Noname manuscript No.
(will be inserted by the editor)

Dirichlet prior for cascade SDE with Markov regime-switching

A. TOSSA · D. BERNARD · R. EMILION · S. K. IYER

Received: date / Accepted: date

Abstract A Stochastic Differential Equation appearing in the statistical theory of turbulence is extended in random environment by assuming that its two parameters are switched by an unobserved continuous-time Markov chain whose states represent the states of the environment. A Dirichlet process is placed as a prior on the space of the sample paths of this chain, leading to a hierarchical Dirichlet model whose estimation is done both on simulated data and on real data of wind speed measured at the entrance of a mangrove ecosystem.

Keywords Cascades · Dirichlet process · Dissipation · Mangrove · Markov regime switching · Random environment · Stochastic differential equation

1 Introduction

Models in which parameters move between a fixed number of regimes with switching controlled by an unobserved stochastic process, have been heavily used in many disciplines including Finance (Hamilton and Susmel, 1994), Meteorology (Zucchini and Guttorp 1991), Computational biology (Durbin et al., 1998), Networks or Speech recognition (Rabiner 1989) to name but a few, notably because they take into account random regime changes in the environment. We will consider here a model described by a stochastic differential equation (SDE) with Markov regime-switching (MRS), i.e., with parameters controlled by a finite state continuous-time Markov chain (CTMC) as done, for example, in Ghosh and Deshpande (2008) . In such a setting, parameters estimation problem is a real challenge, mainly due to the fact that the paths of the CTMC are unobserved. A standard approach consists in using the celebrated EM algorithm (see Dempster et al. 1977) as proposed for example in Hamilton (1990).

Adaté TOSSA
CEREMADE, Université Paris Dauphine, 75775 Paris, France
E-mail: oliviertossa@gmail.com

Didier BERNARD
LPAT, Université Antilles-Guyane, B.P. 250, 97157 Pointe-à-Pitre, Guadeloupe
E-mail: didier.bernard@univ-ag.fr

Richard EMILION
MAPMO, Université d'Orléans, B.P 6759, 45067 Orléans Cedex 2, France
E-mail: richard.emilion@univ-orleans.fr

Srikanth K. IYER
Department of Mathematics, Indian Institute of Science, Bangalore 560 012 India
E-mail: skiyer@math.iisc.ernet.in

In the present paper, our estimation approach is Bayesian, the aim being to find a pair (parameters, CTMC path) with likelihood as large as possible. Standard priors are placed on the parameters space but, as the CTMC paths are unobserved, a large number of paths are drawn from a Dirichlet process placed as a prior on the path space of the CTMC. The complete model then appears as a Hierarchical Dirichlet Model (HDM) (see Ishwaran and Zarepour 2000 ; Ishwaran and Lancelot 2002, 2003) whose estimation procedure requires some rather nontrivial computations of posterior distributions due to the temporal level induced by the specific SDE and the CTMC. Using the well-known stick-breaking approximation, each set of iterations selects the pair with largest likelihood and then the Dirichlet process is updated in order to look for other paths which can eventually improve the likelihood.

The considered SDE, called here cascade SDE, is the one introduced by F. Schmitt (Schmitt 2001) in the statistical theory of turbulence. This model depends on two parameters, one representing the energy dissipation and the other being a scaling parameter. As will be shown for the data set considered in this paper, these parameters change considerably with time. We model these changes by assuming that the system operates in a finite number of regimes, each characterized by different values of the parameters. The evolution of these regimes is modeled by a CTMC, thus extending the cascades SDE to a MRS setting.

Our paper is structured as follows. In Section 2 we present the cascade SDE with MRS and the complete HDM. Section 3 is devoted to the posterior computations. The algorithm used for estimation is described in Section 4. Section 5 contains numerical results. The proposed algorithm is first tested on a simulated data set. Then it is applied to a real data set of wind speed measured at the entrance of a Mangrove ecosystem. Thus the CTMC transitions correspond to micro-meteorological regime changes. We conclude in the last Section. The proofs of the results are given in the Appendix.

2 Cascade SDE with Markov regime-switching and Dirichlet prior

2.1 Cascade SDE with Markov regime-switching

F. Schmitt and D. Marsan (see Schmitt and Marsan 1998, Schmitt 2001) introduced a *stochastic dissipation process* ε, with a continuum of scale levels, which satisfies a stochastic differential equation (SDE), called the *cascade SDE*, which depends on an *intermittency* parameter μ and a *scale* parameter λ. This SDE is a continuous version of the Yaglom multiplicative cascade model (Kolmogorov 1962, Yaglom 1966). The process $\gamma_\lambda(t) = \log(\varepsilon_\lambda(t))$, called the singularity process satisfies the SDE,

$$\gamma_\lambda(t) = -\frac{\mu}{2}\ln\lambda + \mu^{1/2}\int_{t+1-\lambda}^{t}(t+1-u)^{-1/2}dB(u). \tag{1}$$

where $B(.)$ is a standard Brownian motion.

We propose an extension of this cascade SDE to a model in random environment by defining a dissipation process subject to *regime* changes. In order to take into account random regime changes in the environment, we assume that the parameters μ and λ are themselves stochastic processes driven by a continuous time Markov chain. This idea is well-known in mathematical finance when modelling regime switching markets with stochastic volatility (see e.g. DI-MASI et al. 1994 ; Ghosh and Deshpande 2008).

We consider the following cascade stochastic differential equation

$$\gamma(t) = -\frac{\mu_{X(t)}}{2}\log(\lambda_{X(t)}) + \mu_{X(t)}^{1/2}\int_{t+1-\lambda_{X(t)}}^{t}(t+1-u)^{-1/2}dB(u), \tag{2}$$

where $B(.)$ is a standard Brownian motion, $X = (X(t))_{t\geq 0}$ is a continuous time Markov chain taking values in a finite set $S = \{1,2,\ldots,M\}$ that represents different *regimes* of the environment. If, at a given time, t $X(t) = i$ $(i \in S)$, then the environment is in regime i so that $\mu_{X(t)} = \mu_i$ and $\lambda_{X(t)} = \lambda_i$. Observe that the distribution of γ appears as a mixture of distributions of cascade SDE. The chain $X(t)$ spends an exponential amount of time with parameter β_i in state $i \in S$, and then jumps to state $j \in S, j \neq i$ with probability p_{ij}. The transition rate matrix $Q = (q_{ij})_{i,j \in S}$ is such that off-diagonal entries q_{ij} are nonnegative real values, and the diagonal element q_{ii} is constrained to be $q_{ii} = -\sum_{j \neq i} q_{ij}$, so that the row sums of Q are zero. We have $q_{ii} = -\beta_i$ for all $i \in S$, and $q_{ij} = \beta_i \times p_{ij}$ for $j \neq i$.

