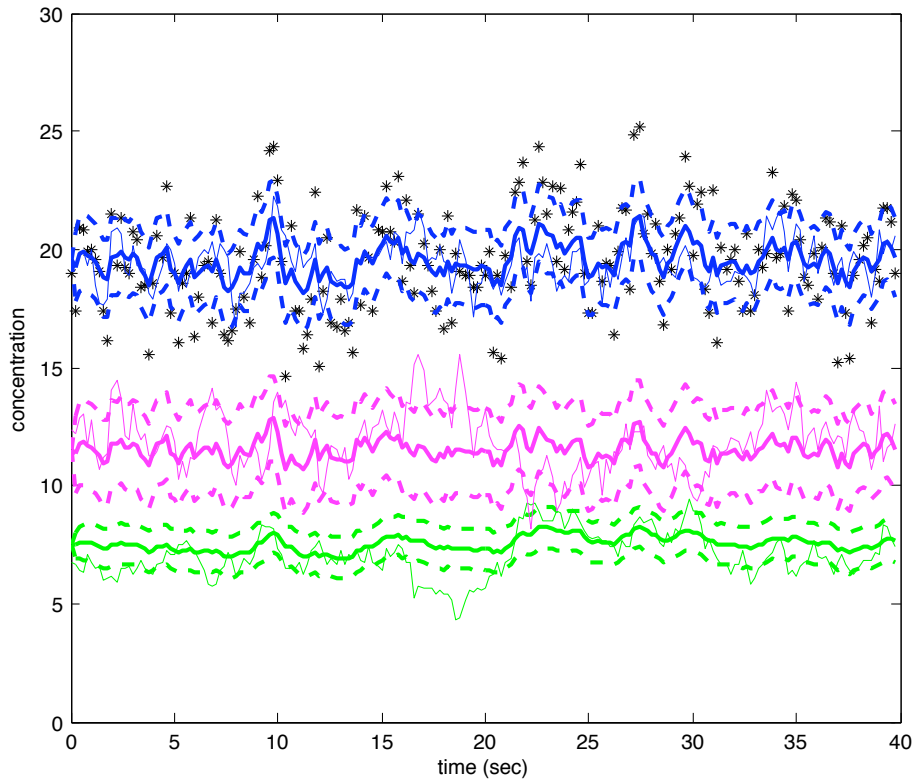


Figure 1: Noisy observations of one simulated data set ( $n = 200$ ,  $\Delta = 0.2$ ,  $\sigma^2 = 3$ ) are plotted with stars. True simulated trajectories (thin solid lines), mean estimated trajectories (thick solid lines) and estimated 95% confidence intervals (dotted lines) obtained with the Kalman algorithm are plotted with dark lines for  $S(t)$ , light lines for  $P(t)$  and very light line for  $I(t)$ .

$\sigma^2 = 1$							
$n = 200, \Delta = 0.2$				$n = 1000, \Delta = 0.2$			
Par.	true value	EM (SE)	MLE (SE)	Fisher SE	EM (SE)	MLE (SE)	Fisher SE
$\theta_1$	0.60	0.62 (0.11)	0.64 (0.17)	(0.15)	0.62 (0.05)	0.68 (0.11)	(0.11)
$\theta_2$	0.90	0.89 (0.06)	0.79 (0.15)	(0.04)	0.92 (0.02)	0.87 (0.07)	(0.11)
$\theta_3$	0.70	0.85 (0.20)	0.78 (0.26)	(0.25)	0.86 (0.09)	0.76 (0.15)	(0.37)
$\theta_4$	0.20	0.10 (0.01)	0.12 (0.16)	(0.14)	0.10 (0.01)	0.10 (0.09)	(0.36)
$\theta_6$	20.00	20.00 (0.39)	20.00 (0.38)		20.01 (0.17)	20.01 (0.17)	
$\sigma_1^2 + \sigma_2^2$	0.32	0.35 (0.09)	0.34 (0.14)		0.34 (0.04)	0.30 (0.07)	
$\sigma^2 = 3$							
$n = 200, \Delta = 0.2$				$n = 1000, \Delta = 0.2$			
Par.	true value	EM (SE)	MLE (SE)	Fisher SE	EM (SE)	MLE (SE)	Fisher SE
$\theta_1$	0.60	0.58 (0.13)	0.59 (0.20)	(0.17)	0.57 (0.06)	0.59 (0.13)	(0.24)
$\theta_2$	0.90	0.89 (0.06)	0.76 (0.19)	(0.05)	0.92 (0.02)	0.88 (0.07)	(0.04)
$\theta_3$	0.70	0.96 (0.27)	0.85 (0.43)	(0.53)	0.96 (0.10)	0.88 (0.23)	(0.22)
$\theta_4$	0.20	0.10 (0.01)	0.15 (0.23)	(0.11)	0.10 (0.01)	0.14 (0.09)	(0.13)
$\theta_6$	20.00	20.00 (0.41)	20.01 (0.39)		19.98 (0.18)	19.98 (0.18)	
$\sigma_1^2 + \sigma_2^2$	0.32	0.41 (0.14)	0.42 (0.31)		0.39 (0.05)	0.38 (0.12)	