We complete the above model by placing a Dirichlet process (DP) as a prior on the path space of the Markov chain X. This leads us to the following Bayesian modeling framework:

$$\begin{cases} (X|P,\alpha) \sim P \\ (P|\alpha) \sim \mathscr{D}(\alpha H) \\ (\alpha) \sim Gamma(\eta_1, \eta_2) \end{cases} \tag{3}$$

where H is the distribution of a specific Markov chain determined by the transition rate matrix Q and an initial distribution π_0. In practice, process (2) is sampled at times $t_1, \dots, t_n$, providing a finite random vector $\gamma = (\gamma_1, \dots, \gamma_n)$ of observations, where γ_i is the value recorded at time t_i. We prove that the conditional distribution of the vector γ is Gaussian so that we arrive at the following hierarchical model that we want estimate from the observed data:

$$\begin{cases} (\gamma|X,\mu,\lambda,P) \sim \mathscr{N}_n(m(X),\Sigma) \\ \mu_1,...,\mu_M \overset{iid}{\sim} \Gamma_1 \\ \lambda_1,...,\lambda_M \overset{iid}{\sim} \Gamma_2 \\ (X|P,\alpha) \sim P \\ (P|\alpha) \sim \mathscr{D}(\alpha H) \\ (\alpha) \sim Gamma(\eta_1, \eta_2) \end{cases} \tag{4}$$

The next section recalls some notions that are important for the estimation procedure.

2.2 Stick-breaking representation of Dirichlet process

A constructive approach to the Dirichlet process is the so called stick-breaking scheme. Let H be a probability distribution on some measurable space $\mathscr{V}$ and let $N \geq 2$ be an integer. Let $V_k \overset{i.i.d}{\sim} Beta(a_k, b_k)$, $k = 1 \dots N-1$, with shape parameters $a_k, b_k > 0$, and let $V_N = 1$. Let $p_1 \dots p_N$, be defined by

$$p_1 = V_1, \quad p_k = (1-V_1)(1-V_2)...(1-V_{k-1})V_k, \ \ k \geq 2. \tag{5}$$

Note that $\sum_{k=1}^N p_k = 1$.

Let $Z_k \overset{i.i.d.}{\sim} H$, $k = 1, \dots, N$, be independent of $(p_k)_{k=1,\dots,N}$. The random measure $\mathscr{P}_N$ defined by

$$\mathscr{P}_N = \sum_{k=1}^N p_k \delta_{Z_k}(.) \tag{6}$$

is said to be a *stick-breaking* construction. The Ferguson Dirichlet process $\mathscr{D}(\alpha H)$ (Ferguson 1973) is the best known example of an infinite stick-breaking prior. This is made explicit by the following proposition due to Sethuraman (1994).

Proposition 1

$$\mathscr{P}_N(.) = \sum_{k=1}^N p_k \delta_{Z_k}(.) \xrightarrow{a.s.} \mathscr{D}(\alpha H).$$

This proposition yields an efficient approximation of a Dirichlet process that is very useful in Bayesian nonparametrics statistics. The following Lemma will be crucial in our estimation procedure.

Lemma 1 *Let $X : \Omega \to \{1,\dots,N\}$ be a r.v. with conditional distribution $Pr(X \in .|P) = \sum_{k=1}^{N} p_k \delta_k(.)$, where P is defined by the stick-breaking construction (5). Then the conditional distribution of P given X is also defined by (5), where $V_N = 1$ and the V_k, $k = 1,\dots,N-1$ are independent $Beta(a_k^*, b_k^*)$ r.v.'s with*

$$a_k^* = a_k + I_{\{X=k\}} \tag{7}$$

$$b_k^* = b_k + \sum_{j=k+1}^{N} I_{\{X=j\}}. \tag{8}$$

3 Posterior computations

The estimation procedure of the various parameters of the model is based on the Gibbs sampling scheme. Implementation of this scheme requires the computation of the following conditional distributions:

$$(\gamma|X,\mu,\lambda,p),\ \ (\mu|\gamma,\lambda,X),\ \ (\lambda|\gamma,\mu,X),\ \ (X|p,\alpha),\ \ (\alpha|P),\ \ (p|\alpha)\ \ \textit{and}\ \ (H|X)$$

For simplicity, $\gamma(t_i)$ and $X(t_i)$ will be denoted below by γ_i and $X(i)$, respectively.

3.1 Conditional for γ

Proposition 2

$$(\gamma|X,\mu,\lambda,p) \sim \mathcal{N}_n(m,\Sigma),$$

$$\textit{with}\quad m = \left(-\frac{1}{2}\mu_{X(1)}\log(\lambda_{X(1)}),\dots,-\frac{1}{2}\mu_{X(n)}\log(\lambda_{X(n)})\right)\ \ \textit{and}\ \ \Sigma = (\sigma_{st})_{s,t=1,\dots,n},$$

$$\textit{where}\quad \sigma_{st} = (\mu_{X(s)}\mu_{X(t)})^{1/2} COV\left(\int_{s+1-\lambda_{X(s)}}^{s} (s+1-u)^{-1/2} dB(u), \int_{t+1-\lambda_{X(t)}}^{t} (t+1-u)^{-1/2} dB(u)\right).$$

The coefficients σ_{st} can be given explicitly using a computation of F. G. Schmitt (see Schmitt 2003, pages 89-90).

Proposition 3 *Let s and t be integers, let $w = min(s,t)$, $a = max(s+1-\lambda_{X(s)}, t+1-\lambda_{X(t)}$ and $\tau = |s-t|$. Then*

$$\sigma_{st} = 2(\mu_{X(s)}.\mu_{X(t)})^{1/2} \log\left(\frac{\sqrt{w+1-a}+\sqrt{w+1-a+\tau}}{1+\sqrt{1+\tau}}\right).$$

If j is a state reached by the Markov chain, let $t_{1j},\dots,t_{n_j j}$ be the times at which j is reached and let

$$\gamma_j = (\gamma_{t_{1j}},\dots,\gamma_{t_{n_j j}}). \tag{9}$$

Corollary 1

$$(\gamma_j|\mu,\lambda,X) \sim \mathcal{N}_{n_j}\Big(-\frac{\mu_j}{2}\log(\lambda_j)(\underbrace{1,1,\dots,1}_{n_j\ times}),\Sigma_j\Big),$$

where $\Sigma_j = (\sigma_j(s,t))$, is a $n_j \times n_j$ matrix with

$$\sigma_j(s,t) = 2\mu_j \log\left(\frac{\sqrt{w+1-a}+\sqrt{w+1-a+\tau}}{1+\sqrt{1+\tau}}\right).$$

In the following sections, the probability densities of $\mathcal{N}_n$ and $\mathcal{N}_{n_j}$ will be denoted by f_n and f_{n_j} respectively, that is

$$f_n(\gamma,\mu,\lambda) = \frac{\exp\left(-\frac{1}{2}(\gamma - m)\Sigma^{-1}(\gamma - m)^T\right)}{(2\pi)^{n/2}\,(det(\mu\Sigma))^{1/2}}. \tag{10}$$

$$f_{n_j}(\gamma_j,\mu_j,\lambda_j) = \frac{\exp\left(-\frac{1}{2}(\gamma_j - m_j)(\mu_j\Sigma_j)^{-1}(\gamma_j - m_j)^T\right)}{(2\pi)^{n_j/2}\,(det(\mu_j\Sigma_j))^{1/2}}.$$

where $m_j = -\frac{\mu_j}{2}\log(\lambda_j)(\underbrace{1,1,\ldots,1}_{n_j\ times})$

3.2 Conditional for X.

Proposition 4 *Let N be a positive integer.*

$$\left(X|p = \sum_{i=1}^{N} p_i\delta_{X_i},\gamma,\mu,\lambda\right) \propto \sum_{i=1}^{N} p_i^*\delta_{X_i}$$

where $p_i^* = p_i f_n(\gamma,\mu,\lambda)$, where f_n is as defined in (10).

3.3 Conditional for p

Proposition 5

$$(p_1|X) \sim V_1^*,\ \ (p_k|X) \sim V_k^*\prod_{i=1}^{k-1}(1 - V_k^*) \qquad k = 2,3,\ldots,N, \tag{11}$$

where $V_k^* \sim\ Beta\ \left(a_k^*, b_k^*\right)\ for\ k = 1,\ldots,N-1$, with a_k^*, b_k^* given by Lemma 1 and $V_N^* = 1$.

Remark 1 *The fact stated in this proposition is that the current value of parameter a_k increases by 1 each time the path k is chosen. What this does is that as we repeat the process of generating from the conditional within each iteration, the Beta distributions will gradually concentrate on paths that best explain the data.*

3.4 Conditional for α

Proposition 6

$$(\alpha|p) \sim Gamma\left(N + \eta_1 - 1, \eta_2 - \sum_{i=1}^{N-1}\log(1 - V_i^*)\right),$$

where the V_i^* are the same as those obtained in the conditional for p.

3.5 Conditional for μ

Let $j \in S$ be a state of X and π_1 be the prior for μ_j.

Proposition 7 *If for all* $t \in \{1, \ldots, n\}$, $X(t) \neq j$, *then* $(\mu_j | \gamma, \lambda_j, X, \alpha) \sim \pi_1$, *otherwise*

$$(\mu_j | \gamma, \lambda_j, X, \alpha) \sim \Gamma_1 \propto \mathcal{N}_{n_j}(m_j, \mu_j \Sigma_j) * \pi_1,$$

where $m_j = -\frac{\mu_j}{2} \log(\lambda_j) (\underbrace{1, 1, \ldots, 1}_{n_j \text{ times}})$.

3.6 Conditional for λ

Let $j \in S$ be a state of X and let π_2 be the prior for λ_j. As in the case of μ, if for all $t \in \{1, \ldots, n\}$, $X(t) \neq j$, then

$$(\lambda_j | \gamma, \mu_j, X, \alpha) \sim \pi_2$$

otherwise, for each of the other values of $j \in S$,

$$(\lambda_j | \gamma, \mu_j, X, \alpha) \sim \Gamma_2 \propto (\gamma | \lambda_j, \mu_j, X, \alpha) * \pi_2. \tag{12}$$

3.7 Conditional for H

Recall that the distribution H is determined by the transition rate matrix Q and the initial distribution π_0. The conditional distribution of H is given by the following proposition which is nothing but the standard Maximun Likelihood Estimation for a CTMC. Refer to Yin, G. G. and Zhan Q. (1997) or Noris J. R. (1997) for more details.

Proposition 8 *Let* $X = (X(1), \ldots, X(n))$ *be the values of the CTMC* X *at times* $1, \ldots, n$. *Let* i *and* j *be two distinct states in* S, *and* $P = (p_{ij})$ *the transition probability matrix of* X.

$$\pi_0(i) = \frac{1}{n} \sum_{k=1}^{n} \delta_i(X(k)), \quad p_{ij} = \frac{1}{n} \sum_{k=1}^{n} \delta_{ij}(X(k), X(k+1)) \quad \textit{and} \quad Q_{ij} = \beta_i p_{ij},$$

where $\delta_{ij}(X(k), X(k+1)) = 1$ *if* $X(k) = i$ *and* $X(k+1) = j$, *and* 0 *otherwise.* β_i *is the reciprocal of the average durations spent in state* i.