Table 1: Mean estimated values (with empirical standard errors in bracket) obtained with the exact MLE and the EM algorithms and exact standard errors obtained from the Fisher information matrix, evaluated on 1000 simulated data with  $n = 200$  and  $n = 1000$  observations and  $\sigma^2 = 1$  or  $\sigma^2 = 3$  ( $\sigma^2$  and  $\theta_5$  fixed to their true values).

$\sigma^2 = 1$								
$n = 200, \Delta = 0.2$					$n = 1000, \Delta = 0.04$			
Par.	true value	EM (SE)	MLE (SE)	Fisher SE	true value	EM (SE)	MLE (SE)	Fisher SE
$\theta_1$	0.60	0.62 (0.11)	0.64 (0.17)	(0.15)	0.91	0.86 (0.03)	0.87 (0.08)	(0.06)
$\theta_2$	0.90	0.89 (0.06)	0.79 (0.15)	(0.04)	0.99	0.97 (0.01)	0.94 (0.12)	(0.02)
$\theta_3$	0.70	0.85 (0.20)	0.78 (0.26)	(0.25)	0.20	0.26 (0.02)	0.06 (0.26)	(0.23)
$\theta_4$	0.20	0.10 (0.01)	0.12 (0.16)	(0.14)	0.03	0.09 (0.01)	0.03 (0.13)	(0.17)
$\theta_6$	20.00	20.00 (0.39)	20.00 (0.38)		20.00	20.01 (0.56)	20.01 (0.56)	
$\sigma_1^2 + \sigma_2^2$	0.32	0.35 (0.09)	0.34 (0.14)		0.05	0.09 (0.01)	0.02 (0.09)	

$\sigma^2 = 3$								
$n = 200, \Delta = 0.2$					$n = 1000, \Delta = 0.04$			
Par.	true value	EM (SE)	MLE (SE)	Fisher SE	true value	EM (SE)	MLE (SE)	Fisher SE
$\theta_1$	0.60	0.58 (0.13)	0.59 (0.20)	(0.17)	0.91	0.74 (0.06)	0.79 (0.15)	(0.01)
$\theta_2$	0.90	0.89 (0.06)	0.76 (0.19)	(0.05)	0.99	0.96 (0.02)	0.89 (0.17)	(0.01)
$\theta_3$	0.70	0.96 (0.27)	0.85 (0.43)	(0.53)	0.20	0.24 (0.02)	0.20 (1.39)	(0.04)
$\theta_4$	0.20	0.10 (0.01)	0.15 (0.23)	(0.11)	0.03	0.08 (0.01)	0.14 (1.85)	(0.01)
$\theta_6$	20.00	20.00 (0.41)	20.01 (0.39)		20.00	19.99 (0.53)	19.99 (0.54)	
$\sigma_1^2 + \sigma_2^2$	0.32	0.41 (0.14)	0.42 (0.31)		0.05	0.13 (0.01)	0.39 (5.40)	

Table 2: Mean estimated values (with empirical standard errors in bracket) obtained with the exact MLE and the EM algorithms and exact standard errors obtained from the Fisher information matrix, evaluated on 1000 simulated data with  $n = 200$  and  $n = 1000$  observations and  $\sigma^2 = 1$  or  $\sigma^2 = 3$  ( $\sigma^2$  and  $\theta_5$  fixed to their true values).