4 Estimation procedure

4.1 Algorithm

We now describe the estimation procedure. Observe that steps (i), (iv), (v), (viii) and (ix) are new as compared to the procedures in Ishwaran-Zarepour (2000) and Ishwaran-James (2002).

i. Choose a large integer N and generate N paths $X_1, \ldots, X_N$, of the continuous time Markov chain with distribution H.
ii. Draw α from $Gamma(\eta_1, \eta_2)$ and draw $p_1, \ldots, p_N$ according to (5) with $a_k = 1$ and $b_k = \alpha$.
iii. Draw $\lambda = (\lambda_1, \ldots, \lambda_M)$ and $\mu = (\mu_1, \ldots, \mu_M)$ from their priors π_1, π_2 respectively.

iv. Draw one of the paths $X_1,\dots,X_N$, with probability $p_1,\dots,p_N$, respectively.

Iterate over the following steps (v) *through* (ix):

v. - Compute $\sigma_{ij} = COV(\gamma_i,\gamma_j)$ from Proposition 3
- Define $p_j^* \propto p_j f_n(\gamma,\mu,\Sigma)$, using Proposition 2
- Excecute (iv) with p_j replaced by p_j^*.

vi. - Define a_k^* and b_k^* using (7) and (8) where X is the index of the chosen path.
- Compute $p_1 = V_1^*$, and $p_k = (1-V_1^*)\dots(1-V_{k-1}^*)V_k^*$, $k = 2,3,\dots,N$ where $V_k^* \sim Beta(a_k^*, b_k^*)$, and $V_N^* = 1$.

vii. Draw α from $(\alpha|p) \sim Gamma\left(N+\eta_1-1, \eta_2 - \sum_{i=1}^{N-1}\log(1-V_i^*)\right)$.

viii. Given γ,λ and a chosen path X, for each state $j \in S$,
- If $X(t) \neq j$ for all $t \in \{1,\dots,n\}$, then draw μ_j from the prior π_1.
- otherwise, deternime the times $t_{1j},\dots,t_{n_j,j}$ at which the Markov chain takes the value j and compute $\sigma_j(s,t)$ from corollary 1.
- draw μ_j from the conditional distribution of μ_j given by Proposition 7.

ix. - If $X(t) \neq j$ for all $t \in \{1,\dots,n\}$, then draw λ_j from the prior π_2.
- otherwise, determine the times $t_{1j},\dots,t_{n_j,j}$ at which the Markov chain takes the value j and compute $\sigma_j(s,t)$ from corollary 1.
- draw λ_j from the conditional distribution of λ_j given by (12)

Remark 2 *Since n will be very large for the application that we have in mind ($n \approx 72,000$), the algorithm will be computationally infeasible since we have to invert the matrix Σ. Hence breaking the dataset into smaller pieces will help. However we keep the size large enough to estimate the largest significant correlation. We use the posterior obtained for one subset as the prior for the subsequent data set.*

Remark 3 *Another way to simulate λ is to generate a path of the Brownian motion B. For all times $t_{1j},\dots,t_{n_j j}$ for which X takes value j, solve equation (2) in λ_j numerically, using a discretization of $[t_j+1-\lambda_j,\ t_j]$. This gives values $\lambda_{1j},\dots,\lambda_{n_j,j}$. Use the values obtained this way for a large number of Brownian paths to compute the conditional for λ_j and draw value of λ_j from this conditional.*

The algorithm below summarizes the estimation procedure. In practice it will be implemented using the Gibbs sampling technique.

1. ***Initialization***
   - Let γ be the vector of observations
   - Choose the hyper-parameters η_1 , η_2 and N
   - Generate α from Gamma(η_1,η_2)
   - Generate N paths of the Markov chain
   - Draw $p = (p_1,\dots,p_N)$ from stick-breaking(α, N)
   - Choose one of the N paths according to p
   - Generate μ and λ from their priors
2. ***Iterations***
   - Compute $f_n(\gamma,\mu,\lambda)$ and update p_k, k being the index of the chosen path.
   - Choose one of the N paths according to p
   - For each state j, draw μ_j, drawλ_j
   - Draw α
   - Draw a new p

4.2 Truncation error bound

Let $\gamma(t)$ be defined as in (2) and let $\gamma = (\gamma_1, \ldots, \gamma_n)$ be a n sample from the process $\gamma(t)$. Let $m_N(\gamma)$ and $m_\infty(\gamma)$ denote the marginal density of γ subject to $\mathscr{P}_N$ and $\mathscr{D}(\alpha H)$ respectively. Using a result in Ishwaran (2002) it can be shown that

$$\int_n |m_N(\gamma) - m_\infty(\gamma)| d\gamma \leq 4\mathtt{exp}(-(N-1)/\alpha). \tag{13}$$

This result provides an error bound for the truncated Dirichlet process and shows that the sample size n has no effect on the bound. The adequacy of the truncation then depends on N and α. Of course the value of α changes during the different iterations of our Gibbs sampler. However, since the bound decreases exponentially fast, even for a fairy large value, $\alpha = 3$ for example, a truncation with $N = 30$ leads to an error bound of 25×10^{-5}. For the computations in the next section we have chosen a value of $N = 50$.

5 Numerical results

5.1 Simulated data

The present subsection aims at testing the reliability of the model. We perform numerical simulations of the stochastic process $\gamma(t)$. We consider a model with five regimes. The associated Markov chain then has five states and is defined by the following transition probability matrix

$$P = \begin{pmatrix} 0 & 0.2 & 0.7 & 0 & 0.1 \\ 0.5 & 0 & 0.1 & 0.26 & 0.14 \\ 0.6 & 0.18 & 0 & 0.17 & 0.05 \\ 0.55 & 0.13 & 0.1 & 0 & 0.22 \\ 0.08 & 0.43 & 0.16 & 0.33 & 0 \end{pmatrix}$$

and the initial distribution $\pi_o = [0.20 \quad 0.20 \quad 0.20 \quad 0.20 \quad 0.20]$. We also suppose that the parameters of the average durations of time spent in the states of the CTMC are given in table 1.

Table 1 Parameters of the exponential distributions of the duration spent in the states of the CTMC

	Regime 1	Regime 2	Regime 3	Regime 4	Regime 5
β_i	0.05	0.1	0.02	0.2	0.041

So the transition rate matrix of the chain is

$$Q_O = \begin{pmatrix} -0.05 & 0.01 & 0.035 & 0 & 0.005 \\ 0.05 & -0.1 & 0.01 & 0.026 & 0.014 \\ 0.012 & 0.0036 & -0.02 & 0.0034 & 0.001 \\ 0.11 & 0.026 & 0.02 & -0.2 & 0.044 \\ 0.00328 & 0.01763 & 0.00656 & 0.01353 & -0.041 \end{pmatrix}$$

We choose the prior of μ and λ to be independent truncated Gaussian distributions, and simulate the parameters $\mu = (\mu_1, \ldots, \mu_5)$ and $\lambda = (\lambda_1, \ldots, \lambda_5)$, corresponding to the five regimes. We also simulate a path of length $n = 600$ of the Markov chain. Using μ, λ and the Markov chain, we simulate a sample path γ of the stochastic process $\gamma(t)$ (see Fig. 1). Taking the data γ as input, we estimate the parameters

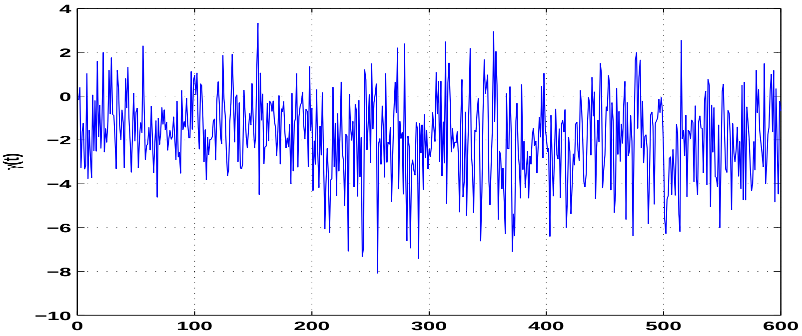

Fig. 1 A sample path of the stochastic process $\gamma(t)$

of the model through the algorithm presented in Section 4. For the purpose, we performed 500 Gibbs sampling runs each with $1,000$ iterations. At the end of each run Gibbs sampling, we choose the path with maximum likelihood and the corresponding values of the parameters μ and λ. This path is used to update the matrices Q and P for the next run of the Gibbs sampling. The summary statistics for μ and λ are given in table 2 and table 3 respectively. The estimated values are the averages of the estimates obtained in the 500 runs.

Table 2 Summary statistics of μ.

	Regime1	Regime 2	Regime 3	Regime 4	Regime 5
Actual Values	0.19	0.33	0.36	0.41	0.45
Estimated Values	0.22	0.31	0.38	0.42	0.46
95% credible interval	[0.189, 0.25]	[0.285, 0.335]	[0.35, 0.40]	[0.405, 0.43]	[0.44, 0.48]

Table 3 Summary statistics of λ

	Regime1	Regime 2	Regime 3	Regime 4	Regime 5
Actual Values	1067	997	1234	1743	1408
Estimated Values	1070	996	1234	1742	1410
95% credible interval	[1064, 1075]	[992, 1000]	[1228, 1240]	[1734, 1749]	[1403, 1416]

Table 4 Regime characteristics of the process $\gamma(t)$

	Regime1	Regime 2	Regime 3	Regime 4	Regime 5
μ	0.22	0.31	0.38	0.43	0.46
λ	1070	996	1234	1742	1410

5.2 Wind speed data

This section is aimed at testing the model on real data. We consider a dataset collected at the entrance of the mangrove ecosystem in Guadeloupe island (D. BERNARD and C. D'ALEXIS 2006). Wind velocity was recorded at a frequency of 20Hz by its $3D$ components v_x, v_y and v_z. As we are interested in the longitudinal velocity, only the components v_x and v_y are considered. Our observation time interval is one hour, providing a $2D$ series of length $n = 72,000$. Let $u = \frac{1}{\sqrt{(v_x)^2+(v_y)^2}}(\bar{v}_x, \bar{v}_y)$ be the mean longitudinal velocity vector and $w = \frac{1}{\sqrt{(v_x)^2+(v)_y^2}}(-\bar{v}_y, \bar{v}_x)$. Let (S_1, S_2) be the new coordinates of (v_x, v_y) in the basis (u, w) and let $S = \sqrt{S_1^2 + S_2^2}$ be the wind modulus. Computing the energy dissipation series $\varepsilon(t) = (S(t+1) - S(t))^2/(1/20)$, the aim is to fit our model (3) to the series $\gamma(t) = \log(S(t))$. Estimates of μ and λ on sliding windows of 60 seconds length, show that these parameters remain stable for random durations of time and then jump to another value. Regimes can be observed in Fig.2.

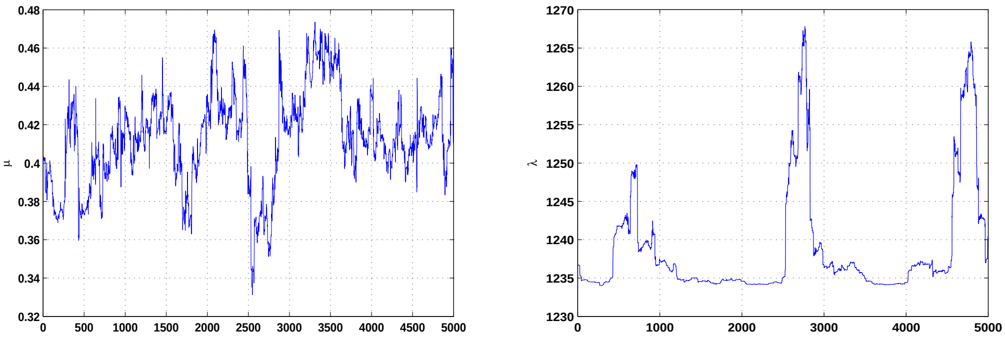

Fig. 2 Regime changes for μ (left) and λ (right).

Considering the histogram of the values of μ (resp. of λ) over the above sliding windows (see Fig. 3), a truncated Gaussian (resp. a *Gamma*) distribution is taken as inital prior for μ (resp. for λ). We

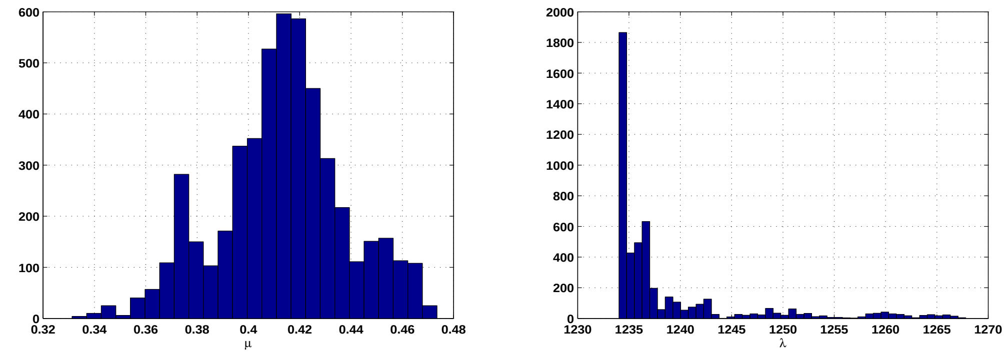

Fig. 3 Histograms of μ (left) and λ (right)

first ran our algorithm with many values of M and statistical comparasion tests lead us to take $M = 4$ states for the Markov chain. We present here the results of our algorithm for 500 Gibbs sampling runs of 25,000 iterations each, including 3,000 burn-in iterations. At each run of the Gibbs sampling, we evaluated the log-likelihood (LLH) of the selected path at each of the $25,000$ iterations and choose the path with maximum likelihood. So, after 500 runs, 500 paths are retained. Among these 500 paths, we consider that the one with the highest likelihood (that is the best of the best in terms of likelihood) is the one that best fits the data. It has the characteristics presented in table 5 and table 6.

Table 5 Characteritics of the path with highest likelihood

	Regime 1	Regime 2	Regime 3	Regime 4
μ	0.3885	0.2703	0.1314	0.2914
λ	1281	1448	1367	1240
% of occupation	17.2%	14%	8.5%	60.3%

Table 6 Sequence of regimes in the highest likelihood path

Regimes	4	1	3	4	2	1
Duration	287	38	51	75	84	65

6 Conclusion

We have proposed a new model for dissipation: cascade SDE with Markov regime switching to represent randomness in the environment, and Dirichlet prior on the path space of the continuous time Markov chain to make the model more flexible. It can be seen that this model is a complex mixture hierarchical model. The numerical results obtained lead us to believe that such mixture model better fit to many real world data sets than usual SDE models. The proposed Bayesian algorithmic method, whose key idea is the simulation of paths, can be extended to many other situations as soon as posterior distributions can be computed or simulated and priors used cautiously. A topic for further research may consist in replacing the continuous time Markov chain by a diffusion process. This requires a deeper study of the behaviour of the parameters.

APPENDIX: PROOFS

Proof (Lemma 1)
Note that $Pr(X=j|p)=p_j=(1-V_1)\cdots(1-V_{j-1})V_j$. Thus, if $A_1,\ldots,A_n$ are measurable subsets of [0,1], and if π is the joint distribution of the V_i's, we have,

$$Pr(V_1\in A_1,\ldots V_N\in A_N,X=j)=\int\prod_{k=1}^{N}I_{\{V_k\in A_k\}}Pr(X=j|V_1,..,V_N)\pi(dV_1,\ldots,dV_N)$$
$$=\int\prod_{k=1}^{N}I_{\{V_k\in A_k\}}(1-V_1)\cdots(1-V_{j-1})V_j\pi(dV_1,\ldots,V_N)$$
$$=\int I_{\{x_k\in A_k,k=1,\ldots,N\}}(1-x_1)\cdots(1-x_{j-1})x_j\prod_{k=1}^{N}x_k^{a_k-1}(1-x_k)^{b_k-1}(dx_1,\ldots,dx_N)$$
$$=\int I_{\{x_k\in A_k,k=1,\ldots,N\}}x_j^{a_j}(1-x_j)^{b_j-1}\prod_{k=1}^{j-1}x_k^{a_k-1}(1-x_k)^{b_k}\prod_{k=j+1}^{N}x_k^{a_k-1}(1-x_k)^{b_k-1}(dx_1,\ldots,dx_N).$$

This implies that

$$V_j|X=j\sim Beta(a_j+1,b_j)$$
$$V_k|X=j\sim Beta(a_k,b_k+1)\quad for\quad 0<k<j$$
$$V_k|X=j\sim Beta(a_k,b_k)\quad for\quad k>j$$

as sumarized in the Lemma.

Proof (Proposition 2)

According to definition (2), each component γ_i of γ is a Gaussian r.v. with mean

$$m_i(X)=E(\gamma_i)=-\frac{1}{2}\mu_{X(i)}\log(\lambda_{X(i)})$$

since

$$E\left(\int_{i+1-\lambda_{X(i)}}^{i}(i+1-u)^{-1/2}dB(u)\right)=0.$$

Therefore

$$E(\gamma)=m(X)=\left(-\frac{1}{2}\mu_{X(1)}\log(\lambda_{X(1)}),\ldots,-\frac{1}{2}\mu_{X(n)}\log(\lambda_{X(n)})\right).$$

Now let $Z=\sum_{j=1}^{J}\alpha_j\gamma_j$ be a linear combination of components of γ.

$$Z=\sum_{j=1}^{J}\alpha_j\left(-\frac{\mu_{X(t_j)}}{2}\log(\lambda_{X(j)})+(\mu_{X(j)})^{1/2}\int_{t_j+1-\lambda_{X(j)}}^{t_j}(t_j+1-u)^{-1/2}dB(u)\right)$$

which can be written in the form

$$Z=A_J+\sum_{j=1}^{J}\alpha_j(\mu_{X(j)})^{1/2}\int(t_j+1-u)^{-1/2}I_{[t_j+1-\lambda_{X(j)},\ t_j]}dB(u)$$

that is

$$Z=A_J+\int(t_j+1-u)^{-1/2}B_JI_{[t_j+1-\lambda_{X(j)},\ t_j]}dB(u),$$

showing that Z is a Gaussian r.v. It follows that γ is a Gaussian random vector. Moreover

$$\sigma_{st}=Cov(\gamma_s(X),\gamma_t(X))=E\left[(\gamma_s(X)-m_s(X))(\gamma_t(X)-m_t(X))\right]$$

$$=E\left((\mu_{X(s)})^{1/2}\int_{s+1-\lambda_{X(s)}}^{s}(s+1-u)^{-1/2}dB(u)\times\mu_{X(t)}^{1/2}\int_{t+1-\lambda_{X(t)}}^{t}(t+1-v)^{-1/2}dB(v)\right)$$
$$=(\mu_{X(s)})\mu_{X(t)})^{1/2}E\left(\int_{s+1-\lambda_{X(s)}}^{s}(s+1-u)^{-1/2}dB(u)\times\int_{t+1-\lambda_{X(t)}}^{t}(t+1-v)^{-1/2}dB(v)\right).$$

Proof (Proposition 3)
The covariance matrix coefficients $\sigma_{st} = COV(\gamma_s, \gamma_t)$ involve two Gaussian stochastic integrals:

$$\sigma_{st} = (\mu_{X(s)}\mu_{X(t)})^{1/2} E\left(\int_{s+1-\lambda_{X(s)}}^{s} (s+1-u)^{-1/2} dB(u) \times \int_{t+1-\lambda_{X(t)}}^{t} (t+1-u)^{-1/2} dB(u)\right).$$

Recall that Gaussian stochastic integrals are zero mean Gaussian r.v.s that have the property

$$E\left(\int_{A_1} F(x)dB(x) \times \int_{A_2} G(x)dB(x)\right) = \int_{A_1\cap A_2} F(x)G(x)dx.$$

So, if

$$I = \int_b^a F(x)dB(x)$$

then

$$\sigma_I^2 = \int_b^a F^2(x)dx.$$

Suppose that $s < t$, then $w = s$ and $t = s+\tau$. Let

$$K = E\left(\int_{s+1-\lambda_{X(s)}}^{s} (s+1-u)^{-1/2} dB(u) \times \int_{t+1-\lambda_{X(t)}}^{t} (t+1-u)^{-1/2} dB(u)\right).$$

It follows that

$$K = \int_a^s (s+1-u)^{-1/2}(t+1-u)^{-1/2} du.$$

As in Schmitt (Schmitt 2003) we see that

$$\begin{aligned} K &= \int_a^s \frac{du}{\sqrt{(s+1-u)(t+1-u)}} \\ &= \int_a^s \frac{du}{\sqrt{(s+1-u)(s+1-u+\tau)}} \\ &= \int_1^{1+s-a} \frac{dx}{\sqrt{x(x+\tau)}} \\ &= 2\log\left(\frac{\sqrt{s+1-a}+\sqrt{s+1-a+\tau}}{1+\sqrt{1+\tau}}\right) \end{aligned}$$

where we have made the variable change $x = s+1-u$ and used the identity

$$\int \frac{dx}{\sqrt{x(x+\tau)}} = 2\log\left(\sqrt{x}+\sqrt{x+\tau}\right).$$

Therefore

$$\sigma_{st} = 2(\mu_{X(s)}\mu_{X(t)})^{1/2}\log\left(\frac{\sqrt{w+1-a}+\sqrt{w+1-a+\tau}}{1+\sqrt{1+\tau}}\right).$$

Proof (Proposition 4)
We know that the conditional density of X is

$$(X = x|p = y, \gamma = g, \mu = m, \lambda = l) = \frac{(X, p, \gamma, \mu, \lambda)(x,y,g,m,l)}{\int (X, p, \gamma, \mu, \lambda)(x,y,g,m,l)dP(x)}.$$

But,

$$\begin{aligned} (X, p, \gamma, \mu, \lambda)(x,y,g,m,l) &= (\gamma = g|X = x, p = y, \mu = m, \lambda = l)(X = x, p = y, \mu = m, \lambda = l) \\ &= (\gamma = g|X = x, p = y, \mu = m, \lambda = l)(X = x|p = y, \mu = m, \lambda = l) \\ &\quad \times (p = y, \mu = m, \lambda = l) \\ &= f_n(g,m,l)(X = x|p = y, \mu = m, \lambda = l)(p = y, \mu = m, \lambda = l). \end{aligned}$$

Since

$$(X = x|p, \mu, \lambda) = \sum_{i=1}^{N} p_i \delta_{X_i}(\{x\}),$$

and since the distribution of (p,μ,λ) does not depends on X, we have

$$(X,p,\gamma,\mu,\lambda)(x,y,g,m,l) = f_n(g,m,l)\sum_{i=1}^{N} p_i\delta_{X_i}(\{x\}).$$

Using that
$f(x)\delta_{X_i}(\{x\}) = f(X_i)\delta_{X_i}(\{x\})$, for any function f, we get

$$\begin{aligned}(X,p,\gamma,\mu,\lambda)(x,p,g,m,l) &= \textstyle\sum_{i=1}^{N} f_n(g,m,l)p_i\delta_{X_i}(\{x\}) \\ &= \textstyle\sum_{i=1}^{N} p_i^*\delta_{X_i}(\{x\})\end{aligned}$$

so,

$$(X|p,\gamma,\mu,\lambda) \propto \sum_{i=1}^{N} p_i^*\delta_{X_i}.$$

Proof (Proposition 5)
By proposition (4), $(X|p) \propto \sum_{i=1}^{N} p_i^*\delta_{X_i}$. The result then follows from Lemma 1.

Proof (Proposition 6)
By Connor and Mosimann (CONNOR and MOSSIMANN 1969), the probability density of p defined by equation (5) is

$$\left\{\prod_{k=1}^{N-1}\frac{\Gamma(a_k,b_k)}{\Gamma(a_k)\Gamma(b_k)}\right\} p_1^{a_1-1}\dots p_{N-1}^{a_{N-1}-1}p_N^{b_{N-1}-1}\times(1-P_1)^{b_1-(a_2+b_2)}\dots(1-P_{N-2}^{b_{N-2}-(a_{N-1}+b_{N-1})}),$$

where $P_k = p_1+\ldots+p_k$. When $a_k = 1$ and $b_k = \alpha$, using $\Gamma(1+\alpha) = \alpha\Gamma(\alpha)$, we get that the conditional density of p given α is

$$f(p|\alpha) \propto \alpha^{N-1}p_N^{\alpha-1} = \alpha^{N-1}e^{(\alpha-1)\log(p_N)}.$$

As $f(\alpha|p) \propto f(p|\alpha)f(\alpha)$ and the prior for α is $Gamma(\eta_1,\eta_2)$, we get

$$f(\alpha|p) \propto \alpha^{N-1+\eta_1}e^{-(\eta_2-\log(p_N))\alpha}$$

So, $(\alpha|p) \sim Gamma\,(N+\eta_1-1,\eta_2-\log(p_N))$, that is

$$(\alpha|p) \sim Gamma\left(N+\eta_1-1,\eta_2-\sum_{i=1}^{N-1}\log(1-V_i^*)\right).$$

Proof (Proposition 7)
Let $t_{1j},\ldots,t_{n_j j}$ be the times at which the Markov chain takes value j.
We know from corollary 1 that,

$$(\gamma_{1_j},\ldots,\gamma_{n_j j})|_{\mu,\lambda} \sim \mathcal{N}_{n_j}(-\frac{\mu_j}{2}\log(_j)(1,1,\ldots,1),\mu_j\Sigma_j).$$

Moreover

$$(\gamma,\mu_j,\lambda_j,X,\alpha) = (\gamma|\mu_j,\lambda_j,X,\alpha)\otimes(\mu_j,\lambda_j,X,\alpha) = (\gamma|\mu_j,\lambda_j,X,\alpha)\otimes(\mu_j\otimes\lambda_j\otimes X\otimes\alpha)$$

since X,μ_j,λ_j and α are independent. It follows that

$$(\mu_j|\gamma,\lambda_j,X,\alpha) = \frac{(\gamma|\mu_j,\lambda_j,X,\alpha)\otimes\mu_j\otimes\lambda_j\otimes X\otimes\alpha}{\int(\gamma,\mu_j,\lambda_j,X,\alpha)dP(\mu_j)}.$$

As λ_j, X and α do not depend on μ_j we have

$$(\mu_j|\gamma,\lambda_j,X,\alpha) \propto (\gamma|\mu_j,\lambda_j,X,\alpha)\otimes\mu_j$$

That is

$$(\mu_j|\gamma,\lambda_j,X,\alpha) \sim \Gamma_1 \propto \mathcal{N}_{n_j}(m_j,\mu_j\Sigma_j)*\pi_1.$$

References

1. Bernard, D. and D'Alexis, C. (2006) Mesures atmosphériques à l'entrée de la mangrove en Guadeloupe. *Projet Atmo-Mangrove*, LPAT, Université Antilles-Guyane, Guadeloupe.
2. Connor, R. J. and Mossimann, J. E. (1969) Concept of indepedence for proportions with a generalization of Dirichlet Distribution. *J. Am. Statist. Assoc* 64: 194-206.
3. Dempster, A.P., Laird, R. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via EM algorithm. *J. Royal Stat. Soc. B* 39: 1-38.
4. DI-MASI, G. B., Kabanov, M. Y. and Runggaldier, W. J. (1994) Mean-variance hedging of options on stocks with Markov volatility. *Theory Probab. App* 39: 173-174.
5. Durbin, R., Eddy, S., Krogh, A. and Mitchison, G. Biological Sequence Analysis. *Cambridge University Press*, Cambridge (1998).
6. Ferguson, T. F. (1973) A Bayseian analysis of some non parametric problems. *Ann. Stat.* 1: 209-230.
7. Ghosh, M. K. and Deshpande, A. (2008) Risk minimizing option pricing in a regime switching market. *Stoch. Analysis and App.* vol 28.
8. Hamilton, J. A. (1990) Analysis of time series subject to changes of regime. *J. Econometrics*, 45: 39-70.
9. Hamilton, J.D.and Susmel, R. (1994) Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64: 307-333.
10. Ishwaran, H. and Zarepour, M. (2000) Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two- Parameter Process Hierarchical Models, *Biometrika* 87(2): 371-390.
11. Ishwaran, H. and Lancelot, F. J. (2002) Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. *Journal of Computational and Graphical Statistics* 11(3): 1-26.
12. Ishwaran, H. and Lancelot, F. J. (2003) Some futher developments for Stick-breaking priors: Finite and infinite clustering and clasification. *The Indian Journal of Statistics* 65(3): 577-592.
13. Kolmogorov, A. N. (1962) A refinement hypothesis concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *Journal of Fleuid Mech.* 13: 82-85.
14. Noris J. R. Markov Chains. Cambridge series in statistical and probabilistic mathematics, *Cambridge U. Press* (1997)
15. Rabinier, L.R. A (1989) tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77: 257285.
16. Schmitt, F. G., Vannitsem, S. and Barbosa, A. (1998) Modeling of rainfall time series using renewal processes and multifractals, *J. Geophys. Res.* 103: 23181-23193.
17. Schmitt, F. G. and Marsan, D. (2001) Stochastic equations generating continuous multiplicative cascades. *The European Physical Journal B* 20: 3-6.
18. Schmitt, F. G. (2003) A causal multifractal stohastic equation and its statistical properties. *The European Physical Journal B* 34: 85-98.
19. Sethuraman, J. (1994) A constructive defnition of Dirichlet priors. *Statistica sinica* 4: 639-650.
20. Yaglom, A. M. (1966) The influence of fluctuations in enegy dissipation on the shape of turbulent characteristics in the inertial interval. *Soviet Physics Doklady* 11: 26-29.
21. Yin, G. G., Zhan Q. Continuous-Time Markov chains and application. A singular pertubationapproach. *Springer Verlag, New York Inc.* 1997
22. Zuccini,W. and Guttorp, P. (1991) A hidden Markov model for space-time precipitation. *Water Resources Research* 27: 19171923.